



**Stay-at-Work/
Return-to-Work (SAW/RTW)
Models and Strategies
Evaluation Design
Options Report**

**Contract # DOLQ129633231
Order # 1605DC-17-U-00095**

September 2020

**Chief Evaluation Office and
Office of Disability Employment Policy**

U.S. Department of Labor
200 Constitution Ave, NW
Washington, DC 20210

Submitted by:

Austin Nichols
Michel Grosz
Sarah Prenovitz
Daniel Gubits
Michelle Wood
Jennifer Bagnell Stuart
Zachary Epstein
Cara Sierks

Abt Associates

6130 Executive Boulevard
Rockville, MD 20852

This report was prepared for the U.S. Department of Labor (DOL), Chief Evaluation Office, in collaboration with the Office of Disability Employment Policy by Abt Associates, under contract number DOLQ129633231 Order # 1605DC-17-U-00095. The views expressed are those of the authors and should not be attributed to DOL nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

Table of Contents

List of Abbreviations iv

Executive Summary v

1. Introduction 1

 1.1 Policy Context 1

 1.2 What Is Stay-at-Work/Return-to-Work? 2

 1.3 Organization of the Report 4

2. Findings from the Knowledge Development Phase and Implications for Evaluation

Design Options..... 6

 2.1 Program Review and Synthesis 6

 2.2 Summary of Evidence 9

 2.3 Early Intervention Pathways and Target Population Profiles 11

3. Option A: Longitudinal Study of Individual Workers..... 16

 3.1 Research Questions..... 16

 3.2 Conceptual Model..... 17

 3.3 Intervention(s) 17

 3.4 Participants 17

 3.5 Data..... 18

 3.6 Analysis Plan 19

 3.7 Practical Considerations 22

 3.8 Contributions and Limitations 24

4. Option B: Test an Informational Intervention for Workers: Offer Workers Advice on Work and Disability 25

 4.1 Research Question(s)..... 25

 4.2 Conceptual Model..... 26

 4.3 Intervention..... 26

 4.4 Participants 30

 4.5 Data..... 32

 4.6 Analysis Plan 33

 4.7 Practical Considerations 35

 4.8 Contributions and Limitations 36

5. Option C: Examine Employer Practices and Test an Intervention that Informs Employers about Resources and Best Practices to Retain Workers with Impairments 38

 5.1 Research Question(s)..... 38

 5.2 Conceptual Model..... 38

 5.3 Intervention(s) 39

TABLE OF CONTENTS

5.4	Participants	42
5.5	Data.....	44
5.6	Analysis Plan	46
5.7	Practical Considerations	47
5.8	Contributions and Limitations	48
5.9	Potential for Combining Options B and C.....	48
6.	Option D: Test the Effects of Informing Providers about Best Practices in Medicine to Facilitate SAW/RTW	50
6.1	Research Question(s).....	51
6.2	Conceptual Model.....	51
6.3	Intervention.....	52
6.4	Participants	53
6.5	Data.....	54
6.6	Analysis Plan	57
6.7	Practical Considerations	58
6.8	Contributions and Limitations	59
7.	Option E: Partial Payments in Temporary Disability Insurance Programs.....	60
7.1	Research Question(s).....	62
7.2	Conceptual Model.....	62
7.3	Intervention(s)	64
7.5	Data.....	68
7.6	Analysis Plan	70
7.7	Practical Considerations	73
7.8	Contributions and Limitations	75
8.	Summary	76
8.1	Summary of Open Questions.....	76
8.2	Summary of Designs	77
8.3	Considerations in Selecting Designs	77
8.4	Conclusion.....	82
	References.....	83
	Appendix A. Sources for Administrative Data Matched in Option A	90

List of Exhibits

Exhibit 1-1. A Stylized Map of the SAW/RTW Concept..... 3

Exhibit 3-1. Proportion of Workers Separated from Work Who Participate in Six Touchpoints in the 16 Months Following Separation from work, and Rates of Return to Work, and Receipt of SSDI/SSI 17-20 Months after Earnings Loss 17

Exhibit 3-2. Sample Sizes to Detect a Given Hazard Rate Compared to 2 Percent Baseline Rate with Nine Follow-Up Time Periods..... 21

Exhibit 3-3. Sample Sizes to Detect a Given Hazard Rate Compared to 0.5 Percent Baseline Rate with 120 Follow-Up Time Periods..... 21

Exhibit 4-1. Comparison of Low-Intensity Model versus Personalized Information Model..... 29

Exhibit 7-1. Income by Earnings under TDI..... 63

Exhibit 8-1. Summary of Evaluation Design Options..... 78

Exhibit 8-2. Evaluation Design Options: Practical Considerations 79

List of Abbreviations

COHE	Centers of Occupational Health and Education
DOL	Department of Labor
IRB	Institutional Review Board
ODEP	Office of Disability Employment Policy
OSHA	Occupational Safety and Health Administration
RETAIN	Retaining Employment and Talent after Injury/Illness Network
RTW	Return-to-Work
SAW	Stay-at-Work
SIPP	Survey of Income and Program Participation
SSA	Social Security Administration
SSDI	Social Security Disability Insurance
SSI	Supplemental Security Income
TDI	Temporary Disability Insurance

Executive Summary

When workers experience a major injury or illness, some may develop a work disability and leave the labor force. Such labor market exits can have far-reaching consequences for the workers, but also for employers and governments. Workers who exit the labor force lose earnings, which affects the well-being of the workers and their families. These workers eventually may apply for federal disability programs, with substantial consequences for the costs of those programs. (Bardos et al., 2015).

Stay-at-Work/Return-to-Work (SAW/RTW) programs aim to help such ill or injured workers remain at work or return to work as soon as medically feasible. Successful SAW/RTW programs would benefit workers, employers, and government budgets. Crafting SAW/RTW programs that are effective and efficient requires understanding the current policy landscape as well as evidence about what kinds of supports are effective and for whom.

This is the final report for the *Stay-at-Work/Return-to-Work (SAW/RTW) Models and Strategies* project. The U.S. Department of Labor’s Chief Evaluation Office, in collaboration with the Office of Disability Employment Policy (hereafter collectively referred to as DOL), funded this project to build evidence about effective SAW/RTW strategies. To carry out the project, DOL contracted with Abt Associates to describe existing SAW/RTW programs, review evidence about these programs, analyze target populations and potential early intervention pathways, and develop evaluation design options.

Abt developed four deliverables to document the results of these activities:

- 1) Synthesis of Stay-at-Work/Return-to-Work (SAW/RTW) Programs, Models, Efforts, and Definitions (Epstein et al., 2020)
- 2) Synthesis of Evidence about Stay-at-Work/ Return-to-Work (SAW/RTW) and Related Programs (Nichols et al., 2020a)
- 3) Early Intervention Pathway Map and Population Profiles (Nichols et al., 2020b).¹

Drawing on the lessons from the first three deliverables, we developed this report:

- 4) Evaluation design options to help DOL plan future research to build the evidence base about SAW/RTW. This final report for the project presents detailed descriptions of five options for new research to build evidence about the target populations for SAW/RTW and to evaluate the effectiveness of SAW/RTW interventions on employment outcomes.

Objectives and Approach

This report provides DOL with five evaluation designs that address two core research questions about SAW/RTW:

- 1) What is the effect of SAW/RTW interventions on employment for workers who experience illness or injury?
- 2) What is the effect of SAW/RTW interventions on application and receipt of federal disability benefits?

To develop the design options we drew on findings from the review of 68 SAW/RTW programs (Epstein et al., 2020) and the evidence review (Nichols et al., 2020a). Those findings revealed gaps in knowledge

¹ All deliverables for this project are available at www.dol.gov/oasp/evaluation/completedstudies.gov

about which types of interventions are most effective in promoting return to work. Our analysis of the pathways that workers take after the onset of a potentially disabling illness or injury (Nichols, et al., 2020b) also pointed to the need for further research to understand which workers are most likely to benefit from SAW/RTW interventions. The findings from these early project activities and expert opinion from the group of outside experts on the project's Technical Working Group suggested several promising areas for further study. We developed evaluation design options to address these promising areas:

- Describing the target populations for SAW/RTW interventions
- Testing interventions related to the health care touchpoint and employer practices
- Testing the effectiveness of particular types of supports for workers, medical professionals, and employers
- Determining which SAW/RTW program models and strategies are effective, for whom.

Evaluation Design Options

The five evaluation design options listed below address critical research questions about workers who experience a major illness or injury and their employers, health care providers, and insurers. These options could assist DOL to prioritize future research and evaluation efforts and to design interventions to help Americans stay at work or return to work after experiencing an injury or illness. We have developed general plans for implementing each option that address:

- (1) **Research questions.** We developed the options to address the two core research questions about the effects of SAW/RTW interventions. Each option would also answer questions about a specific stakeholder—worker, employer, medical provider, or insurer.
- (2) **Conceptual model.** For each option we describe the rationale for the study or the intervention to be tested and the mechanisms by which the intervention might achieve the outcomes.
- (3) **Intervention.** Options B through E would test new interventions, directed to workers, employers, or medical professionals. We describe the type of intervention to be offered in each option.
- (4) **Participants.** We discuss the target population for each option, the setting for the study, and possible methods to identify and recruit the target population.
- (5) **Data.** For each option we discuss possible data sources and data collection considerations.
- (6) **Analysis plan.** Before implementing any of these options, a research team would need to develop a detailed analysis plan. We discuss the outcomes to be measured and preliminary assumptions and considerations for conducting the analysis, including plausible effects, sample sizes, and statistical power.
- (7) **Practical considerations.** We discuss requirements for developing and implementing the intervention and evaluation. We also discuss implications for the timeline, cost considerations, and geographic scope.
- (8) **Contributions and limitations.** We discuss the policy relevance of the findings each option could produce, the contributions each could make to policy, and limitations on internal and external validity.

In brief, the five evaluation design options are:

Option A: Longitudinal study of individual workers. This option would examine potential target populations for SAW/RTW efforts and chart patterns of program participation. This option would not test a specific intervention but could produce quasi-experimental evaluation evidence for patterns of program participation that could be used to design future interventions. It would use a nationally representative set of data, matched to administrative records from the Social Security Administration (SSA) to follow workers after injury or illness. This would build on the data analysis conducted for the project analyzing pathways workers take after illness or injury (see Nichols et al., 2020b, available at <https://www.dol.gov/agencies/oasp/evaluation/completedstudies>).

Option B: Test an informational intervention for workers: Offer workers advice on work and disability. This option would investigate whether providing targeted information about SAW/RTW to workers who have recently experienced an injury or the onset or worsening of an illness improves their employment outcomes. This design would provide a random sample of workers with information and advice on a range of topics, such as work limitations for those with particular conditions, employer policies, and local or national resources. Information might be delivered in a low-intensity format, such as a letter, text message, or through a meeting with a counselor who provides personalized information.

Option C: Examine employer practices and test an intervention that informs employers about resources and best practices to retain workers with impairments. This option would generate information on current employer SAW/RTW practices. The option would explore the extent to which providing information about accommodations and best practices improves employers' use of these strategies. This option involves two components, a survey of employers (Component 1) to learn about SAW/RTW practices and an experimental evaluation (Component 2) to test whether providing information about SAW/RTW practices affects the behavior of employers. Component 1 could be implemented as a standalone study. Component 2 would require Component 1, because the intervention tested in the experimental evaluation would be delivered via the survey.

Option D: Test the effects of informing providers about best practices in medicine to facilitate SAW/RTW. This option would identify occupational medicine best practices then estimate the impact of an informational intervention to promote those practices. This option would use an experimental evaluation design to test whether an informational “nudge” changes the behavior of medical professionals to promote SAW/RTW among their patients. Medical professionals would be sent messages about the best practices that would contain one or more recommended activity to encourage continued employment among their patients. We describe two interventions: 1) a checklist containing multiple activities or 2) advice about a single activity, to ask patients about returning to work. We outline evaluation designs that randomize by large geography or by individual medical professional.

Option E: Evaluate the effects of partial payments in temporary disability insurance programs.

This option would examine the effects of partial-payment provisions within Temporary Disability Insurance (TDI) on return to work and SSDI application. Partial-payment provisions allow workers to receive part of their TDI benefit if they are able to return to work but at a lower number of hours or lower wages. We outline a research design that would implement a new partial-payment provision in an existing state or private TDI program. The study would randomly assign TDI users to eligibility for the benefit. Then it would collect data on program operations, program costs, and worker outcomes. We also suggest two alternatives to this design: (1) conducting an evaluation of existing partial-payment benefit programs using existing data and a quasi-experimental analysis; and (2) developing, implementing, and testing the effect of a new TDI program designed to incentivize and support SAW/RTW.

Option A would provide a systematic analysis of the target population for SAW/RTW interventions but would not test a specific intervention. Options B-E would test interventions focused on workers, employers, and medical professionals to learn more about what encourages higher employment rates. Options B-D would test the effect of providing additional information to workers, employers, or medical professionals. Option E would test the effects of partial payment provisions in temporary disability benefits on employment and disability application outcomes and program costs.

Considerations in Selecting Design Options

There are several factors DOL would need to consider in selecting and pursuing these design options. The options vary in the questions they answer, the level of effort required, the methods they use, and the time required. Before implementing any of these evaluation designs, DOL would need to develop a detailed analysis plan, schedule, and cost estimate. Each option also involves technical and practical requirements that DOL would need to address. DOL might also consider whether to combine options, using some research to inform other research, or making use of partnerships or permissions more than once.

Study Design

The quality of evidence provided by any research project depends in part on the evaluation design used, the sample size and thus statistical power of the study, and the generalizability of results. We considered a range of evaluation designs to address each of the options. We believe that the evaluations in Options B through E could be conducted using random assignment, which would provide causal evidence about effects with high internal validity.² In order to reliably detect the effect of an intervention, or differences between groups, sufficiently large samples are needed. Sample size needs depend on the expected size of an effect, the variance of outcomes, the data used, and the level (individual or cluster) at which treatment status is determined. In the descriptions of Options B, C, D, and E we present general assessments of potential sample sizes, statistical power, and preliminary estimates of minimum detectable effects. Before implementing any of the designs a research team would need to examine these preliminary assumptions carefully and develop a more detailed analysis plan

Practical Considerations

Practical considerations include a study's timeline, implementation complexity, data collection requirements, and the associated costs. We estimate that Options A and B might be completed in roughly 4 years, Options C and E might require up to 5 years, and Option D would likely require more than 5 years. If DOL chose to evaluate the effect of any of Options B, C, D, or E on long-term employment or benefit applications, a longer follow-up period would be needed.

² If executed well, studies with these designs would be eligible for top rankings (e.g. 'high' or 'meets standards without reservations' from systematic clearinghouses such as CLEAR or the What Works Clearinghouse).

Several factors influence the potential costs of the evaluation design options. Options B-E would require the design and implementation of a new intervention. The study designs that use random assignment would require a detailed randomization plan and sample recruitment strategy. Our experience shows that implementing those strategies and sustaining them over the period needed to enroll an adequate sample could also pose substantial costs. Delivering the intervention and monitoring to make sure it is delivered with fidelity are other sources of costs. Another cost consideration is data collection. Option A would use existing data, but Options C and E might involve developing and administering new surveys.

Conducting new surveys would require Paperwork Reduction Act approvals which would influence both costs and the studies' timelines. Options B, C, D, and E would also likely rely on new arrangements to collect administrative data to measure outcomes and these arrangements would require negotiating data access and permissions. All options would require partnerships with other organizations or agencies, but the options vary in the number of partnerships needed as well as their extent. Chapters 3 through 7 discuss the practical considerations and challenges associated with each option in detail.

We also discuss the challenges to implementation. These are hurdles that DOL and its contractors would need to address in order to pursue the options. Some are common issues that have been successfully overcome in other research studies, while others are likely to be more challenging. For example, Option A would require data sharing agreements and permissions to use the data. Depending on the exact data used, this could be addressed using standardized procedures or could require more extensive coordination within and across agencies. Designs that implement an intervention with the employer or medical professional (C and D) may face particular challenges with recruiting the study sample and gaining their cooperation, as well as with obtaining individual workers' employment outcomes.

All options would require partnerships, but the options vary in the number of partnerships needed as well as their extent. Option E-3, which would entail establishing a new TDI program, would require a particularly high level of coordination with partners. Option C, however, might only require partners to provide letters of support, which would involve much less coordination and likely be easier to arrange. Chapters 3 through 7 discuss the practical considerations and challenges associated with each option.

Complementary Research Questions

The evaluation design options are not mutually exclusive. Combining options strategically might allow DOL to generate more information from two (or more) research options than could be achieved by pursuing both individually. Combining options might also achieve efficiencies by using some of the same partnerships and data use arrangements more than once.

For example, information from Option A could help determine the workers most likely to be on the margin of staying at work or returning to work, and these workers could be targeted for Option B. Findings from Option C could be used to customize the information provided to workers in Option B to reflect the options that were available at their employer or to provide data on the prevalence of certain policies within their industry. Combining options might also allow DOL to obtain information on the synergies that would be achieved if multiple new policies were adopted more widely. For example, Options B, C, and D all involve providing information to stakeholders in the SAW/RTW process. It is possible that the effect of informing both the worker and the employer is larger than the effect of informing the worker plus that of informing the employer, because the two parties are better able to coordinate and work together, or less effective if the informed party would have otherwise shared information with the other. Finally, combining options could allow DOL to achieve efficiencies for some of the administrative tasks by working with some of the same partners or using some of the same data sources more than once.

1. Introduction

When workers experience a major injury or illness, some may develop a work disability and leave the labor force. Such labor market exits can have far-reaching consequences for the workers, but also for employers and governments. Workers who exit the labor force lose earnings, which affects the well-being of the workers and their families. These workers eventually may apply for federal disability programs, with substantial consequences for the costs of those programs. (Bardos et al., 2015).

Stay-at-Work/Return-to-Work (SAW/RTW) programs aim to help such ill or injured workers remain at work or return to work as soon as medically feasible. Successful SAW/RTW programs would benefit workers, employers, and government budgets. Crafting SAW/RTW programs that are effective and efficient requires understanding the current policy landscape as well as evidence about what kinds of supports are effective and for whom.

The U.S. Department of Labor (DOL) Office of Disability Employment Policy (ODEP) and the Chief Evaluation Office is undertaking a rigorous effort to build evidence about effective Stay-at-Work/Return-to-Work (SAW/RTW) strategies. This project, *Stay-at-Work/Return-to-Work (SAW/RTW) Models and Strategies*, is one part of that effort.

To carry out the project, DOL contracted with Abt Associates to describe existing SAW/RTW programs, review evidence about these programs, and analyze target populations and potential early intervention pathways.

Drawing on lessons from these activities, we developed evaluation design options to help DOL plan future research to build the evidence base about SAW/RTW that will help guide policy. This final report for the project presents detailed descriptions of five options for new research to build evidence about the target populations for SAW/RTW and to test the effects of interventions on SAW/RTW outcomes.

1.1 Policy Context

When workers experience a major injury or illness, some leave the labor force, temporarily or permanently. Such labor market exits can have far-reaching consequences for the workers, but also for employers and government entities. SAW/RTW programs aim to help workers remain at work or return to work as soon as medically feasible. Successful SAW/RTW programs would benefit workers, employers, and government budgets (Bardos et al., 2015; Waddell & Burton 2006).

- For **workers and their families**, successful SAW/RTW programs would help maintain workers' productivity and standard of living.
- For **employers**, successful SAW/RTW programs could save the costs of hiring and training new workers. If the injury is work-related, successful SAW/RTW programs could keep their workers' compensation insurance premiums down.
- For the **federal government**, successful SAW/RTW programs could potentially reduce spending on federal disability benefits.
- For **all levels of government**, successful SAW/RTW programs that keep workers employed would help generate tax revenues.

Given these potential benefits, identifying effective SAW/RTW strategies—especially early interventions that precede a worker's receiving federal disability benefits—is an important policy priority. Ideally, such early interventions begin as soon as possible after a worker is injured or falls ill, or after a worker's

condition worsens, threatening to become a work disability. Identifying individuals early in the injury/illness process can be challenging, particularly before such individuals have interacted with an employer, medical professional, or other source of SAW/RTW assistance. On the other hand, interventions might be most effective at improving employment when they begin soon after the onset of a medical condition. The evaluation design options we consider in this report recognize this tradeoff and help DOL generate knowledge about the appropriate target populations for SAW/RTW interventions and how best to design policy that improves employment outcomes.

1.2 What Is Stay-at-Work/Return-to-Work?

Currently, when illness or injury threatens a worker's ability to work, no single, coordinated service delivery system exists to help them remain in the labor force. Instead, injured or ill workers must navigate a range of service systems—health care, insurance, employers' initiatives, vocational rehabilitation, workforce assistance—all with different goals and rules (Ben-Shalom et al., 2017). This section discusses how SAW/RTW programs might operate within these service systems.

If an injury occurs on the job or the illness is work-related, workers typically are eligible for medical care and cash benefits through the state's workers' compensation program. Workers' compensation insurance pays for the cost of medical care to treat the illness or injury. Workers' compensation insurance also pays a portion of the employee's salary, while the employee is unable to work, within minimum and maximum limits set by state law (McLaren et al., 2018). Benefit provisions and workers' compensation laws differ substantially across states. Workers who experience a work-related injury or illness that is compensated through the workers' compensation program may receive services from their employers to help them remain at work. These workers may also seek treatment from their private health care providers, in addition to the assistance they receive from the workers' compensation program.

An illness or injury that is not work-related is not covered by workers' compensation. Workers with non-work-related illness or injuries:

- Sometimes can receive services from their **employers** to help them remain at work.
- Can receive treatment from their **private health care providers** that may or may not aim to help them stay at work.
- Might have **private disability insurance**, either provided by their employers or purchased on their own, that can pay a portion of lost wages if they are unable to work. (About 40 percent of workers have this insurance; DOL/Bureau of Labor Statistics, 2019).³
- Sometimes seek assistance from **state Vocational Rehabilitation programs or workforce agencies** to help them remain at work or return to work at the same or different job.

The stylized map in Exhibit 1-1 illustrates how an ill or injured worker might seek services and potentially interact with multiple service providers. The exhibit starts at the far left with a **healthy individual**. Just to the right, an injury occurs or an illness develops or worsens. In the context of a fragmented service delivery system, Stay-at-Work/Return-to-Work programs have emerged as a promising policy direction.

³ Often, private disability insurance uses less stringent definitions of disability than do federal disability benefits programs (namely Supplemental Security Income (SSI) and Social Security Disability Insurance (SSDI), and private insurance may offer larger benefits).

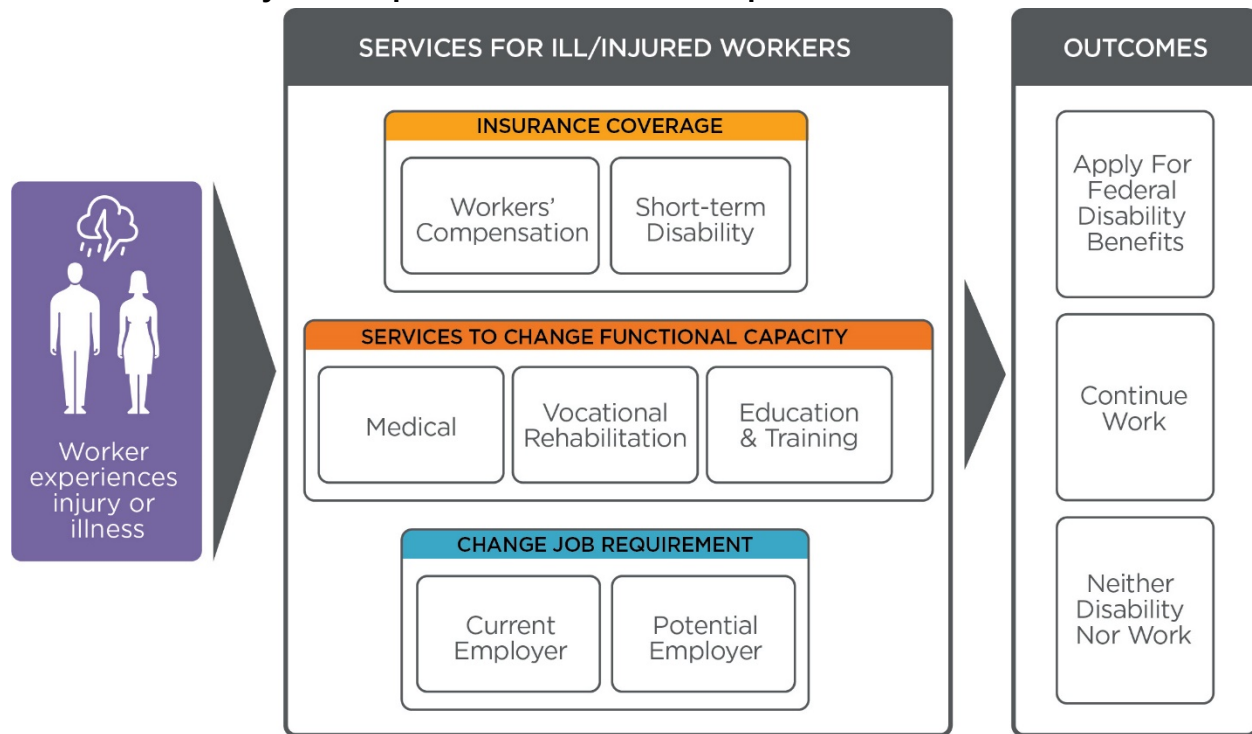
Exhibit 1-1. A Stylized Map of the SAW/RTW Concept

Exhibit 1-1 shows that workers might be able to access a range of supports and services.

- **Insurance coverage**—Benefits offered to the individual through coverage under workers’ compensation, short-term disability insurance (public or private), or some other provider.
- **Change in functional capacity**—Services to improve the individual’s functional capacity for work, including medical treatment, vocational rehabilitation, or new education and training.
- **Change in job requirement**—Services to change what work is performed, how, and/or where including accommodations made for the individual’s medical condition that allow him or her to continue or resume work.
- **A SAW/RTW program**—Provides or coordinates one or more service or and activity in the three areas described above.

Any of the service providers shown in the exhibit—insurers, employers, health care providers, or vocational rehabilitation service providers—can operate SAW/RTW programs. Effective SAW/RTW program may engage with multiple service providers and adopt several strategies to encourage workers to remain in the labor force.

These programs may alter incentives or information provided to workers or to any of the service providers that interact with them. Altered incentives or information might influence employers, health care providers, or other service providers to change how they operate. Changed incentives might also influence worker behavior. For example, employers may be more likely to make workplace accommodations if they have new information to help them. Providers might also forge new linkages. For example, faced with a new incentive, an insurer might refer workers to a novel SAW/RTW program instead of encouraging them to apply for disability benefits.

These changes in how employers and other providers operate and how workers behave could result in improved outcomes. If successful, the ill or injured individual could be more likely to resume or continue to work, represented by the **Continue Work** box on the far right in the Exhibit 1-1. However, if the changes are not successful, the individual may not continue working and eventually apply for federal disability benefits, represented by the **Apply for Federal Disability Benefits** box. Other workers will neither return to work nor apply for federal disability benefits, represented by the **Neither Disability nor Work** box.

1.3 Organization of the Report

The balance of this document proceeds as follows. **Chapter 2** summarizes the distinctive features and available evidence of existing SAW/RTW programs and preliminary conclusions about the pathways workers take from illness or injury to federal disability benefits. Chapter 2 also describes the ***Retaining Employment and Talent after Injury/Illness Network (RETAIN)*** demonstration that DOL is conducting in conjunction with the Social Security Administration. The RETAIN demonstration will test the impact of early-intervention strategies on the employment outcomes of individuals who experience work disability while employed.⁴ The RETAIN demonstration will address several research questions about the impacts of a return-to-work coordinator who facilitates the early provision of health care and employment-related supports and services, as well as physician education about occupational health best practices, and incentives for adopting those best practices. RETAIN will produce evidence on the impacts of these interventions on employment outcomes and application for federal disability benefits. Because the RETAIN demonstration is important context for any new evaluation design option DOL might pursue, this report discusses how the evaluation design options presented here would build on lessons learned from RETAIN.

Chapters 3-7 of this report describe the five evaluation design options in detail, focusing on eight factors:⁵

- (1) **Research questions.** We developed the options to address the two core research questions about the effects of SAW/RTW interventions. Each option would also answer questions about a specific stakeholder—worker, employer, medical provider, or insurer.
- (2) **Conceptual model.** For each option we describe the rationale for the study or the intervention to be tested and the mechanisms by which the intervention might achieve the outcomes.
- (3) **Intervention.** Options B through E would test new interventions, directed to workers, employers, or medical professionals. We describe the type of assistance to be offered in each of the interventions.
- (4) **Participants.** We discuss the target population for each option, the setting for the study, and possible methods to identify and recruit the target population.
- (5) **Data.** For each option we discuss possible data sources, and data collection considerations.
- (6) **Analysis plan.** Before implementing any of these options, a research team would need to develop a detailed analysis plan. We discuss preliminary assumptions and considerations for conducting

⁴ A work disability is a medical condition that limits a person's ability to work.

⁵ For the intervention strategies in Options B, C, D, and E, we outline evaluation options using random assignment. Evaluations designs that use random assignment are typically eligible for top rankings (e.g., "high" or "meets standards without reservations" from systematic evidence review clearinghouses, such as DOL's Clearinghouse for Labor Evaluation and Research (CLEAR) or the What Works Clearinghouse.

the analysis, including outcomes to be measured, plausible effects, sample sizes and statistical power.

- (7) **Practical considerations.** We discuss requirements for developing and implementing the intervention and evaluation. We also discuss implications for the timeline, cost considerations and geographic scope, and threats to implementation as designed.
- (8) **Contributions and limitations.** We discuss the policy relevance of the findings each option could produce, the contributions each could make to current knowledge, and limitations on internal and external validity.

Chapter 8 concludes by summarizing the five evaluation design options and providing a framework of both technical and practical considerations to facilitate comparison of their relative merits. We include references at the end of the report.

2. Findings from the Knowledge Development Phase and Implications for Evaluation Design Options

To generate the evaluation design options presented in Chapters 3-7, the Abt team conducted a program synthesis and evidence review.⁶ Then, we analyzed publicly available data to examine potential target populations and pathways for early interventions.⁷ The lessons from these activities helped us identify open research questions and promising areas for interventions and evaluation design options.

This chapter summarizes the findings of these activities. The chapter also discusses the evidence we anticipate DOL will develop through the *Retaining Employment and Talent after Injury/Illness Network* (RETAIN) demonstration and how we used information about RETAIN to develop the evaluation design options.

2.1 Program Review and Synthesis

To understand the service delivery systems in which new interventions or studies of existing ones would operate, we conducted a systematic, structured search that identified 68 SAW/RTW programs operating in the United States. These 68 include programs that were active in 2018 or early in their implementation, as well as demonstrations that have concluded.⁸

We found that SAW/RTW programs vary in five dimensions:

- **Program components or types of services offered**, including employer-provided job accommodations, medical treatment innovations, case management or information sharing, and counseling and training of various kinds.
- **Entities that administer the program**, including employers, state workers' compensation systems, state labor departments, state vocational rehabilitation agencies, and insurance companies.
- **Timing of intervention** relative to application for federal disability benefits: before application, after application but before award of benefits, or after benefit award.⁹
- **Target population**, whether targeted to workers with specific types of health conditions or to workers with work-related illnesses or injuries.
- **Service providers or stakeholders that engage with the program**, including health care providers, return-to-work coordinators or case managers, and Employee Assistance Programs.

⁶ The review of existing programs is reported in Epstein et al. 2020. The review of existing evidence is reported in Nichols et al. 2020a. We limited the scope of the evidence review to studies published from 2008-2018 (the 10 years prior to the review). We did so to consider the most relevant evidence—in which findings are still applicable, and where the technologies, laws and public policies examined are still relevant.

⁷ The results of the analysis of potential target populations and early intervention pathways is reported in Nichols et al. 2020b.

⁸ The methods we used to search for and categorize SAW/RTW programs are described in Section 1.2 and Appendix A of Epstein et al. 2020.

⁹ For this study, we defined early-stage SAW/RTW programs as those that assist a worker prior to application for federal disability programs. Such interventions may not necessarily be early relative to the onset of the injury or illness. Information available about the programs was insufficient to determine timing of the intervention relative to onset of the illness or injury.

2. FINDINGS & IMPLICATIONS

Our synthesis identified five main types of SAW/RTW program components:

1. **Employer-provided job accommodations:** Of the programs we identified, 38 percent offered employer-provided job accommodations. Common examples included physical changes to the employee's workplace or workstation to promote accessibility such as automatic door openers, ramps, or wider doorways, and assistive technologies such as speech-recognition software, screen readers, automatic page-turners, and book holders. Other types of accommodations included modifications to work policies and procedures such as allowing flexible work times or altering job duties.
2. **Financial incentives to employers and workers:** Altogether, 37 percent of the programs we identified offered financial incentives to employers or workers. We found that financial incentives targeted to employers are typically implemented either through WC programs or through the tax code. Some WC agencies have explicit policies that incentivize employers to hire or retain workers with disabilities. Examples include "preferred worker programs," in Oregon, North Dakota, and Washington, in which employers are offered a wage reimbursement and relief from WC premiums or future claims costs in return for hiring a qualifying worker (Council of State Governments, 2018). Another example is Washington's Stay at Work Program which provides wage reimbursement for WC claimants when employers provide light duty or transitional work to a claimant not ready to return to work. That program also provides other types of assistance to encourage employers to promote return to work, including vocational and return to work consulting services, funds for training, and work site modifications (Stapleton & Christian, 2016). Second injury funds, which cover benefits for future WC claims, may also lessen employers' concerns that a claimant is particularly prone to future work-related injuries or illnesses, and thus encourage employers to retain workers who have filed a WC claim.

We also found that six states (Delaware, Iowa, Louisiana, Maryland, New York, and Tennessee) offer tax credits to employers to encourage hiring workers with disabilities. At the federal level, through the Work Opportunity Tax Credit, employers may claim a tax credit if they hire and retain individuals who have significant barriers to employment and are referred from state Vocational Rehabilitation programs. The credit applies to all qualifying employees hired in a tax year and is generally worth 25 percent or 40 percent of a new employee's first-year wages, up to a maximum.

We also found examples of policies operated through workers' compensation and short-term disability insurance programs to lessen disincentives to work that arise for workers through the wage replacement benefits the programs offer. These measures include not allowing vacation and sick time to accrue during the absence, holding the job open for a defined period of time, setting proactive return to work policies, and communicating with workers during the absence. Insurers can also implement programs to improve the attractiveness of returning to work. Some of these programs operate through incentives to employers to offer the worker an easy return to work.

Private disability insurers can also offer temporary partial disability benefits for workers who are able to return to light duty or part-time work, at lower earnings than the pre-injury job. Similar benefits are mandated by WC regulations in many states and might help to maintain the worker's attachment to the job and make returning to work more financially beneficial to the worker than remaining out of work altogether (McLaren et al., 2018).

3. **Information or navigation help:** We found that more than half (60 percent) of all programs we identified offered information or navigation help to workers. Examples include technical assistance, case management, and case coordination. Current SAW/RTW practice emphasizes

2. FINDINGS & IMPLICATIONS

helping workers to navigate post-injury services or sharing information across providers to promote better coordination of services.

4. **Medical management:** SAW/RTW program staff may directly engage with medical providers or encourage workers to engage more directly with their medical providers about return to work. We found that about one-quarter (26 percent) of the programs we reviewed offered this type of assistance. SAW/RTW program staff might ask attending physicians to approve accommodation plans, but our review identified only one program that attempts to modify health care provider practices. The Centers for Occupational Health and Education (COHE) program operates through Washington’s monopolistic workers’ compensation system, and provides health care providers received payments for timely completion of reports of accident forms and use of occupational best practices.
5. **Employment services and training:** About one quarter (26 percent) of the programs we identified include an employment services and training component. This program component is not commonly included among SAW/RTW initiatives implemented in the workers’ compensation system. The finding is consistent with that system’s emphasis on transitional and light-duty work. Such transitional assignments may require workers to use their existing skills in a modified work assignment rather than applying new skills to alternative work. We also found that employment services and training are more common in late-stage interventions, targeted to individuals who are already receiving federal disability benefits.

The lessons from our program review and synthesis underscored two main points that influenced our evaluation design options.

First, SAW/RTW programs vary substantially. We do not find that any one particular service delivery system is commonly used to help workers who experience an illness or injury. Multiple stakeholders—employers, health care providers, insurers (workers’ compensation programs and private or state disability programs), and other service providers—may operate SAW/RTW programs. The stakeholders may face different goals and incentives which influence the types of services they may offer to workers. For example, insurers have strong incentives to encourage workers to remain at work because return to work lowers the costs of insurance programs. Workers compensation and other insurance providers may offer incentives to workers and employers to encourage return to work, but decisions about whether to return to work are made by the worker. The insurers have little control over worker outcomes. Encouraging workers to remain at work can benefit employers by reducing costs of turnover and training that occur when a worker leaves the job, but employers may perceive costs of providing accommodations to help workers stay at work. We think that the variation that we found in the types of programs reflects these varying incentives. We found a wide range of services, such as employer-provided job accommodations, including job modifications or assistive technologies, medical treatment innovations, case management or information sharing, and counseling and training of various kinds (see Epstein et al 2020).

Second, SAW/RTW strategies differ in administrative contexts: work-related programs versus non work-related programs. Lessons learned regarding the administration of workers’ compensation initiatives may not directly apply to the administration of programs for workers whose illness or injury is not work-related. Non-work related illnesses and injuries are more frequent than work-related injuries, and the rates at which these illnesses and injuries lead to federal disability claims do not seem to differ, though individuals receiving SSDI are *less* likely to have received workers’ compensation than those with a work disability and not receiving SSDI (Reville & Schoeni, 2003). Identifying and testing the impact of interventions available to help workers who experience an illness or injury that is not work-related are needed to inform policy on a broad scale.

2.2 Summary of Evidence

The Abt team conducted a review of the evidence on the effects of SAW/RTW or related programs on two key outcomes: employment and the receipt of federal disability benefits. This section summarizes findings from the evidence review. Nichols et al. (2020a) provides more details about the evidence review.

2.2.1 Methods

We first searched for evidence about the 68 programs we found in the program review and synthesis described in Section 2.1. We found that only 11 of the 68 programs have been evaluated in studies reporting direct evidence of impacts. Of these, six are not early interventions, because they involve current disability beneficiaries. For the evidence review, we reviewed evidence available for the five programs that are early interventions. We also conducted a comprehensive search for additional evidence. The search for additional evidence yielded 377 sources, including journal articles, reports, and websites.

Of the 377 sources, we removed 90 that did not include an evaluation, and 35 that did not evaluate outcome measures that can generate relevant findings. We also eliminated 111 studies with publication dates prior to 2008, two studies of programs that are not early interventions,¹⁰ one study that did not involve a comparison group, and 17 that did not study a return-to-work intervention. We also identified 28 review articles that we analyzed separately.

We included the 87 remaining studies in a meta-analysis of evidence. We coded each study according to:

- the program model examined (employer-provided accommodations, financial incentives for employers or workers, information, medical management, employment services and training)
- type of disability (mental health, musculoskeletal, or other)
- other study features such as the quality and relevance of evidence presented

Of the 87 articles, 72 provided “high-quality” evidence (16 judged more relevant,¹¹ and 56 less relevant) and 15 provided “low-quality” evidence (5 judged more relevant and 10 less relevant).¹² To ascertain how estimated effects vary with program model or other study characteristics, we analyzed that larger body of evidence using meta-analytic regression, which is the most efficient way to synthesize findings (DerSimonian & Laird, 1986).

2.2.2 Results

The evidence reviewed varies widely in its quality, that is, its credibility to identify causal impacts. The evidence also varies in the breadth of its applicability or relevance for the U.S. workforce. Many studies with high-quality evidence have limited relevance because of a non-U.S. geographic scope. Because the

¹⁰ We excluded studies on the effects of six late-stage interventions that we had identified in the program synthesis. Our search for this report also found studies about two other late-stage interventions and we excluded those from the evidence review.

¹¹ Our external validity ratings range from 0 (not generalizable to U.S. contexts) to 5 (nationally representative). We label studies with external validity scores of 2 or higher as “more relevant” and studies with external validity scores of 0 or 1 as “less relevant” (most of these are from settings outside the U.S.).

¹² Our internal validity ratings range from 0 (no evidence) to 5 (most rigorous evidence). We rate studies with internal validity scores of 4 and 5 as having “high-quality evidence” and studies with internal validity scores of 1, 2, and 3 as “low-quality evidence.” We developed these standards to be compatible with review standards used in the What Works Clearinghouse Standards 4.0 ([IES 2017](#)) or the Department of Labor’s Clearinghouse for Labor Evaluation and Research ([CLEAR 2015](#)).

2. FINDINGS & IMPLICATIONS

rest of the social welfare system is expected to interact with early intervention effects, and foreign settings typically have very different institutional context, we judge the external validity of non-U.S. studies for inferences about effects in the U.S. to be low.

On balance, the available evidence on SAW/RTW programs tends to be low-quality, and findings from high-quality studies are difficult to interpret. Most studies do not include numerical information on outcomes, and others do not include sufficient information to estimate program impacts. Most studies that do provide estimates of program impact either have:

- high internal validity (credibility to identify causal impacts in a given setting) but are not generalizable to a broad U.S. context (many experiments are conducted outside the U.S.), or
- low internal validity (less credibility to identify causal impacts in a given setting) but are more relevant to a broad U.S. context.

A further consideration for assessing the quality of evidence is precision (statistical power). In general, studies with larger sample sizes have greater statistical power and can conclude with high confidence that estimated impacts fall within a narrow range. One way to assess adequate statistical power is to ascertain whether any impacts differ statistically from zero or equivalently whether confidence intervals are large relative to estimated effects, since we may expect that in a large group of interventions with generically nonzero effects, we should see effects that are statistically significant if the studies are adequately powered. Another way to judge power is to ascertain whether confidence intervals are large relative to effects that would be policy-relevant.

The evidence about SAW/RTW programs in existing reviews is often limited to one type of program or type of condition (for example musculoskeletal conditions, including low back pain, or mental illness). For those conditions, we identified numerous randomized control trials with high internal validity (credibility as to the identification of causal impacts). For low back pain, evidence reviews show statistically significant positive impacts, suggesting that there is sufficient power to detect impacts. Evidence reviews also point to statistically significant positive impacts for individuals with mental illness. Mental illness (40 percent) and musculoskeletal conditions (20 percent) were the most commonly targeted disability types in the articles reviewed, so the majority of the evidence is relevant for understanding impacts on SAW/RTW for those conditions.

In contrast to the findings from the literature reviews, the results of our meta-analytic review of all individual studies finds few stable patterns in how impacts vary with disability type or program model. That is, patterns of results differ across different classes of studies, and even when pooling all studies most characteristics of studies exhibit wide confidence intervals, meaning effects could be much larger or much smaller comparing across studies with or without that feature. We do not detect any statistically significant advantage in impacts by disability type. There is insufficient sample size to detect whether programs tend to produce larger or smaller impacts when targeting mental illness, musculoskeletal, or other conditions.

We also find that medical interventions, accommodation, informational, and financial incentive models have no average advantage or disadvantage relative to the other models, though the estimates are again imprecise. However, our review of individual studies indicates that SAW/RTW programs that include employment services and training tend to have larger impacts on average, with effects about 1.1 larger on the log odds scale, a finding driven by high-quality evidence using experimental designs.

This review also finds that measured impacts are systematically different depending on the type of evidence, when we examine interactions with program models, which is indicative of selection bias¹³ in quasi-experimental evidence, though there is no clear pattern as to the nature of that bias.

The review of evidence highlights open questions for further study and two key factors that influenced our choices of evaluation design options.

- There is surprisingly little evidence about program effects, given claims often made in guidelines or policy briefs (e.g., ACOEM, 2006) and the narrative reviews we discuss further in our review. There is also little evidence about what types of programs produce larger impacts, for whom. This means that evaluation design options that examine specific services or interventions could make an important contribution to policy.
- Most individual studies suffer from low statistical power, in the sense that the standard error is larger than policy-relevant effects. For example, of 82 studies in our meta-analysis with employment outcomes, the standard error on the effect size (log odds of employment) is larger than one tenth in 79 out of 82 (96 percent) of them. This is usually due to small sample sizes; 74 of the 82 studies referred to have sample sizes under a thousand, and all 74 have large standard errors. Future research designs that provide sufficiently large sample sizes to detect program effects are thus a high priority. The tradeoffs between a large sample size and a cost-effective evaluation design imply that one attractive design would randomize an intervention across a very large sample and use administrative or nationally representative survey data to measure outcomes.

2.3 Early Intervention Pathways and Target Population Profiles

To design and test SAW/RTW programs, it is important to be able to identify the target population of workers to serve. It is equally important for program operators to be able to reach that target population and provide SAW/RTW services and supports.

A major challenge for SAW/RTW programs is identifying workers at risk of exiting employment or the labor force because of an illness or injury, but *before* they have applied for SSDI. Another challenge is to determine when to intervene. To enable policymakers to target resources efficiently thus requires an understanding of which workers are most likely to leave the labor force after an illness or injury, what incentives they face, and which services or systems they engage with after becoming ill or injured but before they have applied for SSDI (Hollenbeck, 2015, Stapleton et al., 2015).

We analyzed publicly available data from three panels of the Survey of Income and Program Participation (SIPP), covering the period from 2001 to 2013 to identify SIPP respondents who separated from a job for health reasons or who report a disability, and who were not receiving SSDI at the time of the job separation. For this sample, we measured subsequent program participation, return to work, or receipt of federal disability benefits in the 16 months after the job separation. We compared some of these statistics

¹³ Selection bias refers to the fact that individuals who choose to participate in an intervention are different from those who choose not to. When an intervention is not randomly assigned, or characteristics are not otherwise adjusted for, we cannot infer that any difference in outcomes arises from the intervention rather than the intrinsic difference between individuals who choose to participate and those who choose not to. Selection bias can also arise from a selection mechanism that involves someone else making choices about who participates, if a similar difference in characteristics arises. This bias is the major challenge that studies with high internal validity overcome, i.e., a high internal validity study or strong evidence has low risk of selection bias.

2. FINDINGS & IMPLICATIONS

to SSDI and SSI caseload measures from SSA's Annual Statistical Supplement. The project report Nichols et al. 2020b provides details of the methods used and findings.

This analysis identifies interactions or touchpoints that workers have with various services or systems after a job separation and prior to application for federal disability benefits. These touchpoints constitute the pathways a worker may experience from illness/injury to SSDI/SSI. The touchpoints are occasions when workers might experience an early intervention that could keep them in the labor force, thus preventing their applying for federal disability benefits. The analysis that we conducted considered participation in six such touchpoints: (1) Unemployment Insurance,¹⁴ (2) workers' compensation, (3) public assistance programs, (4) private disability insurance, (5) job training or educational enrollment, and (6) health care utilization (i.e., visits to doctors or hospitals).¹⁵

The key findings showed that:

- Eight in 10 workers were in a pathway that involved health care utilization.¹⁶ Of workers in a pathway with health care, 54 percent interacted with another touchpoint.¹⁷
- Some 24 percent of workers received public assistance, including assistance from the Supplemental Nutrition Assistance Program and Temporary Assistance for Needy Families, during the 16 months following separation from work.
- A substantial fraction of workers interacted with none of these touchpoints (a “no-touchpoint” pathway). Some 11 percent of all workers reported no participation in Unemployment Insurance, workers' compensation, public assistance, private disability insurance, job training, or health care during the 16 months following separation from work.

The key findings also showed that with respect to likelihood of beginning to receive federal disability benefits:

- Compared to the no-touchpoint pathway, pathways with public assistance, private disability insurance, and health care touchpoints were all associated with higher rates of federal disability benefit receipt 17 to 20 months after separation from work, adjusting for other types of participation. Relative to the no-touchpoint pathway, pathways with health care and public assistance were associated with about 19 percent higher rates of disability receipt (10 percent

¹⁴ Unemployment insurance (UI) provides an alternate source of income support for individuals with a work disability that lose their job. It is only available to those who are not working or working at low levels, but are looking for work in some capacity. Although UI has traditionally focused on dislocated workers, UI modernization efforts in recent years have made the program more relevant to workers who experience work-limiting illnesses and injuries. Specifically, many states offer UI benefits to workers who leave their previous jobs for health reasons, and three states (Illinois, Massachusetts, and Montana) allow workers to claim UI while seeking part-time work if they have documented health reasons for doing so (Callan, Linder, & Nichols, 2015; Lindner & Nichols, 2014; McHugh et al., 2002).

¹⁵ Data limitations prevented us from examining four other touchpoints—(7) employee assistance programs, (8) case coordination, (9) workforce services, and (10) state vocational rehabilitation—that we identified in the evidence review (Nichols et al., 2020a). Public assistance programs include Supplemental Nutrition Assistance Program, Temporary Assistance for Needy Families, and other federal and state means-tested transfer programs.

¹⁶ Unfortunately, data limitations prevented the analysis from identifying the specific types of interactions with medical professionals and their precise timing.

¹⁷ Of all workers, 37 percent follow a path with health care only, and 43 percent follow a path with health care and another type of interaction, together accounting for 80 percent of all workers. Of those 80 percent, 43 percent with another type of interaction constitutes 54 percent.

2. FINDINGS & IMPLICATIONS

higher for public assistance and nine for health care, as shown in Exhibit B-1). Pathways with private disability insurance had federal disability benefit receipt rates 23 percent higher than the no-touchpoint pathway.

- Limiting our attention to three touchpoints of public assistance, private disability insurance, and health care, we defined unique pathways according to the receipt and timing of PA or PDI (or both at the same time). We analyzed outcomes relative to the no-touchpoint pathway and found that, compared to the no-touchpoint pathway: :
 - Five of the six pathways that involve health care are associated with substantially and statistically significantly higher rates of federal disability benefit receipt 17 to 20 months after separation from work. The pathway with only health care has seven percent higher rates of federal disability benefit; pathways with health care and other touchpoint have 20 to 50 percent higher rates.
 - Relative to the no-touchpoint pathway, the only pathway with lower rates of federal disability benefit receipt 17 to 20 months after separation from work was a pathway with no health care visit, and private disability receipt beginning after public assistance (with rates nearly 20 percent lower than the no-touchpoint pathway).
- Worker and job characteristics seem to matter little, except that older workers are more likely to be awarded federal disability benefits.

Together these analyses provide three insights that influenced our choice of evaluation design options.

- Any intervention that relies on recruiting workers at a single touchpoint (health care, public assistance, workers' compensation, unemployment insurance, or private disability insurance) is unlikely to be successful in reaching the entire desired target population. Workers who are likely to exit work because of illness or injury, who may go on to enter the federal disability system, and who would benefit from a SAW/RTW intervention, interact with several touchpoints. However, combining outreach or recruitment via health care and public assistance programs might be promising.
- By definition, all of the workers in our analysis sample leave employment. This suggests that all these workers could be recruited to an intervention that targets the employer. Employer-focused interventions seem a potentially promising avenue for further study. An option of incentivizing employers to report when workers leave employment, especially when due to a non-work-related injury or illness, might offer a mechanism to identify workers at risk of labor force separation and long-term disability.
- Because this analysis relies on publicly available data, it faces some limitations for identifying potential target populations for SAW/RTW programs. Additional analysis with survey data matched to individual-level administrative data on receipt of disability benefits is needed to more fully understand potential target populations. Potential target populations might include workers with a specific impairment, or workers who interact with specific services or supports. Evaluation design Option A (described in Chapter 3) would expand on the previous analysis to provide more detailed information about potential target populations and SAW/RTW pathways.

2.4 The RETAIN Demonstration

Lessons from the program review and synthesis, evidence review, and early intervention pathways analysis point to unanswered questions and suggest important considerations for developing evaluation

2. FINDINGS & IMPLICATIONS

design options. In addition, any evaluation design options that DOL considers should complement a major new early-intervention demonstration that DOL is undertaking with the Social Security Administration: *Retaining Employment and Talent after Injury/Illness Network (RETAIN)*.¹⁸ DOL is funding the RETAIN grants to state entities and a technical assistance contractor and SSA is funding the evaluation contract.

RETAIN is modeled after promising practices operating in Washington State's Workers' Compensation program, particularly its Centers of Occupational Health and Education (COHE) model (Wickizer et al., 2001). However, unlike the Washington State programs, RETAIN will not be limited to workers with work-related injuries and illnesses. RETAIN is intended to provide evidence about interventions that address the broader target population of workers who experience medical conditions that inhibit their ability to work, whether work-related or not. Each state that operates RETAIN will develop its own demonstration design, and while DOL expects that the interventions will incorporate key elements from the COHE model (e.g., focus on early coordination of health care services and physician education), each will be adapted to the state's labor market conditions and other unique factors.¹⁹ The states will also develop partnerships and data systems needed to carry out RETAIN. RETAIN will test a package of interventions in specific state systems that combine early provision of health care and employment-related supports and services, coordinated by a return-to-work coordinator.

In September 2018, DOL awarded 18-month planning grants to state departments of labor, workforce agencies or other state agencies in eight states (California, Connecticut, Kansas, Kentucky, Minnesota, Ohio, Vermont, and Washington). During the planning grants, the states will develop RETAIN demonstration projects and conduct a pilot. At the end of the planning phase, DOL and SSA will select up to five states to expand their RETAIN interventions to serve a larger population of workers.

SSA's request for proposals for the RETAIN evaluation defined two primary research questions to be addressed:

- What is the impact of RETAIN on labor force participation and workforce retention among the target population?
- What is the impact of RETAIN on applications and receipt of SSDI and SSI?

The evaluation may also address other research questions, such as these:

- What interest in program participation is there among providers, workers, and employers?
- What are the differential effects of RETAIN based on participant demographics (subgroups)?

¹⁸ The RETAIN Demonstration Projects are a collaborative effort led by ODEP in partnership with DOL's Employment and Training Administration (ETA) and SSA. RETAIN projects will test the impact of early intervention strategies that improve stay-at-work/return-to-work (SAW/RTW) outcomes of individuals who experience work disability while employed. <https://www.dol.gov/odep/topics/SAW-RTW/how-to-apply.htm>

¹⁹ The Funding Opportunity Announcement specified that the states' interventions should include seven components: 1) Return to work (RTW) coordinators to coordinate health and employment service delivery; 2) Training for participating health providers in occupational health best practices and alternatives to opioids for pain management; 3) Incentives for participating health care providers to utilize the best practices; 4) Early communication to all stakeholders to return the worker to the workplace as soon as possible; 5) Workplace-based interventions (including accommodations such as lighter and/or modified duties, and adjustments to work schedules, tasks, and the physical worksite, if necessary); 6) training/rehabilitation for workers who can no longer perform their prior job or other available suitable alternate work; and 7) Tracking and monitoring the medical and employment progress of participating workers.

- What are the costs and benefits of RETAIN?
- What are other effects of RETAIN?

In developing evaluation design options for this project, we have considered the RETAIN demonstration design and the type of evidence the evaluation will produce. We believe that the design options we have developed complement what DOL will learn from RETAIN. We also believe that our options would address questions that RETAIN will leave unanswered. Examples include:

- **Who is the target population for SAW/RTW?** The states that implement RETAIN will identify workers to target for the demonstration, within the grant participant eligibility criteria, along with mechanisms to recruit the workers. Some states may focus on large employers; other states may target claimants for private or state disability insurance. We anticipate variation across the RETAIN states in the target populations and recruitment methods.

For that reason, we believe that a SAW/RTW evaluation design that uses national data to characterize the size of potential target groups would be useful. To develop a national SAW/RTW policy response, DOL and other agencies would benefit from broad, representative information on which workers are most likely to exit the workforce because of medical conditions, and of those, which factors predict who is likely to enter the SSDI program and who is likely to benefit from SAW/RTW supports. An empirical analysis of the national population will yield valuable insight on those questions.

- **What are the effects of specific interventions that influence the behavior of employers, health care providers, and workers?** SAW/RTW evaluation design options could focus on one particular aspect of the RETAIN intervention, such as information provided to workers, and test the aspect on a large scale. Such options can offer complementary evidence about how best to deliver such assistance. Particularly for informational models, these more targeted tests would likely identify which mediums and messages are more effective at changing behavior (Contreary and Perez-Johnson, 2016; Juras et al., 2018). This insight, as well as findings that extend beyond one state, might not be available from the RETAIN evaluation. Finally, evaluation options that test one specific approach can help identify which elements of RETAIN are most cost-effective. For example, if a low-cost informational and an intensive intervention combined with more costly new incentives has similar impacts, the first option would be more efficient. RETAIN might only test the second option.
- **What is the effect of temporary partial disability benefits?** RETAIN will examine a package of services and supports centered around a return-to-work coordinator to facilitate the early provision of health care and employment-related supports and services, physician education on occupational health best practices, and incentives for adopting those best practices. RETAIN also will examine workplace interventions that employers can adopt. RETAIN will not test the impacts of temporary or partial disability benefits. We believe that tests of this type of intervention, that could be offered by many states to reach a broad population of workers with non-work illnesses and injuries, would augment lessons from RETAIN.

In Chapters 3 through 7 that follow, we describe each evaluation design option in detail, including how it addresses the two core research questions (effects on employment and disability benefit receipt) and any additional research questions each addresses. The descriptions also include how the option would produce new knowledge relative to the existing literature and ongoing demonstrations including those anticipated as part of RETAIN.

3. Option A: Longitudinal Study of Individual Workers

This option uses a nonexperimental design to address the core research question of how program participation is associated with employment and receipt of disability benefits. This option would not test any specific intervention. Instead, this option builds on the data analysis conducted previously for the project to analyze the steps workers take after illness or injury. It would use a nationally representative dataset, matched to administrative records from the SSA, or an alternative data source of comparable quality, to follow workers after injury or illness. It would identify potential target populations for SAW/RTW interventions and examine the types of experiences that are helpful in maintaining workers' attachment to the labor force. The goal of this option would be to increase understanding of potential target populations for SAW/RTW efforts, and to chart patterns of program participation.

The findings from this analysis could be useful in planning other future research evaluations, such as Options B through E in this report, and could provide useful context to guide ongoing research projects, including RETAIN. This analysis might also reveal differences between subgroups or relationships between variables that would suggest new research questions about why groups differ, whether relationships are causal, or how to change existing patterns. Major challenges to executing this research include obtaining access to high-quality data, with sufficient sample size to detect policy-relevant differences. We discuss potential challenges to the study and limitations in the data in the sections that follow.

This chapter discusses the research questions that this option would address, the conceptual model for the study and participants. The chapter includes details about the data sources and analysis plan. Appendix A provides additional details about the data sources and constructing the analysis data files. Because this option does not test a new intervention, it would not require intervention design or a plan for recruiting study participants.

3.1 Research Questions

More information is needed about who might benefit from SAW/RTW interventions—both the size of the populations and how they can be reached. This study would address the following questions:

- **Which workers will leave the labor force after an illness or injury?**
- **Which of them will eventually apply for SSDI?**

Answers to these questions will help policymakers develop better approaches for targeting interventions and developing research designs. The study would examine employment and disability application outcomes.

As summarized in Section 2.3 of this report, Abt conducted preliminary analysis using publicly available data to address these questions. Exhibit 3-1 shows the proportion of workers in the sample who interacted with each of six touchpoints, and the proportion who returned to work or received SSDI or SSI. However, using only public data prevented us from observing outcomes for a large fraction of the sample, primarily because a large fraction does not reach either a return-to-work or a permanent labor force exit in the 16-month observation window we chose. Option A's longitudinal study of individual workers would improve on that previous analysis. The study would use individual-level survey data matched to SSA administrative data to identify which individuals eventually apply for and receive SSDI or SSI, yielding larger sample sizes, and much longer follow-up periods. With this superior sample and data, researchers could, for example, determine which workers returned to work and which workers applied for federal disability across many years. This design examines nationally representative samples of workers, and will

provide broader information about potential target populations than the state demonstrations tested in RETAIN are anticipated to provide.

Exhibit 3-1. Proportion of Workers Separated from Work Who Participate in Six Touchpoints in the 16 Months Following Separation from work, and Rates of Return to Work, and Receipt of SSDI/SSI 17-20 Months after Earnings Loss

Touchpoint	Participation Rate	Return to Work	DI/SSI Receipt
Health care	80%	48%	21%
Job training	8%	71%	8%
Private disability insurance	9%	31%	37%
Public assistance	24%	42%	25%
Unemployment Insurance	16%	61%	16%
Workers' compensation	10%	38%	19%
None	11%	57%	5%

Source: Authors' analysis of publicly available SIPP data, 2001-2013.

3.2 Conceptual Model

The conceptual model for this research design is that individuals vary in which services, systems, and programs they interact with after an injury or illness (Contreary & Perez-Johnson, 2016). We refer to these as interactions with a “touchpoint” in Nichols et al. (2020b). Examining this variation in interactions may inform us as to whether interactions with specific touchpoints are associated with higher rates of return to work.

A number of early interventions could affect the employment trajectory of someone who experiences an injury or illness that could develop into a work disability. Early interventions to encourage labor force retention could be implemented through a variety of programs such as (1) Unemployment Insurance, (2) workers' compensation, (3) public assistance programs, (4) private disability insurance benefits or paid leave, (5) job training or educational enrollment, (6) health care system, (7) employee assistance programs, (8) workforce or employment services such as those offered by American Job Centers, or (9) state vocational rehabilitation.

The existing evidence does not tell us if any of those touchpoints, or any combination of them, is associated with substantially higher rates of return to work or lower rates of application for disability benefits.

3.3 Intervention(s)

This design would not test any specific intervention. Rather, the research project is an investigation of patterns of participation in a wide variety of potential programs where early interventions might be implemented.

3.4 Participants

Target population. The target population is individual workers who have experienced an injury or illness that threatens their ability to work. In the data, this could be identified as a new report of such a disability or health condition, among those in the labor force.

Setting/venue. This design would investigate program use across many different settings, using a nationally representative sample.

Recruitment plan (if applicable). No recruitment would be needed, as this design would use existing data.

3.5 Data

Data needed/data access—permissions, matching as applicable. Data for this evaluation design option would come from matched survey and administrative data sources. Survey data allow for the identification of the target group of workers and also provide detailed descriptions of demographic factors (often administrative data are missing education, family structure, and other factors that may be important determinants). The administrative data make it possible to track workers' application for federal disability benefits and their use of a variety of supports and services with accuracy for a long period of time (from the date of the survey to the most recent year covered in administrative data), depending on the source of the administrative data. Combining the two forms of data means that the study would take advantage of the best features of each source.

There are a number of publicly available **survey data** sources that can measure different patterns of participation, including the U.S. Census Bureau's Survey of Income and Program Participation, Current Population Survey, and American Community Survey; the National Health Interview Survey; the Medical Expenditure Panel Survey; and the National Survey of Drug Use and Health. Privately operated surveys include the Panel Study of Income Dynamics, the RAND American Life Panel, and the Health and Retirement Study. Many of these sources can be linked to **administrative data** from the SSA, Treasury's Internal Revenue Service (IRS), or the Centers for Medicare and Medicaid Services, using various data use agreements and mechanisms for linking, including via the Census or the National Directory of New Hires (we discuss options in Appendix A).

Alternative data sources. An alternative to using existing matched survey and administrative data is to field a new longitudinal survey that is representative of workers who experience an injury or illness. Following these individuals over time, and collecting information on their workplaces and use of services could provide a fuller picture than any existing data source, because the survey could be designed to answer the specific questions of interest. Sampling and recruiting for such a survey is very expensive, but it is possible in theory to combine the two approaches, and recruit for a new panel survey from an existing survey. For example, a topical module in the final wave of the SIPP might measure whether individuals had experienced an injury or illness. Researchers could then recruit these individuals for subsequent follow-up. Likewise, a sample could be built from a private survey such as the RAND American Life Panel or Ipsos Online Access panels.

In principle, data from the Centers for Medicare and Medicaid Services could also be merged to these files, to identify specific medical encounters, procedures, and diagnoses. Access to the universe of Medicare and Medicaid beneficiaries' data is licensed to contractors for a variety of research projects, under very strict data use agreements by the Centers for Medicare and Medicaid Services, and often requires a substantial payment to its contractor ResDAC to acquire an extract of the data. Merging to medical records, either from the Centers for Medicare and Medicaid Services, private health insurers, or both, would likely be difficult. The advantage of merging encounter data from the Centers for Medicare and Medicaid Services and/or private health insurers is that use of medical services would be subject to much lower rates of measurement error (Gaskell et al., 2000; Wolinsky et al., 2007; Stull et al., 2009, Bhandari & Wagner, 2006; Kjellson et al., 2014; Brusco & Watts, 2015; Dalziel et al., 2018). Classification errors in who uses specific services tend to introduce bias in assessments of whether the use of services predict higher rates of return to work and lower rates of application for disability. Measurement error can also reduce statistical power, which we discuss below in the context of Exhibit 3.1.

IRB, security, logistical concerns. A full Institutional Review Board (IRB) review would be required. Because the data would be used at a secure Census or SSA facility, data security would not be a concern. However, logistical concerns would include obtaining and maintaining access to the data and travel required to the secure facility for analysts.

3.6 Analysis Plan

Outcomes and hypotheses. Option A would investigate the association between the programs individuals participate in, or touchpoints interacted with, on the one hand, versus employment and disability application outcomes, on the other. This study would use **hazard models** that predict the rates of return to work, application for federal disability benefits, award of benefits, or labor force exit. The use of hazard models allows one to predict times to eventual changes in state, or the chances of winding up in a state, even when the final outcome is not observed. Since these models allow both the estimation of probabilities of reaching different final states, and the time needed to reach these states, the models are often referred to as “duration,” “time-to-event,” or “survival” models as well. We anticipate that models that predict the probability of return to work would also be used to estimate the time out of work until that return occurs, since little additional effort is required to obtain this richer estimate.

A pervasive problem when estimating durations of time out of work or final outcomes is that we do not observe the final outcome for large fractions of the individuals in the sample. Instead, when we observe someone out of work for 13 months, we know only that the time to eventual re-employment is at least 13 months (i.e., the duration is censored at 13). This is known as “right censoring.” Such right censoring not only introduces bias in estimates of duration of non-employment spells, but also lowers the statistical power of comparisons (Clark et al., 2003). Power is lower because we compare outcomes such as rates of return to employment or receipt of disability benefits by participation pattern, but many people may not have either outcome recorded at a single point in time in follow-up surveys. Hazard models can generate estimates of rates, and determinants of those rates, that are not biased by right censoring and allow for the use of all available information to maximize power. They can also be made robust to unobserved heterogeneity, using appropriate distributional assumptions.

Model(s) or analysis methods. The basic hazard model using discrete time, used to avoid bias in estimating duration or effects on likelihoods of eventual outcomes, can be expressed in a simple regression format (Jenkins, 1995) using the equation:

$$Y_{it} = f(X_{it}b + W_{it}c + T_{it}d + e_{it})$$

where Y is an outcome measure that is zero in any month where an event (a transition) does not take place, one in any month where the event does take place, and is missing in every month after an event takes place. We anticipate the outcomes used in Y would include employment and application for or receipt of federal disability benefits. X represents any time-varying covariates of interest, such as interaction with various touchpoints, and W represents characteristics that do not vary over time, such as gender or educational attainment at the time of injury or illness onset (such variables W would be dropped in a model that is robust to unobserved heterogeneity). T is a vector of time indicators, which capture the “baseline hazard” of transitions; that is, the hazard when $X=0$ and $W=0$. The error term e is assumed to be uncorrelated with X and W . The function $f()$ is the link function, such as the complementary log-log; that is, $f(z) = 1 - \exp[-\exp(z)]$.

By including any prior exposure variables among the variables in W , we can also measure whether past interactions have effects at later dates. By multiplying these past exposure variables by contemporaneous touchpoints, we can measure the association of a particular sequence of touchpoint interactions with higher hazard of return to work or labor force exit. For example, we could measure whether use of public assistance prior to exiting employment was associated with higher or lower relative hazards of return to

work associated with subsequent use of public assistance by interacting an indicator of past public assistance in W , call it W_I , with interaction with a public assistance touchpoint in X , call it X_I . The number of possible interactions here is combinatorically large, so a feature selection method might need to be employed (Zare et al., 2013). A variety of related machine learning methods could be used to investigate the likely functional form of these relationships (Mullanaithan & Spiess, 2017)

Expected methodological issues and responses. Suppose we wish to determine whether workers' compensation or public assistance beneficiaries are more likely than non-beneficiaries to eventually exit from the labor market and apply for federal disability benefits. This implies using the model described in the prior section to estimate the impact of participation in workers' compensation on application hazards for federal disability benefits. There are three primary issues: modeling functional forms correctly (described in the last section), power and sample size, financial and time costs of data acquisition, and modeling selection into participation. The last problem is the most difficult, and we discuss that last, but note here that we use "difference" in outcomes by participation pattern and "impact" interchangeably below, though that will typically not be justified when participation is not controlled by the researcher.

Power and sample size. A pervasive problem is the small samples in surveys who may leave the labor force and apply for federal disability benefits. With a rare event, samples must be much larger to detect impact reliably. We illustrate the kinds of calculations to be undertaken with a single example. Assume the study aims to estimate a hazard model where the underlying risk of an outcome event (such as application for disability benefits) across each four-month period is 2 percent, has at least nine waves of follow-up data, and 10 percent of the sample participates in the particular program of interest.

There are more than 50,000 adults in each Survey of Income and Program Participation panel, but we know from the report Nichols et al. 2020b (described in Section 1.2) that only 1 to 2 percent of survey participants will report a new work-limiting condition from wave to wave. Combining the five most recent panels (which began in 1996, 2001, 2004, 2008, and 2014) could yield at least 2,500 adults with an onset of a work-limiting condition, in one of the early waves. In Exhibit 3-1, we illustrate required sample sizes to detect a difference at various hazard rates among participants. We compute the sample size required to achieve a power of 80 percent to detect a difference between hazard rates, using iterative methods and formulas described in Stata (2019, volume PSS, page 575), Lachin, J. M., (1981), Lachin, J. M., & Foulkes, M. A., (1986), and Lakatos, E., & Lan, K. G. (1992).

This shows that if we assume the hazard to be half as large in the participating group as in the comparison group, even a sample of 1,500 will suffice to detect a difference. However, when the hazard rate in the participating group is three-quarters the comparison group's hazard (1.5 percent compared to 2 percent), more than five times as large a sample is required to detect a difference.²⁰ The anticipated sample of more than 2,500 adults with an onset of a work-limiting condition in the Survey of Income and Program Participation data alone would allow researchers to detect a difference between a hazard of 2 percent and 1.2 percent, roughly. This is a large minimum detectable impact, indicative of low power. However, it might be the case that such an impact is the policy target needed to be achieved to justify the cost of an intervention; if so, such a large effect size could still be relevant. In particular, we can never detect an arbitrarily small difference in outcomes; we must decide what the smallest detectable difference of interest is, and then design a study with a good chance of detecting that impact.

By matching to administrative data on outcomes, researchers could extend to onset in any wave, and about a 10th of the sample will report onset during the life of each panel. With 5,000 adults across five

²⁰ This analysis assumes a comparison of hazards—that is, sample size is for hazard analysis comparing two exponential survivor functions by using parametric tests for the difference between hazards, with unequal allocation between the two groups. Losses to follow-up or censoring are assumed to be exponentially distributed. Effects are not measured as differences in percentage points, but on a hazard ratio scale.

panels reporting the onset of a work-limiting condition, the study could detect a difference between a hazard of 2 percent and 1.4 percent, a much smaller minimum detectable impact (see Exhibit 3-2). But with administrative data, researchers can also extend beyond nine follow-up time periods; for example, 120 months of follow-up (moving to monthly data implies we should assume a lower baseline hazard of half a percent per month). As illustrated in Exhibit 3-3, with 120 follow-up time periods, a hazard rate half or three-quarters the assumed baseline rate needs about one-third the sample size. The anticipated sample of more than 5,000 adults with an onset of a work-limiting condition would be more than enough to detect a difference between hazards of 0.5 percent and 0.4 percent, which correspond to annual hazards of 1.6 percent and 2.0 percent, a substantially smaller minimum detectable impact—implying substantially higher power.

Depending on whether such an impact is smaller than the policy target needed to be achieved to justify the cost of an intervention, this reduction in detectable effect size could be worth substantial increases in the cost of research. Furthermore, if survey responses have 90 percent the statistical reliability as the administrative records, the sample sizes would need to be increased by about 11 percent, and if survey responses have 70 percent the statistical reliability as the administrative records, the sample sizes would need to be increased by about 44 percent (Kanyongo et al., 2007).

Exhibit 3-2. Sample Sizes to Detect a Given Hazard Rate Compared to 2 Percent Baseline Rate with Nine Follow-Up Time Periods

Participant Hazard Rate	Total Sample Size Required	Participant Sample Size	Nonparticipant Sample Size
1.0%	1,476	135	1,341
1.1%	1,925	175	1,750
1.2%	2,567	234	2,333
1.3%	3,523	321	3,202
1.4%	5,024	457	4,567
1.5%	7,566	688	6,878
1.6%	12,340	1,122	11,218
1.7%	22,858	2,078	20,780
1.8%	53,502	4,864	48,638
1.9%	222,304	20,210	202,094

Exhibit 3-3. Sample Sizes to Detect a Given Hazard Rate Compared to 0.5 Percent Baseline Rate with 120 Follow-Up Time Periods

Participant Hazard Rate	Total Sample Size Required	Participant Sample Size	Nonparticipant Sample Size
0.250%	511	47	464
0.275%	669	61	608
0.300%	896	82	814
0.325%	1,236	113	1,123
0.350%	1,773	162	1,611
0.375%	2,683	244	2,439
0.400%	4,400	400	4,000
0.425%	8,199	746	7,453
0.450%	19,301	1,755	17,546
0.475%	80,676	7,335	73,341

Looking across multiple types of participation patterns to identify significant effects, researchers would have to make adjustments for multiple comparisons (adjusting significance tests to account for the number of tests conducted to achieve the correct error rates). To fix ideas, we illustrate sample size needs for a single comparison only. We also have not accounted for clustering, though all of the potential data sources are cluster samples. However, preliminary calculations indicate the design effect is approximately 1.15, meaning standard errors adjusted for clustering are about 7 percent larger and effective sample sizes are 87 percent as large as the count would suggest. A design effect twice as large, at 2.3, would make standard errors adjusted for clustering about 52 percent larger, and effective sample sizes would be 43 percent as large as the count would suggest. Conservatively, we might assume an effective sample size closer to 2,000 than 5,000, and a minimum detectable difference of 0.15 percentage point in the hazard rate (0.5 compared to 0.35 percent). Whether smaller effects would be important policy-relevant differences is an open question, but substantially smaller differences would be difficult to detect without dramatically larger sample sizes.

Researchers could also add additional panels to the dataset; for example, nearly doubling the number of panels by including 1990, 1991, 1992, and 1993 panels, at the cost of including older data that reflect a labor market even further removed from the present.

Financial and time costs of data acquisition. We have described sources of matched data that we know exist already in Census Research Data Centers, and that can be accessed via an application process that is low cost but we estimate would take approximately one year to finalize. Even these matches depend on receiving approval from Census and Treasury, and demonstrating a benefit to those agencies of the proposed research. Researchers using the data or examining results prior to the Disclosure Review process also need Special Sworn Status, which can take a year to obtain (presumably, DOL could require some research staff to have this status prior to contract award). DOL might also consider matching other sources of administrative data to the population earnings and disability application histories from administrative data. These sources might be workers' compensation administrative records, medical records, or records from several large employers. Doing so might hugely expand the sample size and lower the minimum detectable differences, but such novel data matching comes with substantial uncertainty in time required (or even likelihood of eventual success).

For this reason, we suggest that such speculative data matches should be combined only with a research project that uses known data sources, such as the matched Survey of Income and Program Participation data described here. This research project could certainly access matched Survey of Income and Program Participation data within the typical timeline of a multi-year project, whereas access to other prospective matched data cannot be assured.

Selection into patterns of participation. The estimation of differences in outcomes associated with different patterns of participation does not allow us to conclude that these differences are due solely to the patterns of participation, since the researcher does not control patterns of participation and cannot randomly assign them. It is possible that in the course of this study, a plausible regression discontinuity or instrumental variables approach could be developed to allow causal inference using this observational data (Nichols, 2007), but that is far from guaranteed and would rely on finding an appropriate "natural experiment" that generates naturally occurring exogenous variation in participation. As described below in Section 3.8, this is a limitation of the study, but the study would still make important contributions to our knowledge about potential target populations for SAW/RTW interventions.

3.7 Practical Considerations

Implementation timeline. Because this research design uses existing data, the time required to generate findings is relatively short, once data use agreements and access permissions are in place. In order to construct the data set needed for this option, researchers would require access to the SSA administrative

data files described in Section 3.5 above. SSA typically makes these data extracts available but requires several steps to grant access. DOL would need to factor in time for data sharing agreements, and DOL might be able to help researchers facilitate such agreements.

We estimate that obtaining data access might take 12 months or longer. This implies a three- to four-year project, assuming that finalizing the research design and some preparatory research takes place while data access is being negotiated. To make efficient use of time waiting for data access, DOL would benefit if its researchers were to be already familiar with the matched data. Research teams familiar with the matched data could prepare analytic programs and table shells in advance of being granted access to data. Based on previous guidance from SSA, we estimate that it would take at least one to two years to obtain access to data for the study. We estimate that it would take approximately one year to analyze the data and prepare reports.

Cost. Because the longitudinal study design in Option A uses existing data to examine the association between program participation and employment outcomes after an illness or injury, it does not require a research team to design and operate a new intervention. This option also does not require sample recruitment. The primary cost implications for Option A are related to establishing an analysis plan; securing data use agreements; and collecting, managing, and analyzing the existing data. As described in Section 3.5, the Census Bureau has linked all past years of the Survey of Income and Program Participation to several other datasets.

One important cost element would be acquiring permission to use these linked data from the Census Bureau, SSA, and Treasury. Our understanding is that these permissions often are negotiated in one step via SSA, which would simplify negotiations and limit associated costs. Adding to the Census’s linked dataset—for example, including data from the Centers for Medicare and Medicaid Services to identify specific medical encounters—would require additional negotiation and data matching, adding cost.

The next largest cost component—data management and analysis—involves reweighting the matched data to make the dataset nationally representative and analyzing the data using a hazard model to predict rates of return to work.

Geographic scope. This study would use nationally representative data, but sample sizes would likely preclude producing estimates for small areas, including state-level estimates for lower-population states.

Threats to implementation as designed. The primary threat to implementation is timely permissions to access data. Partnerships with other agencies would not be a prerequisite for DOL to carry out such a study, but permission to merge data or to use existing matched data sources would be required. Contractors who used such matched data would require security clearance, the exact nature of which would depend on the files used and the facility in which data were accessed. We anticipate researchers would need suitability clearance from SSA, Special Sworn Status, and to provide Title 13 and perhaps Title 26 justifications²¹ for the data use. Researchers would need to work either on site at SSA or in a

²¹ Title 13 USC [<https://www.govinfo.gov/content/pkg/USCODE-2007-title13/pdf/USCODE-2007-title13.pdf>] authorizes the Census to share information for research purposes, and Title 26 U.S.C.A. § 1 et seq. [1986] authorizes IRS to do so; Title 26 is also known as the Internal Revenue Code of 1986 (Pub. L. No. 99-514, § 2, 100 Stat. 2095 [Oct. 22, 1986]). In both cases there must be some demonstrable benefit to the agencies. See e.g. sample research proposal information from the Census [https://www.census.gov/content/dam/Census/library/publications/2017/adrm/Research_Proposal_Guidelines.pdf attachment 3] for further information; note that only benefit numbers 5 to 12 are valid for projects requesting economic (Title 26) data. Example benefit relevant for both purposes include “Identifying shortcomings of current data, collection programs and/or documenting new data collection needs,” and “Understanding and/or improving the quality of data produced through a Title 13, Chapter 5 survey, census, or estimate.”

secure data center such as a Census Research Data Center. Working in a Census Research Data Center is often a low-cost solution, both in terms of time and resources, but would not afford access to all administrative data. In particular, commingling survey data with income and Social Security data is frequently done in the Census Research Data Centers, but Medicare/Medicaid data typically cannot be matched in the same environment. Adding Social Security data to a survey is typically easiest, followed by income or employment data, with health data being the hardest to merge on. Experience with permissions may vary, however, and there may be subtle complementarities, as owners of data may be willing to authorize a single merge but not multiple merges, depending on other data to be merged.

3.8 Contributions and Limitations

Policy relevance. This study would provide valuable information to policymakers to help them target SAW/RTW services to those workers who may be most likely to benefit. For example, if the risks of exiting the labor market were found to be very small among certain groups in the absence of any novel SAW/RTW interventions, policy designed to reach those groups would become less appealing. Alternatively, if a large fraction of those exiting the labor market were found to be in a particular group of workers, that finding would justify designing novel SAW/RTW interventions for that group.

Contribution. This study would fill gaps in current knowledge regarding the experiences individual workers consider helpful in returning to work or the barriers that they identify as problematic. The use of matched data would improve on the analysis Abt conducted previously for this project by using a longer follow-up period and richer and more accurate information on the touchpoints with which individuals interact. This new knowledge could help employers or public agencies such as vocational rehabilitation or American Job Centers to design or redesign programs, which might represent improvements that would lead to better outcomes for workers. These new programs could then be tested against the status quo in a later study, since the exploratory work in this option does not support strong causal inference.

Internal and external validity. This study would have low internal validity, in that it would not provide rigorous evidence of effectiveness of specific strategies for the targeted populations. However, as described, it would have high external validity, if it is representative of workers who experience an injury or illness and it measures their subsequent experiences well.

4. Option B: Test an Informational Intervention for Workers: Offer Workers Advice on Work and Disability

This option would investigate whether providing targeted information about SAW/RTW to workers who have recently experienced an injury or the onset or worsening of an illness improves their employment outcomes. The option would address the core research question of how to increase employment.

This design would provide a random sample of workers with information and advice on a range of topics, such as advice about work limitations for those with particular conditions, information about employer policies, and local or national resources. Information might be delivered in a low-intensity format, such as a letter, text message, or through a meeting with a counselor who provides personalized information.

Outcomes could be tracked over time through follow-up data collection or with administrative data on employment and wages. In addition to the main intervention, a brief survey to gauge workers' beliefs and knowledge about work, disability, and resources, could be administered.

To our knowledge, no existing research directly addresses the value of such an informational “nudge” for workers (Contreary & Perez-Johnson, 2016). We anticipate RETAIN and other research will provide findings about the effects of multi-component interventions. However, it does not seem likely that the RETAIN evaluation will be able to separately identify the impact of providing newly injured or ill workers with targeted information. Thus, this design option fills an important gap in knowledge.

This chapter discusses the research questions and conceptual model for this study. It then addresses intervention design and the key questions DOL would need to consider when designing the information intervention. The chapter discusses considerations for recruiting a study sample and for collecting data to measure outcomes. A major challenge is recruiting a sufficient number of individuals to detect small but policy-relevant impacts. The mechanisms to recruit study participants discussed in this option also imply a sample that does not represent all workers; this lack of generalizability would affect the policy relevance of the findings.

4.1 Research Question(s)

This evaluation design option would answer one main research question:

- **How does advice about working, and information about employer policies and resources affect employment outcomes for workers who have recently experienced an injury or the onset of an illness?**

The evaluation would examine outcomes such as:

- labor force attachment and employment;
- rate of return to work and duration out of work;
- earnings and/or work hours; and
- workers' perceptions of their likelihood of returning to work and their recovery process.

We treat *employment* as the primary outcome hereafter.

4.2 Conceptual Model

This evaluation design option aims to generate evidence about the effects of providing information to workers on their employment outcomes. As shown in the conceptual map in Exhibit 1-1, workers must often navigate policies and practices across various sectors in order to make a successful return to work. These sectors include their current employer, health care providers, insurance providers, workforce assistance programs, and other service providers. Workers may not know about the choices they have. This lack of knowledge may inhibit their ability to take advantage of benefits that are available to them, which may affect their ability to stay employed (Contreary and Perez-Johnson, 2016). Workers may also not know about how long they might expect to be away from work due to their condition, or the benefits of returning to work. This study would test whether providing individuals with information can improve employment outcomes.

4.3 Intervention

This design option would test an intervention that provides information to workers who have recently experienced an injury or are at the onset of an illness to determine whether information affects employment outcomes.

4.3.1 Intervention Details/Approaches

The specific details of how the intervention is administered may have important consequences for the outcomes of the participants. Those details are information content, information specificity, information delivery medium, information source, and extent of contact. Any intervention would need a specific approach to each of them. This section discusses alternatives for each of these intervention details.

Information content. Potentially useful information for study participants falls into four broad categories.

1. *Information about how their particular injury or illness may affect their ability to work, either currently or in the future.* Workers with chronic back pain, for example, will be limited in certain types of physical activities, which may not be relevant information for workers with mental health conditions.
2. *Information about internal resources to help them return to work.* This could include details about any benefit programs that the employer might provide, potential workplace accommodations or job modifications programs, or the contact information for program representatives associated with the particular employer who could provide further resources.
3. *Information about external resources.* These materials could inform workers about local organizations and programs that might help them to maintain employment, workers' rights under the Americans with Disabilities Act, resources to resolve disputes, or national advocacy organizations.
4. *Information about SSDI and SSI benefits.* These materials could inform workers about their potential benefit amounts, about the likely waiting time between application and the start of benefits, the likelihood of being awarded benefits at each stage of the application process, and the limitations that the SSDI and SSI programs place on earnings and assets.

Information specificity: If researchers decide to include information about how particular injuries and illnesses affect work abilities (the first type of content in the list above), they would need to determine whether to provide specific or general information. The information could be tailored to the worker's specific illness or medical condition. Tailored information could be based on individual work history and

the results of the Work Disability Functional Assessment Battery (WD-FAB).²² Alternatively, researchers might simply send information applicable to a wide range of health conditions, regardless of the one the worker currently faces. While highly specific information may be more useful to the worker, it may also be more difficult and costly to administer through the intervention. The other types of information content described above (items 2 through 4) are general information that would be applicable to all workers.

Information delivery medium. Information content could be delivered through a variety of mediums. Options include a letter, text message, email, phone call, or in-person conversation. The most effective medium depends on the content and specificity of the information itself. For example, general information about a workers' rights under the Americans with Disabilities Act may be better delivered as a letter than as a text message. Personal information on the medical and employment implications of a worker's illness may be more effective through a more personal medium such as a text message or phone call. Different delivery media may be interpreted as more or less personal, and this in and of itself could be an aspect of the intervention that researchers might measure.

The choice of medium will also have implications for the cost and reach of the intervention. For example, depending on the desired scale of the intervention, it might not be affordable to provide information through phone calls, though these might be a quite effective medium. On the other hand, mailing printed materials to a worker's home at a lower cost may not be cost-effective, if workers are likely to ignore them or discard them as junk mail (Sinclair et al., 2012).

Information source. Researchers will need to decide who will be identified as sending the information to workers. Options include the worker's employer, a specific employee within the employer's firm (e.g., human resources officer or a supervisor), a medical professional, an insurance agency, a union or other trade organization, a state agency, DOL, or the research group itself. Though the particular choice of the information's source might not be an aspect of the intervention that researchers wish to explicitly assess, it is nevertheless an important part of designing the intervention. The trust that workers place in the source of the information may influence their likelihood of acting on the information provided.

Extent of contact. The extent of contact may include combinations of the number, frequency, and duration of interactions. One extreme option would be a purely informational intervention consisting of a single text message or mailing. Other interventions might be more complex such as a text-message-based intervention that consists of several such messages delivered over a few months. More in-depth interactions could consist of face-to-face meetings similar to benefit counseling or case management. The extent of contact will have consequences for the potential impacts of the intervention as well as its cost and size (Dillman et al., 2014). Another related consideration is whether the worker can interact with the information. The worker may not be able to respond to an automated text message; in a face-to-face or phone interaction, however, the worker could engage in a conversation.

4.3.2 Study Design

An informational intervention is well-suited to a randomized control design, the most rigorous design possible to identify the impact of information. To design the random assignment study, researchers need

²² SSA has funded research to develop the Work Disability Functional Assessment Battery (WD-FAB) to explore the feasibility of integrating a new assessment tool within the context of work disability determination processes. WD-FAB is a computerized self-report assessment that includes a battery of 8 domains of physical and mental health (Basic Mobility; Mood and Emotion; Upper Body Function; Community Mobility; Mood and Emotions; Resilience and Sociability; Self-Regulation; and Cognition/Communication). The test can be administered in less than 15 minutes and was developed with Item Response Theory (IRT) and Computer Adaptive Testing (CAT).

to consider several choices. The first is whether to test a single treatment or multiple treatments. An informational intervention is naturally amenable to a design with multiple treatments. For example, the same information can be delivered to participants in different ways, or different types of information can be delivered using a single medium.

If the study involves testing more than one treatment, researchers will need to decide how to structure the randomization. In a *factorial design*, different components of the intervention would be randomly assigned separately.²³ An individual might be in the treatment group for one component but in the control group for another (or any combination of those). This design would allow DOL to recruit one large sample and use it to examine several questions simultaneously, without losing power to detect significant effects. A factorial design is different from a *multi-arm design*, where several treatment groups are compared to a comparison group and to one another. The larger sample sizes needed in a factorial design or multi-arm study relative to a single treatment arm design would have implications for the study cost.

Another issue to consider is the *unit of randomization*. For the most part, the informational intervention is conducive to individual-level randomization. However, group-level randomization may be necessary in order to avoid contamination: workers in a comparison group may learn the treatment information from their treatment group coworkers at the same employer. This would be unlikely unless there were many participants at the same employer. If an employer²⁴ with many establishments across multiple states could be recruited, randomization could be conducted by establishment instead of by employer. Workers could also be randomized by geography, such as by Census tract.

Below we discuss in more detail two models of Option B that DOL could implement in order to study the effects of information on employment and other outcomes.

4.3.2.1 Design B-1

The first model, Design B-1, is a low-intensity intervention conducted through one or several employers (we discuss recruiting employers in Section 4.7), or via a partnership with an insurance company or state agency. The design involves the delivery of information through electronic or mail methods, without personal interaction. Participants would be identified for the study in various possible ways. They could have missed a large number of days of work, for example, or have applied for workers' compensation or temporary disability benefits. Study Design B-1 seeks to identify whether a low-intensity informational intervention can have effects on employment outcomes.

4.3.2.2 Design B-2

The second model, Design B-2, provides a higher-intensity intervention than the first. The design includes detailed, personalized information provided in-person at least once. Participants would be identified when they apply for temporary disability insurance. In conjunction with one or more private temporary disability insurers, the participants would be provided information specifically tailored to their particular

²³ For example, there might be one dimension of interventions that tests different kinds of messaging (formal and authoritative with either a loss-based appeal or gains-based appeal, or story-based with either a loss-based appeal or gains-based appeal), and another dimension that varies the specificity of information (information tied to the general labor market and rules, information tied to a specific geography, or information tied to a specific employer in a specific place). With four levels in the first dimension, and three in the second, there would be 13 arms in a multi-arm study, but the factorial design would compare four levels to control, then three levels to control.

²⁴ We do not use the concept of "large employer" required under the Affordable Care Act, of at least 50 full-time-equivalent workers, but instead highlight that the key requirement here is operating multiple establishments in many states. A franchise operation, such as fast food restaurant or hotel chain, would not be as useful, as that would require two layers of recruitment: first the parent company, then the franchisees.

condition, as well as the ability to talk to an advisor about their employment and job options both within and beyond their current employer.

4.3.2.3 Comparison of B-1 and B-2 design options

Exhibit 4-1 below summarizes the key elements of the proposed models.

Exhibit 4-1. Comparison of Low-Intensity Model versus Personalized Information Model

	Design B-1: Low-Intensity	Design B-2: Personalized Information
Touchpoint	Several possibilities: (a) workers who miss a certain number of days of work; (b) workers' compensation applicants; (c) applicants for temporary disability benefits or vocational rehabilitation services, (d) Veterans transitioning to the civilian labor force, or (e) participants in a public assistance program.	Applicants for private temporary disability insurance are asked by the insurance provider if they consent to participating in a study.
Collaborating institutions	Employer or agency. Health care providers could also be involved.	<ul style="list-style-type: none"> • Temporary disability insurance provider provides the randomization and implements the treatment. • Community organizations may participate in developing information materials. • Doctors and medical professionals could be involved. • Employers could be involved.
Intervention description	<p>The intervention tests the effect of information content on employment.</p> <p>The intervention will provide groups with information on (a) job accommodations, modifications, and other benefits, either in general terms or specifically available to employees at their organization; (b) work abilities and limitations for workers with several types of health conditions; and (c) external organizations and programs that might help them to maintain employment, including local organizations, workers' rights under the Americans with Disabilities Act, resources to resolve disputes, or national advocacy organizations.</p> <p>A multi-arm or factorial design could include two varieties of information to test whether the content or design of the intervention produces different effects, or could vary the information delivery medium (e.g., could test information sent by (a) text message, (b) email, and/or (c) paper mailings to the worker's home).</p>	The intervention tests the effectiveness of an in-depth and personalized information session for applicants to temporary disability insurance. Applicants will meet with a counselor and receive information on their particular physical or mental condition, the options available through their employer, and outside resources. The intervention may also include follow-up meetings or additional resources through the disability insurance provider.
Treatment conditions	Treatment group participants receive information (which could take more than one form, or be delivered in more than one way as described above). Comparison group participants do not receive any new information.	Treatment group participants receive the informational intervention. Comparison group participants still can apply for benefits but do not receive the additional information resources.
Study design	Randomized control trial with one treatment group, or more than one in a hybrid design.	Randomized control trial with one treatment group.

	Design B-1: Low-Intensity	Design B-2: Personalized Information
Data sources	Employer or agency records on participant’s occupation, job duties, wage, hours, leave-taking, and full-time status. National Directory of New Hires or Unemployment Insurance data could be merged onto the data for administrative measures of employment and earnings.	Applicant information from the insurance provider on applicant’s occupation, employer information, and other individual characteristics. National Directory of New Hires or Unemployment Insurance data could be merged onto the administrative data to observe employment and earnings. All applicants (treatment and comparison) are administered a survey prior to random assignment, or prior to the point in which the treatment group receives the intervention.

4.3.2.4 Design B-3

A third potential model for studying the effect of information on employment would be to combine Designs B-1 and B-2 in such a way as to compare whether a low-intensity or personalized information intervention is more effective. This “hybrid” mode, Design B-3, would build off Designs B-1 and B-2, but instead of one treatment arm there would be two treatment arms and a comparison group. A single population of workers, perhaps identified through an insurer or employer (both options discussed for Designs B-1 and B-2), would be assigned to these three conditions. One treatment arm would include personalized information, as specified in Design B-2. The other treatment arm would include a low-intensity delivery of information, as in Design B-1. The comparison group does not receive an intervention. Design B-3 would allow DOL to understand whether information to workers affects employment relative to no intervention, but also which informational format is more effective.

4.4 Participants

Target population. This design would target a group of workers who experience an illness or injury. The power of the design would be improved if the intervention were targeted at workers with a relatively high likelihood of applying for federal disability benefits. A key challenge is that it is difficult to identify these workers (this is precisely the question that Option A seeks to address). The ideal target population will depend on the specific type of intervention to be studied. Participants would be recruited when they interact with a touchpoint—a service or support—in a way that identifies them as a worker who might benefit from information on work and disability.

A further challenge is recruiting participants into the study. If the intervention occurs with the cooperation of one or several employers,²⁵ then the target population would include **workers who file new workers’ compensation claims, experience a high number of health-related absences, or apply for temporary disability benefits**. A benefit of using these groups as the pool of eligible participants is that they are easily identifiable, but there are some drawbacks. Women on maternity leave make up a large share of workers using temporary disability benefits, but also may not have a higher risk of applying for permanent benefits (Donovan, 2019; Vahratian & Johnson, 2009; Waldfogel, 1999). Because maternity leave claims are easy to identify, researchers would exclude these and potentially other ineligible cases prior to the evaluation. Regardless, most workers do not have short-term disability benefits, so this might not be the best way to identify a sample of sufficient size.

A potential target population could be the set of **workers who make their first contact with a state vocational rehabilitation agency**. This point of entry is relatively late on the pathway between injury or illness and long-term disability benefits, but these workers are relatively easy to identify.

²⁵ For a recent example of such a design, at “a large employer with over 12,000 employees,” see Jones et al. (2018).

Another potential group consists of **Veterans who have recently separated from service with low disability ratings**. The Veterans Affairs Disability Compensation system assigns ratings of between 0 percent and 100 percent, reflecting a Veteran's residual functional capacity, with higher ratings yielding larger benefits. Though Veterans with lower ratings may not qualify for SSI/SSDI at the time of their separation, they do have a high probability of eventually applying, and thus they would be an appropriate target population for the intervention. These are individuals who may have trouble finding work or retaining it once they are hired (Stern, 2016). Identifying this group would require the partnership of the U.S. Department of Defense or the Department of Veterans Affairs.

Still another possibility is recruiting **recipients of public assistance benefits** who may have irregular or tenuous labor force attachment. Participants in a program such as the Supplemental Nutrition Assistance Program, Medicaid, or Temporary Assistance for Needy Families may benefit from targeted information, especially if they find themselves subject to new work requirements due to a state policy change.

An important consideration for Option B's study design is the touchpoint for random assignment, which has consequences for the resulting sample of participants. The research question assumes that participants are workers who have recently experienced an injury or onset of an illness. However, this definition of the study sample is broad and there are several ways it can be operationalized during the study itself. Ideally, participants would be identified for inclusion in the study as soon as the health condition appears, but this is not feasible: there are delays between when an illness or injury occurs and when the individual notices or decides to take action (there may be further delays before an observable event, such as missing work, occurs). Instead, the study would choose from among several touchpoints at which individuals with potentially disabling medical conditions find themselves early on in the SAW/RTW process.

The decision of which of these touchpoints to choose will define the sample, and thus has implications for what the counterfactual to the treatments would be. The choice of touchpoint also has more pragmatic implications on how to establish consent and collect individual-level data.

Several touchpoints are worthy of consideration. Some examples include:

- *A worker's first contact with a primary care physician to report an illness or injury.* In a manner similar to the COHE model in Washington State, the physician could refer patients with a particular set of health conditions to services within the intervention, conditional on their consent.
- *Partner with employers to help identify workers at risk.* Employers could identify workers at different points in time. These include workers with a high rate of health-related absences, workers filing workers' compensation claims, or workers applying for temporary disability insurance benefits. Recruitment would depend on the ability and cooperation of partnering employers.
- *New participants in vocational rehabilitation.* Participants in vocational rehabilitation already receive information on how to adapt their work to their health condition. So, to be effective, an intervention using vocational rehabilitation as the touchpoint might be limited in the range of the content of information it would provide. For example, such an intervention could provide resources specific to the worker's employer.
- *Application or re-certification for public assistance benefits.* Participants in public assistance face work requirements but may not know how to work with their health condition. As with other touchpoint recruitment strategies, this group would not be representative of all workers.

Setting/venue. The intervention would occur in at least two distinct settings: participants will need to be recruited, and they will need to receive an intervention. The setting of the intervention may determine

who the participants are, and different choices of setting will result in samples of participants with varying characteristics. DOL would partner with one or more organizations to recruit sample members. This might be the Department of Defense, state vocational rehabilitation agencies, state workers' compensation agencies, public assistance agencies, or state or private temporary disability insurance providers. DOL might also work with one or many employers.

Similarly, DOL might coordinate with local organizations or state vocational rehabilitation agencies to develop the information that will be provided to participants. It is not necessarily the case that the two types of settings would be different. The intervention might occur through a state vocational rehabilitation agency, with the same agency providing the materials. On the other hand, workers might be recruited through an employer and receive information from a local organization.

Recruitment plan. This design would require recruiting a partner organization that has contact with the target population and would be willing to facilitate recruitment. Another possibility would be to mount an expensive recruitment process via a geographic sample. Potential partner organizations include the Department of Defense, the Department of Veterans Affairs, workers' compensation agencies, public assistance agencies, vocational rehabilitation agencies, private disability insurers, and large employers.

4.5 Data

Data needed. This option requires three types of data:

1. *Baseline data on study participants*, including basic information on them (e.g., demographics), their employer, job characteristics, and nature of injury or illness. These data would likely come from administrative records. Employer records, for example, can provide baseline demographic and job characteristics.²⁶ The study could also use a survey to obtain baseline data.
2. *Process details and/or implementation records*, including information on each participant's treatment condition and on the receipt of the intervention itself. For example, the number of text messages or mailings received and—if the participant was in the higher-intensity treatment arm—the number of meetings with a counselor and perhaps case notes from the conversations. These data would come from the records of the entity implementing the evaluation.
3. *Outcome data, primarily employment*. Employer records can provide employment outcomes for workers who remain with the same employer. Other outcome data, such as earnings and employment, could come from administrative sources such as the National Directory of New Hires. Data on disability applications could come from SSA.

This option would require linking participant information to administrative data from the National Directory of New Hires, SSA, or one or more state Unemployment Insurance programs to measure employment and earnings, or collecting follow-up data expressly for the evaluation.

Permissions, matching, and consent. This evaluation design option would necessitate collection of existing data that include personally identifiable information. Most important among them would be administrative employer records, which would include information about the workers' salary/wages, job duties, and tenure with the company. In order to provide a full picture of the employment trajectories of the workers, employer data could be matched to administrative state or federal earnings records, such as from state unemployment insurance agencies or the National Directory of New Hires. These data, providing quarterly information on each worker's earnings and industry of employment, could be matched to the employer records by Social Security Number. If workers were recruited outside of an

²⁶ It can be very challenging to collect information from employers, but barriers would be lower if working with a smaller number of larger firms. We discuss this further in 4.7 below.

employer link, for example via an insurer, data on employment and earnings would need to be obtained from these other sources alone. In either case, the research team would need permission from individual participants to match to administrative data, via Social Security Number if available or name/address/date of birth if not. Informed consent would not be obtainable for all individuals, leading to some attrition. Further, inevitably some participants would not be matched in administrative data, and there could be some false matches, though experience indicates rates of true matches are well above 90 percent (Layne et al., 2014).

Data collection plan. No new data collection via a survey is required. Existing data would come from employer records and other administrative sources such as National Directory of New Hires or state Unemployment Insurance agencies. DOL would collect the process or implementation data from the entity implementing the evaluation itself. Depending on the setting chosen for the intervention, this might be a set of employers, the Department of Veterans Affairs, or a state vocational rehabilitation agency.

Still, DOL might want to consider a survey of eligible participants at baseline. The goal of the survey would be to gain information on participants' education, occupation, job activities, and experience with their current employer. The survey could also ask some questions about the nature of the injury or illness and how it affects their ability to work. The survey would also focus on questions related to the outcomes of interest, such as asking about participants' perceptions of their future employment, earnings, and hours of work, job transitions, and job duties.

The survey should not ask about perceptions and knowledge about SAW/RTW strategies or resources, or workers' planned course of action in response to their medical condition. Such survey questions themselves might provide new information to respondents about SAW/RTW resources, and thus be a type of treatment in and of itself.

A follow-up survey could collect key outcome data, if administrative data proved infeasible as an option, and mounting a follow-up survey would allow for a broader scope of outcomes to be measured. However, we expect a follow-up survey would result in much higher expenditures on data collection.

IRB, security, logistical concerns. Obtaining administrative data from one or several employers would require establishing a data use agreement with the employers, approval by an IRB, and appropriate security measures to protect personally identifiable information. Matching records across different administrative data sources would require establishing multiple data use agreements with, for example, state agencies collecting data on employment and earnings, the Office of Child Support Enforcement which administers the National Directory of New Hires, or the Social Security Administration for disability applications. Maintaining these matched records would require a secure data environment, presumably at least "Moderate" per the Federal Information Security Management Act of 2002. Developing data agreements and matching procedures could occupy the first two years of the study, leaving less time to recruit individuals and follow their outcomes over time; we describe timeline considerations in 4.7 below.

4.6 Analysis Plan

Outcomes and hypotheses. The informational intervention in evaluation design Option B may have impacts on the following outcomes (we list primary and secondary outcomes as well as a hypothesis of their direction):

- **Employment.** This is the primary outcome of interest of the intervention. Employment outcomes can be defined over a variety of timelines. Given a study timeline of approximately 3 to 4 years, the employment outcome will necessarily be constrained to short term, with follow-up one or two

years following the intervention. The aim of the intervention is to improve employment, so we expect positive effects, if any, on this outcome.

- **Application for disability benefits.** This is a primary outcome of interest for this option, if the needed data can be secured. The aim of the intervention is to improve employment, so we expect negative effects, if any, on this outcome.
- **Work Absence.** A secondary outcome related but distinct from employment is the time spent away from work before returning, or before applying for disability benefits, if available. These work absences may be through paid sick leave or unpaid sick leave, so collecting information on the nature of income sources during work absences would be useful if feasible. Data on participation in workers' compensation or temporary disability programs, if available, could provide additional secondary outcomes of benefits paid. The expected sign of this outcome is ambiguous, because workers may be more likely to stay attached to the labor force and put off applying for federal disability benefits, but stay out of work longer by seeking out additional supports. Alternatively, participants may be induced to apply immediately for federal disability benefits, and leave work entirely. Other participants could return immediately to work, and never apply for federal disability benefits, and the relative size of these different responses is an open empirical question.
- **Earnings.** This secondary outcome would be defined either through self-reported earnings from a survey or with administrative data from Unemployment Insurance agencies or the National Directory of New Hires. The impact of the intervention on earnings relative to a comparison group should be positive in the long term if the intervention leads to a positive employment outcome. However, earnings could be lower for the treated group in the short term if the intervention influences workers to take on job accommodations or modifications that reduce their hours or earnings but the comparison group remains at work through an unhealthy overextension.
- **Intensity of employment.** This is another secondary outcome. Similar to earnings, we expect the impact of the intervention on weeks worked per quarter, the number of hours worked, or some other measure of intensity in the long term to be positive relative to the comparison group, but potentially negative in the short term. Data on intensity may not be available from administrative data sources, so this should be regarded as an optional outcome measure to be supplied by an employer or a follow-up survey.
- **Job duties/occupation.** This is another secondary outcome. Information on a worker's occupation can be linked to administrative datasets such as O*NET that include more details on the occupation characteristics. Following an informational intervention, workers in the intervention groups might change occupations. A change in job duties or occupation is not necessarily a positive or a negative outcome for the worker. Some workers may take on a lower-paying job if they cannot resume their prior tasks. Others may gain new training for a new job with similar or higher pay. Thus, this outcome could be measured together with earnings and intensity of employment. Given the data sources, this outcome would likely only be available for workers who remain with their same employer, or who respond to a follow-up survey.

Model(s) or analysis methods. Any intervention of this type would be studied using a randomized control trial (experimental) design. The analysis would include comparing outcome means across the treatment and control arms, if randomization were conditioned on many factors, such as age and job type. This comparison of means can be viewed as a regression of outcomes y_i (y for outcome measure, i for unit of analysis) on a single indicator A of assignment to treatment with the excluded category in the control arm shown in the equation below (or for a multi-arm design, on a vector of indicators A of treatment types). The coefficient(s) α represent the impact(s) of treatment (relative to control) in this model.

$$y_i = \delta + A_i\alpha + \varepsilon_i \quad (\text{Equation 1})$$

Covariates can be included in the model if randomization were not conditional, to improve precision (Lin 2013); this flexibility is one advantage of formulating a comparison of means as a regression. Other advantages include flexibility in standard error calculations and hypothesis testing. All estimates would take the design of the study into account, including adjusting for clustered assignment or differing probabilities of assignment across locations or subgroups if needed.

Expected methodological issues and responses. We conducted a simple calculation of the sample size needed for a straightforward single treatment arm design, such as Design B-1 or B-2 on its own. For these calculations, we seek to detect an effect size of 0.1 (i.e., a 10th of a standard deviation of the employment outcome for the control group). This amounts to an employment increase of around 5 percentage points for the treatment group relative to the comparison group which is an appropriate size because the sample would include individuals at high risk of labor market exit.²⁷ To detect this effect would require a sample of 2,200 participants individually randomly assigned in equal proportions to treatment and comparison groups, with 1,100 in each group.²⁸ To detect a larger impact, of 6 percentage points, would require a total sample of only 1,500 (750 each in treatment and comparison arm), and to detect a smaller impact, of 4 percentage points, would require a sample of 3,400 (1,700 each).

In order to test a hybrid model, the two samples would be added together, with just one comparison group. Thus, to detect an effect of 5 percentage points would require 1,100 in the comparison group, 1,100 in the low-intensity arm, and 1,100 in the personalized information arm, for a total sample of 3,300. Similarly, the sample required to detect a 6 percentage point effect in the hybrid model would be 2,250, and 5,100 to detect a 4 percentage point effect.

4.7 Practical Considerations

Implementation timeline. We estimate that this evaluation design option would require a timeline of approximately four to five years, including time to design the intervention. An initial evaluation and analysis design phase might last several months, and might require several revisions as recruitment and data agreement successes unfolded. Securing data use agreements and data sharing procedures could take between one and two years. Enrollment in the study would be on a rolling basis and would likely last at least a year. The enrollment period length would be a function of the enrollment rate and the required sample size. The intervention itself can be quite short (from a day to several weeks). We would recommend including an analysis of outcomes after a delay of at least a year, as our analyses of existing data suggest that a substantial share of those who leave work due to a health condition have neither returned to work nor been awarded SSDI or SSI 16-20 months after onset. The timeline would also require time for analysis and reporting. With two years for planning and data/participation agreements, a year of recruitment, a year and a half of follow-up, and half a year of analysis and reporting, the total time required for the study would be five years. These rough time periods are based on our experience in mounting studies of similar scope, but the details depend crucially on design details, so should not be taken as accurate estimates of required time unconditional on the specific of the evaluation.

²⁷ Hollenbeck (2015) finds that approximately a third of individuals aged 25-61 with work-limiting conditions were employed. The variance of the outcome is thus 0.22, the standard deviation is 0.47, and thus the one tenth of the standard deviation is approximately 0.05, or 5 percentage points.

²⁸ The calculation is based on a power level of 80 percent and an alpha of 0.10, corresponding to a *t*-statistic of 1.68. This calculation assumes no regression adjustment and an even split between treated and control arms.

Cost. This design would require the planning and infrastructure to conduct a randomized control trial, but delivering the intervention itself could be low cost. The cost of data collection would depend in large part on the data used. For example, administrative records such as quarterly wage records drawn from the National Directory of New Hires would be much less costly than survey data collection with participants.

Implementing Option B requires (i) designing a new informational intervention for workers, either a low-intensity general information campaign or personalized information; (ii) developing the content of the information and plan for delivering it; (iii) a detailed evaluation plan with a randomization strategy that specifies the unit of randomization and the number of treatments to be tested. If the design included more specific information or more frequent and personalized contact, then the option's cost would be higher.

Sample recruitment represents an important source of costs for this study design. DOL would need to establish partnerships with the organization identifying (and providing the names of) potential participants. Possible partners to identify the sample of workers include employers, the Department of Defense, the Department of Veterans Affairs, workers' compensation agencies, public assistance agencies, vocational rehabilitation agencies, or private disability insurers.

Once the treatment is implemented, data collection would be required to measure participant baseline characteristics and outcomes, but primarily of existing data. Whether or not DOL undertakes a participant survey to measure baseline data, this design option requires data from employers to measure work history and baseline employment status. Costs accrue from coordinating with each employer to establish a data use agreement and ensure appropriate security measures to protect personally identifiable information. Employment and earnings outcomes could be measured from administrative data sources such as the National Directory of New Hires. DOL could also choose to administer a survey via structured interviews to measure participant responses, which would increase costs.

Geographic scope. The geographic scope of this intervention depends on the partnering institutions. There is no particular state policy that the intervention would be evaluating, so the scope of the intervention could be national. However, state policies or changes in policy could moderate the effect of the information, so it might be useful to choose several types of states in which to operate the intervention (e.g., states that impose new work requirements on public assistance participants and states that do not).

4.8 Contributions and Limitations

Policy relevance. This design has the potential for immediate policy relevance, as the intervention tested could be easily adopted by existing programs if found to be effective. The degree to which the results can be generalized to other populations will vary by population and depend on how the sample is selected. For example, a sample that includes workers from only one industry might be informative about other workers in that industry, but less informative about workers in another.

Contributions. This design option would generate information on whether delivering targeted information to workers who might leave the workforce due to their health improves employment outcomes. By randomly assigning workers to different arms of treatment, this design option could also generate information on the delivery method, topics, or timing of information that produce the best results (e.g., text message vs. letter, information on the returns to work vs. information on resources in the local community, immediate vs. delayed). The control group could also be exposed to different treatments in a stepped wedge design, so that eventually all participants receive the full complement of information. This research design could lay the groundwork for larger-scale programs that could be adopted at relatively low cost, if the low-cost variants of treatment produce large impacts. While it is frequently the case that higher cost interventions produce large impacts, many of the highest value-for-money interventions are informational interventions (Dhaliwal et al., 2012).

Internal and external validity. Because the study design is a randomized control trial, the study will have high internal validity. The sample of participants will not be a representative sample of all workers, because they would be sampled based on employment at particular firms, be insured by particular insurers, or otherwise selected for inclusion in the study. Depending on the degree to which the workers in the sample are similar to the full population of workers, generalizability may be limited, as it is in many randomized control trials. This would affect the policy relevance only insofar as we believe impacts in other populations differ substantially, and this could be investigated using methods described by Olsen et al. (2013).

5. Option C: Examine Employer Practices and Test an Intervention that Informs Employers about Resources and Best Practices to Retain Workers with Impairments

Employers play a major role in the choices workers make to stay at work or return to work following injury or illness, but little is known about employers' decision-making. This option seeks to understand what SAW/RTW practices (if any) employers are currently using and to determine whether providing information would increase or improve the use of SAW/RTW practices. This option would involve two components. Component 1 would collect information on current practices through a survey to employers, perhaps targeting medium to small enterprises with fewer than 500 employees. This survey might be supplemented with a mixed-methods study of a small subset of the employers in the survey sample. In Component 2, a randomly selected sample of employers would receive information, delivered at the end of the Component 1 employer survey, in a brochure, or as part of a consulting visit. Outcome data would be collected through a follow-up survey.

In this chapter we discuss the research questions and conceptual model for this option. We then discuss the study design for Component 1 and the intervention to be tested in Component 2. We also discuss the target population, recruitment strategies, data sources, and analysis plans. Option C overlaps little with existing research, and it is unlikely to duplicate any findings in RETAIN. Though the RETAIN evaluation includes some contact with employers, evaluating the effects of employer informational interventions alone is not a focus. Securing employer engagement, perhaps the largest challenge in implementing this option, would need to be addressed in the planning phase of this study.

5.1 Research Question(s)

This option would seek to answer three research questions:

- (1) **What practices, if any, do employers use to retain workers with impairments?**
- (2) **Does providing employers with information about SAW/RTW best practices improve employer practices?**
- (3) **If so, do improved practices lead to better employment and disability outcomes for employees?**

The first question is primarily descriptive. Answering this question would give a comprehensive view of SAW/RTW practices across a broad range of employers. The second and third research questions assess the effects of an information intervention on employer practices and employee outcomes.

5.2 Conceptual Model

An employer's SAW/RTW and accommodation policies or lack thereof may change the likelihood that a worker's injury or illness develops into a work disability. Understanding the extent to which employers have explicit policies to encourage SAW/RTW would inform policymakers. Component 1 of this evaluation design option would document the practices employers use, which ones they are aware of, and which they might take up. Component 2 would evaluate whether receiving information about SAW/RTW practices affects employer practices and, potentially the employment and outcomes for their employees.

5.3 Intervention(s)

The research questions lead to a study with two components. Component 1 uses an employer survey to describe the SAW/RTW practices employers use. Component 1 could be implemented as a standalone study or as the baseline data collection for Component 2. Component 2 could only be implemented in combination with Component 1. Component 2 would use an experimental evaluation design, randomly assigning employers to a treatment group that receives information on SAW/RTW practices, or to a control group that does not receive the information. This section discusses options for conducting Component 1 and the intervention details and study design for Component 2.

5.3.1 Component 1

Component 1 would seek to gather information about employers and their practices regarding workers with recent injuries or onset of illness. Ideally, the evaluation would gather information about a broad range of employers, but could be targeted businesses with fewer than 500 employees. These are the types of employers that are less likely to have adopted any type of SAW/RTW program or strategy.²⁹

First, the study would survey employers to collect four types of information:

1. *Details about the business itself*, such as the characteristics of the workforce, the occupations and tasks completed by workers, and information on earnings.
2. *Types of benefits programs*, including health insurance, leave policies, and disability insurance.
3. *Prevalence and types of illnesses and injuries* experienced by the employer's workers. We note that employers may not be able to answer parts of this question.
4. *Current, prior, and planned practice regarding SAW/RTW*. This would also include gauging the employer's knowledge and interest in existing resources and best practices. Not only would researchers learn about whether employers know about or implement a particular practice, but they would seek, to the extent possible, information about employers' barriers and incentives regarding the practices.

This survey would consist of multiple-choice and open-ended questions, and it would be administered electronically as an internet survey. A key challenge would be to identify the survey respondents—representatives of the firm who can answer questions about the firm's knowledge and opinions about various SAW/RTW practices and approaches. These representatives are likely the officers of the firm, and in past work, including work ODEP has commissioned to collect information about employers' practices, contacting these individuals and securing their cooperation has proven difficult. Thus, a key design decision involves the nature of outreach at the initial survey stage to improve response rates. A mix of financial incentives to participate, official outreach from authorities, and partnerships with business groups would likely be necessary to improve response rates. It might also be possible to combine the survey with an existing survey, such as the Employee Benefits Survey administered by the Bureau of Labor Statistics in DOL. This could improve response rates and reduce respondent burden, but might require extensive up-front negotiations within and across federal agencies.

DOL could consider supplementing the Component 1 survey with a mixed-methods study of a subset of the employers who respond to the survey. This subset could be a random sample of employers, in order to gain a broad range of insights, or it could focus on employers that either already have some kind of

²⁹ A recent survey finds that 43 percent of companies with 25 to 99 employees have formal SAW/RTW protocols in place as do 29 percent of companies with 100 to 499 employees (Standard Insurance Company, 2019, p. 27). Of companies with 500 or more employees, 61 to 64 percent have formal SAW/RTW protocols in place.

SAW/RTW strategy in place or have tried one in the past. For the mixed methods study, researchers would conduct site visits to interview a range of employees of the firm and to conduct focus groups. These employers might also provide firm-specific data, providing a deeper understanding of the human resources at the firm.

5.3.2 Component 2

Component 2 would use an experimental evaluation design to test the effects of providing information on SAW/RTW practices on employer practices and (potentially) on employee outcomes. If Component 2 were implemented, then Component 1 would serve as the baseline information collection for the randomized trial. Component 1 might also serve as the delivery mechanism for the information intervention (see below for more detail).

5.3.2.1 Intervention Development

The first step of conducting the randomized trial would be to develop the informational intervention. The goal of this development process would be to create a package of information about the most promising SAW/RTW practices and to determine the most effective way to deliver this information. DOL could draw upon a range of individuals with expertise in SAW/RTW (e.g., disability accommodation specialists and employer representatives) to guide this development process. DOL or researchers could apply a Delphi method that collects guidance from the experts in two rounds. The Delphi method (Dalkey & Helmer, 1963) uses a series of open-ended interviews with experts in a first round; then in the second round, the experts comment on and review a collated analyzed version of that feedback that includes the intersection of suggested types of guidance and notes on each from experts. The first round expert interview would ask about the content of the informational intervention, the delivery mode of the information (including the extent of follow-up repetition), and who best in the employer's organization to receive the information.

In addition to the Delphi method, the intervention development could involve a few focus groups with employers. The focus groups would have two main questions: (1) what do employers want to know about SAW/RTW practices? And (2) to what extent do employers find the experts' recommended informational interventions useful? Developing the intervention would also involve cognitive testing to understand how employers process the information in order to improve on the intervention. The principles of behavioral economics elaborated in Kahneman (2013) and in Thaler and Sunstein (2009) suggest that the way an informational intervention is designed is crucial to its success. It would be important to pilot any intervention, however, before implementing the demonstration at scale, to avoid any costly redesign.

5.3.2.2 Intervention Details/Approaches

There are a variety of ways to deliver the informational intervention to employers. The most straightforward method would be to provide the information to treatment group employers at the end of the Component 1 survey. Researchers could include reading material at the end of the Component 1 survey, and follow it with a mailed brochure containing the same information. Information on the survey would contain no value judgments about different practices, and it would include a wide list of practices.³⁰

The informational treatment could be more comprehensive, however. For example, treatment group employers could be offered a three-hour consulting visit with the possibility of follow-up consulting as desired by the employer. These two types of follow-up—post-survey repetition versus consulting or technical assistance—could also be used in combination or tested as separate treatment arms. Because

³⁰ A respondent's act of thinking and reflecting on SAW/RTW strategies and the current practices of the firm could be considered an information treatment in a sense, but the impact of this would likely be minor, and would be common to both the treatment and control groups.

further dividing the treatment group would reduce statistical power, we describe a single treatment version.

In the remainder of this chapter, we assume the informational intervention in Component 2 will be delivered online at the end of the Component 1 survey, followed by a mailed brochure with information on best practices in SAW/RTW and the benefits of adopting best practices. The brochure (and online version) could adapt existing resources from DOL or its partners, subject to findings from the Delphi method described above. The treatment information should address both the “business case” for adoption to promote the benefits of considering a change in practice and concrete steps that could be immediately taken by an employer. For example, the brochure might highlight that in a survey conducted for the Job Accommodation Network,³¹ employers “report the direct and indirect benefits of accommodating people with disabilities:

- 90 percent of employers reported higher retention of valued employee
- 72 percent report increased employee’s productivity
- 60 percent report eliminated costs associated with training a new employee
- 38 percent reported saved workers’ compensation or other insurance costs
- 45 percent report increased workplace safety
- 28 percent report increased profitability
- 17 percent report [an] increased customer base.”

Such marketing of benefits of accommodating people with disabilities could then be followed up with a set of “Business Strategies That Work.”³² For example:

- Develop and communicate statements of the company’s commitment to inclusion of workers with disabilities.
- Encourage workers with disabilities and other employees to identify barriers and individual and systemic concerns without fear of reprisal, and provide mechanisms to allow them to provide this information anonymously or confidentially.
- Provide training opportunities, including apprenticeship programs, on-the-job training, job shadowing, and tuition reimbursement for employees with disabilities.
- Develop and communicate written procedures for processing requests for reasonable accommodations.

³¹ The Job Accommodation Network is a technical assistance center funded by DOL and serves customers across a wide range of employers. For more information on the survey results, see: <https://askjan.org/modules/valueproposition/upload/valuepropJANTranscript.docx>

³² These examples are adapted from “Business Strategies That Work: A Framework for Disability Inclusion” (DOL/ODEP, 2012), with a focus on strategies that are applicable to small businesses. Thus, “assigning a full-time director of disability services or workplace supports to coordinate accommodations strategies” is not included, as a firm with 10 employees probably cannot support a full-time director of disability services.

- Provide training for staff about new strategies and devices, such as telework and assistive technologies.
- Ensure that both managers and employees are aware that they may contact the Job Accommodation Network to receive confidential and free advice and technical assistance on workplace accommodations.

Follow-up Survey. To collect outcomes of the informational intervention, a second (or “follow-up”) survey would be administered to all employers that responded to the Component 1 survey. Treatment and control group employers would be administered the same follow-up survey. This follow-up survey would repeat many of the same questions as the Component 1 survey, and it would include the same battery of questions on the employer’s current and planned SAW/RTW practice and experiences with ill or injured workers. Comparing the change in these responses from the Component 1 survey to the follow-up survey, across the treatment and control groups would provide an estimate of the effectiveness of such information on employer practice. It would also give a sense of what types of SAW/RTW practices employers find to be worth implementing.

5.4 Participants

Target population. The target population for this evaluation design option consists of employers. Because Component 1 collects information about current SAW/RTW practices, a wide range of employers—both those with formal policies and those without—is desirable. However, because Component 2 tests whether information affects employers’ use of SAW/RTW practices, Component 2 should be aimed at employers that do not already have such practices in place. Employers with fewer than 500 employees fit this description, as they are less likely to have robust SAW/RTW practices (Standard Insurance Company, 2019). We also anticipate that employers that hire more workers in occupations that do not require a college degree should be selected with higher probability, as these workers are more likely to apply for disability benefits in the absence of a SAW/RTW program (Center on Budget and Policy Priorities, 2019).

Within the sample of employers, there are a number of different potential respondents. The Component 1 and follow-up surveys would ask questions about an employer’s basic operations, its SAW/RTW policies, and its knowledge or interest in adopting new SAW/RTW policies. Managers and human resources personnel can adequately answer questions about the existing policies. However, because the study is about knowledge and beliefs about various SAW/RTW policies and practices, it is important that the survey be completed by individuals with relatively similar sets of responsibilities or positions across employers.

Because Component 2 is a test of providing new information to a random subset of the employers responding to the Component 1 survey, the personnel responding to the survey should be in a position to make decisions about the employer’s policies. These individuals could include the head of human resources or others with executive roles; in a small employer with fewer than 50 employees, there might only be one or two individuals fitting this description. Securing their participation will be the primary challenge in this design. Several approaches may be necessary, including official outreach with incentives to participate and careful design of the outreach to emphasize potential benefits.

If Component 1 implemented a mixed-methods approach, one aim would be to gather information and beliefs from individuals with a wider range of experiences and responsibilities within the firm. To that end, researchers would assemble a group representing various types of employees within the firm. These might include, depending on the size and type of the firm, executives, human resources managers, managers, entry-level workers, and workers who have previously experienced injury or illness affecting their job. Component 2 would exclude the employers selected for the in-depth mixed methods sub-study

from the randomized trial to avoid confounding the impact estimates of information delivered at the end of the Component 1 survey.

Setting/venue. To the extent possible, the study could have a national scope, aiming to cover as broad a representation of employers across the country as possible. Alternatively, it could focus on a specific industry or set of industries. In any case, the study should focus disproportionately on relatively small employers (under 500 employees), which are less likely to have implemented substantial SAW/RTW policies. To accomplish the goal of broad representation with an oversampling of small- to medium-sized employers, the study sample for Component 1 should be selected using stratified sampling, using employer size categories as strata.

Small employers are an important part of the SAW/RTW landscape. Firms with fewer than 50 employees are generally exempt from requirements under the Affordable Care Act and the Family and Medical Leave Act, so will be less likely to have implemented policies around leave or flexibility.³³ They may also be exempt in some cases from provisions of the Americans with Disabilities Act.³⁴ Smaller employers are thus an important target group for implementing new SAW/RTW policies. Small firms also make up the majority of employers nationwide: of all firms, approximately three in four have fewer than 10 employees, and 96 out of 100 firms have fewer than 100 employees.³⁵ Most employees are at larger firms, but because small firms are more likely to lack the policies and programs of interest, a random sample of firms stratified by firm size and industry could more effectively target the firms least likely to have policies in place.

An alternative approach would be to target a specific industry. This approach would aim the informational intervention at the specific needs of this industry. The endorsement of a relevant trade association or employer organization could aid the implementation of the study.

An industry well-suited for this approach might have high marginal turnover costs and a range of occupations with varying requirements. Industries with high marginal costs of employee turnover would be most likely to respond to the intervention. Research on employee turnover shows that associated costs are about one-fifth of the annual salary for lower-wage earners and perhaps an increasing proportion as salaries rise (Boushey & Glynn, 2012). Therefore, the marginal cost of employee turnover within an industry depends upon the mix of salaries and occupations within that industry. Selecting an appropriate industry might require an analysis of the occupational composition of several industries to better understand the employee turnover costs. Also, some industries with high marginal cost of turnover might face stronger incentives to retain workers, and so might already have SAW/RTW practices in place because it is in their interest to do so.

Another consideration is that in some industries, such as retail sales, a worksite might include a range of positions with varying requirements, making it possible for employers to make alternative assignments to retain a worker who experiences an illness or injury. Marginal cost of turnover in this industry might be relatively low compared to industries requiring more specialized skills, but there might be greater potential to study the effects of information than for an industry with little variation in job requirements.

³³ See websites from the National Conference of State Legislatures (https://www.ncsl.org/documents/health/aca_requirements_for_employers.pdf) and DOL (<https://www.dol.gov/sites/dolgov/files/WHD/legacy/files/whdfs28.pdf>).

³⁴ Employers with fewer than 15 employees are exempt from compliance with the ADA (<https://www.eeoc.gov/laws/guidance/fact-sheet-disability-discrimination>)

³⁵ Per the Business Dynamics Statistics, a public-use dataset of annual aggregate statistics describing establishment openings and closings, firm startups, job creation and destruction by firm size, age, industrial sector, and state [https://www.census.gov/ces/dataproducts/bds/data_firm.html].

Recruitment plan. An employer survey could be based on an existing sample frame. For example, researchers could purchase lists of employers by certain criteria, including firm size and industry, from third-party vendors such as Dun and Bradstreet. Basing the sample on an existing sample frame would assist the study in establishing national representativeness. In order to address the tremendous challenge of securing the cooperation of firm decision-makers, the study could partner with state agencies and/or business groups. The study team should also attempt to recruit business groups, such as chambers of commerce, trade or industry groups, or fraternal organizations, to write letters that endorse the study as a way to improve the bottom line of firms while simultaneously improving the lot of employees.

It would be possible to attempt to recruit a sample directly through partnerships with these organizations, but partnerships on their own are unlikely to generate a large enough or representative enough sample of employers that have not yet adopted SAW/RTW policies and programs. We recommend, therefore, that a sample frame be purchased from a vendor, and that recruitment involve partnerships with state agencies and business groups.³⁶

An alternative approach to recruiting would be to partner with a single, very large employer. Such an employer would need at least two to three thousand worksites. Examples of such employers would be the federal government and Walmart. The very large employer would need to be willing to maintain different policies across worksites for the duration of the study and would need a worksite-level respondent with authority to implement SAW/RTW policies. Partnering with a very large employer is a possible approach for both the Component 1 survey and the Component 2 impact study. For the results of the impact study to be useful, the large employer would need to have not already implemented a set of SAW/RTW practices.

5.5 Data

Data needed. For Component 1, the primary data would consist of a survey of employers. This survey would ask for basic information about the employer, as well as questions specifically tailored to the employer's need, knowledge, and interest in SAW/RTW policies and practices. The survey should also measure aggregate employee outcomes, such as turnover rates, reasons for separations (i.e., voluntary versus involuntary), and use of leave or flexible work schedules. Data on employer characteristics, such as size, industry, and occupational mix, might be drawn from sample frame information that is used to select the sample, or collected as part of the survey itself. We would model employer nonresponse as a function of these characteristics and the probabilities of selection to construct sample weights.

If DOL decided to conduct the mixed-methods approach in Component 1, data could come from the existing programs studied and information from interviews or focus groups with employers conducted as part of site visits. Key questions would include take-up, what resources employers believe they need, and what practices they put into place, as well as the effect of program access on practices, hiring, and retention. Researchers might also ask some representatives of the employer to reflect on their perceptions of the actual or potential costs and benefits, or return on investment, of adopting certain SAW/RTW

³⁶ Other types of partners might be valuable for identifying employers for the optional in-depth mixed methods sub-study of Component 1. DOL might consider partnering with existing employer organizations, such as the Disability Management Employer Coalition (www.dmec.org) or the Burton Blatt Institute's Disability Case Study Research Consortium. DOL might also consider the sample constructed for a survey conducted on behalf of the Job Accommodation Network, in which employers report the direct and indirect benefits of accommodating people with disabilities. DOL might also partner with existing organizations that provide information and technical assistance, such as the Americans with Disabilities Act National Network (adata.org). To improve the chance of identifying employers that are willing to participate, another type of recruitment strategy would be to use lists of employers from other DOL-funded projects, such as the American Apprenticeship Initiative.

practices. These questions would be fielded to a variety of sources at the firm, such as key leadership, human resources representatives, and workers engaged in a variety of job tasks. They might also be asked to reflect on specific examples of when the SAW/RTW policies were enacted for a particular employee.

Component 2 would require baseline data on the sample, implementation data, and outcome data:

- *Baseline data* for Component 2 consists of the Component 1 employer survey.
- *Implementation data* for Component 2 would add two new elements—treatment status of the employer and what treatment was received—to the Component 1 data. If DOL chose to have an additional treatment such as in-person consultations with experts, then data on the extent and content of the additional treatment would also be included in the Component 2 implementation data.
- *Outcome data.* DOL or researchers would field a follow-up survey. It would ask similar questions as in the baseline employer survey, to measure any changes in practice and aggregate employee outcomes.

The surveys in Components 1 and 2 would focus on employer practices and policies. These surveys would collect employer-level information on the number and types of employees and their average outcomes. If employers provide permission, information on employment outcomes for workers could be matched from administrative data to validate employer responses. However, we anticipate that asking for worker-level data would be likely to substantially reduce employer participation rates, which are already the key challenge of evaluation design Option C. Employers would also be invited to share any administrative data that would be helpful, such as benefit plan details.

Data collection method(s). Both the Component 1 survey and the Component 2 follow-up survey would be administered to employers. Each survey would be tested first with a pilot at a number of employers not to be included in the analytic sample. We anticipate that the surveys would be administered online with phone follow-up. The mixed-methods element of Component 1, if included, would include recordings of interviews and focus groups. This study would potentially also include administrative records from the employer itself, provided electronically, and Form 5500 data on benefit plans, using information obtained from the employer.³⁷

IRB, security, logistical concerns. The major logistical concern is employer response rates, and this can be addressed via a two-pronged approach of official cover letters attached to the survey and participation incentives. The most effective version of this approach would involve extensive high-level partnerships. We anticipate financial incentives to cover the time cost of participation would also involve large expenditures. Still, given response rates to prior employer surveys and responses to voluntary DOL programs, it is likely that the response rate will be quite low. The primary data collection (a survey) would need to meet OMB Paperwork Reduction requirements. DOL would need to obtain OMB approval prior to conducting either component of this option.

Low response rates might harm the generalizability of the Component 1 results. Even if the Component 1 respondent sample is not representative of all employers, the findings of Component 2 could be useful for future policy development if DOL is comfortable generalizing from the selected sample. In other words, because Component 2 is randomized, it will have high internal validity even if it is not representative of all employers.

³⁷ Form 5500 is used to file an employee benefit plan.

5.6 Analysis Plan

Outcomes and hypotheses. Component 1 would provide a descriptive picture of the use of SAW/RTW practices across employers. The primary outcome for Component 2 is the formal adoption of recommended SAW/RTW practices. Researchers would examine whether the informational intervention affected the adoption of any type of SAW/RTW practice, and whether it affected the adoption of particular types. We hypothesize that providing information about SAW/RTW best practices will increase and/or improve the use of these practices.

Model(s) or analysis methods. The Component 1 employer survey would involve standard sampling and weighting techniques to adjust for oversampling of employers with fewer than 500 employees. The mixed-methods element of Component 1 would require qualitative analysis of focus groups. Component 2 would use a randomly assigned sample, so the analysis could leverage this random design and compare regression-adjusted means. Using information from the Component 1 survey as baseline covariates would boost the statistical power of the estimation model in Component 2.

Expected methodological issues and responses. Components 1 and 2 rely primarily on a sample of employers. If both components are implemented, each employer would answer two surveys. Component 2 requires a sufficiently large sample size in order to detect effects of an expected size. Randomization would occur at the employer level, with equal numbers of treatment and comparison group employers. To calculate the required sample size, we seek to detect an effect size of 0.1 (i.e., a 10th of a standard deviation of the take-up rate of a set of practices in the control group). Assuming that half the employers have a SAW/RTW practice currently in place, this amounts to a 5 percentage point increase. If only 20 percent of employers have a practice in place, the desired effect becomes 4 percentage points. To detect a 5 percentage point increase in the share of employers with SAW/RTW policy, the sample size would need to be 2,400, split evenly between the two groups.³⁸

Evaluations that rely on employer surveys usually have very low response rates, and this evaluation design option requires respondents to participate twice. For a survey of employers about the Family and Medical Leave Act, Daley et al. (2012) found a response rate of 21 percent. Component 2 of this evaluation design option would rely on two survey waves, so response rates must be factored in for both waves. Only respondents to the Component 1 survey would be sent a follow-up survey. Employers that answered the Component 1 survey would likely be more willing to respond to the follow-up survey, and thus follow-up response rates would be higher than Component 1 response rates. Assuming a response rate of 20 percent for the Component 1 survey and a response rate of 80 percent for the follow-up survey, this leads to a required initial pool size of 15,000 firms.³⁹ If the Component 1 survey response rate was 10 percent instead of 20 percent, the required number of firms to be sampled would double, to 30,000. A survey with Component 1 survey response rate of 20 percent and a follow-up survey response rate of 50 percent would require an initial pool of 24,000 firms. To detect an increase of 4 percentage points, the final sample size would need to be 2,600; assuming a Component 1 survey response rate of 20 percent

³⁸ The calculation is based on a power level of 80 percent and an alpha (size) of 10 percent. This calculation assumes no regression adjustment and an even split between treatment and comparison arms.

³⁹ Under these assumptions, an initial pool of 15,000 would yield 3,000 first-wave responders, of which 2,400 would respond to the second wave.

and a follow-up survey response rate of 80 percent, the initial pool size would need to be 16,250 firms.^{40, 41}

The mixed-methods part of Component 1 would require a much smaller sample of employers, such as 10 to 20 firms.

5.7 Practical Considerations

Implementation timeline. Option C includes research components that could be implemented separately or in combination. DOL might choose to conduct Component 1 as a standalone study. We believe that Component 1 could be completed in approximately three years. This timeline would include time to identify the sampling frame, design the survey, obtain OMB approval, administer the survey and analyze results. We estimate that adding the mixed methods study to Component 1 might add approximately six months to the timeline, to conduct site visit interviews and focus groups with a subset of employers that complete the Component 1 survey.

If DOL pursues Component 2, it would be conducted in conjunction with Component 1. We estimate that the timeline for conducting Component 2 would be approximately five years. Component 2 would require time to design the information intervention with input from outside partners or through the Delphi method with a number of experts. Researchers would also need to design the randomization plan, and plan for recruiting employers. Administering the Component 2 intervention itself could be done relatively quickly, over the period when the Component 1 survey is administered, perhaps 4 to 6 months. However, outcomes would be measured with a delay of at least a year. Employers that successfully implement a new SAW/RTW practice or policy would presumably only do so after a certain amount of internal discussion and planning.

Cost. If DOL were to conduct only Component 1, costs would derive from developing and administering the first-wave employer survey and designing and executing the mixed-methods study (if included). If DOL were to conduct Component 2, costs would derive from designing the new informational intervention, designing a randomization plan, administering the intervention, and collecting outcome data from employers. In either Component 1 or Component 2, costs would include conducting analysis and reporting. Both components would involve costs to identify a wide variety of medium to small businesses, coordinating with firms to identify representatives who could reliably answer the survey's questions, and securing their participation.

DOL might choose to partner with existing employer organizations for recruitment and with technical assistance providers to access a national sampling frame of businesses to survey. We anticipate researchers would purchase contact information for the sample frame from a third-party vendor. Another component of the cost of data collection might stem from supplementing the survey data with the employers' administrative records or conducting focus groups or other qualitative data collection. The cost of data collection would depend on the size of the sample, but it would consist primarily of the cost

⁴⁰ The calculation of 2,600 assumes 20 percent of employers have a practice already in place, a power level of 80 percent, an alpha (size) of 10 percent, no regression adjustment, and an even split between treatment and comparison arms.

⁴¹ The required sample of 2,600 is roughly two to three times as large as the Survey of Employers in the Low-Skill Labor Market of 2007 (N=1,060; response rate = 54 percent; Acs & Loprest, 2009), jointly funded by the U.S. Department of Health and Human Services and the Ford Foundation. The sample is roughly one and a half times the size of the worksite surveys in the 2012 Family and Medical Leave Act survey funded by DOL (NN=1,812, response rate = 21 percent; Daley et al., 2012). All of those surveys used computer-assisted telephone interviewing, for which response rates have lately dipped below 10 percent.

of administering two employer surveys and negotiating partnerships and delivering incentives to improve response rates.

Geographic scope. This design option seeks to gain a national perspective of the SAW/RTW policies employers use and their likelihood of adopting new ones. A stratified random sample across many states could ensure that any findings apply to a variety of state policy contexts. The only exception to the national geographic scope is the mixed-methods element of Component 1, which is not meant to be nationally representative and could focus on a particular geographic region in order to minimize costs.

5.8 Contributions and Limitations

Policy relevance. Any findings would be of immediate policy relevance to DOL. If a low-cost informational intervention were found to increase SAW/RTW outcomes for injured or ill workers, DOL could mount a large-scale informational campaign to employers. Such a policy could be modeled along the lines of the existing marketing effort for the On-Site Consultation Program at the Occupational Safety and Health Administration (OSHA). For that model, sending marketing brochures to the treated group almost doubled the six-month consultation request rate relative to the comparison group (Juras et al., 2017).

Contribution. Component 1 would generate information on the understanding and existing practices of employers that are not especially active in the SAW/RTW arena. A mixed-method element would provide additional insight into how employers operate, beyond what can be obtained in a simple survey. Component 2 would test the potential of an informational intervention to improve practices and increase knowledge.

Internal and external validity. The Component 1 employer survey would generate results that are representative of those who respond to the survey, and it would be nationally representative with weights used to adjust for nonresponse. The mixed-methods element within Component 1 would not generate rigorous causal estimates, but it would result in detailed descriptive information. Component 2, would use a randomized design, resulting in high internal validity. Component 2 would also be nationally representative when weighted appropriately.

5.9 Potential for Combining Options B and C

Combining Options B and C into a single randomized trial would allow DOL to compare the effects of the employee informational intervention (Option B) with the effects of the employer informational intervention (Option C). While in principle it might be possible to combine these options, we think that the challenges involved make it essentially infeasible. We discuss the challenges here.

The most straightforward way to combine these options would be to join them only at the analysis phase. If some administrative data source (such as the Census' Longitudinal Employer-Household Dynamics [LEHD] data⁴²) were available to provide employee-level outcomes for both randomized trials, then effects could be compared, perhaps after some adjustment (e.g., weighting on employer size) to make the samples as comparable as possible.

Another approach to combine the options would be to join the employee and employer interventions into a single evaluation design. As currently structured, Option B examines individual level outcomes and

⁴² The Census' LEHD dataset is described at <https://lehd.ces.census.gov/>.

Option C examines employer or worksite level outcomes. In order to combine these into a single design, both options would need to collect data on the individual level.⁴³

A single evaluation design might add a second treatment group to Option C. Employers recruited for the study would then be randomly assigned to one of three groups: (1) a treatment group that is provided an employee informational intervention, (2) a treatment group that is provided an employer informational intervention, or (3) a control group with no intervention.

Another way to combine the options into a single evaluation design would be to use a factorial design. Such a design could start with Option C, and conduct a second random assignment within both treatment and control employers to select which employers would offer the employee intervention to workers with recent injuries or illness onset. This would result in four assignment groups: (1) a treatment group that is provided an employee informational intervention, (2) a treatment group that is provided an employer informational intervention, (3) a treatment group that is provided both an employee informational intervention and an employer informational intervention, and (4) a control group with no intervention.

While a single evaluation design would allow for the direct comparison of the impact of the approaches, it has major drawbacks. First, all recruited employers would need to be willing to provide access to individual level information (if not to data from the employer then to individual level identifiers that could be used to match to other data sources). As mentioned above, asking for worker-level data would likely substantially reduce the employer participation rate, which is already the key challenge of evaluation design Option C. Second, the employee intervention would still target only those employees with recent injury or illness onset. Therefore, the evaluation might need to identify such employees in all assignment groups in order to maximize statistical power. Third, performing random assignment for the *employee* intervention at the *employer* level would require more employees than the 1,500 – 2,200 range described in Option B to achieve equivalent statistical power to individual random assignment. All these factors would greatly increase the effort required to conduct the trial.

⁴³ Since the employee informational intervention of Option B is not designed to influence employer practices, it would not make sense to examine effects of the Option B intervention on employer level outcomes.

6. Option D: Test the Effects of Informing Providers about Best Practices in Medicine to Facilitate SAW/RTW

This option addresses the core research questions of how to increase employment and reduce application and receipt of disability benefits. This option would identify occupational medicine best practices, and then estimate the impact of an informational intervention to promote those practices. This option would use an experimental evaluation design to test whether an informational “nudge” changes the behavior of medical professionals to promote SAW/RTW among their patients. Medical professionals would be sent messages about the best practices that would contain one or more recommended activity to encourage continued employment among their patients.

To our knowledge, the effect of an informational “nudge” to medical professionals on workers’ employment is not addressed in existing literature, and is unlikely to be answered as part of ongoing demonstrations such as RETAIN.

This chapter describes the research option in detail. We describe the research questions, conceptual model, and two informational interventions; a “checklist” (labeled intervention *a*) or “advice” (labeled intervention *b*). We suggest two possible evaluation designs to study the effects of these interventions.⁴⁴

- Design D-1 would randomly assign geographic regions to treatment *a*, treatment *b*, or a control group. In the treatment group regions, every medical provider would receive either treatment *a*, or treatment *b*. Medical professionals in the control regions would not receive information. Outcomes would be analyzed at the geographic region level.
- Design D-2 would randomly assign individual medical professionals to treatment *a*, treatment *b*, or a control group. Providers assigned to the treatment groups would receive either treatment *a* or treatment *b*. Medical professionals assigned to the control group would not receive information. Outcomes would be analyzed at the individual level.

Both designs would measure the effects of the interventions on the employment and disability benefit outcomes of the medical professionals’ patients. Both designs would also measure effects on provider practices and behavior. Compared to Design D-2, the geographic design (D-1) minimizes recruitment costs because it would not require individual-level contact and consent. Design D-1 also minimizes the potential for treatment group crossover. Because broad geographic labor market data could measure employment outcomes, Design D-2 also involves lower data collection costs compared to Design D-2. However, the individual random assignment design (D-2) greatly improves statistical power relative to Design D-1, but would be more complex to implement and would require individual-level data to measure outcomes. Challenges to implementing either design include obtaining contact information for medical professionals and obtaining sample sizes large enough to detect relatively small effects. We discuss the participants, data requirements, analysis plan, practical considerations and contributions for each design.

⁴⁴ Variants of Design D-1 and D-2 are also possible. For example, researchers could test three-arm studies with treatments *a*, *b*, and control group, or a two-arm study testing either treatment *a* or *b* and a control group.

6.1 Research Question(s)

This option would address two research questions:

- **Would changes to medical practice promote individuals staying at work or returning to work?**
- **Does an informational nudge induce changes to medical practice?**

This option would measure two types of outcomes:

- Employment and disability benefits for the patients of the medical providers in the study.
- Medical professionals' practices.

To carry out the study, researchers would identify the key activities that experts in occupational medicine believe promote SAW/RTW outcomes and conduct a test of a low-cost intervention to promote those activities. A promising way to identify best practices is a Delphi method that draws on expert guidance in two rounds. Experts on occupational health best practices would provide unstructured responses to researchers in a first round. Then researchers would collate the information and send it back to experts with others' commentary attached for their endorsement or rejection. Researchers would then deem a subset of responses as the widely accepted best practices. This set would be either the universally endorsed responses or the responses endorsed by the majority of the experts.

Literature describes which key activities and medical practices experts in occupational medicine believe promote SAW/RTW outcomes, but there is little evidence to support specific recommendations (Wickizer et al., 1999; Wickizer et al., 2001).⁴⁵ We are not aware of any randomized control trial that measures the impact of specific practices on SAW/RTW outcomes, nor of the effect on SAW/RTW outcomes of advising medical professionals on the use of these practices. Further, we are not aware of any studies that compare using a checklist of best practices to advice to talk to injured or ill workers about the supports they need to return to work, the interventions designated for this evaluation design option.

6.2 Conceptual Model

As described in Section 2.3, about 8 in 10 workers who experience a separation from work due to injury or illness seek medical attention. However, medical professionals who provide care do not have return to work or staying at work as a primary focus (Jurisich et al., 2017; Black, 2012; Denne et al., 2015). As a result, they may provide treatment that does not maximize the likelihood that an individual worker returns to work. Perhaps this is because they are unaware of practices that could help, and perhaps because they are focused on the alleviation of symptoms without regard to impact on work disability.

⁴⁵ Per Wickizer et al. (1999): "The peer-review literature does not present information regarding the efficacy of occupational health centers of excellence. Participants in this project's MD expert panel, however, strongly advocated for the development of centers of excellence."... Guidance provided to medical professionals as part of the Washington Centers for Occupational Health and Education (COHE) model was based on Medical Management Guidelines and Indications and the Return to Work protocols (developed by Johns Hopkins physicians) and the Program to Prevent and Treat Work-Related Musculoskeletal Disorders and The Lead-in-Construction Project (from the Mount Sinai Irving J. Selikoff Center for Occupational Medicine).

6.3 Intervention

This design option would test a low-cost informational intervention that sends messages to medical professionals, using a web- and telephone-based survey. A control group would receive no information. We suggest testing two interventions:

- **Treatment (a)** would receive a list of best practices or “checklist” to encourage treatment that promotes SAW/RTW.⁴⁶ We suggest using a checklist because this format has been shown to have large impacts in other settings—as with avoiding central line infections in intensive care units (e.g., see Pronovost et al., 2006) and improving surgical safety (e.g., see Haynes et al., 2009). These checklists would be organized around types of injury or illness that might develop into a work disability and would suggest treatment options. For example, workers with back injuries would be encouraged to pursue a physical therapy regimen and limit prescription use; workers with either musculoskeletal or mental impairments would be encouraged to participate in a progressive goals attainment program.
- **Treatment (b)** would receive advice to ask patients about what treatment options would impair their ability to work and to suggest to patients that barriers to work could be removed by different treatment options. This advice would take the form of a very simple statement to the medical professionals that talking to patients about return to work can be beneficial, as can offering to help them return to work. For example, the message could quote ACOEM guidelines to say “Unnecessary prolonged work absence [from] work can cause needless but significant harm to a person’s well-being,” which would motivate medical professionals to think of return to work as a health outcome. Medical professionals could also be advised, optionally, that employers or benefits claims administrators may ask medical professionals “precise questions and elicit particular language that later becomes the basis for benefit, claim, or employment determinations,” but “nonadversarial participation by impartial physicians may be helpful” in promoting return to work (ACOEM 2006) This optional additional language could provide an impetus for medical professionals to offer to patients any needed help in communicating with employers. Treatment *b* medical professionals would not receive the checklist distributed to treatment *a*.
- **Control group (c)** would receive no communication from researchers.

The intervention(s) would be designed to reach workers at an early point after injury or illness, when they first seek medical attention. The potential impact of this approach rests on findings that 80 percent of workers seek medical attention when they experience an injury or illness, whereas only small fractions interact with other potential touchpoints (Nichols et al., 2020b).

These types of interventions are similar to some of the practices of the Washington Centers for Occupational Health and Education (COHE) model that provide training for physicians on occupational health best practices. Thus, they may overlap with parts of RETAIN. They also differ from the COHE model in two ways. First, in the COHE model, physicians are incentivized to undertake these best practices because they are able to bill for added services, and thus they receive financial compensation for adopting them. Second, in the COHE model, training in occupational health best practices is part of a large bundle of interventions that also includes new financial incentives and substantial changes in the

⁴⁶ Researchers might also choose to generate several variations on the checklist, in order to test the effects of different messages. We do not describe designs with additional treatment arms further, except we note that to conduct a multiple-arm trial, each of the treatment *a* communications would contain a randomly selected subset of best practices. The results of a multiple-arm trial could identify specific items on the checklist that are more effective, but the study design would require larger sample sizes.

health care delivery system. In contrast, in evaluation design Option D, the low-cost informational intervention is not bundled with other interventions or incentives. By separating out the informational component, DOL could measure the impact of just the “nudge.” This strategy complements what policymakers can learn from RETAIN and is also distinct from the COHE program currently operating in Washington State.

6.4 Participants

Target population. This option (with either Design D-1 or D-2) targets medical professionals (physicians, physicians’ assistants, nurses, and the like) that most individual workers who have experienced an injury or illness will visit near the time of injury or illness. For Design D-2, researchers also might consider recruiting individual patients of the medical professionals in order to measure the providers’ patients’ outcomes.

Organizations and Agencies. Ideally, the Federation of State Medical Boards (Young et al., 2017) or the American Medical Association (Robinson et al., 2002) could provide contact information for all medical professionals in the United States who are members, but the availability of national sample frames is not guaranteed. The study could also use data from individual state licensing agencies for medical professionals, in order to identify a sample frame of medical professionals to be assigned the intervention.

Licensing agencies for physicians, physicians’ assistants, and other medical professionals have served as sample frames for other studies of medical professionals. Studies often report using a licensing board to find eligible clinicians; for example, in Alabama (MacLennan et al., 2014), Colorado (Robinson et al., 2002), and Oregon (Cox et al., 2012). These sample frames would likely have to be collected from each state individually, over as short a period of time as possible to reduce the chance of migration across frames. Other national medical professional organizations include the American Occupational Therapy Association and the American Physical Therapy Association which represents more than 100,000 member physical therapists, physical therapist assistants, and students of physical therapy.

Setting/venue. We discuss the setting for this option in the next section.

Recruitment plan (if applicable). Design D-1 would use geographic randomization so that all providers in an area receive a consistent message. This design might help reinforce the message and would minimize contamination. In Design D-1, we propose to randomize across the 709 “commuting zones” (groups of counties with minimal labor market overlap between them) identified in the 2000 Census. Because it would not require researchers to recruit individual providers, randomization by geography would be much lower cost, and would impose lower burden on individuals than would randomization by individual. Randomizing by geography would still require obtaining contact information for medical professionals, but these might be obtained from state licensing boards. Compared to Design D-2, randomization by geography would have much lower power at the same overall sample size.

Design D-2 would involve sending messages to individual medical professionals. This would require recruiting a large medical system, an insurer, or a national professional organization (e.g., of physicians or occupational or physical therapists) to obtain contact information and to help with outreach to improve response and consent rates. Ideally, the organization(s) recruited would agree to directly provide some of its associated medical professionals with the treatment *a* checklist and encourage its use (and to provide the advice in treatment *b*). This design would increase the salience of the intervention, as it would come from an authority in the eyes of the medical professionals, and so might increase impact.

Compared to Design D-1, D-2 would involve much higher costs for recruitment because of the need to contact potential providers, explain the study, obtain consent, and conduct random assignment. Individual random assignment might also increase the potential for spillover from treated medical professionals onto

untreated professionals. Spillover could occur if medical professionals in the same practice share information and/or patients, and if one provider in a practice was assigned to the treatment group and others to the control group this could increase the potential for contamination. Randomizing groups of medical professionals by geographic area, as in Design D-1, avoids this problem, though it requires much larger sample sizes to achieve the same power, due to clustering by geography.

6.5 Data

Data needed/data collection method(s). In this section we describe the data needed for randomization and for measuring outcomes, then describe implications for minimum detectable impacts and practical considerations.

6.5.1 Data for Randomization

Design D-1. To conduct the geographic randomization, this design would require a list of medical professionals sorted by geographic area. To implement the design requires choosing a geography and fraction assigned to treatment. We assume the aggregate statistics on employment and disability outcomes are measured at a lower level than the level of randomization. For example, the research team could randomize treatment status by commuting zone (large groups of counties), but collect data on employment and disability outcomes by county. Many commuting zones overlap multiple states and there are many possible data sources for different types of medical professionals, so collecting this list of individuals to contact is a large data collection exercise.

Design D-2. To randomize individual medical professionals, this design would also require a list of individual medical professionals. The major barrier would be obtaining an adequate sample frame. Even assuming the cooperation of a national medical professional organization, their list of members is not representative of all medical professionals in the nation, and contact information may not be reliable. Some selectivity of the sample will necessarily be introduced. For example, the American Medical Association has more than 200,000 members, but many of them are students or retired physicians, and there are nearly a million physicians in the country.⁴⁷ It might be preferable to recruit medical professionals through affiliated institutions, such as clinics and hospitals, but this makes sampling and weighting much more difficult.

6.5.2 Data for Measuring Employment and Disability Benefit Outcomes of the Patients of Medical Professionals

Design D-1. To measure employment outcomes of the medical professionals' patients in the geographic randomization design, data sources would depend on the size of the geographies. If the geography selected corresponds to data collection undertaken by Census or BLS, employment outcomes could be measured using data already being collected periodically by the government. Similarly, if the geography selected corresponds to data collection already undertaken by SSA, applications for and receipt of federal disability benefits could be measured using data already being collected by the government.

Design D-2: To measure employment and disability benefit outcomes of the medical professionals' patients, Design D-2 would require individual patient data. To obtain these data, researchers could recruit individual patients of the medical professionals in the treatment group, and obtain consent to collect data

⁴⁷ Per Collier (2011), about 15 percent of physicians are member of the AMA, and membership declines are due in part to competition from the American College of Physicians, the American College of Surgeons, the American Academy of Family Physicians, and other "state and specialty medical organizations." Currently active physicians number over one million per KFF [<https://www.kff.org/other/state-indicator/total-active-physicians/>] which cites as a source a data request for information on active state licensed physicians from Redi-Data, Inc, March 2020.

on employment, and applications for and receipt of federal disability benefits. The individual-level data could be matched to employment and disability records already collected by the government, with appropriate consent and data use agreements in place. If researchers were to pursue this design and recruit individual medical professionals and their patients, then it would be valuable to consider additional data collection to answer ancillary research questions on health care use.

6.5.3 Data for Measuring Medical Professionals Practices

To measure provider outcomes, in either design D-1 or D-2, DOL may want to conduct a survey of medical professionals. DOL could use the survey to test whether provider practice and behaviors change after receiving the information. Response rates from surveys of medical professionals are low. Cho et al. (2013) show average response rates declining from 80 percent to 40 percent between 1958 and 2012 with a steady downward trend, and suggest mail surveys with financial incentives and multiple follow-ups tend to perform better. Such surveys would substantially add to the cost of the study, because both the survey and sampling plan would require additional development time, and achieving high response rates would likely require multiple follow-ups and financial incentives for participation. However, mounting such a survey would allow DOL to determine the ways in which medical professionals respond to advice and potentially indicate which practices that are associated with improved labor market outcomes. A series of focus groups in both treatment groups and the control group could provide insights about the ways in which providers respond to the informational intervention, without adding the cost of a large-scale follow-up survey.

6.5.4 Minimum Detectable Impacts

Design D-1. For Design D-1, if the research team randomizes treatment status by commuting zone (large groups of counties), but collects data on employment and disability outcomes by county, D-1 is a clustered design even with aggregate data on outcomes. Choosing 20 commuting zones to be in each treatment group and 669 in the control group,⁴⁸ we estimate a minimal detectable impact for this design of 10 percentage points (e.g., an increase in employment rate from 33 to 43 percent), with an assumed intracluster correlation of 10 percent. The more treatment commuting zones, the smaller⁴⁹ is the minimum detectable impact, but the higher the cost of delivering the treatment. With an intracluster correlation of 5 percent, the minimal detectable impact would be 7.2 percentage points with 20 treatment clusters, 5.3 percentage points with 40 treatment clusters, and 4.1 percentage points with 80 treatment clusters. In general, the number of clusters in the treatment group strongly affects the detectable impact, and the number of cases per cluster only weakly affects the detectable impact, unless the intracluster correlation approaches zero.

⁴⁸ Because the optimal sampling rate depends on the cost of each treatment, the control group should constitute a larger fraction of all 709 commuting zones than does either treatment group. If we assume an optimal allocation where the sample sizes are proportional to the square root of variance divided by cost, and a cost of data collection in treatment areas about 1,000 times the cost in a control area, with variances comparable, there should be about 20 commuting zones in each treatment group, leaving 669 in the control group. If researchers sample commuting zones at rates proportional to size, they would still get large populations in each group. In one test randomization, we assigned 20 commuting zones to treatment group 1 containing 125 counties with 12 percent of the total U.S. population, and we assigned 20 commuting zones to treatment group 2 containing 107 counties with more than 10 percent of the total population.

⁴⁹ Doubling the number of commuting zones to 40 in each treatment group, with 629 in the control group, would reduce the minimal detectable impact to 7.3 percentage points (e.g., an increase in employment rate from 33 to 40.3 percent). Doubling again the number of treatment commuting zones (to 80 in each group), with 549 control commuting zones, would reduce the minimal detectable impact to 5.5 percentage points (e.g., an increase in employment rate from 33 to 38.5 percent).

Design D-2. For Design D-2, we assume a simple random sample from a given frame, such as a list of physicians, even if that sample is not perfectly representative. Further assume that a research team sampled one patient per physician and obtain employment outcomes over subsequent months via a match to administrative data (so there is no attrition and no reporting error). Under that assumption, supposing researchers randomly assign 35,000 physicians to each treatment, the minimum detectable increase in employment among their patients is one percentage point (from 33 to 34 percent). This is roughly one-seventh the size of the detectable impact in the first design, where randomization is done at broad geographies. But for this sample size to be attainable, we assume no consent or recruitment is needed from the sample frame, i.e. all member of the frame can be randomly assigned. To obtain a minimum detectable impact of six percentage points (an increase in employment from 33 to 39 percent), only one thousand physicians would need to be randomized to each group. This large gap in required sample sizes (or detectable impacts) highlights the tremendous advantage of individual random assignment. If individual-level recruitment and consent is needed for physicians, the much smaller sample size of a thousand per group might be preferred on cost and logistical grounds, and the gap in detectable impacts between the designs would shrink.

For Design D-1, the minimum detectable effects are large, and they vary far less by the number of individual people in each group rather than the number of clusters randomized. If a single confirmatory impact is designated for each research question, limiting attention to a more responsive subgroup of individuals affected by the behavior of medical professionals would maximize the chance of detecting impacts. Option A findings would be useful in designating the target group, but we suggest limiting attention to individuals without a college degree, older than age 47, and/or who receive public assistance (all factors associated with higher chances of exit from the labor force, per the analysis described in Section 2.3), if a single confirmatory impact is specified for a design like D-1.

The tradeoff here is giving up generalizability to the whole population versus gaining a higher likelihood of detectable effects. Mathematically, limiting to a 10th or even a 20th of the total sample would cost little in terms of power if the likely effects are larger in that smaller group. Practically, however, the smaller sample is simply not the whole population of interest. That is, suppose individuals under age 47 have approximately zero change in employment rates due to induced changes in the behavior of medical professionals and the impact over age 47 averages 12 percentage points. We can detect the latter impact, per our preliminary calculations, but could not detect an overall impact in the pooled groups. If the research plan did not require specifying a single confirmatory impact for each research question, this concern dissipates.

In sum, in Design D-1, where individual medical professionals are selected for *different messages targeted by broad geography* (e.g., the 709 commuting zones), then only aggregate labor market outcome data need to be collected to determine whether the messages affect labor market outcomes. The tradeoff here is clear: with individual-level outcome data in Design D-2, the power to detect small differences is much higher, but it seems difficult at best to obtain the requisite permissions to match individual patient data and labor market outcomes. So the individual-level design D-2 is more powerful, but more challenging and costly.

As discussed under Option A, the minimum detectable impact of interest is the smallest impact that would justify a policy change. If effects on work are smaller than a few percentage points, it might be that an informational component alone cannot successfully compete against one of the comprehensive interventions to be studied as part of RETAIN. Thus, we anticipate that a minimum detectable impact of 5 or 6 percentage points would still be small enough to justify pursuing this study.

IRB, security, logistical concerns. In Design D-1 if only publicly available data are used to construct outcome measures, data security is straightforward, with no disclosure risk. There are still substantial logistical concerns in obtaining the needed sample frames to deliver the intervention to medical

professionals in the treatment group geographies. For either Design D-1 or D-2, securing permission to contact members of professional societies is likely to require substantial investments of time, and the participating organizations may also require input into the design of messaging, or even measurement of outcomes, which would have implications for the study's timeline.

6.6 Analysis Plan

Outcomes and hypotheses. The primary outcome of interest is employment and application for and receipt of disability benefits for patients seen by medical professionals in the study, in both treatment and control arms. DOL might also investigate the effect of intervention(s) on the length of spells out of the workforce due to illness or injury. One hypothesis is that in areas where medical professionals receive checklists of occupational health best practices, more injured or ill workers will stay at work or return to work and employment will be higher over subsequent years, whereas disability benefit application and receipt rates will be lower. If multiple treatment arms were used, a second hypothesis would be that where medical professionals receive advice to ask patients about treatment options that would help them stay at or return to work, more injured or ill workers will do so, and employment rates will be higher over subsequent years in the treatment group *b* areas than in the control group, but lower than in the treatment group *a* group.

Model(s) or analysis methods. Evaluation design Option D would require only straightforward analysis of the alternative messages using standard regression methods for an experimental design, with a cluster-robust standard error calculation for the first design. If randomization (of clusters in the design D-1, or individual medical professionals in design D-2) were conditioned on many factors, such as population, expected underlying injury/illness rates, disability application rates, age structure, and economic conditions, a simple comparison of means can be estimated. This comparison of means can be viewed as a regression of outcomes y_i (y for outcome measure, i for unit of analysis) on a single indicator A of treatment a with the excluded category in the control arm (Equation 1), or for a three-arm design, on two indicators A and B of treatment a and treatment b (Equation 2).

$$y_i = \delta + \alpha A_i + \varepsilon_i \quad (\text{Equation 1})$$

$$y_i = \delta + \alpha A_i + \beta B_i + \varepsilon_i \quad (\text{Equation 2})$$

Covariates can be included in the model if randomization were not conditional, to improve precision (Lin, 2013); this flexibility is one advantage of formulating a comparison of means as a regression. Other advantages include flexibility in standard error calculations and hypothesis testing.

If several dimensions of advice can be identified from expert recommendations, the messages could be designed as a factorial experiment.⁵⁰ That is, if advice seems to come in three different types in each of two different dimensions, there are nine (three types in one dimension multiplied by three types in the other) intervention options to be tested, and a factorial design can measure the advantage of each type of advice using fewer degrees of freedom. We assume a single treatment a checklist, but this variation should be revisited based on the results of the checklist construction process. Even with a factorial design, introducing any additional treatment variation would increase the minimum detectable impact. Effectively, introducing the factorial design would be similar to moving from a two-arm design to a three-arm design as above, and would complicate analysis in order to address additional research questions. This would be accompanied by needs to address multiple comparisons (and adjustments to p-values), and larger sample size needs for the same power (Grayling and Wason, 2020). A design that adapts over time

⁵⁰ In a factorial design, different components of the intervention would be randomly assigned separately. An individual might be in the treatment group for one component but in the control group for another (or any combination of those).

to incoming findings is also possible, using response-adaptive randomization, or arm dropping or early stopping for efficacy or futility (if early results show dramatic differences, one arm can be dropped from further recruitment or data collection); see Ryan et al. (2020) for one recent example.

Expected methodological issues and responses. Because the optimal sampling rate depends on the cost of intervention, the control group should constitute a larger fraction of the sample (i.e. all 709 commuting zones, or a sample frame list of medical professionals) than should the treatment group(s). Small numbers of treatment clusters results in much larger minimum detectable impacts in the first design, however. Aggregating public use data sources will capture much of the month-to-month and year-to-year variation in labor markets that constitutes background noise. This variation reduces the chance of detecting an impact of the treatment. It also could involve narrowing the research focus to a group in which we are more likely to observe a response, as described above. Administrative data and individual randomization achieves much greater statistical power (much smaller minimum detectable impacts), but at higher cost.

6.7 Practical Considerations

Implementation timeline. We estimate that conducting this evaluation option would take at least five years. We believe it might take approximately one year to design the study, including a survey or focus group of occupational medicine experts and collection of a sample frame of medical professionals. Messages might be delivered to those professionals for approximately one year, and then outcomes collected for approximately two years following the intervention. The final year would be taken up by analyzing, reviewing, and writing up results.

Cost. This study would require a research team to design and implement a new intervention. The first step would be to consult with experts to determine occupational medicine best practices. The study would then require work to develop the specific intervention—the checklists and message content to be tested and a plan for disseminating the information. Researchers would need to identify a sample of medical professionals for the study and a plan for randomizing by geographic area. Another cost component would be establishing partnerships with the Federation of State Medical Boards, the American Medical Association, or state licensing boards to obtain contact information for eligible participants. If a national frame is not available, the study could partner with state licensing agencies for medical professionals; negotiating data use agreements each state agency individually could increase expenses. After recruiting the medical professionals, providing financial incentives to encourage their participation could be a significant cost of this evaluation design option. As described above, Cho et al. (2013) find declining participation by medical professionals is only partly offset by financial participation incentives.

We assume that to measure outcomes for the medical professionals' patients, either the study would involve linking to existing data (public use or administrative), or DOL might consider developing and administering a new survey. If DOL chose to conduct a survey of medical professionals, to measure provider outcomes, that choice would substantially add to the cost of the study. Adding a survey component Design D-1 would involve constructing sample frames for each commuting zone in the control group and designing and fielding the survey. Collecting data from individual physicians and patients would increase the costs of this option. We anticipate data collection costs could be reduced by using aggregate labor market outcome data and public-use survey data.

Geographic scope. This study could be designed to be representative of the entire U.S. population of medical professionals, by drawing an appropriate sample of geographic areas and obtaining lists of contact information for all licensed medical professionals in design D-1, or a national sample frame for D-2. However, these lists may not be complete, so the representativeness of the study will only be as good as the representativeness of the sample frames obtained for the study, in general. Because we expect state licensing boards to maintain updated information on medical professionals authorized to operate in states, we expect design D-1 to have a high degree of representativeness, assuming all state lists can be obtained.

Threats to implementation as designed. The major threat to implementation is obtaining high-quality sample frame data for medical professionals practicing in multiple states, with up-to-date contact information. We anticipate substantial effort should be allocated to hand-checking contact information, calling or otherwise researching specific medical professionals (at least for a validation sample), and negotiating sharing of contact lists from licensing boards or professional membership organizations.

Alternatively, a representative sample of the population could be used to identify medical professionals who could then be recruited to receive different types of messages from the study. However, this sample would need to be extremely large. By way of comparison, the 2008 Survey of Income and Program Participation (SIPP) in wave 1, month 4, has 105,303 individuals, including only 189 physicians, 15 physician assistants, 63 physical therapists, 33 physical therapist assistants, 29 occupational therapists, 14 occupational therapist assistants, and 998 nurses. Assuming all medical professionals are recruited (not just physicians, physician assistants, and occupational therapists), the initial sample would have to be several times as large as the SIPP to obtain even three thousand medical professionals and a six percentage point minimum detectable impact. Only the American Community Survey is large enough to support an approach based on recruiting from a representative sample of the population, and we are not aware of any instance where it was used to recruit study participants in this way.

6.8 Contributions and Limitations

Policy relevance. The policy relevance is immediate, as the findings would be of interest to medical professionals and government agencies. If large impacts of a simple checklist were detected, medical practice could rapidly change, as happened in surgical and intensive care units (Pronovost et al., 2006; Haynes et al., 2009). Alternatively, if a checklist has a small or null impact, the importance of altering incentives of medical professionals would become more salient for future policy.

Contribution. We think this evaluation option would produce evidence in three areas. First, DOL could learn what occupational medicine experts believe is sound advice for medical professionals to give injured and ill workers to improve their SAW/RTW outcomes. Second, policymakers could learn whether medical professionals change their practice in response to messaging about best practice. Third, policymakers would measure the impact of any such change in practice on patient return to work outcomes and application for federal disability benefits. This study would fill gaps in current knowledge about the role that medical professionals play in the SAW/RTW outcomes of workers who experience a potentially disabling injury or illness.

Internal and external validity. This study would have high internal validity, in that it would provide rigorous evidence of effectiveness of specific strategies for the targeted populations. It would also have high external validity, if it is representative of medical professionals who see individual workers who experience an injury or illness. However, as we note above, the representativeness of the study will only be as good as the representativeness of the sample frames obtained for the study, and this is a major threat to implementation of this option.

7. Option E: Partial Payments in Temporary Disability Insurance Programs

This option addresses the core research questions of how to increase employment and reduce application and receipt of federal disability benefits. The option would use an experimental design to examine the effects of temporary disability benefits on return to work and SSDI application. In particular, it would look at partial-payment provisions within Temporary Disability Insurance (TDI).

Partial-payment provisions allow workers to continue to receive part of their TDI benefit if they are able to return to work but at a lower number of hours or lower wages. We outline a research design that would involve implementing a new partial-payment provision in an existing state or private TDI program. The study would randomly assign TDI users to eligibility for the benefit. Then it would collect data on program operations, program costs, and worker outcomes. We also suggest two alternatives to this design: (1) conducting an evaluation of existing partial-payment benefit programs using existing data and a quasi-experimental analysis; and (2) developing, implementing, and testing the effect of a new TDI program designed to incentivize and support SAW/RTW. These alternatives vary in the level of resources, types of partnerships, and precise policies analyzed, but both would address the question of how the availability of partial payment affects return to work.

Workers whose injury or illness is not work-related may receive benefits from short-term disability insurance, also known as temporary disability insurance (TDI). Five states (California, Hawaii, New Jersey, New York, and Rhode Island) and Puerto Rico offer TDI or require employers to provide it. Family and Medical Leave programs passed by Massachusetts, Washington, and the District of Columbia in recent years also provide TDI.⁵¹ Workers in other states may be eligible for TDI benefits through their employer, or can purchase coverage individually. Roughly 40 percent of private-sector workers have TDI coverage from some source (DOL/Bureau of Labor Statistics, 2019).

When workers are able to return to work, some may wish to do so gradually or begin with modified work tasks that do not pay as well as previous work. This kind of partial return to work both allows workers to maintain a connection to their job and employer while they are not yet able to make a full return and allows them to “test” their ability to return to work. These benefits can either increase or decrease program costs, depending on whether workers who receive partial payments would receive the full benefit or no benefit in the absence of the partial payment provision. Most workers’ compensation programs accommodate this kind of gradual return by offering partial benefits for workers who return to work but have not yet returned to their previous level of earnings (Ashley et al., 2017). These provisions also exist

⁵¹ In Hawaii and New York, nonexempt employers are required to provide state TDI benefits for their employees by either purchasing private insurance or self-insuring. California, New Jersey, and Rhode Island operate public state TDI programs that cover the vast majority of the workforce. State-provided and state-mandated short-term disability insurance programs cover workers for medical conditions that are not due to a workplace event. Where state requirements are lacking, employers can choose to provide TDI to their employees. Rhode Island first paid benefits in 1943, California in 1947, New Jersey in 1949, New York in 1950, Hawaii in 1970 (McLaren and Scherur, 2019).

in some state temporary disability programs, but are far from universal.⁵² Twenty eight states and the District of Columbia offer partial benefits, known as “Work Sharing,” as part of their UI programs.⁵³

Design E-1 under evaluation design Option E would test the effect of adding a partial-payment provision to a TDI program on employment outcomes using a random assignment design,⁵⁴ as we do not know of any settings in the U.S.⁵⁵ that would allow measuring such effects in a quasi-experimental design aside

⁵² Rhode Island’s TDI Partial Return to Work Program is explicitly designed for this purpose (Bourbonniere & Mann, 2018). California’s program provides support for returning employees working shortened schedules, but not those working in lower-paying tasks (State of California Employment Development Department, n.d.). Family and Medical Leave laws in Washington State and the District of Columbia allow for intermittent work, but not partial days of work: The District requires that leave be taken in units of days (DC Law 21-264 § 101.9, 2017); Washington State does not provide benefits for days on which a person worked for pay (Washington Substitute Senate Bill 5975, §5d, 2017.) Massachusetts, however, allows for partial payments in a similar framework to that used by Rhode Island (Mass Gen Laws ch.175M § 3c, 2018).

⁵³ These benefits are tied to a slowdown in economic activity necessitating layoffs, not to an individual worker’s injury or illness, but could also serve as a platform for developing a new TDI option. Rhode Island’s program is described at <http://www.dlt.ri.gov/ui/ws.htm>. The senior Senator from Rhode Island, Jack Reed, also served on the bipartisan taskforce that negotiated provisions of the Coronavirus Aid, Relief, and Economic Security (CARES) Act. The CARES Act provides \$100 million through Reed’s short-time compensation (STC) provision for work sharing, and “includes 100 percent federal funding of work share programs for states with programs already in place, and 50 percent federal funding of work share programs for states that work with the U.S. Department of Labor to develop a new work-share plan”

[<https://www.reed.senate.gov/news/releases/from-1200-stimulus-payments-to-enhanced-unemployment-to-500-per-child-reed-outlines-coronavirus-economic-rescue-benefits-for-ri-ers>]. Further, the “Pandemic Unemployment Assistance program expands unemployment insurance to cover individuals who are not currently covered by traditional unemployment assistance, including: Individuals who are unable to work because of coronavirus,” including due to illness [<https://www.reed.senate.gov/news/releases/-us-senate-passes-22-trillion-emergency-coronavirus-economic-rescue-agreement>]. We do not recommend studying the expansions of programs occurring in 2020, however, as any quasi-experimental design would capture not just the effect of a new program, but also the health and economic conditions that accompanied the new program. The large expansions of existing programs undertaken in 2008 and 2020 often accompany a crisis, but can also serve to highlight new models that should be tested in a future demonstration, before widespread adoption.

⁵⁴ We outline this design option as a randomized controlled trial because such designs have a high degree of internal validity and have the potential to receive high ratings under review standards such as DOL’s CLEAR. Different designs, using quasi-experimental methods, would also be possible, but would generally lack the high degree of internal validity offered by randomization. Nichols (2007) classifies these methods into different groups, including regression adjustment/matching/reweighting, panel methods including difference-in-difference designs, instrumental variables, and regression discontinuity, ranging from lower to higher expected internal validity. The internal validity of these designs depends crucially on finding a plausible “natural experiment” where participation in programs is driven by known factors, and we have not found an instance of a setting where TDI effects could be measured using instrumental variables or regression discontinuity in the U.S. Design E-2 is a difference-in-difference design, which can incorporate regression adjustment/matching/reweighting as well.

⁵⁵ A Canadian policy of short-term private TDI was studied by Stepner (2019) using a difference-in-difference design. These programs operate in a different setting (“all Canadian workers have short-term disability insurance, with a mix of public and private provision”), and are more generous than the typical program in the U.S. (“the average employer-provided plan had a replacement rate of 70” percent). Autor and Duggan (2010) proposed that universal private short-term disability insurance would result in more workers with work limitations receiving assistance and returning to work, reducing long-term disability rates and expenditures. Stepner (2019) claims that the opposite is true, and that expanding private short-term disability insurance would increase public long-term disability rolls and spending. One unique feature of the Canadian program landscape is that there is a gap in time between short-term disability insurance of several months, which the program Stepner studied eliminates and he is “unable to isolate the effect of this specific mechanism.”

from the one proposed as Design E-2. The partial-payment provision would be broadly modeled after one adopted by Rhode Island in 2006. Two alternative designs would evaluate the effect of Rhode Island's partial-payment option using a nonexperimental design (Design E-2) and a new TDI program explicitly designed to facilitate return to work (Design E-3), respectively. We do not anticipate that any state will propose to test a TDI partial-payment option under RETAIN or any other ongoing demonstrations. Major challenges to implementing the main design, Design E-1, include recruiting sufficient sample and obtaining participation of necessary partners (e.g., a state agency, multiple states, or large private entities).

7.1 Research Question(s)

What effect do partial payments in TDI, and TDI in general, have on work, benefits, and program costs? Specific outcomes might include:

- Timing of return to work.
- Timing of return to the hours and earnings level experienced before the injury or illness.
- Likelihood of returning to work.
- Stability of work or long-term work and earnings.
- Application for and receipt of SSDI/SSI in the long term.
- Cost of providing TDI.

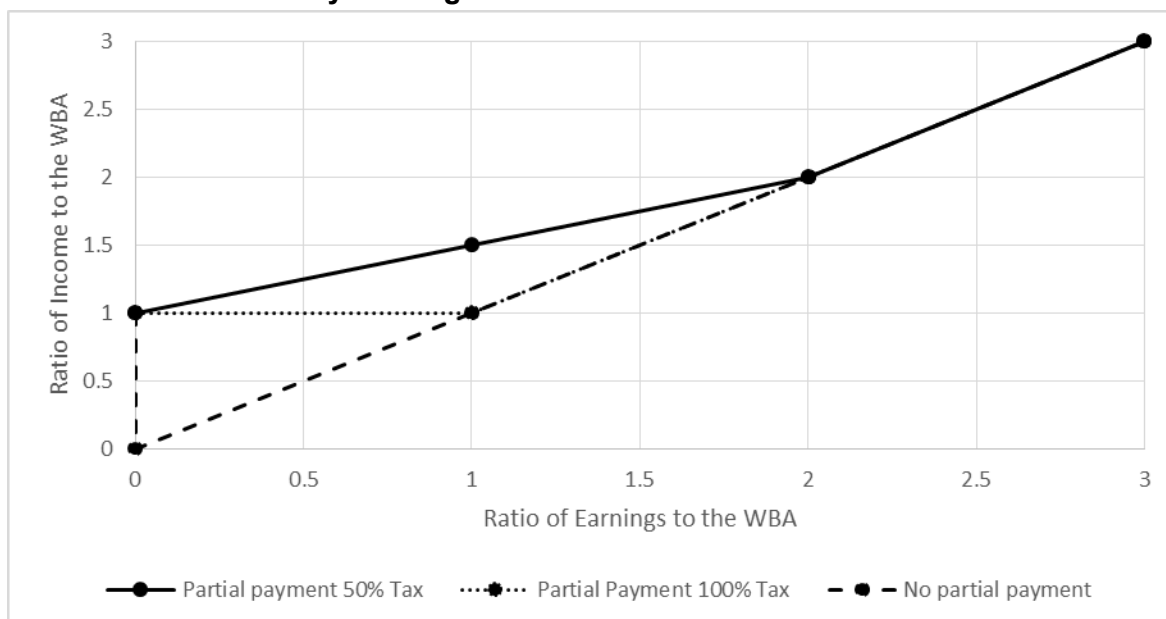
7.2 Conceptual Model

Partial payments in TDI have been suggested as a potential mechanism for promoting ongoing workforce attachment and transitions back to work, by providing incentives that encourage workers who are able to work, but at low levels of earnings, to do so (Bourbonniere & Mann, 2018).

Consider a worker who experiences an injury or the onset of an illness and receives TDI while recovering. TDI benefits are generally calculated as a percentage of pre-injury wages, often up to some maximum value. Among private-sector employees in 2019, the median salary replacement is about 60 percent, and three-quarters had a maximum benefit amount, with a median value of \$637 weekly (US DOL/Bureau of Labor Statistics, 2019). State TDI programs each use their own benefit calculations and offer levels of wage replacement roughly similar to the national average for private-sector workers (McLaren & Scherur, 2019).

When that worker is ready to resume work, they may find themselves in one of three general situations. Each is shown in Exhibit 7-1. The horizontal axis describes earnings in terms of the ratio of earnings to the worker's weekly benefit amount (WBA), the maximum weekly TDI payment to which the worker is entitled if they are not able to work. The vertical axis describes income, including TDI payments and earnings, also as a ratio to the WBA. Actual benefit formulas are, in general, more complicated than those below, and might take into account whether earnings are more or less than pre-injury earnings, or the specific loss of earnings. These are presented as simple examples for exposition.

Exhibit 7-1. Income by Earnings under TDI



Notes: This figure displays total income from TDI benefits and earnings, at different levels of earnings in terms of the WBA, under three benefit structures. Under no partial payment, marked with the dashed line, individuals either receive their full WBA when not working, or no benefits. Under partial payments with a 100% tax, marked with the dotted line, individuals who earn less than the WBA receive a benefit equal to the difference between their earnings and the WBA. Under partial payments with a 50% tax, marked with the solid line, the benefit is reduced by \$1 for every \$2 earned, until the benefit is phased out at twice the WBA.

- In the first situation, there is no partial payment provision.*** This situation is illustrated with the dashed line. A worker who is unable to work has earnings of zero, and collects his or her WBA, but a worker who works at all receives no benefit. A worker who is able to do some work, but at an earnings level below that of his or her WBA, has a lower income while working than if he or she did not work and continued to receive TDI. Many of the workers in such a situation who would like to work at a level between zero and their WBA will choose not to, as doing so would make them strictly worse off in the short term.⁵⁶
- In the second situation, there is a partial payment provision with a 100 percent effective tax.*** This is illustrated by the dotted line in Exhibit 7-1. Here a worker who has no earnings receives the WBA. During some form of transitional phase, while the worker is able to work but is earning less than the WBA, TDI supplements his or her earnings by making up the difference between those earnings and the WBA. In this scenario, earnings are taxed at an effective 100 percent rate, so that income is flat for all those earning less than the weekly benefit amount. Workers who earn more than their WBA receive no benefit, as they would with no partial payments.
- In the third situation, there is a partial payment provision with a 50 percent effective tax.*** In other contexts, such as Unemployment Insurance, effective taxation rates are often lower, on the order of 50 percent. This situation is illustrated by the solid line in Exhibit 7-1. Workers have an incentive to return to part-time work when able to do so, as they have higher income when working part time and receiving partial benefits than when not working and receiving the full benefit. Depending on the specifics of the plan, benefits may drop to zero when earnings reach

⁵⁶ Workers may choose to work anyway, if they enjoy work or are forward looking and believe that doing so is sufficiently beneficial to long-term employment and earnings.

the WBA or some other threshold (e.g., 80 percent of pre-injury earnings), rather than decreasing at the effective taxation rate until they reach zero; this feature is not presented on the graph above.

Partial payments are an incentive to take temporary or lower-paying jobs—by removing the strong disincentive created by benefits dropping to zero with any work in a system without partial payments.⁵⁷ This may lead to more substantial employment in the future or at least allow the worker to maintain an attachment to the labor force. In the TDI setting, such jobs may include assignments to light-duty work or shortened schedules at the existing employer. These jobs might be temporary, to allow for recovery, or transitional, to allow workers to learn new roles that fit their abilities. Maintaining connections to the previous employer and the workforce may be particularly valuable for TDI users, who might otherwise struggle to do so due to their inability to resume previous job tasks and hours. However, it is also possible for workers who could return fully to restrain the amount they work, particularly if they are able to find a medical professional willing to certify them as unable to return to full work status.

In the first 20 months of Rhode Island’s partial-payment RTW program, 1,023 workers used it, and their payments were \$976,701 less than if they had received their full WBA instead (Rhode Island Department of Labor and Training, 2007). However, it is unclear whether, in the absence of the program, those workers would have remained out of work and received their full WBA or returned to work full-time and received no benefit.

Although relatively new as a part of TDI, partial payments have a long history in Unemployment Insurance benefits (e.g., Munts, 1970) and are extremely common in workers’ compensation programs. In UI, partial payments supplement the earnings of workers who are eligible for Unemployment Insurance and have some earnings, but are earning at relatively low levels. These designs attempt to balance the insurance function of these programs—providing income support to those without a job—with incentives to return to work.

Partial payments are also a key component of the SSA’s Benefit Offset National Demonstration. In the Benefit Offset National Demonstration, participating SSDI beneficiaries lost \$1 in benefits for every \$2 earned above the annualized “substantial gainful activity” level, rather than losing all benefits when earning at that level after completing the nine-month trial work period. Interim analyses found evidence that the incentives in the Benefit Offset National Demonstration led to small increases in employment and in the percentage of participants earning above the annualized substantial gainful activity threshold, but not in the level of annual earnings. However, the amount of benefits paid out also increased, suggesting that allowing SSDI beneficiaries to keep \$1 of every \$2 earned will not improve program finances (Gubits et al., 2018; Hoffman et al., 2017; Croake et al., 2017). The lack of an effect on annual earnings, and the patterns in earnings that produced the increased benefit payments, may reflect the fact that few SSDI beneficiaries return to work, and those who would like to face many barriers to doing so. The Promoting Opportunity Demonstration also tests the effect of a \$1 for \$2 benefit offset model, but uses a lower disregard level and different work incentive rules (Promoting Opportunity Demonstration, 2018).

7.3 Intervention(s)

In Option E, we describe one main design, E-1, as well as two alternative designs, E-2 and E-3. Design E-1 would test the effect of adding a partial-payment provision to an existing state or private TDI system on employment outcomes and program costs. Design E-2 would test the effect of Rhode Island’s

⁵⁷ There is evidence on both short-run response and longer-run effects of various incentives that subsidize wages, e.g. from the Earned Income Tax Credit, summarized by Nichols and Rothstein (2016). But the incentive described here differs substantially from those incentives, and is more akin to findings from the Unemployment Insurance literature.

existing partial-payment provision, implemented in 2006. Design E-3 would involve the design of a new TDI program, crafted to incentivize and support return to work.

7.3.1 Intervention Details

In Designs E-1 and E-2, the intervention would be the addition of a partial payment provision to an existing TDI program. In the case of E-2, the provision would be that adopted by Rhode Island in 2006. In E-1, DOL would collaborate with the administrator of the existing program (a state or private insurer) to determine the specifics of the provision. This flexibility might be particularly important if working with a state, as program details might need to be set through legislation.

In Design E-3, the intervention is a new TDI program which includes a partial payment provision. It would likely be modelled after existing programs, but would be intentionally designed to improve return to work outcomes, potentially at the expense of other program goals such as maintaining income or ease of administration. As in E-1, the benefit would be designed collaboratively by DOL and the administrator of the new program.

7.3.2 Intervention Study Design

7.3.2.1 Design E-1

The first design would involve the design, implementation, and evaluation of a new partial-payment provision. We anticipate that DOL would partner with a state that currently operates or is planning to implement a TDI system but does not currently have a partial-payment provision. New Jersey might be a particularly promising partner, as the state has come close to implementing such a provision in recent years, but other states that do not already offer partial payments might also be approached. If DOL is unable to recruit a state TDI system that is interested in implementing and testing partial payments, it might pursue a partnership with one or more private TDI insurers—partial-payments are common among private insurers but not universal (Ashley et al., 2017). Though private insurers operating in states without TDI mandates cover a selected sample of workers, some of these samples are quite large, and it might be possible to test the impact on workers across states. Moreover, private providers may not face the political pressures that sometimes hamper state efforts to change policies.

Working with either a state or a private TDI program, DOL's contractor would randomly assign new TDI participants to one of at least two conditions—*partial payments* and *no partial payments*. If the sample is large enough, participants could be assigned to three conditions—*partial payments with a 100 percent effective tax rate*, *partial payments with a lower (perhaps 50 percent) effective tax rate*, and *no partial payments*. These three conditions correspond to the three illustrated in Exhibit 7-1. Testing more than one version of partial payments requires a larger sample size, as well as the ability and willingness to administer a more complicated program, but would generate additional information about the most effective design of the provision.

If randomization were conducted at the individual level, workers filing a TDI claim would be notified that they could be eligible for a partial payment as part of a demonstration. Those that consented to participate would be randomly assigned to eligibility for the provision. If randomization by geography were used, new TDI claimants could be notified of their region's status at the time a claim was made. If DOL were able to locate a state or insurer who was willing to implement a partial payment provision, but not one willing to participate in random assignment, DOL might pursue this research topic using quasi-experimental methods.

Communicating program rules clearly, so that all participants understand their options, is one of the major challenges of this research option. Despite concerted attempts to inform study participants of the benefit rules that applied to them, many SSDI beneficiaries participating in the Benefit Offset National Demonstration did not understand the offset rules (Geyer et al., 2018). Confusion about program rules is

common—in a 2010 survey of SSDI and SSI beneficiaries, fewer than half reported knowing about the work support with the highest level of awareness (Wright et al., 2012). TDI rules are generally simpler than those in SSDI, but substantial attention to this issue is still needed. Depending on the state or private TDI system used, it may be possible to randomize by locality, so that participants in one experimental group are less likely to interact with those in the other. Though this would simplify the messaging and administration of the program, it would come at the cost of power. This issue is explored in more detail in Section 7.6.

We expect that DOL would be able to locate a state or private insurer willing to take up partial payments, and that such a program would be appealing to other states and insurers if it were shown to lower expenditures or improve outcomes at very low cost. States or private insurers would be interested in offering partial payments primarily in the hope that partial payments would decrease program payments, as would be the case if most people who took up partial payments would have otherwise received full payments. Rhode Island’s experience suggests that the resulting drop in payments could be substantial, but it does not offer definitive proof that a drop will occur (Rhode Island Department of Labor and Training, 2007).

States might also hope to improve the well-being of ill and injured workers by offering them more flexibility in their return to work, or to collect additional tax revenue if workers experience better long-term employment outcomes. Six states currently offer tax credits to incentivize employers to hire persons with disabilities, and four offer tax credits to subsidize barrier removal and employment supports.⁵⁸ Only one of these states currently offers or mandates TDI, and one other has enacted legislation to do so in the future. This suggests that many states that are not currently offering TDI have at least some willingness to spend state funds on efforts to improve employment outcomes for persons with disabilities.

It is an empirical question whether decreased program costs and other benefits would occur, and how large they would be.

7.3.2.2 Design E-2

The second design is a quasi-experimental study of the effect of Rhode Island’s adoption of partial payments in its state TDI system. For this design, the intervention is the partial-payment provision as implemented in Rhode Island. Under the Partial Return to Work Program, implemented in 2006, workers who are receiving TDI benefits can receive a partial benefit if they work but earn less than their WBA, and their medical provider certifies that they are able to work, but not to return to their full hours or customary tasks (Bourbonniere and Mann, 2018). These workers keep the first 20 percent of their earnings, after which benefits are reduced by one dollar for every dollar earned (Rhode Island Code of Regulations, 260-RICR-40-05-1). Partial benefits are generally available for up to 8 weeks, but the period can be extended to 12 weeks with agency approval.

7.3.2.3 Design E-3

The third design involves the design, implementation, and evaluation of a new TDI program, crafted to support workers’ return to work by providing financial and other supports. The evaluation would randomly assign some workers to be covered by the new TDI program while others were not, and compare outcomes of the two groups. The new program might include partial payments for those returning to work, physical or occupational therapy for those with injuries or illnesses that require it, case coordination, or referral to resources. In order to implement this intervention, DOL would need to identify

⁵⁸ Delaware, Iowa, Louisiana, Maryland, New York, and Tennessee offer tax credits (Epstein et al., 2018). Washington, Oregon, Montana, and Iowa offer reimbursement for accommodation costs (Epstein et al., 2018, Ashley et al., 2017).

a partner—perhaps a state or large employer—that was interested in establishing a TDI program and wanted support from DOL and its contractors to design and evaluate the program.

DOL would work with partner organizations to determine the exact parameters of the benefit and how to administer it, as well as the most appropriate research design to evaluate it. A state would likely need to make legislative changes, while a large employer might need to self-insure or negotiate with a private insurer. Depending on the expected sample and the ability and willingness to administer a more complicated evaluation, it may be possible to randomize individuals across more than one version of the program in order to develop evidence on the optimal design.

7.4 Participants

7.4.1 Target Populations

Design E-1: The randomized control trial evaluation of a partial-payment benefit targets workers who have experienced an injury or the onset of an illness and are applying for TDI for income support while out of the workforce. We expect the effects of the intervention to be concentrated among those who have both a motive and opportunity to use partial payments. In order to have a motive, workers must experience a period of time when they are able to do some work but not their customary tasks or hours. In order to have an opportunity, they must be able to locate a position with reduced hours or pay, at either their previous employer or another firm. Both hours and pay will likely vary based on occupation and industry, as well as demographic factors such as education. Based on Rhode Island’s experience in the first 20 months of its program, we might expect 1 to 2 percent of TDI claimants to use the partial benefits.⁵⁹ This means both that the addition of a partial-payment provision could be expected to have a reasonably small effect on the overall operation of a TDI program and that a large program is needed in order to detect an effect.

Design E-2: The Rhode Island TDI Partial Return to Work Program benefit targets workers who experience an injury or illness that makes them eligible for TDI and experience a period of time during which they are able to engage in some work but not earn more than the WBA.

Design E-3: An evaluation of a new TDI benefit would target a broader array of workers—those in a given state or organization who experience an injury or illness that makes them unable to work for a period of time. Depending on the design of the program, the program might provide targeted services to particular groups, such as workers who experience musculoskeletal injuries and require physical therapy.

7.4.2 Settings/Venues

Design E-1: This design would take place in the state(s) that were recruited to participate, or in the state(s) where a private TDI provider operates. The study would be primarily managed by the TDI program to which the partial-payments intervention was added, with assistance from DOL, an evaluator, and other technical assistance providers as necessary.

Design E-2: The Rhode Island TDI Partial Return to Work Program takes place in Rhode Island and is managed by the Rhode Island Department of Labor and Training.

Design E-3: The randomized control trial evaluation of a new TDI benefit would take place in the state or organization recruited to participate. DOL and its partner(s) would hold primary responsibility for designing and evaluating the benefit, in consultation with the state or organization offering the TDI

⁵⁹ Total includes those who take TDI due to pregnancy, but not those who receive Temporary Caregiver Insurance while bonding with a child.

benefit. The state or organization offering the TDI benefit would hold primary responsibility for implementing the benefit, with support from DOL and its partners.

7.4.3 Recruitment Plan

Design E-1: To study the effect of a new partial-payment benefit program, recruitment could work in one of two ways. First, all workers filing new TDI claims could be randomly assigned to either be eligible for partial payments or be eligible only under the pre-existing payment rules. This case is possible across a representative sample of workers if the existing state or private TDI system allows such a change. If not, the randomization would have to be done in a recruited sample of volunteers only, or across firms or establishments.

Second, workers filing new TDI claims could opt into the study. In this case, workers would be asked whether they were interested in entering a lottery to have access to the partial-payment benefit, and those who consented to participate would be randomly assigned. We anticipate the first method would generate the largest sample and would be easier to administer, but randomizing only volunteers could reduce that sample by a factor of 20 (Gubits et al., 2018). So, we focus in subsequent discussions on randomizing across establishments. These could be establishments of a single large employer, such as Walmart or Subway, or across divisions of a less geographically dispersed employer, such as a large university. As an example, Jones et al. (2018) randomly assigned program eligibility and financial incentives of a novel workplace wellness program across more than 12,000 employees, at the individual level, at the University of Illinois at Urbana-Champaign.

Design E-2: No recruitment is needed to study Rhode Island's TDI Partial RTW Program.

Design E-3: To study the effect of a new TDI program, recruitment of individuals is not necessary, but may be desirable, either to obtain consent if required for the evaluation (e.g., if individuals must consent to data collection or matching, or individual random assignment to treatment or control statuses), or to collect baseline information. We assume that any evaluation of a novel program could be designed along with the program and use random assignment, but the nature of the random assignment could depend on the partners in program design and any legal restrictions in states included.

7.5 Data

Designs E-1 and E-3: An evaluation of a new program or a new partial-payment provision would use a combination of existing administrative data, program data collected to administer the new program, and perhaps survey data to measure indicators of participant experiences. DOL would need to be able to measure labor force outcomes such as employment and earnings, both shortly after program use and after a lag. Ideally, DOL would also obtain information on longer-term outcomes such as employment stability and applications to SSDI and/or SSI. Information on employment and earnings could be obtained from the National Directory of New Hires or state Unemployment Insurance data, as described in detail below.

Design E-2: An evaluation of Rhode Island's TDI Partial RTW Program would involve comparing changes in Rhode Island to those in other states that did not adopt a partial-payment provision in 2006. The comparison and construction of this group are discussed below in the section on the analysis plan. This requires information on workforce outcomes for a large sample of workers spread across the United States. The evaluation would be richer if this data could be supplemented by information on program implementation and operations.

7.5.1 Data Collection Plan

Design E-1: An evaluation of a new partial-payment benefit would include the collection of data as part of program operations. These data might include information on the number of individuals using partial

payments and the amounts of those payments. In order to better understand the implementation of the program and how it was used, DOL might consider conducting structured interviews or focus groups.

Design E-2: An evaluation of Rhode Island’s partial-payment system would not require the collection of new data. It would rely on existing data, discussed below.

Design E-3: Data collection for an evaluation of a new TDI program would be similar to that collected under design E-1. Data on implementation and experiences would be more important, as the scope of the program being tested is much broader.

7.5.2 Existing Data—Permissions, Matching as Applicable

Designs E-1 and E-3: An evaluation of a new partial-payment benefit or new TDI program could be accomplished using Unemployment Insurance records from the state(s) that implements the program, which could be obtained from that state. With cooperation from SSA, it would be possible to match those records directly to SSA administrative data on applications and benefit awards in order to analyze longer-term effects on SSDI and SSI applications and awards.

Design E-2: An evaluation of Rhode Island’s partial-payment program would require data on labor force outcomes for a large number of people across many states. Two data sources might fit this need.

The Current Population Survey is a large, nationally representative survey of households. Sample households are surveyed for four months, then for another four months after an eight-month break. This structure, as well as the survey’s emphasis on labor force information, makes it possible to reasonably identify workers who are temporarily out of work for health reasons and then to track returns to work over short time horizons. However, such analyses would not be able to determine whether individual workers were eligible for, or used, TDI. An analysis of the Current Population Survey would ideally be paired with an analysis of SSA administrative data, which would allow researchers to measure applications to SSDI and SSI as well as benefit awards.

Another option would be to use data from the Longitudinal Employer-Household Dynamics program, which aggregates data from workers and employers in 49 states, the District of Columbia, Puerto Rico, and the Virgin Islands. Although the Longitudinal Employer-Household Dynamics currently contains data only through 2011, this would be suitable for analyzing the introduction of Rhode Island’s policy. The major downside to this dataset is that it does not contain information on injuries or illnesses that would allow for the identification of potential TDI users. That shortcoming would make it more difficult to confirm that differences experienced in Rhode Island were results of the change to TDI policy rather than other factors.

The analysis of the Current Population Survey or Longitudinal Employer-Household Dynamics would ideally be paired with analysis of administrative records from Rhode Island, to determine the extent to which the partial-payment option was used and the workforce outcomes of those who took up the benefit. This would be particularly important for the Longitudinal Employer-Household Dynamics data, which offers less information that could be used to identify potential TDI users, but would also greatly improve analyses using the CPS. The linkage could be best achieved using linked data on Rhode Island residents that is currently housed at a secure facility at the Research Improving Peoples’ Lives facility at Brown University (see ripl.org for more information). This data includes information on TDI use, information on work and earnings from the state Unemployment Insurance system, participation in programs such as Temporary Assistance for Needy Families and Medicaid, and education records. Alternately, it would be possible to obtain TDI and Unemployment Insurance data directly from state agencies. Although the dataset at Research Improving People’s Lives is incredibly rich, much of the detail it contains is not directly relevant to this project.

7.5.3 IRB, Security, Logistical Concerns

Designs E-1 and E-3: An analysis of a new partial-payment provision or TDI program would require IRB approval to collect and analyze data, as well as appropriate data security measures. It would also require permission from the implementing state(s) or organization to obtain its data.

Design E-2: Using Current Population Survey data to analyze workforce outcomes would not require special permissions, but using that data linked to SSA records or using Longitudinal Employer-Household Dynamics data would require permission, with a relatively long process to obtain clearance (the process would likely take approximately a year). Obtaining administrative data from Rhode Island would likely require establishing a data use agreement with state agencies, approval by an IRB, and appropriate security measures to protect personally identifiable information.

7.6 Analysis Plan

7.6.1 Outcomes and Hypotheses

Designs E-1 and E-2: Partial-payment schemes could affect many aspects of the RTW process. We expect that an evaluation under either design option would generate information on how the presence of a partial-payment scheme affects:

- Timing of return to work.
- Timing of return to the hours and earnings level experienced before the injury or illness.
- Likelihood of returning to work.
- Stability of work or long-term work and earnings.
- Application for and receipt of SSDI/SSI in the long term.
- Cost of providing TDI.

We expect that the presence of a partial-payment program will increase the speed with which workers make their initial return to work, as those who are able to work but not at full levels would be incentivized to do so. However, partial payments may decrease the speed of return to full work, or even work at more than the WBA. This is because some of those who could work above the WBA will prefer to earn less if they are able to work less. The exact magnitude of this effect depends on how much income the worker is willing to give up to work less, as well as the program rules that determine how difficult it is to receive partial payments when one could work more.⁶⁰

Partial-payment options would most likely improve long-term earnings and employment stability, if workers use them to support needed recovery time, search for a job that is a better fit, or maintain contact with their employer or the labor market during recovery and adjustment. However, it would also be possible for workers to return to partial work before they are ready to do so, or to hurt labor market prospects by taking too long to return to full work. We would expect any effects on applications for and receipt of SSDI and SSI to depend at least in part on these long-term labor market effects.

⁶⁰ There is also some possibility of induced entry—if TDI with partial payments is more attractive to workers than TDI without, some who would otherwise not have taken up TDI may do so. Induced entry would increase program expenditures, but its effects on other outcomes are unknown.

The effect of partial payments on TDI costs depends on how many of those receiving partial benefits would have otherwise received full benefits, versus no benefits. If partial payments allow for better recovery and/or job matches, they could potentially also decrease future TDI claims.

Design E-3: Access to TDI almost certainly improves the short term financial well-being of individuals who experience a non-work-related injury or illness that makes them unable to work, in comparison to those who do not have TDI coverage. However, the longer-term effects of the program are less clear and likely depend on the structure of the TDI benefit. In addition to verifying that the TDI benefit improved financial well-being while workers are unable to work, this design would generate information on how the presence of well-designed TDI system affects:

- Timing of return to work.
- Timing of return to the hours and earnings level experienced before the injury or illness.
- Likelihood of returning to work.
- Stability of work or long-term work and earnings.
- Application for and receipt of SSDI/SSI in the long term.

If the design tests more than one version of TDI, it could also offer evidence on the relative merits of different benefit structures, or whether adding supplemental services such as targeted health care improves outcomes.

If TDI works well, it can help workers who experience the onset of an illness or injury recover from their condition or adapt to a new normal. This can give workers an alternative to applying for SSDI benefits. If it facilitates connections and conversations among workers, medical professionals, and employers, TDI can support ongoing attachment to the workforce and transitions to new job roles. However, TDI could also harm long-term outcomes through one of two channels.

The first is that by making time out of work less unpleasant, TDI can encourage workers to stay out of work for longer than necessary. Depending on the structure of the benefit, this could be addressed by offering partial payments for those able to do some work, providing targeted treatment plans and making continued benefits contingent on adherence to them, and investing in a compliance mechanism to ensure that only those who are unable to work are identified as such.

The second concern is that workers who use TDI benefits may receive information about their eligibility for SSDI and/or SSI benefits, or that the program may actively encourage application in an attempt to shift costs to the federal government (see e.g., Stepner, 2019). It may be possible to combat this risk by designing the benefit to be of relatively short duration, so that this cost-shifting is of limited value, or by offering front-line benefits staff bonuses for each of their clients who returns to work.

7.6.2 Model(s) or Analysis Methods

Designs E-1 and E-3: For the design options using randomized control trial designs, the main analyses would use straightforward regressions to compare conditional averages of the outcome variables (e.g., returned to work). For outcomes that describe duration (e.g., time until return to work), hazard models might be used to account for the fact that the outcome would be right censored (missing for those with durations longer than the period observed). If clustered assignment or differing probabilities of assignment were used, these would need to be taken into account. Analyses of implementation data would depend on the data collected, but might include descriptive statistics on take-up and user characteristics or qualitative methods to analyze interview responses.

Design E-2: When evaluating the effect of existing partial-payment schemes, the challenge is disentangling the causal effects of the benefits from other factors that are correlated with both eligibility and outcomes. These other factors include differences between workers who are and are not covered by private TDI programs of different types, as well as state-level differences in labor markets and industrial mix. This is not an issue in designs E-1 and E-3, as those samples are randomly assigned, so that in expectation there is no correlation between these other factors and treatment status.

In order to disentangle the causal effects of the program, we suggest focusing on Rhode Island's adoption of partial payments in 2006, comparing changes in labor force outcomes in Rhode Island to those in other states that did not experience such a policy change. Rhode Island's TDI program has been in operation since 1942, so was well established at the time. This analysis approach assumes that were it not for Rhode Island's adoption of partial payments in 2006, trends in the outcomes examined—return to work for workers who experience illness or injuries, for example—would have followed the same trend as prevailed in the locations that did not experience the policy change. Instead of using an unweighted traditional difference-in-differences estimation, we suggest using a generalized synthetic control method, which uses a combination of other states to create a “synthetic” Rhode Island, constructed based on pre-policy trends in the outcome variable.⁶¹ This “synthetic” Rhode Island can provide a better match to Rhode Island's characteristics than any individual state can. The approach also facilitates more reliable inference about policy effects.⁶²

7.6.3 Expected Methodological Issues and Responses

Designs E-1 and E-3: The need for adequate sample size and the collection of data on outcomes where the program could be expected to have relatively large effects are important considerations in the design of a study of a new partial-payment provision or new TDI plan. DOL would also need to be careful that those randomized to either intervention receive clear and consistent information about their eligibility. It may be possible to randomize by locality, to minimize confusion, but such clustered assignment substantially increases the sample size needed to detect an effect of a given size. For example, in an analysis without covariates where randomization is conducted at the individual level, in which the treatment and comparison groups are the same size, 3,142 sample members would be needed to detect an effect size of 0.1 standard deviations, and we assume this represents the approximate sample size for a design that uses a single large employer to recruit study participants.

A similar analysis, but in which assignment is conducted at the cluster level, with clusters of 20 individuals, would require a total sample of 14,120 to detect an effect of the same size if the intracluster correlation is 0.2, or 9,140 if the intracluster correlation is 0.1. More limiting than the number of sample members is the number of clusters. Rhode Island is a very small state, but it had 47,075 TDI claims in 2017 and would easily have a sufficient sample size for such an analysis (Rhode Island Department of Labor and Training, 2017). However, dividing a given state into a sufficient number of clusters may be impossible while ensuring that the clusters are large enough to decrease the administrative burden. For example, Rhode Island has only five counties, 241 populated Census tracts, and 812 populated block groups. Randomization by Census tract or block group would provide a sufficient number of clusters in most states, but might not substantially simplify the administration of the program or lessen the

⁶¹ See Abadie et al. (2010) for an introduction, or Ben-Michael et al. (2018) for an example of recent extensions collectively referred to as “generalized” synthetic control methods.

⁶² Traditional difference in difference estimates do not account for uncertainty in the degree to which the control describes the trends in outcomes the treatment would have experienced were there no treatment (Abadie et al. 2010). However, synthetic control methods do not account for outcome models and are difficult to conduct inference with, whereas generalized synthetic control methods have advantages of each parent model (O'Neill et al. 2020).

possibility of confusion. Census tracts approximate neighborhoods, with an average of 4,000 people each. DOL and its partners must weigh the costs and benefits of assignment at the cluster or individual level.

Design E-2: Rhode Island is a very small state, the population who use TDI is a small subset of the state population, and partial RTW users are a small subset of that group. As a result, if adjustments to the introduction of partial payments are small, we may be unable to detect effects even if they exist. One approach to this issue would be to collect data on outcomes where we expect larger adjustments, such as speed of return and speed of full return, and to use particularly large datasets, in order to have the largest samples possible. The synthetic control method also allows for the inclusion of more pre- and post-reform sample members than would be included under designs focused on the discontinuity in program rules.

7.7 Practical Considerations

7.7.1 Implementation Timeline

Design E-1: Adding a new benefit to an existing program would require time to recruit a partner, design and implement the benefit and randomization scheme, and collect and analyze data. This would take perhaps three to five years. Ideally the evaluation would analyze longer-term outcomes, such as employment stability and SSDI/SSI applications and receipt. Doing so would require waiting several years, perhaps 5 or more, for these longer-term outcomes to occur. The analysis might be conducted in two or more rounds in order to produce faster information on short-term outcomes as well as information on long-term outcomes.

Design E-2: Analyzing the effect of Rhode Island's TDI Partial Return to Work Program would take a fairly short amount of time, perhaps 1-2 years once data access has been negotiated. The program has already been implemented, and enough time has passed that information on medium-term outcomes is already available.

Design E-3: The steps in creating and evaluating a new TDI program would be similar to that in design E-1. However, because the design E-3 would require the design of a full TDI program, rather than one additional component, many of the steps would take longer. In particular, more time would be needed to craft the program parameters, plan for implementation, and coordinate partners. We estimate that this process might take five to seven years. The design of the evaluation would operate on the same timeline as the design of the new program, and the evaluation would be built in to the rollout of the new program. Analysis of longer-term outcomes would require several additional years, to both collect and analyze the relevant data.

7.7.2 Cost

Design E-1: This option requires designing and operating the partial-payment intervention to test through existing state or private TDI programs. Designing and monitoring the intervention would require DOL to partner with a state that currently operates or is planning to implement a TDI system but does not currently have a partial-payment provision. If DOL is unable to find such a state, an alternative would be to establish a partnership with one or more private TDI insurers. We anticipate finding this partner and working with its TDI system to design up to three treatment conditions would require fairly lengthy negotiations, an important cost implication for this option. Additionally, costs accrue from implementing the treatment and monitoring use of the benefit. This option may require researchers to recruit a sample of workers from TDI claimants, following a recruitment schedule and plan to be agreed on with the study partner.

Once the treatment is implemented, data collection and analysis would be similar to a typical randomized controlled trial. Data collection would be required to measure participant baseline characteristics and outcomes. We assume that the study could use a combination of data collected to administer the new

program and existing administrative data. DOL may choose to partner with SSA to match state partner Unemployment Insurance records with SSA administrative data. DOL could also choose to administer a survey via structured interviews to measure participant outcomes, which would increase costs.

Unlike Options A-D, an important cost component of Option E-1 is the cost of partial payments themselves and the administration of the benefit payments. The net cost of the partial benefit is unclear and depends on whether and what sort of behavior change the partial payments would produce—workers could receive the partial benefit either when they would have otherwise received no benefit or when they would have otherwise received the full benefit. The TDI system to which the partial-payment benefit was added would either decrease in cost from lower benefits, or be responsible for higher benefit costs. An empirical study would be needed to determine the response to the offer of partial benefits.

Design E-2: Analyzing Rhode Island’s existing program would not require researchers to design and operate a new intervention or to recruit sample members. They would need to design a quasi-experimental evaluation and analysis plan, identify a comparison group, and negotiate agreements for obtaining data to measure participant outcomes.

Design E-3: The cost of creating a new benefit would include the full cost of the benefit provided, including monetary payments and insurance coverage, as well as the costs of administering a new insurance scheme and designing and evaluating the intervention. If implemented in a state, this cost would likely be funded at least in part by state taxes levied on workers and/or employers, as is the case in existing state TDI systems. It also might be possible for DOL and other federal partners to offer grant funding to partially offset this cost. This design involves multiple costs and would be more costly than either design E-1 or E-2.

7.7.3 Geographic Scope

Design E-1: An analysis of a new partial payment system would likely take place in a single state. As a result, conclusions would be specific to that state but would likely generalize to others that have or are implementing state-wide TDI programs. If DOL were able to recruit additional states or private TDI providers, it might be possible to compare experiences across settings, or to describe program effects in more than one location. Implementing a larger demonstration would, however, be more costly, involve more administrative challenges, and might take a longer period of time.

Design E-2: A study of Rhode Island’s existing partial-payment benefit program would compare experiences in Rhode Island to those elsewhere. Its conclusions would in some ways be specific to Rhode Island. It might generalize better to other states that have or are considering state-wide TDI programs, which serve most workers in a state but in a different state context, than to private TDI programs, which serve only workers whose employers offer these programs.

Design E-3: An analysis of a new TDI program would likely take place in a single state or perhaps through a large employer with locations in many states. Conclusions would be specific to that state or the locations in which the employer was located, but with appropriate care could be generalized to generate predictions for other states or locations, particularly those with similar health care or income support landscapes. To enable this generalization, the analysis might reweight the sample to better match the demographics of the national population (Olsen et al., 2013), or the program might be designed so that it could be implemented in other settings with minimal changes.

7.7.4 Threats to Implementation as Designed

Design E-1: Implementing a new benefit as part of an existing TDI system would require a close partnership with the agency or company administering the insurance scheme.

Design E-2: The main threat to an analysis of Rhode Island's partial-payment TDI benefit is the small sample of workers who take up the benefit. Before embarking on this analysis, a careful power calculation would need to be conducted. DOL and its partners would acquire information on the total number of users to date, as well as average benefit duration for users and non-users, in order to estimate plausible effect sizes and determine whether the sample is large enough to detect them.

Design E-3: Implementing a new stand-alone benefit would require close, continuing collaboration with the state or employer adopting the new program.

7.8 Contributions and Limitations

Designs E-1 and E-2: These two designs would conduct a rigorous test of a promising component of TDI programs about which, to our knowledge, there is no rigorous causal evidence in this setting. Partial payments can potentially encourage and enable workers who experience illness or injury to transition back to work earlier than would otherwise be the case, maintaining their connections to the workplace. If these connections are strengthened considerably, or if the temporary arrangements enable workers to transition to roles that suit their new functional capacity, they could improve long-term labor force outcomes. Design E-1 would use a randomized controlled trial design, so it would generate estimates with high internal validity. It would also allow for implementation data to be collected as the program is designed and put into place. Design E-2 would use a synthetic control design, so have moderate internal validity, and implementation data would be limited to retrospective information or data that Rhode Island is willing to share. It may also be difficult to detect effects of the partial-payment provision, depending on the available sample size. However, design E-1 involves more time, greater cost, and more complications than E-2.

Design E-3: This design would conduct a rigorous test of a TDI program intentionally designed to incentivize and facilitate return to work. Although 39 percent of civilian worker have access to some form of short-term disability benefits, and many have speculated that programs may improve workforce outcomes for those who experience illness or injury, little research has been conducted, particularly in the U.S. setting, on the effect of these programs or their components (DOL/Bureau of Labor Statistics, 2019; Autor & Duggan, 2010). By generating evidence on the effectiveness of TDI programs and their components, this design would inform the choices of states that may be considering adopting or altering their programs, employers who are deciding whether to offer employees TDI and what program components to pursue, and private TDI insurers crafting policies to offer the coverage employers desire at the lowest possible price. The evaluation would use a randomized control trial design, so it would have high internal validity. However, Design E-3 would be a large undertaking and require close cooperation with the state or organization implementing the program as well as other partners.

8. Summary

Programs and policies that help ill or injured workers stay at work or return to work can benefit the workers themselves and their families, as well as employers and governments. Successful SAW/RTW programs can increase worker productivity, generate tax revenue, and reduce the costs of federal disability benefits. These important benefits make a strong case for policymakers to invest in rigorous research to test interventions, even interventions that might require substantial resources to implement. In recognition of this potential, researchers, advocates, and policymakers have devoted considerable energy in recent years to understanding what kinds of supports workers need, and to considering how best to develop and deliver those supports. Despite this attention, considerable gaps remain in our understanding of SAW/RTW programs and their effects.

This report has described five evaluation design options that would address some of these gaps. Each option would allow DOL to build evidence about potential target populations for SAW/RTW programs and potentially to learn more about the effects of specific program components.

To help DOL compare and contrast the five options, this chapter summarizes the open research questions in SAW/RTW that the options would address. The chapter also discusses the technical and practical factors that might influence the choice of which option to pursue.

8.1 Summary of Open Questions

The SAW/RTW process involves multiple stakeholders—workers, employers, health care providers, insurers, workforce agencies, and vocational rehabilitation providers. Their incentives to encouraging workers to remain attached to the labor force after a potentially disabling illness or injury often misalign. For example, private disability insurers might lower their costs of assisting a worker if that worker applies for federal disability benefits (Liebman & Smalligan, 2013). Stakeholders may also lack accurate and timely information about resources and best practices. For example, health care providers might not view labor force retention as relevant to their health care mission, or they may not know how to facilitate labor force retention. Two core research questions guide new evaluation design options:

- 1) What is the effect of SAW RTW interventions on employment for workers who experience illness or injury?
- 2) What is the effect of SAW RTW interventions on application and receipt of federal disability benefits?

Evaluation design options might focus on any of the many stakeholders to address these core research questions. New policies and programs might address the incentives for stakeholders to engage in SAW/RTW activities and the information needed to do so. Our evidence review, administrative data analysis, and expert opinions from the Technical Working Group pointed to several promising areas for further study that would inform the core research questions:

- Describing the target populations for SAW/RTW interventions
- Testing interventions related to the health care touchpoint and employer incentives
- Testing the effectiveness of particular types of supports for workers, health care providers, and employers.
- Determining which SAW/RTW program models and strategies are effective, for whom.

8.2 Summary of Designs

Exhibit 8-1 summarizes the five evaluation design options presented in this report. Exhibit 8-2 summarizes implementation and practical considerations. Each option addresses one or more of the open research questions discussed above and focuses on one or more of the stakeholders involved in the SAW/RTW process.

- **Option A** would produce information about the potential target populations for SAW/RTW programs, and patterns of program participation. This foundational research would increase our understanding of barriers workers face and promising strategies for designing SAW/RTW programs.
- **Option B** would generate evidence on the effects of information on workers' outcomes.
- **Option C** would answer research questions about employers. It would address the practices they use to retain workers with impairments and would test the effects of information on the use of those practices.
- **Option D** would address research questions about medical professionals. It would identify the best practices that experts in occupational medicine believe promote SAW/RTW outcomes and test a low-intensity intervention to promote those practices.
- **Option E** would look at the critical role of insurers. It would examine the effects of temporary disability benefits, either by performing a quasi-experimental analysis of an existing program, or through designing and testing a new temporary benefit demonstration program. This option would generate evidence on the effectiveness of a particular intervention, as well as its cost. It would also provide information on program take-up.

We believe implementing any of these ideas would help DOL move the research agenda forward and complement other ongoing research to build evidence about SAW/RTW. Policymakers will need to take into account the feasibility and practical considerations to decide which options are most promising. We discuss some of those considerations below.

8.3 Considerations in Selecting Designs

In addition to selecting designs that answer research questions of interest, DOL should consider technical and practical considerations.

8.3.1 Technical Considerations

8.3.1.1 The quality of evidence provided by any research project depend in part on the experimental design used, the sample size and thus statistical power of the study, and the generalizability of results

We considered a range of evaluation design methods to address each of the options. After reviewing possible methods, we believe that each of the four options that would test effects of SAW/RTW program components (information to workers, employers, and health care providers and partial benefit payments to TDI claimants) could be implemented using random assignment. Such random assignment studies would provide causal evidence about effects with high internal validity. If executed well, studies with these designs would be eligible for top rankings (e.g. 'high' or "meets standards without reservations") from systematic clearinghouses such as CLEAR or the What Works Clearinghouse.

Exhibit 8-1. Summary of Evaluation Design Options

Evaluation Design Option	Research Question	Target Population & Partnerships	SAW/RTW Strategy To Be Studied	Contribution to Evidence Base
A. Longitudinal study of individual workers	<ul style="list-style-type: none"> Which workers will leave the labor force after an illness or injury? Which of them will eventually apply for SSDI? 	<p>Target population: workers who have experienced an injury or illness that threatens their ability to work</p> <p>Partnerships: Data sharing agreements</p>	Patterns of program participation after illness or injury	The study would provide valuable information to policymakers to help them target SAW/RTW services to those workers who may be most likely to benefit.
B. Test an Informational Intervention for workers	How does advice about working, and information about employer policies and resources affect employment outcomes for workers who have recently experienced an injury or the onset of an illness?	<p>Target population: Workers at risk of applying for federal disability benefits</p> <p>Partnerships: Employers, state or federal agencies or insurance providers to identify sample of workers</p>	Information, advice and resources to workers about working with a disability	This option would produce evidence about the effect of providing targeted information on the employment outcomes of workers who have recently experienced an injury or the onset or worsening of an illness.
C. Examine employer SAW/RTW practices and test an intervention that informs employers about best practices	<ul style="list-style-type: none"> What practices, if any, do employers use to retain workers with impairments? Does providing employers with information about SAW/RTW best practices improve employer practices? 	<p>Target population: Employers</p> <p>Partnerships: Employer organizations, federal-state partnerships</p>	Information to employers on strategies and resources to retain workers with impairments	This option would generate knowledge of the existing practices and knowledge of employers. The option would also produce evidence about the effect of providing information on employer practice and potentially, on employment and disability outcomes of workers.
D. Inform medical professionals about best practices to facilitate SAW/RTW.	<ul style="list-style-type: none"> Would changes to medical practice promote individuals staying at work or returning to work? Does an informational nudge induce changes to medical practice? 	<p>Target population: Medical professionals</p> <p>Partnerships: Licensing agencies for medical professionals</p>	Information to medical professionals on best practices to promote work	This option would identify what occupational medicine experts consider to be best medical practices for promoting work. The option would also produce evidence on effects of informing medical professionals about those practices on provider behavior and patients' employment outcomes.
E. Evaluate the effects of partial payments in temporary disability insurance programs.	What effect do partial payments in TDI, and TDI in general, have on work, benefits, and program costs?	<p>Target population: Individual workers</p> <p>Organizations & Agencies: Private or state TDI insurer</p> <p>Partnerships: Private or state TDI insurer</p>	Partial temporary disability benefits or insurance program with partial payments for those who return to work at a lower number of hours or wages	This option would produce evidence about the effects of offering workers partial temporary disability payments on work, benefits, and program costs.

Exhibit 8-2. Evaluation Design Options: Practical Considerations

	Option A Longitudinal Study of Individual Workers	Option B Test an Informational Intervention for Workers	Option C Examine employer SAW/RTW practices and test an intervention that informs employers about best practices	Option D Inform medical professionals about best practices to facilitate SAW/RTW	Option E Evaluate the effects of partial payments in temporary disability insurance programs.
Design a new intervention?	No	Yes	Yes (component 2)	Yes	Yes (design E-1 and E-3)
Develop randomized control trial design?	No	Yes multiple treatments	Yes (component 2) multiple treatments	Yes multiple treatments	Yes (design E-1 and E-3)
Identify and recruit a study sample?	No	Yes (workers)	Yes (employers)	Yes (medical professionals)	Yes (TDI claimants)
Data collection	<ul style="list-style-type: none"> Existing data Obtain permissions to link national surveys to SSA or other restricted access administrative data 	<ul style="list-style-type: none"> Employer data to measure baseline characteristics Administrative data to measure employment outcomes (with permissions) 	<ul style="list-style-type: none"> New employer survey Employer administrative records Site visits, interviews, program data collection 	<ul style="list-style-type: none"> Aggregate labor market to measure outcomes Public-use survey data (e.g., Current Population Survey and American Community Survey) 	<ul style="list-style-type: none"> TDI claims data for baseline Administrative data to measure employment outcomes Program data from state/insurer Consider participant survey to measure additional outcomes
Potential timeline (estimated)	3-4 years	4-5 years	Component 1: 3 years Component 2: 5 years	5+ years	3-5 years
Cost considerations	<ul style="list-style-type: none"> Analysis plan 	<ul style="list-style-type: none"> Intervention design Partnerships Sample recruitment Data agreements with employer Data collection 	Component 1 Sample recruitment Survey design Component 2 Intervention design Evaluation design Data collection	<ul style="list-style-type: none"> Sample recruitment Intervention design Data collection 	<ul style="list-style-type: none"> Intervention design Evaluation design Cost of partial payments Data collection

8.3.1.2 Sample size and Minimum Detectable Effects (MDEs) for evaluations

Another consideration is sample size. In order to reliably detect the effect of an intervention or differences between groups, larger samples are needed. Sample size needs depend on the expected size of an effect, the variance of outcomes, the data used, and the level (individual or cluster) at which treatment status is determined. For example, in a study attempting to test the effect of a program on employment, a larger sample size is needed to detect a smaller expected effect.⁶³

In the descriptions of Options B, C, D, and E we have presented general assessments of potential sample sizes, statistical power, and preliminary estimates of minimum detectable effects. Before implementing any of the designs a research team would need to examine these preliminary assumptions carefully and develop a more detailed analysis plan. In particular, DOL will need a more detailed analysis of the plausible effects of the interventions, and of feasible sample sizes, based on specific study recruitment and implementation plans. Statistical power could be a challenge in all of the options given the magnitude of potential effects one might expect from the interventions to be tested, and the feasibility of enrolling samples of sufficient size. Each evaluation design option would study a set of comparisons that probably would involve relatively small differences in employment rates, or disability application rates, on the order of five to ten percentage points. Much larger sample sizes would be required to detect effects for specific impairment types, for example.

8.3.1.3 Generalizability and operation at scale

Even the most rigorous study will be limited in its ability to inform policy if its conclusions are not generalizable or if studied interventions cannot be operated at scale. In order to be generalizable, we have drafted the designs in Chapters 3-7 to emphasize nationally representative samples to the extent possible. Where nationally representative samples are not feasible, samples might be drawn from populations of interest (e.g. TDI users, or veterans), or from groups with broad similarities to the national population, such as the employees of a large employer or the resident of a state.

Another important feature of each option is its feasibility of scaling up, and any barriers to adopting policy to expand operations. We have not included options that seemed infeasible due to legal or logistical barriers, but some would be easier to implement widely than others. For example, any checklist or other advice that was found to improve physician practice and thus employment outcomes could be provided to all medical professionals who had not already received it at fairly low cost. Changing the parameters of TDI programs to incorporate partial payments is possible, but only with the cooperation of many stakeholders, so scale-up would be far more difficult.

8.3.2 Practical Considerations: Cost, Complexity, Data, and Timeframe

Practical considerations include a study's timeline, implementation complexity, data collection requirements, and the cost implications associated with these factors. While we have not developed detailed timelines and schedules for each of the options, we think that Options A and B might be completed in roughly 4 years, Options C and E might require up to 5 years and Option D would likely require more than 5 years. If DOL chose to evaluate the effect of any of Options B, C, D, or E on long-term employment or benefit applications, a longer follow-up period would be needed.

Before undertaking any of the options, DOL would need to conduct a detailed assessment of estimated costs. As discussed in Chapters 3 to 7, we have determined the essential requirements for implementing each evaluation design and the complexity of each of the designs, which in turn influence the potential costs of each option.

⁶³ The presence of data on factors that predict employment, such as demographics, can allow for a somewhat smaller sample than would otherwise be needed to detect a given impact. And while it may be simpler or more palatable to randomize treatment status to counties or employers, rather than to individual workers, this increases the number of individual participants needed to detect a given effect, often by orders of magnitude.

Cost considerations relate to several factors. Option A would require DOL or its contractors to obtain restricted-use data, perhaps from several sources, but would otherwise be fairly straightforward to complete. The other options would require the *design and implementation of a new intervention* and this would be an important consideration for the costs of the options. In Option D this would be based on input from occupational health experts, while in Option E intervention design would be done in consultation with the state or TDI program implementing the partial payment provision. All options would require *partnerships with other organizations or agencies*, but the options vary in the number of partnerships needed as well as their extent. For example, Option B would require a partnership with a large employer, or the Department of Defense, or a workers' compensation agency to assist with recruiting a target sample of workers. Option E would require partnering with an existing State TDI program or developing new partnerships to design and implement a new partial payment intervention.

Options B, C, D, and E all include variants that would use random assignment to analyze the effects of a particular intervention. All would require a *detailed randomization plan* and *sample recruitment strategy*. Our experience shows that implementing those strategies, and sustaining them over the period needed to enroll an adequate sample could also pose substantial costs. Delivering the intervention and monitoring to make sure it is delivered with fidelity is another source of costs. Another cost consideration is data collection. Option A would use existing data, but Options C and E might involve developing and administering new surveys. Conducting new surveys would require Paperwork Reduction Act approvals which would influence both costs and the studies' timelines. Options B, C, D, and E would also likely rely on new arrangements to collect administrative data to measure outcomes and these arrangements would require negotiating data access and permissions. Option D would measure aggregate employment outcomes, not individual-level outcomes, but would still require data use agreements to assemble the needed data. Exhibit 8-2 summarizes key practical considerations for each of the options to provide preliminary information DOL can use to compare and contrast the complexity and potential costs of the options.

In the descriptions in Chapters 3 to 7 we also discuss several variations on the main designs. Some of these are additional analyses, while others are stand-alone designs that might be pursued on their own or in combination with others. In many cases there is a continuum of designs, and DOL can determine how intensive an intervention is desired or what level of detail or length of follow-up is needed to inform policy.

8.3.3 Complementary Research Options

In addition to selecting among individual options, DOL may want to consider combining options. Combining options strategically might allow DOL to generate more information from two (or more) research options than could be achieved by pursuing both individually, or to achieve efficiencies by using some of the same partnerships and data use arrangements more than once.

For example, DOL could choose to pursue Option A followed by another option. Information from Option A could help determine the workers most likely to be on the margin of staying at work or returning to work, and these workers could be targeted for Option B in order to generate the largest effect for a given sample size. Information from Option A could also be used to identify subgroups of interest or touchpoints that seem particularly important; these could be investigated in any of the other options by drawing oversamples of particular subgroups or including targeted questions on any surveys administered. If DOL pursued both Options B and C, it might be able to use the findings from Option C to customize the information provided to workers in Option B to reflect the options that were available at their employer, or to provide data on the prevalence of certain policies within their industry. Section 5.9 discusses considerations for combining Options B and C either for analysis or in a combined evaluation design.

Combining options might also allow DOL to obtain information on the synergies that would be achieved if multiple new policies were adopted more widely. For example, Options B, C, and D all involve providing information to stakeholders in the SAW/RTW process. It is possible that the effect of informing both the worker and the employer is larger than the effect of informing the worker plus that of informing the employer, because the two parties are better able to coordinate and work together. It is also possible that it will be less effective, if the one informed party would have provided information to the party that was not provided with information. Whether either of these is the case is an empirical question that could be tested by pursuing more than one of Options B, C, and D.

Finally, combining options could allow DOL to achieve efficiencies for some of the administrative tasks of the options, by working with some of the same partners or using some of the same data sources more than once. For example, if DOL chose to work with a state or private TDI provider to identify sample members for Option A, it could also work with that partner in Option E. If DOL decided to pursue more than one of Options B, C, and E, it could negotiate the use of the National Directory of New Hires or state unemployment insurance records to track employment outcomes for all these projects simultaneously.

8.4 Conclusion

In this report we have provided detailed summaries of five evaluation design options, each of which addresses the two core questions: how can policymakers increase employment and reduce receipt of disability benefits among workers who experience injury or illness. Each option would address one or more additional open questions about SAW/RTW. These options vary in the nature and rigor of the information they would generate, as well as the cost, effort, and time required. While we have chosen five main designs that we believe to be particularly promising, the report notes several ways in which the designs could be altered to address slightly different questions, or to conform to limitations to the partnerships, time, or resources that might be available. DOL can consider these factors when setting its future SAW/RTW research agenda.

References

- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490), 493-505.
- Acs, G., & Loprest, P. (2009). *Job Differences by Race and Ethnicity in the Low-Skill Job Market*. Washington, D.C.: Urban Institute.
- American College of Occupational and Environmental Medicine. (2006). Preventing needless work disability by helping people stay employed. *Journal of Occupational and Environmental Medicine*, 48(9), 972-987.
- American College of Occupational and Environmental Medicine. (2008). The Personal Physician's role in helping patients with Medical Conditions Stay at work or Return to Work. Position Statement. https://stagesd.acoem.org/acoem/media/PDF-Library/Public_Affairs/Policies_And_Position_Statements/Guidelines/Position_Statements/The_Personal_Physician_s_Role_in_Helping_Patients-21.pdf accessed on May 19, 2020.
- Ashley, J., Cashdollar, W., Etcheverry, R., & Magill, K. (2017). *Transition Back to Work: Policies to Support Return to Work after Illness or Injury*. Columbia, MD: IMPAQ International, LLC.
- Autor, A., & Duggan, M. (2010). *Supported Work: A Proposal for Modernizing the U.S. Disability Insurance System*. Washington, D.C.: The Center for American Progress and The Hamilton Project.
- Bhandari, A., & Wagner, T. (2006). Self-reported utilization of health care services: improving measurement and accuracy. *Medical Care Research and Review*, 63(2), 217-235.
- Bardos, M., Burak, H., Ben-Shalom, Y. (2015). *Assessing the Costs and Benefits of Return-to-Work Programs*. Washington, D.C.: Mathematica Policy Research.
- Ben-Michael, E., Feller, A., & Rothstein J. (2018). The Augmented Synthetic Control Method. *Cornell University arXiv.org*, arXiv:1811.04170. Retrieved from <https://arxiv.org/abs/1811.04170>.
- Ben-Shalom, Y., Bruns, S., Contreary, K., & Stapleton, D. (2017). *Stay-at-Work/Return-to-Work: Key Facts, Critical Information Gaps, and Current Practices and Proposals*. Washington, D.C.: Mathematica Policy Research.
- Black, D. C. (2012). Work, health and wellbeing. *Safety and health at work*, 3(4), 241.
- Bound, J., Burkhauser, R.V., & Nichols, A. (2003). Tracking the Household Income of SSDI and SSI Applicants. *Research in Labor Economics*, 22, 113-158.
- Bourbonnierre, A.M. & Mann, D.R. (2018). Benefit Duration and Return to Work Outcomes in short term disability insurance programs: evidence from temporary disability insurance programs. *Journal of Occupational Rehabilitation*, 28(4), 597-609.
- Boushey, H., & Glynn, S.J. (2012). *There Are Significant Business Costs to Replacing Employees*. Washington, D.C.: Center for American Progress.
- Brusco, N. K., & Watts, J. J. (2015). Empirical evidence of recall bias for primary health care visits. *BMC health services research*, 15(1), 381.

- Callan, T., Lindner, S., & Nichols, A. (2015). *Unemployment insurance modernization and eligibility*. Washington, DC: The Urban Institute.
- Center on Budget and Policy Priorities. (2019). Chart Book: Social Security Disability Insurance. Available at: <https://www.cbpp.org/research/social-security/chart-book-social-security-disability-insurance>
- Cho, Y.I., Johnston, T.P., & VanGeest J.B. (2013). Enhancing Surveys of Health Care Professionals: A Meta-Analysis of Techniques to Improve Response. *Evaluation & the Health Professions*, 36(3), 382-407.
- Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival analysis part I: basic concepts and first analyses. *British journal of cancer*, 89(2), 232-238.
- Collier, R. (2011). American Medical Association membership woes continue. *Canadian Medical Association. Journal*, 183(11), E713.
- Contreary, K. & Perez-Johnson, I. (2016). *Behavioral Interventions to Promote Job Retention After Injury or Illness*. Washington, D.C.: Mathematica Policy Research.
- Council of State Governments. (2018). Stay-at-Work/Return-to-Work Toolkit. Lexington, KY. Retrieved from https://www.csg.org/seed_report.pdf
- Cox, S.L., Zlot, A.I., Silvey, K., Elliott, D., Horn, T., Johnson, A., & Leman, R.F. (2012). Patterns of Cancer Genetic Testing: A Randomized Survey of Oregon Clinicians. *Journal of Cancer Epidemiology*, 2012(294730), 1-11.
- Croake, S., Anand, P., Jones, C., Morrison, K., Orfield, C., Stapleton, D., ... Wittenburg, D. (2017). *BOND Implementation and Evaluation: 2017 Stage 1 Interim Process, Participation, and Impact Report*. Cambridge, MA: Abt Associates.
- Daley, K., Kennedy, C., Schalk, M., Pacer, J., Ackermann, A., Pozniak, A., & Klerman, J. (2012). *Family and Medical Leave in 2012: Methodology Report*. Cambridge, MA: Abt Associates.
- Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management science*, 9(3), 458-467.
- Dalziel, K., Li, J., Scott, A., & Clarke, P. (2018). Accuracy of patient recall for self-reported doctor visits: Is shorter recall better?. *Health economics*, 27(11), 1684-1698.
- Denne, J., Kettner, G., & Ben-Shalom, Y. (2015). The role of the physician in the return-to-work process following disability onset. *Mathematica Center for Studying Disability Policy*.
- DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3), 177-188.
- Dhaliwal, I., Duflo, E., Glennerster, R., & Tulloch, C. (2012). Comparative cost-effectiveness analysis to inform policy in developing countries: a general framework with applications for education. *Education policy in developing countries*, 285-338.
- Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, NJ: John Wiley & Sons.
- Donovan, S. A. (2019). *Paid family leave in the United States*. Washington D.C.: Congressional Research Service.

- Epstein, Z., Wood, M., Grosz, M., Prenovitz, S., and Nichols, A. (2020). *Synthesis of Stay-at-Work/Return-to-Work (SAW/RTW) Programs, Models, Efforts, and Definitions*. Prepared for the U.S. Department of Labor. Cambridge, MA: Abt Associates
- Favreault, M.M., & Nichols, A. (2011). *Immigrant Diversity and Social Security: Recent Patterns and Future Prospects*. Washington, D.C.: Urban Institute.
- Gaskell, G. D., Wright, D. B., & O’Muircheartaigh, C. A. (2000). Telescoping of landmark events: implications for survey research. *The Public Opinion Quarterly*, 64(1), 77-89.
- Geyer, J., Gubits, D., Bell, S., Morrill, T., Hoffman, D., Croake, S... Stapleton, D. (2018). *BOND Implementation and Evaluation 2017 Stage 2 Interim Process, Participation, and Impact Report*. Bethesda, MD: Abt Associates.
- Grayling, M. J., & Wason, J. M. (2020). A web application for the design of multi-arm clinical trials. *BMC cancer*, 20(1), 1-12.
- Gubits, D., Stapleton, D., Bell, S., Wood, M., Hoffman, D., Croake, S., Mann, D.R., Geyer, J., Greenberg, D., Nichols, A., McGuirk, A., Carroll, M., Kattel, U. and Judkins, D. (2018). *BOND Implementation and Evaluation: Final Evaluation Report, Volume 1*. Submitted to the Social Security Administration, Office of Research, Demonstrations, and Employment Support. Cambridge, MA: Abt Associates, and Washington, DC: Mathematica Policy Research.
- Haynes, A. B., Weiser, T.G., Berry, W.R., Lipsitz, S.R., Breizat, A., Dellinger, E.P., ... Moorthy, K. (2009). A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. *The New England Journal of Medicine*, 360, 491-499.
- Hoffman, D., Croake, S., Mann, D.R., Stapleton, D., Anand, P., Jones, C., ... Sinclair, M. (2017). *BOND Implementation and Evaluation: 2017 Stage 1 Interim Process, Participation, and Impact Report*. Cambridge, MA: Abt Associates.
- Hollenbeck, K. (2015). *Promoting Retention or Reemployment of Workers after a Significant Injury or Illness*. Washington, D.C.: Mathematica Policy Research.
- Jenkins, S.P. (1995). Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics*, 57(1), 129-138.
- Jones, D., Molitor, D., & Reif, J. (2018). *What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study*. NBER Working Paper No. 24229. Retrieved from <https://www.nber.org/papers/w24229>.
- Juras, R., Minzner, A., Klerman, J., Honnef, P., & Dunton, L. (2017). *The Effect of Marketing on Demand for OSHA’s On-site Consultation Program*. Bethesda, MD: Abt Associates.
- Juriscic, M., Bean, M., Harbaugh, J., Cloeren, M., Hardy, S., Liu, H., Nelson, C., Christian, J. (2017). The Personal Physician’s Role in Helping Patients With Medical Conditions Stay at Work or Return to Work. *Journal of Occupational & Environmental Medicine*. 59(6), 125-131. doi: 10.1097/JOM.0000000000001055.
- Kahneman, D. (2013). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux.
- Kanyongo, Gibbs Y., Gordon P. Brook, Lydia Kyei-Blankson, and Gulsah Gocmen. (2007). “Reliability and Statistical Power: How Measurement Fallibility Affects Power and Required Sample Sizes

- for Several Parametric and Nonparametric Statistics.” *Journal of Modern Applied Statistical Methods*, 6(1): 81-90.
- Kjellson, G., Clarke, P., & Gerdtham, U. G. (2014). Forgetting to remember or remembering to forget: a study of the recall period length in health care survey questions. *Journal of health economics*, 35, 34-46.
- Lachin, J. M. (1981). Introduction to sample size determination and power analysis for clinical trials. *Controlled clinical trials*, 2(2), 93-113.
- Lachin, J. M., & Foulkes, M. A. (1986). Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics*, 507-519.
- Lakatos, E., & Lan, K. G. (1992). A comparison of sample size methods for the logrank statistic. *Statistics in medicine*, 11(2), 179-191.
- Layne, M., Wagner, D., & Rothhaas, C. (2014). Estimating record linkage false match rate for the Person Identification Validation System. *Center for Administrative Records Research and Applications Working Paper*, 2.
- Liebman, J.B., & Smalligan, J.A. (2013). *An Evidence-Based Path to Disability Insurance Reform* (Report No.4). Washington, D.C.: The Hamilton Project, Brookings.
- Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman’s critique. *The Annals of Applied Statistics*, 7(1), 295-318.
- Lindner, S. & Nichols, A. (2014). The Impact of Temporary Assistance Programs on Disability Rolls and Re-employment. *Research in Labor Economics*, 39, 219-258.
- MacLennan, P.A., McGwin, G., Searcey, K., & Owsley, C. (2014). A survey of Alabama eye care providers in 2010–2011. *BMC Ophthalmology*, 14(44), 1-10.
- Mc Hugh, R., Segal, N. E., & Wenger, J. B. (2002). Laid Off and Left Out: Part-Time Workers and Unemployment Insurance Eligibility: How States Treat Part-Time Workers and Why UI Programs Should Include Them. *Washington DC: Washington College of Law, American University*.
- McLaren, C.F., Baldwin, M.L., & Boden, L.I. (2018). *Workers’ Compensation Benefits, Costs, and Coverage – 2016 Data*. Washington DC: National Academy of Social Insurance.
- McLaren, C. & Schreur, E., 2019. Improving Access to Data for Disability-Related Research. Washington, DC: National Academy of Social Insurance. Available at: <https://www.dol.gov/odep/topics/ImprovingAccessToDataOnDisability-RelatedTopics.pdf>
- Mullanathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87-106.
- Munts, R. (1970). Partial benefit schedules in unemployment insurance: their effect on work incentives. *Journal of Human Resources*, 5(2) 160-176.
- Nichols, A. (2007). Causal inference with observational data. *The Stata Journal*, 7(4), 507-541.

- Nichols, A., Geyer, J., Grosz, M., Epstein, Z., & Wood, M. (2020a). *Synthesis of Evidence about Stay-at-Work/ Return-to-Work (SAW/RTW) and Related Programs*. Prepared for the U.S. Department of Labor. Rockville, MD: Abt Associates
- Nichols, A., Dastrup, E., Epstein, Z., & Wood, M. (2020b). *Data Analysis for Stay-at-Work/Return-to-Work (SAW/RTW) Models and Strategies Project: Early Intervention Pathway Map and Population Profiles*. Prepared for the U.S. Department of Labor. Rockville, MD: Abt Associates
- Nichols, A. & Rothstein J. (2016). “The Earned Income Tax Credit (EITC).” in *Economics of Means-Tested Transfer Programs in the United States*, Ed. Robert Moffitt. Cambridge, MA: NBER.
- Nichols, A., Schmidt, L., & Sevak, P. (2017). Economic Conditions and Supplemental Security Income Applications. *Social Security Bulletin*, 77(4), 27-44.
- O’Neill, S., Noemi K. Sutton, M., & Grieve, R. (2020). “A comparison of methods for health policy evaluation with controlled pre-post designs.” *Health Serv Res*. 55:328–338
<https://onlinelibrary.wiley.com/doi/full/10.1111/1475-6773.13274>
- Olsen, R.B., Larry L. Orr, Stephen H. Bell, & Elizabeth A. Stuart. (2013). “External Validity in Policy Evaluations That Choose Sites Purposively.” *Journal for Public Policy Analysis and Management*, 32(1): 107-121.
- Promoting Opportunity Demonstration. (2018). Accessed at: <http://podssa.org/>
- Pronovost P., Needham, D., Berenholtz, S., Sinopoli, D., Chu, H., Cosgrove, S., ... Goeschel, C. (2006). An Intervention to Decrease Catheter-Related Bloodstream Infections in the ICU. *The New England Journal of Medicine*, 355(26), 2725-32.
- Reville, R. T., & Schoeni, R. F. (2003). The Fraction of Disability Caused at Work. *Social Security Bulletin*, 65(4).
- Rhode Island Department of Labor and Training. (2007). *Temporary Disability Insurance Progress Report*. Cranston, RI: RI Department of Labor and Training. Retrieved from http://www.dlt.ri.gov/tdi/pdf/2007TDI_ProgressReport.pdf
- Rhode Island Department of Labor and Training (2017). *2017 Annual Report*. Cranston, RI: RI Department of Labor and Training. Retrieved from <http://www.dlt.ri.gov/pdf/2017DLTAnnualRpt.pdf>
- Robinson, A.R., Hohmann, K.B., Rifkin, J.I., Topp, D., Gilroy, C.M., Pickard, J.A., & Anderson, R.J. (2002). Physician and Public Opinions on Quality of Health Care and the Problem of Medical Errors. *Archives of Internal Medicine*, 162(19), 2186-2190.
- Ryan, E. G., Lamb, S. E., Williamson, E., & Gates, S. (2020). Bayesian adaptive designs for multi-arm trials: an orthopaedic case study. *Trials*, 21(1), 1-16.
- Sinclair, M., O’Toole, J., Malawaraarachchi, M., & Leder, K. (2012). Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches. *BMC medical research methodology*, 12(1), 132.
- Standard Insurance Company. (2019). *Absence and Disability Readiness Index: U.S. Employers’ Report Card on their Practices and Confidence to Deliver on Absence and Disability Programs*.
<https://www.standard.com/eforms/20503.pdf>

- Stapleton, D., Burns, R., Doornink, B., Harris, M., Anfield, R., Cashdollar, W., ... Ufier, K. (2015). *Targeting Early Intervention to Workers Who Need Help to Stay in the Labor Force*. Washington, D.C.: Mathematica Policy Research.
- Stapleton, D., & Christian, J. (2016). Helping workers who develop medical problems stay employed: Expanding Washington's COHE program beyond workers' compensation. *Washington, DC: Mathematica Policy Research*.
- State of California Employment Development Department. (n.d.). FAQ - Part-time, Intermittent, or Reduced Work Schedule. *CA.gov*. Retrieved from https://www.edd.ca.gov/disability/FAQ_PFL_Part-time_Intermittent_Reduced_Work_Schedule.htm.
- Stepner, M. (2019). The Long-Term Externalities of Short-Term Disability Insurance. Retrieved from https://files.michaelstepner.com/short_term_di_externalities.pdf.
- Stern, L. (2017). Post 9/11 veterans with service-connected disabilities and their transition to the civilian workforce: A review of the literature. *Advances in Developing Human Resources*, 19(1), 66-77.
- Stull, D. E., Leidy, N. K., Parasuraman, B., & Chassany, O. (2009). Optimal recall periods for patient-reported outcomes: challenges and potential solutions. *Current medical research and opinion*, 25(4), 929-942.
- Thaler, R. & Sunstein, C. (2009). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New York, NY: Penguin Books.
- U.S. Department of Labor, Bureau of Labor Statistics. (2017). Insurance benefits: Access, participation, and take-up rates. Retrieved from <https://www.bls.gov/ncs/ebs/benefits/2017/ownership/civilian/table16a.htm>. U.S. Department of Labor, Bureau of Labor Statistics. (2019). National Compensation Survey: Employee Benefits in the United States, March 2019. Bulletin 2791. Available at <https://www.bls.gov/ncs/ebs/benefits/2019/employee-benefits-in-the-united-states-march-2019.pdf>
- U.S. Department of Labor, Office of Disability Employment Policy. (2012). Business Strategies that Work: A Framework for Disability Inclusion. Retrieved from <https://www.dol.gov/odep/pdf/BusinessStrategiesThatWork.pdf>.
- U.S. Department of Labor, Office of Disability Employment Policy. (n.d.). S@W/R2W Research & RETAIN Demonstration Projects. Retrieved from <https://www.dol.gov/odep/topics/SAW-RTW/how-to-apply.htm>.
- Unemployment Insurance and Temporary Disability Programs, Rhode Island Code of Regulations, 260-RICR-40-05-1 (<https://rules.sos.ri.gov/regulations/part/260-40-05-1>)
- Vahratian, A., & Johnson, T. R. (2009). Maternity leave benefits in the United States: today's economic climate underlines deficiencies. *Birth (Berkeley, Calif.)*, 36(3), 177-179.
- Waddell, G., & Burton, A. K. (2006). *Is work good for your health and well-being?*. The Stationery Office.
- Waldfoegel, J. (1999). The impact of the family and medical leave act. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, 18(2), 281-302.

- Wickizer, T.M., Franklin, G., Plaeger-Brockway, R., Cheadle, A., & Kyes, K. (1999). *Workers' Compensation Managed Care Pilot Project: Final Report to the Legislature*. Olympia, WA: Washington State Department of Labor and Industries.
- Wickizer, T.M., Franklin, G., Plaeger-Brockway, R., & Mootz, R.D. (2001). Improving the Quality of Workers' Compensation Health Care Delivery: The Washington State Occupational Health Services Project. *Milbank Quarterly*, 79, 5-33.
- Wolinsky, F. D., Miller, T. R., An, H., Geweke, J. F., Wallace, R. B., Wright, K. B., ... & Ohsfeldt, R. L. (2007). Hospital episodes and physician visits: the concordance between self-reports and medicare claims. *Medical care*, 45(4), 300.
- Wright, D., Livermore, G., Hoffman, D., Grau, E., & Bardos, M. (2012). *2010 National Beneficiary Survey Methodology and Descriptive Statistics*. Washington, D.C.: Mathematica Policy Research.
- Young, A.Y., Chaudhry, H.J., Pei, X., Arnhart, K., Dugan, M. & Snyder, G.B. (2017). A Census of Actively Licensed Physicians in the United States, 2016. *Journal of Medical Regulation*, 103(2), 7-21.
- Zare, H., Haffari, G., Gupta, A., & Brinkman, R. R. (2013, January). Scoring relevancy of features based on combinatorial analysis of lasso with application to lymphoma diagnosis. In *BMC genomics* (Vol. 14, No. S1, p. S14). BioMed Central.

Appendix A. Sources for Administrative Data Matched in Option A

The patterns of participation that can be measured in each of the available data sources differ, as do the administrative data that can be matched to in each. The authority to link to administrative data on earnings and federal disability benefits is granted on a case-by-case basis, typically conditional on the linkage providing the owners of the administrative data with information that can improve their products. For Census products, a proposed use of matched data must provide benefits to data programs the Census Bureau conducts under Title 13.⁶⁴ The Census Bureau also holds a variety of data sources owned by other agencies, and can link to those sources if the agency agrees and there is some benefit to the agency. For example, earnings on tax returns can be linked from IRS data held by Census to its nationally representative surveys if both Census and IRS agree, which typically requires making a case that the linkage will be helpful to both Census and IRS in carrying out their mandated responsibilities.

For Option A, the data we base our discussion on is the **Survey of Income and Program Participation (SIPP)** in which individuals are followed for multiple years, even if they leave the initial sample household. This large survey was fielded for three or more years starting in 1984, 1990, 1991, 1992, 1993, 1996, 2001, 2004, 2008, and 2014 (with shorter panels fielded between 1984 and 1990). Each of these surveys is referred to as the SIPP “panel” for that initial year.

Census has previously linked all past years of the Survey of Income and Program Participation to the SSA’s Master Beneficiary Record, the Payment History Update System, the Supplemental Security Record, extracts from the Master Earnings File, 831 disability determination and application records (also known as the 831 Disability Master File, or 831 DMF), and the Numident⁶⁵ file. Permission to use these linked data would need to be acquired from Census, SSA, and Treasury. All of the necessary permissions are often negotiated in one step via Census⁶⁶ or SSA.⁶⁷

For three reasons, survey data matched to administrative data have substantially greater power to detect differences in outcomes by participation pattern, relative to a survey data source without administrative data linked. First, without administrative data on outcomes, the period of sample accumulation in a panel must be much shorter than the panel, because we need to look over a period of time after injury or illness onset at the outcomes. Using administrative data on outcomes, we can accumulate onset over the entire panel. Second, administrative data allows for many years or sometimes decades of follow-up. Third, the administrative data is less subject to measurement error than recall by survey participants, and the increased reliability of measurement translates into smaller required sample sizes to detect a given true impact.

SSA’s Master Beneficiary Record and Payment History Update System could provide information as to whether an individual was eligible for and received Social Security payments due to retirement, disability, spouse retirement or death, parent retirement or death, or some combination of those reasons. These extracts contain both current payment status and historical information so the file would include information as to in which year an individual started receiving benefits and whether the benefits ever

⁶⁴ See e.g. Census (n.d.) “DS002 - Policy on Title 13 Benefit Statements” https://www2.census.gov/foia/ds_policies/ds002.pdf

⁶⁵ The SSA Numerical Identification System (Numident) contains all transactions ever recorded against each Social Security number, for every number ever issued since 1936. Transactions can include a birth, death, or change to demographic information. The file contains information such as name, date of birth, gender, race, place of birth, citizenship status, and date of death.

⁶⁶ For example, see <https://www.census.gov/about/adrm/linkage/guidance.html>.

⁶⁷ As was done, for example, in obtaining data for Bound et al. (2003); Lindner and Nichols (2014); and Nichols et al. (2017).

stopped. SSA's Supplemental Security Record provides the same information about SSI benefits; other data sources are listed at <https://www.ssa.gov/data/>. These datasets combined give us a very accurate picture of who was receiving federal disability benefits and when they were paid. The 831 DMF tells us the result of every step in the process of application for either SSDI or SSI, after step one and before any appeal to the Administrative Law Judge level. Thus, the 831 DMF can tell us the precise date any application by an insured person began, and the nature of decisions taken on that application for almost all relevant decisions.

Like other Census Bureau files such as the Current Population Survey and American Community Survey, the SIPP is matched to administrative records using a Protected Identity Key. The Bureau's Person Identification Validation System relies on matching name, date of birth, gender, and address to administrative data files to find a Social Security number that is then replaced with a Protected Identity Key. The absence of a Protected Identity Key reflects either refusal of consent to match or incomplete or incorrect information that prevents a match from being successful. Survey of Income and Program Participation match rates have varied across panels. The overall match rate is quite high, averaging 83 percent for the 1996, 2001, 2004, and 2008 panels (Favreault & Nichols, 2011). To make the sample representative of the whole population, we would reweight the matched data using a nonresponse adjustment where lack of matched administrative data is considered nonresponse.

The advantage of the SIPP over other surveys is that it follows individuals longitudinally as they move. In contrast, the Current Population Survey and American Community Survey cannot provide representative information on movers over time. On the other hand, the SIPP does not have information on individuals living in group quarters, such as rehabilitation facilities or prisons, unlike the American Community Survey. The American Community Survey also has substantially larger sample sizes than the SIPP, and the sample size alone could privilege it over the SIPP if all longitudinal data were drawn from administrative sources.

A clear advantage to using an existing survey is that the government is already investing significant resources in deploying a data collection effort, saving substantial resources on developing a sample frame and infrastructure to field a new data collection effort. One disadvantage is that it has proven difficult to modify ongoing data collection efforts. The SIPP redesign undertaken for the 2014 panel was the culmination of a decade-long series of convenings, work across numerous government agencies, and many rounds of public comment. Even for a modest extension to ongoing data collection efforts, many stakeholders may weigh in, imposing a different set of costs in time and money in order to secure agreements.