Synthesis of Evidence about Stay-at-Work/ Return-to-Work (SAW/RTW) and Related Programs

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

July 2020

Prepared for:
Chief Evaluation Office and Disability Employment Policy

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

Submitted by:
Austin Nichols
Judy Geyer
Michel Grosz
Zachary Epstein
Michelle Wood
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.

Contents

Glossary of Terms iii
Executive Summary iv
1. Introduction 1
2. Review Methods 3
2.1 Selection of Evidence for Review 3
2.2 Evidence Review Process 5
2.2.1 Ratings for Evidence Strength 5
2.2.2 Ratings of Evidence Relevance 6
2.2.3 Analysis of Impact Sizes 7
2.2.4 Synthesis 10
3. Evidence Findings—Literature Reviews 11
3.1 Narrative Reviews 11
3.2 Systematic Reviews 13
3.3 Comparing Systematic and Narrative Reviews 16
4. Synthesis of Evidence 18
4.1 High Internal Validity Studies 21
4.2 Lower-quality Evidence 26
5. Discussion and Next Steps 29
5.1 Next steps for developing Evaluation Design Options 30
Appendix A. Meta-analytic Average Across All Studies 32
Appendix B. Meta-analytic Regression Methods 34
Appendix C. References 35

List of Exhibits

Exhibit 2.1. Consolidated Standards of Reporting Trials Diagram 4
Exhibit 2.2. Ratings of Internal Validity (quality of evidence) 6
Exhibit 2.3. Ratings of External Validity (relevance of evidence to U.S. labor market) 7
Exhibit 2.4. Program Model Counts in Individual Studies Reviewed 9
Exhibit 4.1. Overall Effect Size Averages 19
Exhibit 4.2. Meta-analytic Regression on Study Characteristics 19
Exhibit 4.3. Studies with Both Employment and Disability Benefits Effect Sizes (ES) Reported 21
Exhibit 4.4. Characteristics of Studies with High-quality Evidence and Greater Relevance 22
Exhibit 4.5. Characteristics of Studies with High-quality Evidence but Lesser Relevance 23
Exhibit 4.6. Meta-analytic Regression Results for Stronger Evidence 25
Exhibit 4.7. Study Characteristics, Low-quality Evidence 27
Exhibit 4.8. Meta-analytic Regression on Study Characteristics, Low-quality Evidence 27

Glossary of Terms

EBEI	Evidence-Based Early Intervention, an intervention that occurs before an injured or ill worker is receiving federal disability benefits, and is based on evidence of effectiveness
Meta-analysis	A numerical method of combining estimates across studies and ascertaining which features of studies explain variation in estimates
Narrative review	A type of review article that does not attempt to examine a systematic or representative sample of studies
RCT	Randomized Control Trial
Review article	A scientific study that describes other studies, rather than analyzing novel data
RTW	Return to work
SAW	Stay at work
Systematic review	A review article that attempts to examine a systematic or representative sample of studies, and may or may not synthesize the studies using meta-analysis

Executive Summary

Stay-at-Work/Return-to-Work (SAW/RTW) programs help workers who experience an illness or injury to remain at work or, if they have separated from employment, return to work as soon as medically feasible. The U.S. Department of Labor (DOL) seeks to build the evidence base about SAW/RTW programs and develop intervention design options and evaluation strategies. Towards that goal, DOL has contracted with Abt Associates to conduct the ***Stay at Work/Return to Work (SAW/RTW) Models and Strategies*** project. To carry out the study, Abt is:

- conducting a comprehensive review of SAW/RTW programs to describe the initiatives that are operating;
- reviewing evidence about the effects of SAW/RTW programs;
- analyzing publicly available data to examine pathways from illness/injury to federal disability benefits to develop early intervention pathways and target population profiles; and
- developing intervention and evaluation design options.

This document is the ***Review and Synthesis of the Evidence about SAW/RTW and Related Programs***. It summarizes evidence regarding the effects of SAW/RTW programs on employment and disability benefits application. This document builds on the conceptual framework and scan of programs presented in the ***Synthesis of SAW/RTW Programs, Efforts, Models, and Definitions (Epstein, et al, 2020).*** The program synthesis document describes the current context in which states and private entities operate SAW/RTW programs and the types of service models they use. In this document, we review the evidence of effectiveness for the programs identified in the synthesis as well as all other evidence found in a broad scan of the relevant literature, including both literature reviews and individual studies.

The evidence reviewed in this document focuses primarily on *early interventions,* which we define as those that occur prior to application for the Social Security Disability Insurance (SSDI). The rationale for examining early interventions is to identify program strategies that might promote workplace retention and prevent long-term disability and the need for federal benefits. It is important to note that such interventions may not necessarily be early relative to the onset of the injury or illness.[1]

We designed this evidence review to inform DOL about what is known about existing program models for the purpose of developing new evaluation design options. We did not limit the evidence reviewed to particular types of research designs. However, to be included in the scan of evidence, a study had to provide sufficient information to assess whether observed effects were due to an intervention of some kind. In practice, we had to limit the synthesis to studies that had some form of comparison group (not necessarily randomly assigned, or even adjusted for baseline differences), because it is not possible to compute a comparable effect size for a single case or pre-post study design (i.e. the only effect size possible would have to assume no changes could occur in the comparison group over time, which makes these effects sizes not comparable to effect sizes computed with a comparison group).

We present the findings to guide future work to develop evaluation design options. Since evaluations provide evidence about program impacts, one goal of this document is to highlight where gaps exist in the

---

[1] Information available about the programs identified in the program synthesis and the evidence review was not sufficient to classify interventions as early or late relative to the onset of injury or illnesses.

evidence. However, the design of future DOL SAW/RTW demonstration programs should maximize the chances of improving outcomes, hence another goal of this document is to identify which program models may produce substantively meaningful positive impacts.

Methods

This evidence review considers 377 original sources (including journal articles, reports, and websites) including sources about the 68 programs described in Epstein et al. We found that only 11 of the programs discussed in the program synthesis have been evaluated in studies reporting direct evidence of impacts. Of these, six are not early interventions, because they involve current disability beneficiaries.[2] We excluded the six late-stage interventions from the evidence review. In Appendix E of Epstein et al, we discuss evidence available from the other five programs that have been evaluated.

Of the 377 sources, we removed 90 that did not include an evaluation, and 35 that did not 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,[3] 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 in Chapter 3.

We included the 87 remaining studies in a meta-analysis of evidence, discussed in Chapter 4. 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); and
- 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,[4] and 56 less relevant) and 15 provided “low-quality” evidence (5 judged more relevant and 10 less relevant).[5] 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 The excluded programs are: 1) Work Incentives and Planning Assistance (Livermore, Prenovitz, and Schimmel 2011), 2) Ticket to Work (Livermore, Mamun, Schimmel, and Prenovitz 2013), 3) Accelerated Health Insurance/Benefits (Michalopoulos et al. 2011), 4) the Youth Transition Demonstration (Fraker et al. 2014), 5) the Job Corps program, and 6) the Kentucky and Minnesota SGA Project Demonstrations (Sevak, Kehn et al. 2017; Sevak, Martin et al. 2017

3 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.

4 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.).

5 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).

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 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. for inferences about effects in the U.S. to be low.

The results of the meta-analytic regression suggest that the higher-quality evidence shows larger impacts for program models that include employment services, such as the Individual Placement and Support model. Our review of systematic reviews indicates that program models have generally positive results for individuals with musculoskeletal conditions, including low back pain, and some models show positive results for individuals with mental illness. However, this may be because the program models more often used for musculoskeletal conditions and mental illness offer employment services. In contrast, we do not see any average advantage in impacts for musculoskeletal conditions and mental illness in our meta-analytic review of studies.

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, i.e. 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 bias[6] in quasi-experimental evidence, though there is no clear pattern as to the nature of that bias.

Implications for the Next Steps in the Project

This review of evidence highlights several open questions for further study. The review finds a lack of evidence about program effects, in the sense that confidence intervals are large (so we cannot precisely predict the impact a future intervention might have). There is also little evidence on what types of programs produce larger impacts, for whom. The review also highlights the importance of obtaining a large sample size for any future evaluation, as most individual studies suffer from low statistical power. Finally, it suggests that employment services and training may be important program components with potential to increase employment rates of people with disabilities. The tradeoffs between a large sample size and a cost-effective evaluation design imply that one attractive design would randomize a low-cost intervention across a very large sample and use administrative data to measure outcomes.

---

[6] 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.

1. Introduction

This report summarizes recent evidence on the effectiveness of Stay-At-Work/Return-To-Work (SAW/RTW) programs to help injured or ill employees gain or retain attachment to the labor force, and to avoid reliance on Social Security Disability Insurance or other federal disability benefits.

- **The goal of this report is to summarize the strength and relevance of findings from evaluations of SAW/RTW programs, as well as the size of the impacts of those programs on employment and receipt of disability benefits.**

The Abt team recorded key features about each study that provided estimates of program effects. First, the team assessed the credibility (or internal validity) of the evidence. Next, the team evaluated each study as to its relevance (or external validity). The team recorded the type of program evaluated in the study and the type of condition or disability examined in the study. The team then synthesized the evidence by looking across this large number of studies to describe how estimated effects vary on average. These findings will inform evaluation design options that the team will develop in Task 4 of the project.

Epstein et al, ***Synthesis of SAW/RTW Programs, Efforts, Models, and Definitions,*** catalogued the types of programs being implemented by a variety of stakeholders in a wide range of settings—state agencies, private employers, and private insurers. That document deliberately considered a broad range of SAW/RTW programs or strategies and did not impose any restrictions on whether researchers had evaluated the program. The program synthesis describes programs that include demonstrations that have concluded, demonstrations that are currently active, programs with evaluations in early stages, and even programs with no planned evaluations. The team used four sources to identify those programs and strategies: (1) a review of previous work completed under the S@W/R2W Policy Collaborative, supported by ODEP (ODEP 2018), (2) a structured online search for literature from academic and non-academic sources, (3) consultation with experts in SAW/RTW interventions from the private and public sectors, and (4) targeted web searches of six large employers and 10 private disability insurers. Altogether, these methods identified 68 SAW/RTW programs. We excluded six of those programs from this evidence review because they are interventions targeted to SSDI beneficiaries.

The set of programs covered by the evidence we reviewed for this report is larger than the set of programs discussed in Deliverable 2.1. This report includes evidence that pertains to (1) one-time, temporary programs or strategies, (2) smaller-scale efforts focused, for example, on particular medical diagnoses or professions, and (3) programs and initiatives in contexts outside the U.S.

This report provides a comprehensive summary of the evidence available about SAW/RTW programs and evaluation activities currently underway. The findings from the program synthesis and this evidence review pertain to a wide range of programs, programmatic enhancements, and innovative policies that we will consider when developing evaluation design options. In the program synthesis we included both programs with strong evidence and programs with no prior evaluations and those with relatively low-quality evidence. Similarly, we consider all types of programs to develop evaluation design options, not only those with a strong evaluation record.

The remainder of this report proceeds as follows:

- Chapter 2 summarizes the selection criteria and review methods.

- Chapter 3 presents a summary of the evidence available in 28 review articles identified in our search.
- Chapter 4 synthesizes evidence from 87 studies identified in our search using meta-analysis.
- Chapter 5 summarizes conclusions of the review.

Appendix A displays the results of the random effects meta-analyses in figures. Appendix B discusses the technical methods we used in the meta-analysis and Appendix C lists all studies included in the review.

2. Review Methods

Towards developing future intervention and evaluation options, this review seeks to identify evidence of the effect of SAW/RTW initiatives on employment and receipt of federal disability benefits. Further, we want to understand how evidence of the effect of SAW/RTW initiatives might vary by the evaluation design, which we discuss further below.

Section 2.1 explains how we selected evidence for review from the many sources we considered. Section 2.2 explains how we evaluated the strength and relevance of the evidence studies, and how we summarize the impact sizes.

2.1 Selection of Evidence for Review

We identified the studies reviewed in three ways:

1. We searched for and reviewed evidence (articles and websites) for the 68 programs identified during the search process for the program synthesis.[7] However, we found that only 11 programs had direct evidence of impacts, and of these, 6 are not early interventions. We excluded evidence about the six late-stage interventions from this this review. We discuss the evidence about the five early intervention programs in Appendix E of Epstein, et al. For some of these five programs, we found multiple studies with direct evidence of impacts, and we include each in the sample of studies reviewed in Chapter 4 of this report.
2. Because we found scant evidence of effectiveness for the 68 programs identified in the program synthesis, we searched the academic literature for evidence about other programs, much of which focuses on programs operating outside the U.S[8]. When conducting the search for the program synthesis, we found several review articles, research syntheses, and meta-analyses. These types of articles were not relevant to the program synthesis. However, we included all studies cited in these articles and published in 2008 or later in this evidence review. We limited the scope of the review to studies published in the last 10 years to ensure that we considered the most relevant evidence—in which findings are still applicable, and where the technologies, laws and public policies examined are still relevant. We also included articles referenced in those reviews. A small proportion of these articles supplied evidence of effectiveness, which we review in Chapter 4.
3. The review articles cited many articles in the *Journal of Vocational Rehabilitation* and in the *Journal of Occupational Rehabilitation*. Since the most recent articles in those journals had not been included in the already-published reviews, we searched all articles in those two journals from 2016 to March 2018 for evaluations of RTW/SAW programs.

   - **These three search processes identified 377 articles and websites for review.**

---

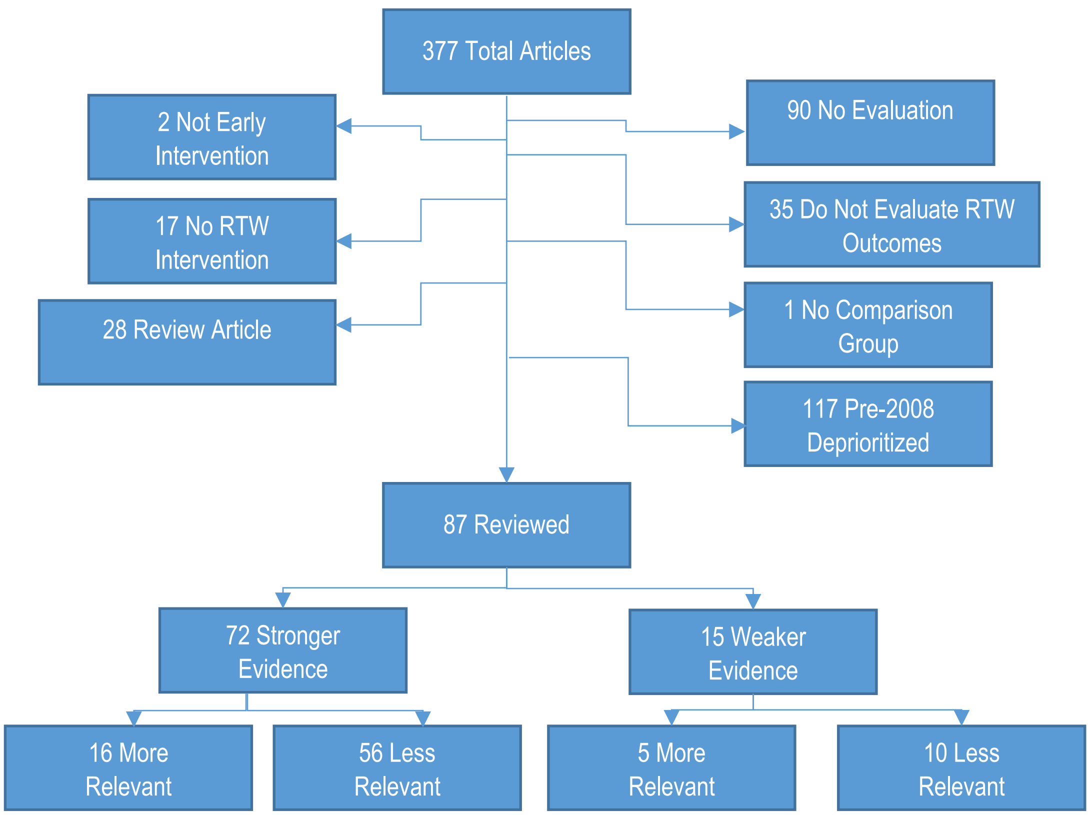
7 We searched for the programs by name, in Google Scholar.

8 We searched Google and Google Scholar to identify the 68 programs using these terms: stay at work, return to work, preferred worker, occupational health, workers compensation, short-term disability, vocational rehabilitation, plan for employement, job accommmodations, and case managementused.

Exhibit 2.1 provides a Consolidated Standards of Reporting Trials (Schulz et al. 2010) diagram that shows the number of articles or websites reviewed and counts of articles or websites discussed in Chapters 3 and 4. Of the 377 identified articles, we removed 90 that did not include an evaluation, and 35 that did not 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[9], 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 discuss in Chapter 3.

We included the 87 remaining articles in the meta-analysis of evidence reported in Chapter 4. Of the 87 articles, 72 provided "high-quality" evidence (16 judged more relevant, and 56 less relevant) and 15 provided "low-quality" evidence (5 judged more relevant and 10 less relevant), using the standards described in the next section.

Exhibit 2.1. Consolidated Standards of Reporting Trials Diagram



Chapter 3 describes findings from the 28 review articles that are themselves reviews of evidence. These reviews of evidence are not included as independent sources of evidence in the systematic review of evidence. We synthesize findings from the 87 individual studies using meta-analysis in Chapter 4.

9 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.

2.2 Evidence Review Process

To review the evidence, the Abt team developed a five-step process and applied it uniformly across all 87 articles reviewed.

- **We labeled each study (1) as having "high-quality or low-quality evidence" and (2) as being "more or less relevant." We then (3) recorded numerical estimates of employment and disability receipt, and coded (4) program model and (5) disability type examined in each study.**

Sections 2.2.1 and 2.2.2 describe how we assign internal validity ratings and external validity ratings. Section 2.2.3 describes how we code the reviewed studies with respect to program impacts on employment and disability benefits (including application or receipt of federal disability benefits). Since these are the twin goals (increasing employment and decreasing benefit receipt) of a SAW/RTW program, we have limited the measures we examine to only these two outcome measures.

2.2.1 Ratings for Evidence Strength

Evidence is "high-quality," i.e., has high "*internal validity,*" if it convincingly establishes that an intervention *causes* the difference in mean outcomes reported (including the case where there is no difference in outcomes); otherwise, it is low-quality.[10] The most rigorous studies use random assignment to a treatment condition that receives the intervention or to a control condition that does not receive the intervention.[11] These studies also maintain low attrition in both assignment groups over the course of the study. Evaluations that use random assignment designs allow us to expect that differences in outcomes for the treatment and control groups reflect only the influence of the intervention, and not any underlying differences in the composition of the two groups.

Several review standards offer researchers methods for evaluating the internal validity of a study. Internal validity refers to how well a study can establish causal effects of the intervention on the outcomes measured, that is, what we refer to as "quality" of the evidence. Our internal validity ratings range from 0 (no evidence) to 5 (most rigorous evidence) and are explained below in Exhibit 2.2. 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).[12]

---

10 Strong evidence does not mean evidence of positive impacts; here it means evidence that, due to the design and implementation of the study, is expected to be free from selection bias.

11 Typically, the control condition is "usual care," or business as usual, but sometimes is a waiting list for treatment or another active treatment.

12 Unlike a full What Works Clearinghouse review, our reviewers did not complete a review workbook for each study reviewed, each study was reviewed by only one reviewer, and reviewers did not evaluate the reliability of the outcome measures of each study. This omission is purposeful: we are searching for all evidence regarding two specific outcomes (employment and disability benefit receipt). Unlike a What Works Clearinghouse or Clearinghouse for Labor Evaluation and Research review, we assign ratings from 0 to 5, and then further label those ratings as "strong," "weak," or "no" evidence.

Exhibit 2.2. Ratings of Internal Validity (quality of evidence)

5	Experimental design, or Randomized Control Trial, with: – Attrition is less than 65% and differential attrition is less than 10%; AND – Intent to treat analysis (no noncompliance, or failures to adhere to RCT protocol); AND – No confounding factors; AND – Symmetric outcome data collection in treatment and control groups; AND – Analysis includes statistical adjustments for selected baseline characteristics on which equivalence between experimental groups was not established; AND – Confirmation that intervention occurred with fidelity to design/description.
4	Experimental design with: – No confounding factors; AND – Symmetric outcome data collection in treatment and control groups; AND – Attrition is less than 65% and differential attrition is less than 10%.
3	Quasi-experimental design, such as matched comparison group, difference-in-difference, or regression discontinuity designs, or experimental designs not included above, with: – Concurrent comparison cases; – Established baseline equivalence of treatment and comparison groups OR adjustment for baseline differences; – Symmetric outcome data collection in treatment and control groups; and – Confirmation that intervention occurred with fidelity to design/description.
2	Quasi-experimental design with poor equivalence of treatment & comparison group (standardized differences exceeding 0.25 on any baseline variables); or QED with unknown baseline equivalence, including time series or pre-post studies with no comparison group.
1	Focus groups, interviews, case studies.
0	No outcomes measured.

Note: This deliverable labels studies with an evidence quality (internal validity) rating of 4 or higher as "high-quality" evidence, and studies with an evidence quality (internal validity) rating of 1 to 3 as "low-quality" evidence. Studies with a rating of 0 are categorized as providing "no" evidence.

2.2.2 Ratings of Evidence Relevance

The "external validity" of an evaluation is the ability to generalize research findings broadly to all types of disabilities and across the U.S. Our external validity ratings range from 0 (not generalizable to U.S. contexts) to 5 (nationally representative) and are explained in Exhibit 2.3 below. 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.).

Exhibit 2.3. Ratings of External Validity (relevance of evidence to U.S. labor market)

5	Nationally representative of workers with a wide variety of work impairments.
4	State-wide sample representative of workers with a wide variety of work impairments (for example, can include Workers Compensation claimants only).
3	State or nationally representative of workers in specific occupation or industry.
2	Injury- or illness-specific, for example lower back pain only or mental illness only.
1	Site, occupation, and injury-specific.
0	International programs, including nationally representative studies in a non-U.S. context, or a review article including many levels above.

This deliverable labels studies with an evidence rating of 2 or higher as "more relevant" and other studies as "less relevant."

2.2.3 Analysis of Impact Sizes

We recorded impacts on the two outcomes of interest: employment and disability benefits. To enable us to characterize how impacts on the two outcomes vary across studies, we also record two other study features: the type of disability examined and program model(s) used in the intervention.

- **We analyzed programs' effects on employment and disability benefit use.**

By Impact. For studies providing evidence, we analyzed programs' effect on employment and disability application or receipt, i.e., the impact estimates. Out of 87 studies that provide evidence, 82 reported impacts on employment, and 13 reported impacts on SSDI application or receipt of disability benefits. Impacts are reported for binary outcomes (e.g., working vs not working), and typical follow-up periods range from six months to two years, though some studies report on follow-up over five years.

Some studies report impacts at multiple follow-up periods; we form a weighted average of each study's reported estimates using a within-study meta-regression (regressing impact on follow-up time, using the methods explained in Appendix B) and assign each study's predicted estimate to the longest follow-up period. This puts studies that report multiple estimates on an equal footing with studies that report a single estimate of impact, and predicts using the best measure of impact and the variance for that estimate (excluding the empirical Bayes or posterior mean estimates of random effects).

Researchers reported impacts on employment and benefit receipt in many different ways: hazard or risk ratios, odds ratios or log odds ratios, percentage point differences, percent increases, and other metrics. We converted all impact estimates on employment or benefit receipt to log odds ratios, so we can interpret estimates (roughly) as percent changes in odds ratios. The log odds ratio is a standard measure of effect size, which puts all effects on a common scale.[13] The log odds ratio also produces symmetric interpretations. For example, an intervention that increases the proportion working from 40 to 50 percent would have an effect size of 0.4, and an intervention that decreases the proportion working from 50 to 40 percent would have an effect size of -0.4 which is not true of other scale-free summary measures. If we instead measured the percentage change in the proportion working in the first case, we would calculate a

[13] The log odds is a good effect size measure because it is approximately normal even when underlying data are very skewed. Also known as the logit transformation, the log odds for a proportion p is the natural log of the odds p/(1-p), or ln(p/(1-p)) and the log odds ratio for two proportions p and q is ln(p/(1-p)) minus ln(q/(1-q)).

25 percent increase (for a 10 percentage point change), and in the second case, we would calculate a 20 percent decrease (for a 10 percentage point change).

The log odds ratio is the only measure in which the numerical estimate is unaffected by how we define the outcome group (e.g., those at work, or those not at work). We also estimated the precision of those estimates (i.e., the standard error of that log odds ratio) using the number of study participants in the treatment and comparison groups.[14]

- **We also coded for disability type and for program model.**

By Type of Disability. In addition to recording impacts for each study, we coded three types of disability: musculoskeletal, mental health conditions, and other. "Other" includes conditions that are neither musculoskeletal nor mental health. "Other" also includes studies that examined a broad range of disability types, possibly including musculoskeletal and mental health conditions among a wider range of disability types. A study that includes a variety of conditions, but explicitly mentions musculoskeletal conditions as being included, would be coded 1 (true) for both musculoskeletal and other.

By Program Model. In addition to recording impacts and specific types of disability examined for each study reviewed, we also recorded five indicators characterizing the programs as belonging to one or more of the five broad program models defined in Deliverable 2.1 (counts of how often each indicator was used in the studies included in subsequent analysis is shown in Exhibit 2.4):

1. ***Employer-provided accommodations***: "Reasonable accommodations," as included in Title I of the Americans with Disabilities Act (ADA), allow individuals with a disability to successfully perform their job duties.[15] Common examples include physical changes to a workplace, the provision of assistive technologies, or modifications to the policies and procedures that determine a worker's duties. Using this broad definition, we coded the SAW/RTW interventions that include employer-provided accommodations.
2. ***Financial incentives for employers and workers***: The accommodation strategies discussed above approach the issue of SAW/RTW only as a matter of an individual's capability to carry out his/her job duties and do not attempt to shift the worker and the employer's financial motivations regarding a return to work. The financial returns to work, for the worker and the employer, are conceptually distinct from functional capacity or capability to work. Therefore, we coded separately those SAW/RTW programs that offered financial incentives for employers or workers to encourage employment. We identified two primary types of employer-targeted incentives: those implemented through WC programs and those implemented through the tax code. We also identified policies operated through workers' compensation and short-term disability insurance programs to lessen disincentives to work that arise for workers. These 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

---

14 For two groups with sample sizes N_1 and N_2, with relevant proportions p_1 and p_2, the standard error of the log odds ratio is $[(N_1p_1)^{-1} + (N_1(1- p_1))^{-1} + (N_2p_2)^{-1} + (N_2(1- p_2))^{-1}]^{1/2}$.

15 "Accommodations," U.S. Department of Labor, accessed February 22, 2018, https://www.dol.gov/odep/topics/Accommodations.htm

incentives to employers to offer the worker an easy return to work. . For more discussion of types of incentives, see Epstein et al, ***Synthesis of SAW/RTW Programs, Efforts, Models, and Definitions***.

3. ***Information***: Information programs offer services delivered by a "Return-to-Work Coordinator," case manager, or other individual that facilitates communication between stakeholders. We also include programs that assist workers to navigate the disability benefits system and other services, such as those available through the health care or workforce development system. This component also includes "technical assistance" services typically delivered by an industry expert to help employees understand their benefits or employers to modify their workplace. In some cases, the programs and interventions identified in our review are comprised only of an information component. These also included technical assistance strategies to assist employers, or technical assistance to American Job Centers.
4. ***Medical Management***: The programs coded in this category attempt to facilitate or improve upon the delivery of medical care with a focus on occupational health to promote employment retention. Examples include providing guidance to physicians on best practices in occupational health, incorporating a physician's input into transitional or alternative work arrangements, or lowering the cost of medical care through subsidized insurance premiums. Services also often include incentives for altering occupational health practices, best practice guidelines, or educational services to change behavior.
5. ***Employment Services and Training***: Employment services facilitate entry into a new occupation/position, likely following functional rehabilitation and typically delivered during a period of unemployment. This component includes re-employment services to assist with navigating the labor market, such as job search, resume writing, and interviewing assistance. These programs can also include higher intensity services like occupational skills training delivered either in a classroom, lab, or an on-the-job setting.

Exhibit 2.4 tabulates the frequency of each component across the 87 studies reviewed. Information components are the most commonly studied, followed by medical management components. Financial incentives are the least commonly studied component. In all, we identified more than 120 types of services in review, which means some programs included more than one type of service. Roughly 40 percent of studied programs included more than one type of service.

Exhibit 2.4. Program Model Counts in Individual Studies Reviewed

Employer-provided accommodations	29
Financial Incentives to employers or workers	3
Information	45
Medical management	34
Employment Services/Training	17

Note: Counts are based on tabulations of results from authors' review of 87 studies. Roughly 40 percent of the programs that were evaluated in the studies we reviewed included more than one type of service.

2.2.4 Synthesis

- **We use a meta-regression to describe how impacts vary with program model, disability type, and evidence quality.**

As we discuss in Chapter 3, it appears that the available evidence may not support the conclusions in many review articles. For this evidence review, we use a formal meta-analytic framework to discipline our conclusions. We chose the framework that is least susceptible to manipulation to synthesize the findings from all identified studies. We combined the impact estimates and the program model information in a meta-analytic regression framework (DerSimonian and Laird 1986; Knapp and Hartung 2003). Specifically, we use a random-effects meta-regression to look for whether certain program components or disability types are associated with greater impacts on employment or SSDI application. For reasons explained in Chapter 3, we split the evidence into stronger and weaker evidence for some of the analysis in Chapter 4, to mitigate the effect of selection bias in each study.

3. Evidence Findings—Literature Reviews

Our search for evidence yielded 28 review articles. In this chapter we discuss evidence available in these review articles. In general, the literature reviews that have examined the body of evidence on SAW/RTS programs find that most studies provide no evidence, and for those that do provide plausible evidence of impacts, the impacts are close to zero in most cases and do not differ statistically from zero. The chapter focuses on two types of literature reviews: narrative reviews and systematic reviews. Narrative reviews, do not attempt to systematically collect and summarize all existing evidence. Systematic reviews carefully document the collection process for evidence. These systematic reviews describe all studies regardless of the findings, though some authors exclude studies with lower quality designs in any quantitative analysis of effects and meta-analyses give less weight to studies with larger standard errors.

Previewing the findings to be presented below, some findings are consistent across narrative and systematic reviews. Each type of review finds evidence that for musculoskeletal conditions; especially those involving low back pain, a variety of interventions are effective. In addition, each type of review finds evidence that, for individuals with mental illness, supported employment models are effective, but not other interventions. Beyond those two findings, the high-quality evidence tends to show that interventions are ineffective or have very small impacts (in the sense of modest estimated effect sizes that tend to be statistically not distinguishable from zero). It is only when averaging across studies that we find statistically significant evidence of modest impacts.

3.1 Narrative Reviews

Narrative reviews are common, partly because there is a paucity of high-quality studies reporting outcome data with a comparison group design. As Johnston et al. (2009) and Johnston, Sherer, & Whyte (2006) point out, there are very few high-quality meta-analytic reviews and most studies do not provide evidence with good internal validity, in which outcomes can plausibly be traced to the intervention being studied. Because such narrative reviews are a common form of evidence, we discuss some of the better-known reviews here. Before doing so, we emphasize that the selection of studies for these reviews is not systematic or representative of all evidence. Further, many of these studies do not connect the numerical estimates to the description of the findings. That is, the stated finding of a narrative review may contradict the numerical evidence in the average impact of cited studies. The balance of this section provides a pair of examples for these claims.

Huber et al. (2017) claim that:

> *A review of the literature in the field published over the past few decades reveals a steady increase in outcomes research demonstrating the positive impact of specific interventions in vocational rehabilitation services (Bolton, Bellini, & Brookings, 2000; Dean et al., 2014; Fleming et al., 2013; Glasner-Edwards& Rawson, 2010; Leahy et al., 2014; O'Neil, Mamun, Potamites, Chan, & Cardoso, 2014; Pruett et al., 2008) and in other rehabilitation-related areas (Bond, 2004; Huber et al., 2009; Marshall & Lockwood, 1998; Workman et al., 2012)"* (Huber et al. 2017; p.12).

However, Bolton, Bellini, & Brookings (2000) did not show impacts, but rather reported a correlation between receipt of rehabilitative services and future employment. Another cited paper, Dean et al. (2014), describes evidence that does not support the claim that the paper is cited to support. First Dean et al. (2014, p.2) state that the:

> *results may suggest that vocational rehabilitation programs can effectively reduce SSDI/SSI roles for certain subgroups, (p.2)*

Then the authors write that they:

> *caution against drawing this type of causal conclusion… [as] this analysis is largely descriptive, and does not formally address the fundamental methodological problem involved in drawing such inferences from observational data.*

- **This highlights a common feature of narrative reviews: that impact estimates in cited articles do not always support the broad claim made in the narrative.**

Stapleton et al. (2015) has similar issues. Many claims in the paper lack evidence to support them. On page 4, the paper claims:

> *Considerable success can be achieved by providing supports to workers with [musculoskeletal] conditions (particularly [low back pain]), mental health conditions, or other chronic conditions for which adherence to treatment is critical.*

However, the paper's summary of the evidence does not necessarily support that statement. Table II.1 cites Hullegie and Koning (2014) as providing evidence of positive impacts for individuals age 40–58, especially males. The paper states:

> *One intriguing finding from the Netherlands: that country's early intervention system appears to have had greater success among male workers between the ages of 40 and 58 who experienced unexpected hospitalizations than for other workers who experienced unexpected hospitalizations—males ages 25–39 or females in either age group. (Stapleton et al. 2015, p.4).*

The reforms in the Netherlands enhanced employer incentives, by making employers responsible for paying sickness benefits and by strengthening monitoring obligations. However, the reforms were universal, leaving no comparison group. Hullegie and Koning (2014) instead compare return to work after an unscheduled hospital admission before and after the reforms, across demographic groups, and conclude older male workers return at relatively higher rates after the reform, but this is driven by falling rates of work among younger workers.[16] Thus, the Hullegie and Koning (2014) paper does not truly provide evidence of positive impacts, or better impacts for male workers between the ages of 40 and 58.

Similarly, Stapleton et al. (2015) show a summary table showing early intervention having positive impacts for musculoskeletal conditions and other low mortality conditions (Williams et al. 2007, Waddell et al. 2008, Van Oostrom et al. 2009, Wickizer et al. 2011, Dibben et al. 2012, Hoefsmit et al. 2012,

---

[16] Hullegie and Koning (2014), "For disabled men aged 25–39 the relative employment rates initially increased, but this was reversed in 2005, the year in which the law that makes employers responsible for the first two years of sickness pay became effective. The relative employment rate of disabled men aged 40–58 seems unaffected by the reforms."

Bevan 2015, Linton et al. 2016, Richmond et al. 2015), mental health conditions (Killackey et al. 2008, Dibben et al. 2012, Hoefsmit et al. 2012, Rost et al. 2004), and chronic conditions requiring adherence to treatment.[17] The paper further claims "Numerous studies point to employer cooperation as a key component of success in staying at work or returning to work" (Stapleton et al. 2015, p.5) though no evidence is adduced. Many of the cited studies do not provide evidence of impact (e.g., Linton et al. 2016), and others review evidence without regard to study quality and conclude with mixed findings, for example, that "the only area that currently offers sufficient evidence for effect-size meta-analysis appears to be in the area of workplace-based interventions for low back pain, where there are several early and recent meta-analyses" (Dibben et al. 2012, p.26).

- **It is difficult to draw a reliable conclusion from narrative reviews, as they neither provide an unbiased selection of studies, nor do they formally synthesize numerical findings across studies.**

In summary, this discussion suggests that the selection of studies in narrative reviews makes it impossible to judge whether the selection creates a bias toward positive or negative findings. In addition, the absence of a formal weighing of evidence on a common metric make it difficult to assess whether claims of the common findings in the literature are in fact supported by the preponderance of evidence. For this reason, we use narrative reviews primarily to identify additional studies that we should include in our review.

3.2 Systematic Reviews

In contrast to narrative reviews, systematic reviews attempt to summarize all evidence on a topic, recent evidence, or a representative sample of that evidence. However, because of the comprehensiveness of systematic reviews, many of them limit their attention to a specific type of disability or intervention. In many of these reviews, a limit is placed on the scope in one way (e.g., only care coordination models are considered, which we consider an "informational" program model), but the studies are further disaggregated in another way (e.g., for cases where the primary condition is chronic low back pain). This makes it difficult to draw conclusions about how average impacts vary as these multiple features of a study vary. Here we consider 10 such reviews, to illustrate these points and highlight the need for the analysis conducted in Chapter 4. Because each study defines the interventions in scope in a different way, the studies examined may overlap, and distinctions across systematic reviews and their conclusions are frequently obscure. We discuss these reviews primarily to motivate the analysis that follows, not to suggest any of these findings should be taken as dispositive.

The two broadest reviews we found were Waddell, Burton, and Kendall (2008) and Kuoppala and Lamminpää (2008). Waddell, Burton, and Kendall (2008) reviewed 450 studies published from 2000–2007 on interventions related to "common health problems that account for two-thirds of long-term sickness (mild/moderate musculoskeletal, mental health and cardiorespiratory conditions)" that could

---

[17] The references regarding chronic conditions requiring adherence to treatment are: Birnbaum et al. (2010) and Burton et al. (2007) on adherence with antidepressant use; Carls et al. (2012) on adherence for diabetes, hypertension, dyslipidemia, asthma/ chronic obstructive pulmonary disease and congestive heart failure; Gifford et al. (2014) review on adherence for diabetes, cardiovascular, depression; Hagen et al. (2014) on oral hypoglycemic agents; Jinnett et al. (2012) on medication adherence and filing short-term disability claims; Loeppke et al. (2011) on adherence to statins in CAD and absenteeism; and Wagner et al. (2012) on adherence to antihypertensive medication regime.

improve "work outcomes (staying at, returning to and remaining in work)." For the purposes of their review, "vocational rehabilitation was defined as whatever helps someone with a health problem to stay at, return to and remain in work," and we may thus interpret this review as including all types of SAW/RTW programs. The authors conclude that interventions can improve work outcomes for those with musculoskeletal conditions. However, they claim that there is little evidence in the study that treatments for mental illness improve work outcomes,[18] and there is no evidence of impact for cardiorespiratory conditions.

Kuoppala and Lamminpää (2008) found "moderate evidence" in 45 studies that "return-to-work programmes decrease long sick leaves" with a risk ratio (RR)[19] of 0.46, and a range of 0.25–1.10. They also conclude that multimodal rehabilitation decreases the risk of disability pension (RR 0.64, range 0.52–1.14). However, the risk ratio confidence intervals include 1, so we should conclude the findings were not statistically significant. That is, while the point estimates indicate promising impacts, we cannot conclude by any reasonable standard of statistical significance that the results were not due to chance.

- **Systematic reviews that disaggregate by condition or type of intervention find some impacts, though estimated impacts tend to be small.**

We examined systematic reviews focusing on interventions targeting musculoskeletal conditions. Van Vilsteren et al. (2015) found, for workers with musculoskeletal disorders, that workplace interventions reduced time to first return to work and cumulative duration of absence from work. The review also examined time to lasting return to work and the authors found only low-quality evidence for workplace interventions. Palmer et al. (2011) conducted a systematic review of community- and workplace-based interventions to reduce sickness absence and job loss in workers with musculoskeletal disorders. Interventions included prescribed exercises, behavioral change techniques, workplace modifications, and provision of extra services. Most interventions appeared modestly beneficial: participants were more likely to avoid job loss related to musculoskeletal disorders, and the median reduction in sickness absence was about one day per month. However, RTW was not significantly higher.

Other systematic reviews examine interventions for workers with mental health conditions. Nieuwenhuijsen et al. (2014) synthesized evidence on interventions to improve return to work among individuals with depression. The study found "moderate quality[20] evidence that a work-directed intervention added to a clinical intervention reduced sickness absence (SMD -0.40; 95% CI -0.66 to -0.14; three studies)."[21] Enhancing primary or occupational care with cognitive behavioral therapy reduced sick leave, per the authors, who "found moderate quality evidence based on three studies that telephone or online cognitive behavioral therapy was more effective in reducing sick leave than usual primary or occupational care (SMD -0.23; 95% CI -0.45 to -0.01)." However, after reviewing the literature for adults

---

18 "There is limited evidence that symptomatic treatments for depression (by medication, psychotherapy or a combination of both, and including Cognitive Behavioural Therapy) in themselves improve occupational outcomes and no clear evidence on the magnitude of any effect" Waddell, Burton, and Kendall (2008; p.21). This conclusion motivates our exploration of all impacts in Chapter 4.

19 The risk ratio (RR) is a measure of effect size commonly used when comparing across studies, which is similar to the odds ratio when both risks are small.

20 Nieuwenhuijsen et al. (2014) "judged 14 studies to have a high risk of bias and nine to have a low risk of bias."

21 SMD refers to Standardized Mean Difference, or difference in means divided by the pooled standard deviation, a summary statistic used in meta-analysis when studies assess the same outcome but measure it in different ways. The SMD allows analysts to standardize the results of studies to a uniform scale.

with adjustment disorders (sometimes called exogenous, reactive, or situational depression, which occurs when an individual is unable to adjust to or cope with a particular stress or a major life event) Arends (2012) found that cognitive behavioral therapy did not significantly speed return to work compared with no treatment.

Some systematic reviews find employment-oriented interventions can produce positive effects on employment for those with severe mental health conditions.

Some systematic reviews have found that employment-oriented interventions (which we may think of as related to the program synthesis concept of Employment Services or Training) can produce positive effects on employment for those with severe mental health conditions. Suijkerbuijk et al. (2017) conclude that supported employment produces better outcomes than medical treatment alone for individuals with severe mental illness.

Studies focused on chronic pain reach a variety of conclusions. Examining Cognitive-Behavioral treatments or cognitive behavioral therapy for low back pain, Richmond et al. (2015, p.14) conclude that in "the majority of studies, when compared to other typical physiotherapy-based treatments, a cognitive behavioral intervention is more effective" at reducing pain, but "when a single study with large effect sizes was removed from [the] analyses, the pooled estimates were reduced and were more consistent with previous meta-analyses of cognitive behavioral versus other active treatments." A similar study by Kamper et al. (2015) found that "Multidisciplinary biopsychosocial rehabilitation interventions were more effective than usual care (moderate quality evidence) and physical treatments (low quality evidence) in decreasing pain and disability in people with chronic low back pain. For work outcomes, multidisciplinary rehabilitation seems to be more effective than physical treatment but not more effective than usual care." Williams et al. (2012) reviewed evidence on psychological therapies for the management of chronic pain and found "Cognitive behavioral therapy but not behavior therapy has small effects on disability associated with chronic pain."

Two recent reviews of return-to-work programs examine coordination of care (which is included in the program synthesis concept of an informational intervention). Schandelmaier et al. (2012) reviewed nine trials from seven countries (eight focused on musculoskeletal conditions and one on mental conditions) that generally followed participants for 12 months or less. The study found that "moderate quality evidence suggested a benefit of return to work coordination on proportion at work at end of follow-up (risk ratio = 1.08, 95% CI = 1.03 to 1.13; absolute effect = 5 in 100 additional individuals returning to work, 95% CI = 2 to 8)." A more recent review by Vogel et al. (2017) of return-to-work programs that incorporate coordination of care searched 4,261 articles, found 183 articles with potentially relevant findings, and analyzed 14 of them that provided experimental findings on these return-to-work programs.

The two review articles on coordination models included the same nine studies, but Vogel et al. (2017) add more studies then separately aggregate by length of follow-up, which leads to (1) finding no significant impacts overall and (2) a pattern of point estimates that shift from favoring RTW programs at short-run (less than one year) follow-up to favoring usual practice at longer run follow-up periods. This may be due to the two studies (Lindh 1997 and Jensen 2012) that have follow-up beyond one year, but also produced negative point estimates at one year.

There are two studies in each of the two reviews (Schandelmaier et al. 2012 and Vogel et al. 2017) of coordination models that find positive impacts. These two studies are not examining programs that clearly differ from those in other studies, except insofar as they focus on musculoskeletal conditions and low back pain. As Vogel et al. (2017, p.21) report, "Eliminating two outliers, both clearly in favour for return-to-work coordination programmes (Donceel 1999; Lambeek 2010), reduced heterogeneity for time to return to work to a moderate level, even though these programmes showed no obvious differences to other studies." On balance, return-to-work programs that rely on case coordination seem to offer no advantage over alternatives, except possibly when treating only musculoskeletal conditions and low back pain. All of the cited studies that meet our inclusion criteria for the evidence review are included in the meta-analysis in Chapter 4.

- **We find many instances of systematic reviews on the same narrow topic reaching contradictory conclusions.**

In short, the evidence is mixed. However, none of the systematic reviews look across all of the available evidence; instead, each focuses on a pre-specified subset of available evidence.

3.3 Comparing Systematic and Narrative Reviews

The previous two sections show strongly contrasting results. Narrative reviews tend to conclude that programs work; most systematic reviews conclude that programs do not work. In the prior section, we illustrated that the systematic reviews that do conclude interventions work often do not find statistically significant impacts, or are overturned by subsequent re-analysis. This section attempts to explain the divergence between the takeaways from narrative and systematic reviews.

To weigh the available evidence, the systematic reviews try to retrieve all studies meeting explicit criteria. The alternative is that researchers draw studies according to some implicit criteria, e.g., what he/she remembers or what he/she considers important. In as much as doing so draws on prior opinion, prior opinion can unconsciously influence the inclusion of evidence (Wason 1960; Kunda 1999). However, systematic reviews tend to exclude weaker evidence that may show promising results for programs not yet evaluated using an experimental design. A useful middle ground is a systematic review that does not exclude nonexperimental evidence, a form of which we report in Chapter 4.

- **There is substantial evidence of selection bias in individual studies, and selected sampling in reviews.**

Some of the weaker evidence referred to in narrative reviews shows impacts that are inconsistent with the stronger evidence. For example, there is almost no evidence that interventions that rely on coordination of care are effective, in the sense of detecting positive impacts, when we examine only high-quality evidence. Yet Wickizer et al. (2011) conclude that the COHE model produces large gains in return to work and reductions in application for disability benefits. This may be because the study uses a difference-in-difference design with substantial baseline differences between the treated and comparison groups. Additionally, Clayton et al. (2011) reviewed 31 studies which evaluated initiatives with an individual focus (improving an individual's employability or providing financial support in returning to work) and found that "many of the studies suffered from selection bias[22] into these programmes of more work-ready claimants."

[22] Note that in this context, selection bias refers to participants having intrinsically better outcomes on average (e.g. there are unobserved factors that promote better outcomes and also lead people to participate in a program), so that a comparison of outcomes makes program effects more positive than the true causal impact.

These findings motivate our choice to divide the evidence we present in Chapter 4 into stronger and weaker evidence to mitigate the effect of selection bias in each study. Even when effects measured via these two different types of studies do not differ statistically, combining studies that suffer from selection bias with those that do not is not strictly justified in a meta-analytic framework, as we cannot assert that the underlying effect being measured is the same in the two kinds of studies. Splitting the studies into high- and low-quality evidence enables us to see whether the underlying effect being measured seems to differ across the two kinds of studies, both qualitatively, and in the sense of a statistical difference.

This is one kind of selection bias that plagues individual studies. Another kind of selection bias is a problem in the selection of studies: that certain kinds of studies are drawn for review, due to implicit cognitive bias.

4. Synthesis of Evidence

To better understand the body of evidence about SAW/RTW programs, we use a random-effects meta-analytic regression to synthesize findings across all the 87 studies we identified that provide any evidence of impact.[23] This is the most efficient means of summarizing numerical estimates across studies and allows us to compare across multiple study features at once.

- **We use a meta-analytic regression to summarize findings.**

We measure all effects on the log odds scale, for both binary outcomes (employment or disability benefit receipt), which ensures estimates are scale-free and can be compared across disparate settings. We first compute a single effect size estimate for each outcome (employment or disability receipt) in each study. Each study reviewed that provides estimates at multiple follow-up points provides one data point, which we construct separately via a within-study random-effects meta-analytic regression. Then we impute the predicted effect size and the standard error of the prediction to the last date of follow-up, so we have one aggregated point estimate per study, per outcome.[24]

We first explore the overall average impacts across all reviewed studies. Because our review of narrative and systemic review articles indicates a high risk of selection bias in individual studies, we next explore the overall average impacts for those with high-quality evidence (Internal Validity rating at least 4, for experimental designs). Third, we explore the overall average impacts for low-quality evidence (lower Internal Validity rating, for studies not using an experimental design). For each, we examine impacts on both employment and disability benefits.

Columns 1 and 3 of Exhibit 4.1 show the average impacts. Specifically, they are the results of meta-analyses (or equivalently, meta-regressions with just a constant). Models that include an indicator for weaker evidence are shown in columns 2 and 4. For neither model is the impact on disability benefits statistically distinguishable from zero. Apparently this is in part because each included study has low statistical power. However, the average impact on employment is large,[25] and the average impact is not statistically distinguishable across stronger and weaker designs in column 2. That we cannot distinguish strong and weak evidence statistically in this model does not imply that we must combine these studies. That weaker evidence tends to have lower magnitude impacts could reflect negative selection into interventions, that is, individuals with worse average outcomes are more likely to participate in the studied interventions. In any case, we believe it is useful to analyze the 72 stronger and 15 weaker studies separately, and we do so in the next two sections.

---

23 The random-effect meta-analytic regression is a generalization of meta-analysis: with just a constant term in the regression, a random-effect meta-analytic regression is a random-effects meta-analysis. The results of the random effects meta-analyses corresponding to columns 1 and 3 in Exhibit 4.1 are displayed as figures in Appendix A. See Appendix B for details on methods.

24 With multiple estimates per study, one can include multiple estimates from each study in the overall meta-regression and adjust standard errors for correlations with study, but it is more straightforward to first aggregate the estimates within a study to a single meta-analytic average, so we have one estimate per study. The random effects meta-analytic average incorporates the chance variation across estimates in its estimate of the overall impact that would be estimated across many hypothetical re-estimations.

25 As an example, an increment to the log odds of one half corresponds to an increase in the probability of working from 50 percent to 62 percent, or from 10 percent to 15 percent.

- **The average impact on employment is positive, and higher quality studies tend to estimate larger impacts.**

Exhibit 4.1. Overall Effect Size Averages

	(1)	(2)	(3)	(4)
	Employment Effect Size		Disability Benefits Effect Size	
LowQualityDesign		-0.181 [-0.539,0.176]		0.0514 [-0.391,0.494]
Constant (Mean effect)	0.459*** [0.313,0.605]	0.499*** [0.334,0.663]	-0.0983 [-0.298,0.101]	-0.117 [-0.385,0.152]
Observations	82	82	13	13
I2	0.626	0.630	0.876	0.877

Note: Not all studies report impacts for both outcomes, so the sample sizes refer to the studies that report on the relevant outcome. 95 percent confidence intervals in brackets, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Exhibit 4.2 reports estimates from models that disaggregate effects on employment. Column 1 disaggregates by program model, column 2 by disability type, and column 3 by length of follow-up data collection. Only program type (column 1) shows statistically significant variation, though point estimates in column 2 suggest positive effects for musculoskeletal conditions could be driven by selection and the point estimate on follow-up time in column 3 suggests impacts may decline with increasing follow-up time. There is no statistically significant difference (the p-value for a joint F test for all terms involving Low-Quality Design appears in the last row of the table) between stronger and weaker evidence in this pooled model, but the only two significant coefficients lead to two different conclusions for the two types of evidence. This leads us to conclude we should analyze these groups of studies separately.

Exhibit 4.2. Meta-analytic Regression on Study Characteristics

	(1)	(2)	(3)
	Employment	Employment	Employment
LowQualityDesign	0.139 [-0.760,1.038]	-0.146 [-0.633,0.340]	-0.105 [-0.715,0.505]
Accommodation	-0.146 [-0.509,0.218]		
Financial Incentive	-0.458 [-1.267,0.351]		
Information	0.165 [-0.285,0.615]		
Medical Management	-0.109 [-0.534,0.317]		
Employment Services, Training	1.170*** [0.638,1.702]		
LowQualityDesign*Accommodation	0.0515 [-0.777,0.880]		

	(1)	(2)	(3)
	Employment	Employment	Employment
LowQualityDesign*FinancialIncentive	0.154 [-0.970,1.278]		
LowQualityDesign*Information	-0.366 [-1.238,0.505]		
LowQualityDesign*MedicalManagement	0.391 [-0.330,1.111]		
LowQualityDesign*EmploymentServicesTraining	-1.243* [-2.188,-0.299]		
Musculoskeletal		-0.496 [-1.501,0.509]	
Mental		-0.354 [-1.407,0.699]	
Other Disability		-0.520 [-1.640,0.601]	
LowQualityDesign*Musculoskeletal		0.736 [-0.981,2.452]	
LowQualityDesign*Mental		-0.155 [-1.045,0.735]	
Followup			-0.0181 [-0.222,0.186]
LowQualityDesign* Followup			-0.0300 [-0.318,0.258]
Constant	0.293 [-0.198,0.784]	0.963 [-0.123,2.049]	0.523*** [0.230,0.816]
Observations	82	82	82
p-value of joint test on all LowQualityDesign Variables/Interactions	0.104	0.725	0.718

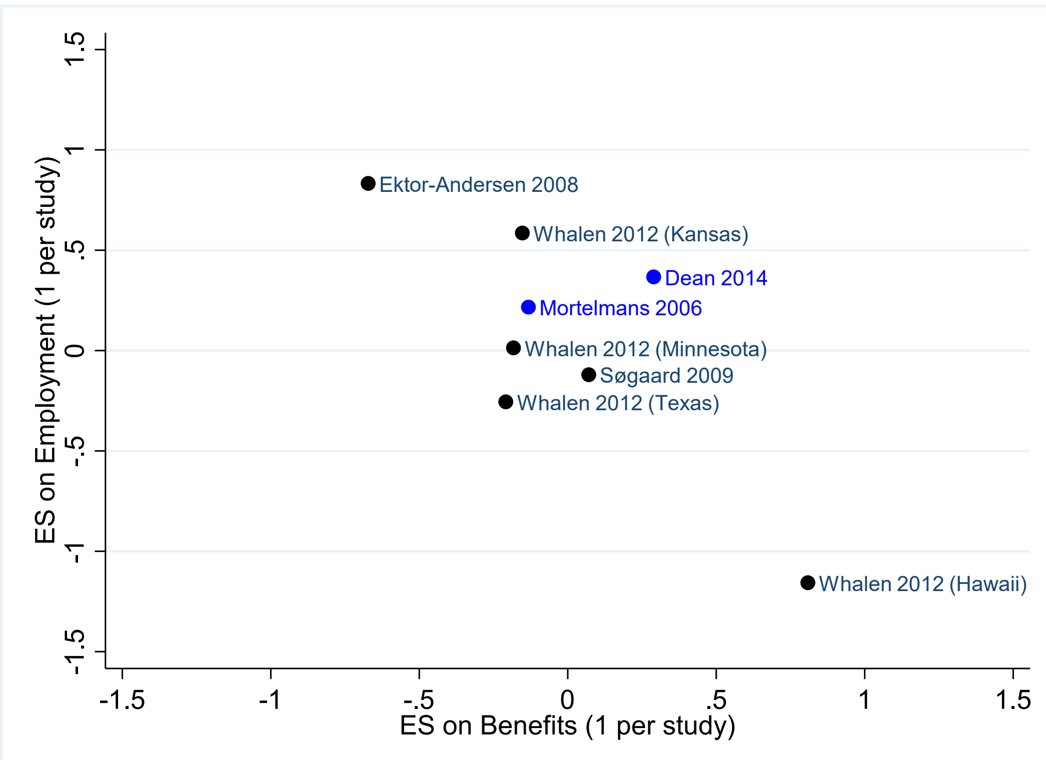
Note: Not all studies report impacts for both outcomes, so the sample sizes refer to the studies that report on the relevant outcome. 95 percent confidence intervals in brackets, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The difference between all high-quality and low-quality evidence is not statistically significant, in the other pooled meta-analytic regressions (final row of columns 2 and 3). Frequently, patterns in the data are overlapping, and we cannot distinguish statistically[26] which of several competing hypotheses we can reject about what explains differences in impact estimates across studies. To explore these patterns further, we separately analyze the high-quality and low-quality evidence in the following two sections of the document.

26 In a meta-analytic regression on all study characteristics, we sacrifice too many degrees of freedom, and confidence intervals become very large; only the coefficient on Employment Services and Training differs statistically from zero. But the regression with all study characteristics included many estimated coefficients, and we should expect at least one to differ statistically from zero based on chance alone. The many-dimensional pooled meta-analytic regression also suffers excess heterogeneity across studies, and we do not report it here.

There are very few studies with both employment and benefits impacts reported, shown on a scatterplot in Exhibit 4.3, where the two weaker designs are shown as blue dots, and they are bracketed by the stronger evidence, shown as black dots. Statistically, one cannot reject that this pattern is a cloud of random noise, with mean zero in both dimensions. A priori, we may expect a negative correlation between effect sizes on employment and benefits, if interventions that are effective will tend to increase employment and lower benefit receipt and other interventions will tend to decrease employment and raise benefit receipt rates. That we observe no statistically significant pattern of this type may suggest that some interventions work better than others or it may simply reflect sampling variation due to the low power of studies reviewed.

Exhibit 4.3. Studies with Both Employment and Disability Benefits Effect Sizes (ES) Reported



Note: High-quality evidence shown in black and low-quality evidence in blue.

In the following two sections, we separately analyze the high-quality evidence first, including only experimental designs, then the low-quality evidence, including only quasi-experimental designs.

4.1 High Internal Validity Studies

This section considers high internal validity studies. Ideally, we would examine a body of evidence that has extremely high internal validity (credibly causal impacts) and external validity (nation- or state-wide representative sample of multiple types of disability), but only one exhibits both. The Kansas Demonstration to Maintain Independence and Employment, described in Whalen (2012), encompasses a wide variety of conditions or disability types, and has a statewide sample, for an external validity rating of 4. In Hawaii, the demonstration study recruited individuals with diabetes, and in Minnesota and Texas, participants all had behavioral health conditions (Whalen 2012), so these studies are rated 2 on the external validity metric (site- or disability-specific interventions). Similarly, the remainder of the high internal validity studies have lower external validity, with 15 rated as 2 on the external validity metric, and 52 others rated zero because they are in settings outside the U.S.

For the 72 studies with stronger designs that include sufficient data to compute effect sizes and standard errors, we list the 16 studies rated with high internal validity and greater external validity in Exhibit 4.4 below and the 56 studies rated with high internal validity and lesser external validity in Exhibit 4.5 below.

Exhibit 4.4. Characteristics of Studies with High-quality Evidence and Greater Relevance

Cite	Location	Measures Employment	Sample Size (Employment Impact)	Measures Disability Applications	Sample Size (Disability Applications Impact)	Musculoskeletal	Mental Health	Other	Accommodation	Financial Incentive	Information	Medical Mgmt.	Employment Serv./Training	Internal Validity	External Validity
Bohman 2011	Texas	Yes	1,480	No	N/A	0	1	0	0	1	1	0	0	5	4
Bond 2015	Illinois	Yes	87	No	N/A	0	1	0	1	0	0	0	1	4	2
Davis 2012	Alabama	Yes	85	No	N/A	0	0	1	1	0	0	0	1	4	2
Godges 2008	California	Yes	36	No	N/A	1	0	0	0	0	1	0	0	4	2
Lerner 2015	USA	Yes	380	No	N/A	0	1	0	0	0	1	0	0	4	2
Matchar 2008	USA	Yes	437	No	N/A	0	0	1	0	0	1	0	0	4	2
Rosenheck 2017	USA	No	N/A	Yes	363	0	0	1	0	0	0	1	1	4	2
Sacks 2008	Pennsylvania	Yes	198	No	N/A	0	1	0	0	0	1	0	0	4	2
Schoenbaum 2001	USA	Yes	1,158	No	N/A	0	1	0	0	0	0	1	0	5	2
Smith 2002	USA	Yes	219	No	N/A	0	1	0	0	0	0	1	0	5	2
Twamley 2012	USA	Yes	58	No	N/A	0	0	1	1	0	0	0	1	4	2
Wang 2007	USA	Yes	604	No	N/A	0	1	0	0	0	1	0	0	5	2
Whalen 2012	Minnesota	Yes	451	Yes	500	0	1	0	0	0	1	1	0	5	2
Whalen 2012	Kansas	Yes	826	Yes	1,155	0	0	1	0	0	1	1	0	5	4
Whalen 2012	Texas	Yes	1,146	Yes	1,585	0	1	0	0	0	1	1	0	5	2
Whalen 2012	Hawaii	Yes	146	Yes	184	0	0	1	0	0	1	1	0	5	2

Note: N= 16 studies.

Exhibit 4.5. Characteristics of Studies with High-quality Evidence but Lesser Relevance

Cite	Location	Measures Employment	Sample Size (Employment Impact)	Measures Disability Applications	Sample Size (Disability Applications Impact)	Musculoskeletal	Mental Health	Other	Accommodation	Financial Incentive	Information	Medical Mgmt.	Employment Serv./Training	Internal Validity	External Validity
Andersen 2015	Denmark	Yes	54	No	N/A	1	0	0	0	0	0	1	0	5	0
Bejerholm 2015	Sweden	Yes	87	No	N/A	0	0	1	1	0	0	0	1	4	0
Bethge 2010	Germany	Yes	169	No	N/A	1	0	0	0	0	1	0	0	4	0
Bethge 2011	Germany	Yes	149	No	N/A	1	0	0	0	0	1	0	0	4	0
Bjorneklett 2013	Sweden	Yes	236	No	N/A	0	0	1	0	0	1	0	0	5	0
Brendbekken 2017	Norway	Yes	284	No	N/A	0	0	1	0	0	1	0	0	5	0
Brouwers 2006	Netherlands	Yes	194	No	N/A	0	1	0	0	0	1	1	0	5	0
Bültmann 2009	Denmark	Yes	113	No	N/A	1	0	0	0	0	1	0	0	5	0
Busch 2011	Sweden	No	N/A	Yes	110	0	0	1	0	0	1	0	0	4	0
Carlsson 2013	Sweden	Yes	33	No	N/A	1	1	0	0	0	0	1	0	4	0
Ektor-Andersen 2008	Denmark	Yes	575	No	N/A	0	0	1	0	0	1	0	0	5	0
Fleten 2006	Norway	Yes	990	No	N/A	1	1	0	0	0	1	0	0	5	0
Hagen 2010	Norway	Yes	238	No	N/A	1	0	0	0	0	0	1	0	5	0
Henchoz 2010	Switzerland	Yes	67	No	N/A	1	0	0	0	0	1	0	0	4	0
Heslin 2011	England	Yes	135	No	N/A	0	1	0	1	0	0	0	1	5	0
Hoffmann 2014	Switzerland	Yes	100	No	N/A	0	1	0	0	0	0	0	1	4	0
Hubbard 2013	Scotland	Yes	18	No	N/A	0	0	1	0	0	1	0	0	4	0
Jensen 2012	Denmark	Yes	351	No	N/A	0	0	1	1	0	0	1	0	4	0
Killackey 2008	Australia	Yes	41	No	N/A	0	1	0	1	0	0	0	1	5	0
Kittel 2008	Germany	Yes	300	No	N/A	0	0	1	0	0	0	1	0	5	0
Krogh 2009	Denmark	Yes	83	No	N/A	0	0	1	0	0	0	1	0	5	0
Lambeek 2010	Netherlands	Yes	134	No	N/A	0	0	1	1	0	0	1	0	4	0
Li-Tsang 2008	China	Yes	63	No	N/A	0	0	1	0	0	1	0	1	5	0
Linton 2016	Sweden	Yes	140	No	N/A	1	0	0	0	0	1	0	0	4	0
Martin 2013	Denmark	Yes	168	No	N/A	0	0	1	1	0	1	0	0	5	0
Michon 2014	Netherlands	Yes	148	No	N/A	0	1	0	1	0	0	0	1	5	0
Myhre 2014	Norway	Yes	406	No	N/A	0	0	1	1	0	0	1	0	4	0
Netterstrom 2013	Denmark	Yes	140	No	N/A	0	1	0	0	0	1	0	0	4	0
Noordik 2013	Netherlands	Yes	143	No	N/A	0	1	0	0	0	0	1	0	5	0
Ntsiea 2015	South Africa	Yes	80	No	N/A	0	0	1	1	0	1	0	0	5	0
Odeen 2013	Norway	Yes	93	No	N/A	1	0	0	0	0	1	0	0	4	0
Oshima 2014	Japan	Yes	37	No	N/A	0	1	0	1	0	0	0	1	4	0
Pedersen 2015	Denmark	Yes	400	No	N/A	0	1	0	0	0	1	0	0	4	0
Rebergen 2009	Netherlands	Yes	240	No	N/A	0	1	0	0	0	0	1	0	5	0
Rossignol 2000	Montreal	Yes	110	No	N/A	1	0	0	1	0	0	1	0	4	0

Cite	Location	Measures Employment	Sample Size (Employment Impact)	Measures Disability Applications	Sample Size (Disability Applications Impact)	Musculoskeletal	Mental Health	Other	Accommodation	Financial Incentive	Information	Medical Mgmt.	Employment Serv./Training	Internal Validity	External Validity
Schene 2007	Netherlands	Yes	59	No	N/A	0	0	1	0	0	1	0	0	5	0
Scholz 2016	Switzerland	No	N/A	Yes	8,050	0	0	1	1	0	0	1	0	5	0
Schultz 2008	Canada	Yes	34	No	N/A	1	0	0	1	0	1	1	1	4	0
Shete 2012	India	Yes	44	No	N/A	1	0	0	0	0	1	1	0	4	0
Streibelt 2014	Germany	Yes	102	No	N/A	1	0	0	0	0	1	0	0	5	0
Søgaard 2009	Denmark	Yes	836	Yes	836	0	1	0	0	0	0	1	0	5	0
Tamminga 2013	Netherlands	Yes	133	No	N/A	0	0	1	1	0	0	1	0	4	0
Thunnissen 2008	Netherlands	Yes	108	No	N/A	0	0	1	0	0	1	0	0	4	0
Tsang 2009	Hong Kong	Yes	111	No	N/A	0	1	0	1	0	0	0	1	4	0
Van Oostrom 2010	Netherlands	Yes	145	No	N/A	0	0	1	1	0	0	0	0	4	0
Van Beurden 2015	Netherlands	Yes	119	No	N/A	0	0	1	0	0	0	1	0	4	0
Varekamp 2011	Netherlands	Yes	115	No	N/A	0	0	1	1	0	0	0	0	5	0
Vlasveld 2013	Netherlands	Yes	126	No	N/A	0	1	0	1	0	1	0	0	4	0
Volker 2015	Netherlands	Yes	220	No	N/A	0	1	0	1	0	0	1	0	4	0
Vonk 2014	Netherlands	Yes	209	No	N/A	0	0	1	0	0	1	0	0	4	0
Waghorn 2014	Australia	Yes	208	No	N/A	0	1	0	1	0	0	0	1	4	0
Willert 2011	Denmark	Yes	102	No	N/A	0	0	1	0	0	1	0	0	4	0
Wong 2008	Hong Kong	Yes	92	No	N/A	0	1	0	1	0	0	0	1	4	0
van der Klink 2003	Netherlands	Yes	153	No	N/A	0	1	0	0	0	1	0	0	4	0
van der Feltz-Cornelis 2010	Netherlands	Yes	60	No	N/A	0	1	0	0	0	0	1	0	4	0
van Vilsteren 2017	Netherlands	Yes	150	No	N/A	0	0	1	1	0	1	0	0	5	0

Note: N = 56 studies

Exhibit 4.6 below summarizes the results of a random-effect meta-analytic regression[27] of effects on employment across all 72 studies. The average impact across studies (weighted by relative precision) is 0.514 on the log odds scale, a large impact.[28] This corresponds to an increase in employment rates from 53.5 percent (average control group employment across studies) to 65.8 percent. Disaggregating this impact by study characteristics—using meta-analytic regression in column 2—implies that programs with an employment services and training component have substantively and statistically significantly higher impacts than programs without this component, and the average program with no employment services and training component has a substantially lower impact. The average program with no employment

27 See appendix B for methods.

28 There is no universally accepted definition of "large" impacts, but an effect size of about 0.20 is often called "small" while 0.50 is called "medium" and 0.80 is considered "large" (Hill et al. 2007).

services and training component has an impact that is not statistically significantly different from zero (since the constant term, and all other coefficients, are not statistically different from zero). Put differently, the overall positive impact of the programs hides the fact that employment programs have a large impact and other programs have a smaller impact that is not distinguishable from zero.[29]

Because there are so few estimates of impacts on application for or receipt of disability benefits, we are not able to estimate the way in which those impacts vary with study characteristics and can only report the average impact (column 3). The result does not differ statistically from zero, but the point estimate is large relative to other programs designed to lower the rates of disability receipt among beneficiaries (-0.0878 on the log odds scale corresponds to a decrease in disability benefit receipt rates from 17 percent to 15.8 percent). The confidence interval includes both relatively large decreases in receipt of disability benefits (e.g., a reduction from 17 to 14.3 percent) and modest increases in receipt of disability benefits (e.g., an increase from 17 to 17.5 percent).

That the estimate for disability receipt has a smaller magnitude than the estimate for employment may indicate that reductions in the fraction of people applying for disability benefits are harder to achieve than increases in the fraction working, for a variety of reasons. It may be that there are more marginal workers than there are marginal applicants for disability benefits. This is consistent with the view that the tradeoff between work and nonwork is a close call for many workers with disabilities (and high disutilities of work given labor market conditions) but that disability benefits are relatively well-targeted at the group facing the most severe barriers to work.

Exhibit 4.6. Meta-analytic Regression Results for Stronger Evidence

	(1)	(2)	(3)
	Employment	Employment	Disability Benefits
Accommodation		-0.143 [-0.552,0.266]	
Financial Incentive		-0.474 [-1.414,0.467]	
Information		0.172 [-0.307,0.651]	
Medical Management		-0.111 [-0.565,0.343]	
Employment Services, Training		1.176*** [0.594,1.759]	
Musculoskeletal		-0.246 [-1.104,0.612]	
Mental		-0.259 [-1.144,0.627]	
Other		-0.257 [-1.212,0.697]	

29 These findings are robust to including indicators for specific levels of internal or external validity, or maximum length of follow-up.

	(1)	(2)	(3)
	Employment	**Employment**	**Disability Benefits**
Constant	0.514*** [0.335,0.693]	0.561 [-0.519,1.640]	-0.0878 [-0.209,0.0330]
Observations	69	69	9
I^2	0.656	0.497	0.253
p-value for test of H_0 of zero heterogeneity	$1.30\text{x}10_{-14}$	0.00000841	0.218

Note: 95 percent confidence intervals in brackets, * *p* < 0.05, ** *p* < 0.01, *** *p* < 0.001

4.2 Lower-quality Evidence

Having considered the studies with high-quality evidence above, we now consider the remaining 15 studies with lower quality evidence (Exhibit 4.8). The impacts (all coefficients in Exhibit 4.8) do not differ statistically from zero, except for the overall impact on employment (without disaggregating by study features). Further, the pattern of point estimates is not even broadly consistent with the findings from the stronger evidence. For example, the overall impact on employment has a confidence interval whose upper bound of 0.431 is lower than the point estimate 0.514 in Exhibit 4.6, column 1. However, we see from Exhibit 4.1 that this difference is not itself statistically significant. No estimates in column 2 are significantly different from zero, but the sign of the point estimates differs, with four of the seven coefficients we can compare differing in sign: information and employment services or training have positive point estimates in high-quality evidence but negative in low-quality evidence, whereas studies examining programs using medical management or targeting musculoskeletal conditions exhibit the opposite pattern. Of course, we can also refer to tests of differences in Exhibit 4.2 to see that programs with employment services and training component are associated with large positive impacts on employment in high-quality evidence but no effect in low-quality evidence, and this is the only statistically significant difference across the two bodies of evidence.

The differences across the two bodies of evidence may be due to a correlation between the quality of implementation of an intervention, and the quality of evidence generated, or it could be due simply to chance. In summary, we can conclude little from the body of low-quality evidence (studies with internal validity rated lower than 4) alone, without reference to the body of high-quality evidence. We might have expected that selection bias would lead us to find uniformly positive impacts in the low-quality evidence, while effects were more muted in the high-quality evidence, but we do not see this pattern. If anything, the high-quality evidence shows larger impacts overall, but there is substantial heterogeneity across studies and types of studies that makes finding any statistically significant differences hard at best. The tremendous variability of results further underlines that the variance of outcomes for any one study is high, and one should avoid cherry-picking results that look favorable.

Exhibit 4.7. Study Characteristics, Low-quality Evidence

Cite	Location	Measures Employment	Sample Size (Employment Impact)	Measures Disability Applications	Sample Size (Disability Applications Impact)	Musculoskeletal	Mental Health	Other	Accommodation	Financial Incentive	Information	Medical Management	Employment Serv./Training	Internal Validity	External Validity
Braathen 2007	Norway	Yes	193	No	N/A	0	0	1	0	0	1	1	0	3	0
Dean 2014	Virginia	Yes	5,163	Yes	5,163	0	0	1	0	1	0	1	1	2	2
Dewa 2009	Canada	Yes	124	No	N/A	0	1	0	1	0	0	1	0	2	0
Grossi 2009	Sweden	Yes	22	No	N/A	0	1	0	0	0	0	1	0	3	0
Hamersma 2008	Wisconsin	No	N/A	No	N/A	0	0	1	0	1	0	0	0	3	4
Howard 2009	Texas	Yes	1,441	No	N/A	0	0	1	1	0	0	0	0	3	2
Jensen 2013	Denmark	Yes	200	No	N/A	1	1	0	0	0	1	0	0	3	0
Karlson 2010	Sweden	Yes	148	No	N/A	0	1	0	1	0	1	0	0	2	0
Karrholm 2008	Sweden	No	N/A	Yes	128	0	0	1	0	0	0	1	0	2	0
Lander 2009	Denmark	Yes	161	No	N/A	0	1	0	0	0	1	0	0	3	0
Langi 2017	USA	Yes	2,521	No	N/A	0	0	1	0	0	0	0	1	3	4
Linden 2014	Germany	Yes	609	No	N/A	0	0	1	0	0	1	0	0	2	0
Mortelmans 2006	Belgium	Yes	1,564	Yes	1,564	0	0	1	0	0	1	0	0	3	0
Stapelfeldt 2011	Denmark	Yes	120	No	N/A	0	0	1	1	0	0	1	0	3	0
Wickizer 2011	Washington	No	N/A	Yes	71,696	0	0	1	0	0	1	1	0	3	4

N= 15 studies with low-quality evidence

Exhibit 4.8. Meta-analytic Regression on Study Characteristics, Low-quality Evidence

	(1)	(2)	(3)
	Employment	Employment	Disability Benefits
Accommodation		-0.0200 [-0.742,0.702]	
Financial Incentive		-0.314 [-0.879,0.252]	
Information		-0.251 [-0.942,0.441]	
Medical Management		0.330 [-0.185,0.844]	
Employment Services, Training		-0.0844 [-0.653,0.484]	
Musculoskeletal		0.458 [-0.667,1.583]	
Mental		-0.0857 [-0.841,0.670]	

	(1)	(2)	(3)
	Employment	**Employment**	**Disability Benefits**
Constant	0.296*** [0.161,0.431]	0.425 [-0.157,1.008]	-0.0400 [-0.725,0.645]
Observations	13	13	4
I^2	0.353	0.277	0.962
p-value for test of H_0 of zero heterogeneity	0.100	0.227	0

Note: Not all studies report impacts for both outcomes, so the sample sizes refer to the studies that report on the relevant outcome. t statistics in parentheses, * p < 0.05, ** p < 0.01, *** p < 0.001. 95 percent confidence intervals in brackets, * p < 0.05, ** p < 0.01, *** p < 0.001

5. Discussion and Next Steps

This document has reported the results of a review of the evidence concerning stay-at-work/return-to-work programs. Specifically, the review has considered evidence about the 28 review articles, and 87 individual studies. The studies that present high-quality evidence of impact do not tend to allow generalization to the entire United States, either because they represent a small geographic area, or take place outside the U.S. This presents challenges to generalizing from the existing evidence.

To address this challenge of synthesizing the evidence, we used meta-analytic regression to review the available evidence in the 87 individual studies formally and objectively. The results of that meta-analytic regression suggest that the stronger evidence shows larger impacts for program models that include employment services, such as the Individual Placement and Support model. Our review of systematic reviews indicates that program models have generally positive results for individuals with musculoskeletal conditions, including low back pain, and some models show positive results for individuals with mental illness. However, this may be because the program models more often used for musculoskeletal conditions and mental illness offer employment services.

In contrast, we do not see any average advantage in impacts for musculoskeletal conditions and mental illness in our meta-analytic review of studies. A caveat to this suggestion is that there is insufficient statistical power due to the small number of studies to rule out quite large systematic advantages or disadvantages for other program types, or by disability type. In addition, longer follow-up tends to be negatively associated with measured impact, though in most groups of studies, we cannot separately identify the difference in measured impacts due to follow-up, program model, and disability type, because there simply are not enough studies to conduct the required meta-analytic regression.

Further, differences between systematic reviews (most of which employ meta-analysis or meta-analytic regression) and narrative reviews in the conclusions drawn indicates a large risk of bias in non-systematic reviews. Differences between findings based on high-quality and low-quality evidence are significant, but we cannot identify any clear pattern, i.e., quasi-experimental designs do not exhibit systematically larger or smaller impact estimates, though the experimental and quasi-experimental designs are associated with different programs models, and each single difference across these designs has a very large confidence interval.

Because of the low statistical power inherent in this synthesis due to the paucity of evidence, we must be careful not to conclude that "absence of evidence" for impacts or differences in impacts is the same as "evidence of absence" of impacts. In some sense, there is at least a very small impact of any program, whether positive or negative, and there are differences across programs (though the differences may be very small), and our failure to detect impacts or differences in impacts means first and foremost that our confidence intervals are too large, primarily because samples are too small to detect the differences that may exist.

However, the highest quality evidence that is most relevant to the U.S. labor market comes from a handful of studies concerning the Individual Placement and Support model for individuals with mental illness, which incorporates extensive employment services and counseling, and the Demonstration to Maintain Independence and Employment evaluations conducted in Kansas, Minnesota, Texas, and Hawaii. While the Individual Placement and Support studies find large positive impacts on employment, on average, the Demonstration to Maintain Independence and Employment studies find little to no impact and have confidence intervals small enough to rule out impacts comparable to those found in the Individual Placement and Support studies.[30]

The range of negative and positive findings could indicate that programs with imperfect targeting or poor implementations yield negative impacts on employment (or increase disability benefit receipt). On the other hand, the wide range of findings may simply result from sampling variation. To be sure, precision is low for most studies, which tends to indicate studies are underpowered, a common design flaw. Meta-regression can overcome this problem only to some extent. Overall, perhaps the most important finding from the current evidence review is that low-quality evaluation designs and small sample sizes have provided little conclusive evidence about the effects of SAW/RTW programs. There is a robust finding from high-quality evidence that programs that involve employment services or training do have large impacts, but most differences across programs could be large in both negative and positive directions. This finding poses challenges for developing evaluation options.

5.1 Next steps for developing Evaluation Design Options

Future evaluation designs must take this challenge seriously and must weigh tradeoffs among quality of research design, necessary sample sizes, and evaluation costs. A substantially larger sample size allows researchers to construct tight bounds on impacts, including in some cases for subgroups. However, the gain in precision per increase in sample falls rapidly as the sample size gets larger, because we need a fourfold increase in sample to achieve a doubling of precision. Estimating impacts for subgroups requires even larger sample sizes. For example, estimating impacts for men and women would require four times as large a sample if men and women are equal fractions of the study sample, but far more than four times as much sample if the fractions are very unequal.

However, carrying out evaluations with sufficiently large sample sizes can be costly—if the researchers must recruit and enroll participants, conduct primary data collection, and program operators must implement the intervention at a large scale. The costs of attaining the large samples needed to detect intervention impacts pose a tradeoff for pursuing evaluation designs that will produce high internal validity. A traditional experimental research design (with participant recruitment, random assignment, and primary data collection) yields high internal validity, but has costs that increase approximately at a linear rate as the sample size increases while precision increases much more slowly. On the other hand a nonexperimental evaluation may be less costly to carry out with large sample sizes because they do not have costs that increase linearly in sample size. However, a nonexperimental design sacrifices internal validity as well as the causal inferences we would like to draw about impacts. As we consider evaluation design options, the team will consider approaches that can reduce the costs of an experimental evaluation to some extent, for example, using administrative data sources instead of surveys to measure outcomes. Another alternative is to exploit digital interventions (such as emails or other zero-cost electronic messaging) or low-cost nudges (such as redesigned forms or application or reporting processes) where the cost does not increase linearly in sample size, but these interventions often have smaller impacts as well.

These tradeoffs imply difficult choices when designing future evaluation options. Given the superior properties of a large sample size and concerns about the cost of an evaluation, we recommend that large-

---

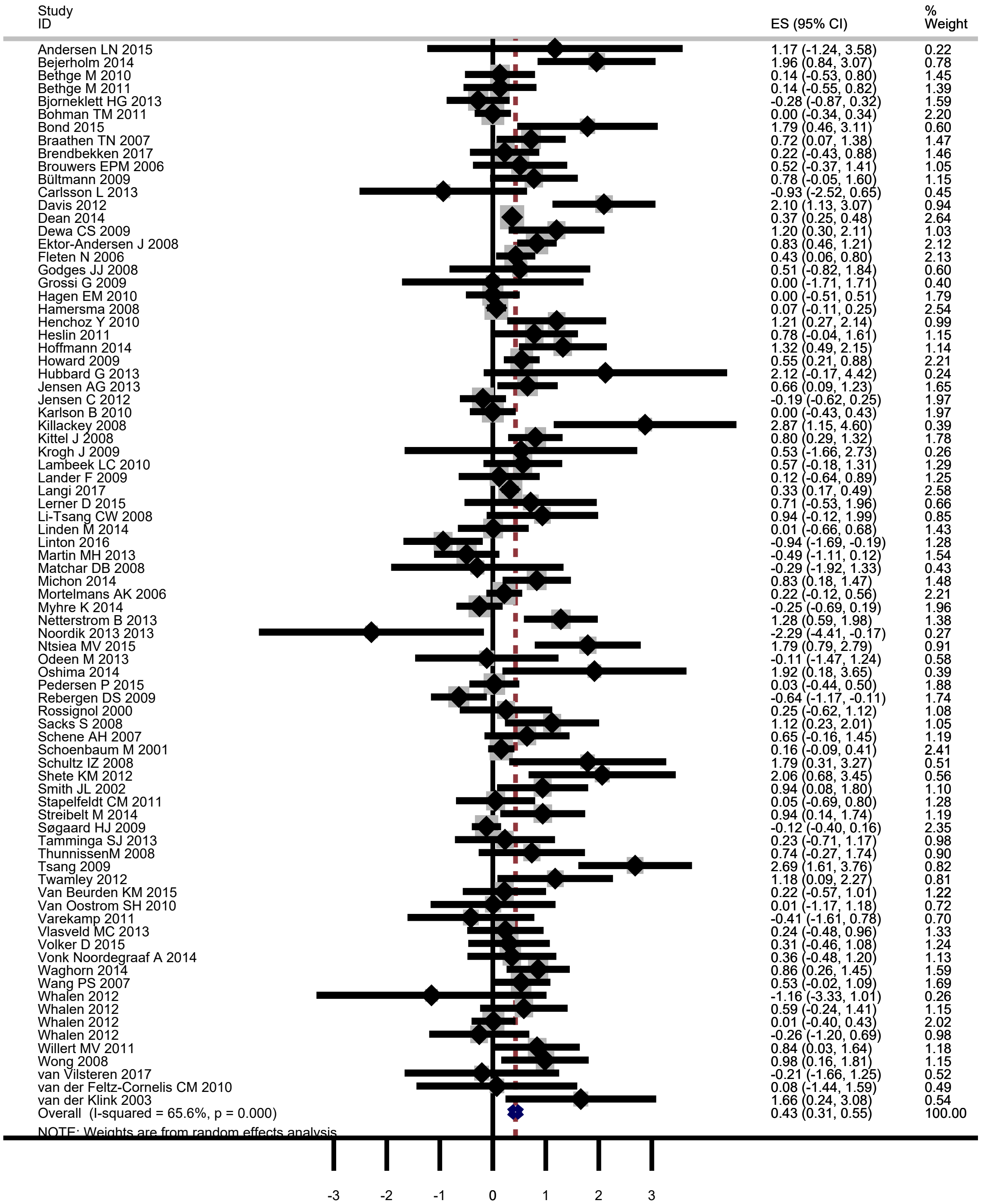
[30] For the effect on employment, a meta-analysis of the 4 state-specific studies in Whalen et al. (2012) gives a confidence interval of (-0.527,0.622) on the log odds scale, while a meta-analysis of the 4 most relevant high-quality studies (Bond 2015; Davis 2012; Twamley 2012; Allaire 2005) gives a confidence interval of (0.704,2.53). No equivalent comparison is possible for effects on disability benefit receipt or application.

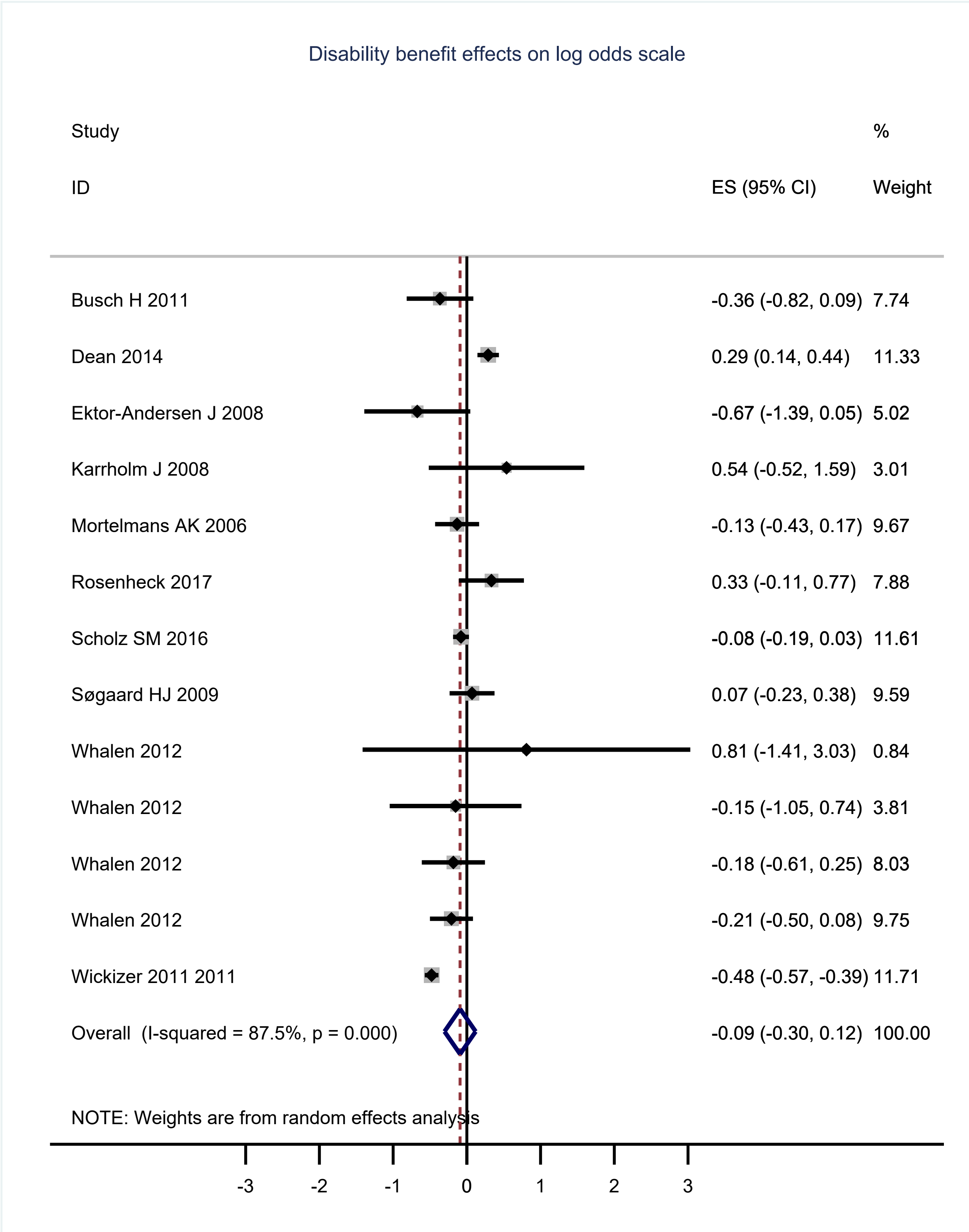
scale low-cost interventions make up at least some of the evaluation designs explored in the future. The most promising impacts from this evidence review concern programs that include employment services and training, but these are also among the most costly evaluations to implement in a large-scale experiment. Recent evidence on low-cost interventions indicates that mailings can effectively change behavior, but that the design of the mailing matters little (Juras et al. 2016), undercutting claims about behaviorally informed design (that is, the style of the messaging may matter less than the message itself, for informational interventions).

With many open questions about the effects of SAW/RTW programs, we also recommend considering design options that would address more than research question. A factorial design would maximize power given that more than one research question exists in each study. Essentially, a factorial design can answer more questions with better power because it exploits extra dimensions in the assignment to treatment arms. For example if we wish to test two dimensions of treatment, for example incentives and information, and if we postulate two directions in which we can alter these relative to the status quo, we can compose a three by three matrix of treatment conditions, for nine conditions (including business as usual). In this example, instead of a sample more than 20 times as large, we would need only a sample a little more than twice as large to measure improvements in both dimensions.

Appendix A. Meta-analytic Average Across All Studies

Employment effects on log odds scale




Disability benefit effects on log odds scale
Study
ID
ES (95% CI)
%
Weight
Busch H 2011
-0.36 (-0.82, 0.09)
7.74
Dean 2014
0.29 (0.14, 0.44)
11.33
Ektor-Andersen J 2008
-0.67 (-1.39, 0.05)
5.02
Karrholm J 2008
0.54 (-0.52, 1.59)
3.01
Mortelmans AK 2006
-0.13 (-0.43, 0.17)
9.67
Rosenheck 2017
0.33 (-0.11, 0.77)
7.88
Scholz SM 2016
-0.08 (-0.19, 0.03)
11.61
Søgaard HJ 2009
0.07 (-0.23, 0.38)
9.59
Whalen 2012
0.81 (-1.41, 3.03)
0.84
Whalen 2012
-0.15 (-1.05, 0.74)
3.81
Whalen 2012
-0.18 (-0.61, 0.25)
8.03
Whalen 2012
-0.21 (-0.50, 0.08)
9.75
Wickizer 2011 2011
-0.48 (-0.57, -0.39)
11.71
Overall (I-squared = 87.5%, p = 0.000)
-0.09 (-0.30, 0.12)
100.00
NOTE: Weights are from random effects analysis
-3
-2
-1
0
1
2
3

Appendix B. Meta-analytic Regression Methods

Looking across studies, there is sampling error in measuring the impact of a program in any one study, which shrinks as the sample size in a study is larger, but there is also variation across studies in true impacts. A random effects meta-regression accounts for both types of error.

Using estimates y_i from N independent studies, each with estimated variance v_i (the square of the standard error of the estimate, s_i), we conduct a random-effects meta-regression using the aggregate-level data with variance-weighted least squares. Meta-regression, also known as meta-analytic regression, extends variance-weighted least squares by estimating an additional component of variance τ^2 characterizing the variance of the error term e_i in equation 1 below.

$$y_i = x_i\beta + u_i + e_i \qquad \text{[Equation 1]}$$

In Equation 1, u_i is assumed to be a normally distributed error term with known standard deviation s_i (the subscript i indicates that the variance will in general vary across units of observation, i.e. studies or estimates), and e_i assumed to be a normally distributed error term with variance τ^2 to be estimated, assumed constant across units. The additional error term e_i represents the random variation of the true impact across studies around an across-study mean true effect. The error term u_i represents sampling variation in the single-study estimate of that study's true mean impact $x_i\beta$+ + e_i (note that u_i does not appear here, as this expression is equal to y_i - u_i).

The matrix x represents characteristics of studies, and also includes a constant. If only a constant is included in x, then the model is equivalent to random-effects meta-analysis, which is why meta-regression is a generalization of meta-analysis. As in other regression methods, if the elements of x are highly correlated, then the variance of each estimate will be appropriately inflated. This implies that collinearity reduces the chance of each element being judged statistically different from zero, but appropriately so.

We use residual maximum likelihood (REML) to estimate the additive (between-study) component of variance τ^2. We use the modification suggested by Knapp and Hartung (2003) to the variance of the estimated coefficients, accompanied by the use of a t-distribution in place of the standard normal distribution when calculating p-values and confidence intervals.

The command used to implement these models is due to Harbord and Higgins (2008). We conducted all estimation in Stata 15.

The percent of total variation that is due to e_i as opposed to u_i is denoted I^2 in tables (and we can test the null hypothesis that this proportion is zero).

Appendix C. References

Abásolo L, Blanco M, Bachiller J, Candelas G, Collado P, Lajas C, et al. A health system program to reduce work disability related to musculoskeletal disorders. *Annals of Internal Medicine* 2005;**143**(6):404–14.

Abásolo L, Carmona L, Hernandez-Garcia C, Lajas C, Loza E, Blanco M, et al. Musculoskeletal work disability for clinicians: time course and effectiveness of a specialized intervention program by diagnosis. *Arthritis & Rheumatology* 2007;**57**(2):335–42.

Alaranta H, Rytokoski U, Rissanen A, Talo S, Ronnemaa T, Puukka P, et al. Intensive physical and psychosocial training program for patients with chronic low back pain. A controlled clinical trial. *Spine* 1994;**16**(12):1339–49.

Alaranta H, Hurme M, Einola S, Kallio V, Knuts LR, Törmä T. Rehabilitation after surgery for lumbar disc herniation: results of a randomized clinical trial. *International Journal of Rehabilitation Research* 1986;**9**(3):247–57.

Allaire SH, Niu J, LaValley MP. Employment and satisfaction outcomes from a job retention intervention delivered to persons with chronic diseases. Rehabilitation Counseling Bulletin. 2005 Jan;48(2):100-9.

Andersen LN, Juul-Kristensen B, Roessler KK, Herborg LG, Sorensen TL, Sogaard K. Efficacy of 'Tailored Physical Activity' on reducing sickness absence among health care workers: a 3-months randomised controlled trial. Manual therapy 2015; Vol. 20, issue 5:666–71.

Anema JR, Steenstra IA, Bongers PM, de Vet HC, Knol DL, Loisel P, et al. Multidisciplinary rehabilitation for subacute low back pain: graded activity or workplace intervention or both? A randomized controlled trial. *Spine* 2007;**32**(3): 291–8.

Arends I, Bruinvels DJ, Rebergen DS, Nieuwenhuijsen K, Madan I, Neumeyer-Gromen A, Bültmann U, Verbeek JH. Interventions to facilitate return to work in adults with adjustment disorders. Cochrane database of systematic reviews. 2012(12).

Arnetz BB, Sjogren B, Rydehn B, Meisel R. Early workplace intervention for employees with musculoskeletal-related absenteeism: a prospective controlled intervention study. *Journal of Occupational and Environmental Medicine* 2003; **45**(5):499–506.

Bakker IM, Terluin B, van-Marwijk HWJ, van der Windt DAWM, Rijmen F, van MW, et al. A cluster-randomised trial evaluating an intervention for patients with stress related mental disorders and sick leave in primary care. *PLOS Clinical Trials* 2007;**2**(6):e26.

Bejerholm U, Areberg C, Hofgren C, Sandlund M, Rinaldi M. Individual placement and support in Sweden—a randomized controlled trial. Nordic journal of psychiatry. 2015 Jan 2;69(1):57-66.

Berg SK, Svendsen JH, Zwisler AD, Pedersen BD, Preisler P, Siersbaek-Hansen L, et al. COPE-ICD: a randomised clinical trial studying the effects and meaning of a comprehensive rehabilitation programme for ICD recipients—design, intervention and population. *BMC Cardiovascular Disorders* 2011;**11**:33. [DOI: 10.1186/1471-2261-11-33

Bergström C, Jensen I, Hagberg J, Busch H, Bergstrom G. Effectiveness of different interventions using a psychosocial subgroup assignment in chronic neck and back pain patients: a 10-year follow-up. *Disability and Rehabilitation* 2012;**34**(2):110–8.

Bernacki EJ, Guidera JA, Schaefer JA, Tsai S. A facilitated early return to work program at a large urban medical center. *Journal of Occupational and Environmental Medicine*. 2000 Dec 1;**42**(12):1172-7.

Bethge M, Herbold D, Trowitzsch L, Jacobi C. Return to work following work-related orthopaedic rehabilitation: a cluster randomized trial. *Rehabilitation* 2010;**49**(1):2–12.

Bethge M, Herbold D, Trowitzsch L, Jacobi C. Work status and health-related quality of life following multimodal work hardening: a cluster randomised trial. *Journal of Back and Musculoskeletal Rehabilitation* 2011;**24**(3):161–72.

Beutel ME, Zwerenz R, Bleichner F, Vorndran A, Gustson D, Knickenberg RJ. Vocational training integrated into inpatient psychosomatic rehabilitation—short and long-term results from a controlled study. *Disability and Rehabilitation* 2005;**27**(15):891–900.

Bevan S. Back to work: exploring the benefits of early interventions which help people with chronic illness remain in work. Fit for Work Europe. Brussels, Belgium: The Work Foundation. 2015 Apr 29.

Birnbaum HG, Ben-Hamadi R, Kelley D, Hsieh M, Seal B, Kantor E, Cremieux PY, Greenberg PE. Assessing the relationship between compliance with antidepressant therapy and employer costs among employees in the United States. *Journal of occupational and environmental medicine*. 2010 Feb 1;**52**(2):115-24.

Bishop A, Wynne-Jones G, Lawton SA, van der Windt D, Main C, Sowden G, et al. Rationale, design and methods of the Study of Work and Pain (SWAP): a cluster randomised controlled trial testing the addition of a vocational advice service to best current primary care for patients with musculoskeletal pain (ISRCTN 52269669). *BMC Musculoskeletal Disorders* 2014;**15**:232. [DOI: 10.1186/ 1471-2474-15-232

Bjorneklett HG, Rosenblad A, Lindemalm C, Ojutkangas ML, Letocha H, Strang P, et al. A randomized controlled trial of support group intervention after breast cancer treatment: results on sick leave, health care utilization and health economy. *Acta Oncologica* 2013;**52**(1):38–47.

Blonk RWB, Brenninkmeijer V, Lagerveld SE, Houtman ILD. Return to work: a comparison of two cognitive behavioural interventions in cases of work-related psychological complaints among the self-employed. *Work and Stress* 2006;**20**(2):129–44.

Bohman TM, Wallisch L, Christensen KB, Stoner B, Pittman A, Reed B, et al. Working Well—the Texas demonstration to maintain independence and employment: 18-month outcomes. *Journal of Vocational Rehabilitation* 2011;**34**(2):97–106.

Bolton BF, Bellini JL, Brookings JB. Predicting client employment outcomes from personal history, functional limitations, and rehabilitation services. *Rehabilitation Counseling Bulletin* 2000;**44**(1):10-21.

Bond GR, Kim SJ, Becker DR, Swanson SJ, Drake RE, Krzos IM, Fraser VV, O'Neill S, Frounfelker RL. A controlled trial of supported employment for people with severe mental illness and justice involvement. Psychiatric Services. 2015 Jun 1;66(10):1027-34.

Bonde JP, Rasmussen MS, Hjøllund H, Svendsen SW, Kolstad HA, Jensen LD, et al. Occupational disorders and return to work: a randomized controlled study. *Journal of Rehabilitation Medicine* 2005;**37**(4):230–5.

Bouwsma EV, Anema JR, Vonk Noordegraaf A, Knol DL, Bosmans JE, Schraffordt Koops SE, et al. The cost effectiveness of a tailored, web-based care program to enhance postoperative recovery in gynecologic patients in comparison with usual care: protocol of a stepped wedge cluster randomized controlled trial. *JMIR Research Protocols* 2014;**3**(2):e30.

Braathen TN, Veiersted KB, Heggenes J. Improved work ability and return to work following vocational multidisciplinary rehabilitation of subjects on long-term sick leave. *Journal of Rehabilitation Medicine* 2007;**39**(6): 493–9.

Brendbekken R, Eriksen HR, Grasdal A, Harris A, Hagen EM, Tangen T. Return to work in patients with chronic musculoskeletal pain: multidisciplinary intervention versus brief intervention: a randomized clinical trial. Journal of Occupational Rehabilitation 2016; Vol. 27, issue 1:82–91. [DOI: 10.1007/s10926-016-9634-5

Brouwers EP, Tiemens BG, Terluin B, Verhaak PF. Effectiveness of an intervention to reduce sickness absence in patients with emotional distress or minor mental disorders: a randomized controlled effectiveness trial. *General Hospital Psychiatry* 2006;**28**(3):223–9.

Buijs PC, Lambeek LC, Koppenrade V, Hooftman WE, Anema JR. Can workers with chronic back pain shift from pain elimination to function restore at work? Qualitative evaluation of an innovative work related multidisciplinary programme. *Journal of Back and Musculoskeletal Rehabilitation* 2009;**22**(2):65–73.

Bültmann U, Sherson D, Olsen J, Hansen CL, Lund T, Kilsgaard J. Coordinated and tailored work rehabilitation: a randomized controlled trial with economic evaluation undertaken with workers on sick leave due to musculoskeletal disorders. *Journal of Occupational Rehabilitation* 2009;**19**(1):81–93.

Burton WN, Chen C, Conti DJ, Schultz AB, Edington DW. The association of antidepressant medication adherence with employee disability absences. *American Journal of Managed Care*. 2007 Feb 1;**13**(2):105.

Busch H, Bodin L, Bergstrom G, Jensen IB. Patterns of sickness absence a decade after pain-related multidisciplinary rehabilitation. *Pain* 2011;**152**(8):1727–33.

Byrne DM, Drury J, Mackay RC, Robinson S, Faranda C, MacAdam DB. Evaluation of the efficacy of an instructional programme in the self-management of patients with asthma. *Journal of Advanced Nursing* 1993;**18**(4):637–46.

Carls GS, Roebuck MC, Brennan TA, Slezak JA, Matlin OS, Gibson TB. Impact of medication adherence on absenteeism and short-term disability for five chronic diseases. *Journal of occupational and environmental medicine*. 2012 Jul 1;**54**(7):792-805.

Carlsson L, Englund L, Hallqvist J, Wallman T. Early multidisciplinary assessment was associated with longer periods of sick leave: a randomized controlled trial in a primary health care centre. *Scandinavian Journal of Primary Health Care* 2013;**31**(3):141–6.

Cheadle A, Wickizer TM, Franklin G, Cain K, Joesch J, Kyes K, et al. Evaluation of the Washington State Workers Compensation Managed Care Pilot Project II: medical and disability costs 639. *Medical Care* 1999;**37**(10):982–93.

Cheng AS, Hung LK. Randomized controlled trial of workplace-based rehabilitation for work-related rotator cuff disorder. *Journal of Occupational Rehabilitation* 2007;**17**(3): 487–503.

Clayton S, Bambra C, Gosling R, Povall S, Misso K, Whitehead M. Assembling the evidence jigsaw: insights from a systematic review of UK studies of individual-focused return to work initiatives for disabled and long-term ill people. BMC public health. 2011 Dec;11(1):170.

Davey CA. The Implementation and Evaluation of a Rehabilitation Co-ordinator Service for Personal Injury Claimants. Edinburgh: University of Edinburgh, 1994.

Davis LL, Leon AC, Toscano R, Drebing CE, Ward LC, Parker PE, Kashner TM, Drake RE. A randomized controlled trial of supported employment among veterans with posttraumatic stress disorder. Psychiatric Services. 2012 May;63(5):464-70.

De Boer AG, Burdorf A, van Duivenbooden C, Frings-Dresen MH. The effect of individual counselling and education on work ability and disability pension: a prospective intervention study in the construction industry. *Occupational and Environmental Medicine* 2007;**64**(12): 792–7.

De Boer AG, van Beek JC, Durinck J, Verbeek JH, van Dijk FJ. An occupational health intervention programme for workers at risk for early retirement; a randomised controlled trial. *Occupational and Environmental Medicine* 2004;**61** (11):924–9.

De Buck PD, le CS, van den Hout WB, Peeters AJ, Ronday HK, Westedt ML, et al. Randomized comparison of a multidisciplinary job-retention vocational rehabilitation program with usual outpatient care in patients with chronic arthritis at risk for job loss. *Arthritis & Rheumatology* 2005; **53**(5):682–90.

Dean D, Pepper JV, Schmidt RM, Stern S. State vocational rehabilitation programs and federal disability insurance: an analysis of Virginia's vocational rehabilitation program. *IZA Journal of Labor Policy*. 2014 Dec;**3**(1):7.

Della-Posta C, Drummond PD. Cognitive behavioural therapy increases re-employment of job seeking worker's compensation clients. *Journal of Occupational Rehabilitation* 2006;**16**(2):223–30.

Demir S, Dulgeroglu D, Cakci A. Effects of dynamic lumbar stabilization exercises following lumbar microdiscectomy on pain, mobility and return to work. *European Journal of Physical and Rehabilitation Medicine* 2014;**50**(6):627–40.

Demoulin C, Grosdent S, Capron L, Tomasella M, Somville PR, Crielaard JM, et al. Effectiveness of a semi-intensive multidisciplinary outpatient rehabilitation program in chronic low back pain. *Joint Bone Spine* 2010;**77**(1):58–63.

DerSimonian, R., and N. Laird. 1986. Meta analysis in clinical trials. Controlled Clinical Trials 7: 177-188.

Dewa CS, Hoch JS, Carmen G, Guscott R, Anderson C. Cost, effectiveness, and cost-effectiveness of a collaborative mental health care program for people receiving short-term disability benefits for psychiatric disorders. The Canadian Journal of Psychiatry. 2009 Jun;54(6):379-88.

Dibbelt S, Greitemann B, Büschel C. Long-term efficiency of orthopedic rehabilitation in chronic back pain—the integrative orthopedic psychosomatic concept (IopKo) [Nachhaltigkeit orthopädischer Rehabilitation bei chronischen Rückenschmerzen—Das Integrierte orthopädisch–psychosomatische Behandlungskonzept (IopKo)]. *Die Rehabilitation* 2006;**45**(6):324–35.

Dibben P, Wood G, Nicolson R, O'Hara R. Quantifying the effectiveness of interventions for people with common health conditions in enabling them to stay in or return to work: A rapid evidence assessment. London: Department of Work and Pensions. 2012.

Donceel P, Du Bois M, Lahaye D. Return to work after surgery for lumbar disc herniation. A rehabilitation-oriented approach in insurance medicine. *Spine* 1999;**24** (9):872–6.

Drews B, Nielsen CV, Rasmussen MS, Hjort J, Bonde JP. Improving motivation and goal setting for return to work in a population on sick leave: a controlled study. *Scandinavian Journal of Public Health* 2007;**35**(1):86–94.

Du Bois M, Donceel P. Guiding low back claimants to work: a randomized controlled trial. *Spine* 2012;**37**(17): 1425–31.

Ektor-Andersen J, Ingvarsson E, Kullendorff M, Orbaek P. High cost-benefit of early team-based biomedical and cognitive-behaviour intervention for long-term pain-related sickness absence. *Journal of Rehabilitation Medicine* 2008; **40**(1):1–8.

Ellis N, Johnston V, Gargett S, MacKenzie A, Strong J, Battersby M, et al. Does self-management for return to work increase the effectiveness of vocational rehabilitation for chronic compensated musculoskeletal disorders? Protocol for a randomised controlled trial. *BMC Musculoskeletal Disorders* 2010;**11**:115. [DOI: 10.1186/1471-2474-11-115

Elvsåshagen H, Tellnes G, Abdelnoor MH. Does early intervention by a specialist in physical medicine and rehabilitation reduce the duration of long-term sick leave among persons with musculoskeletal diseases? [Gir tidlig vurdering og behandling hos spesialist i fysikalskmedisin og rehabilitering reduksjon i langtidsfravær hossykmeldte med muskel– og skjelettsykdommer?]. *Norsk Epidemiologi* 2009; **19**(2):219–22.

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

Eshoj P, Tarp U, Nielsen CV. Effect of early vocational intervention in a rheumatological outpatient clinic—a randomized study. *International Journal of Rehabilitation Research* 2001;**24**(4):291–7.

Feuerstein M, Huang GD, Ortiz JM, Shaw WS, Miller VI, Wood PM. Integrated case management for work-related upper-extremity disorders: impact of patient satisfaction on health and work status. *Journal of Occupational and Environmental Medicine* 2003;**45**(8):803–12.

Fimland MS, Vasseljen O, Gismervik S, Rise MB, Halsteinli V, Jacobsen HB, et al. Occupational rehabilitation programs for musculoskeletal pain and common mental health disorders: study protocol of a randomized controlled trial. *BMC Public Health* 2014;**14**:368. [DOI: 10.1186/ 1471-2458-14-368

Fleten N, Johnsen R. Reducing sick leave by minimal postal intervention: a randomised, controlled intervention study. Occupational and Environmental Medicine. 2006 Oct 1;63(10):676-82.

Fraker T, Mamun A, Honeycutt T, Thompkins A, Valentine EJ. Final report on the Youth Transition Demonstration evaluation. Washington, DC: Mathematica Policy Research. 2014 Nov 11.

Frey WD, Azrin ST, Goldman HH, Kalasunas S, Salkever DS, Miller AL, et al. The mental health treatment study. *Psychiatric Rehabilitation Journal* 2008;**31**:306–12. [DOI: 10.2975/31.4.2008.306.312

Fritz JM, Delitto A, Erhard RE. Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain: a randomized clinical trial. *Spine* 2003;**28**(13): 1363–71.

Gatchel RJ, Polatin PB, Noe C, Gardea M, Pulliam C, Thompson J. Treatment- and cost-effectiveness of early intervention for acute low back pain patients: a one-year prospective study. *Journal of Occupational Rehabilitation* 2003;**13**(1):1–9.

Gifford B, Parry T, Jinnett K. Early Warnings: Using FMLA to Understand and Manage Disability Absence. Integrated Benefits Institute, November, 2014

Godges JJ, Anger MA, Zimmerman G, Delitto A. Effects of education on return-to-work status for people with fear-avoidance beliefs and acute low back pain. Physical therapy. 2008 Feb 1;88(2):231-9.

Greitemann B, Dibbelt S, Büschel C. Multidisciplinary orthopedic rehabilitation program in patients with chronic back pain and need for changing job situation—long-term effects of a multimodal, multidisciplinary program with activation and job development [Integriertes Orthopädisch–Psychosomatisches Konzept zur medizinischen Rehabilitation von Patienten mit chronischen Schmerzen des Bewegungsapparates—Langfristige Effekte und Nachhaltigkeit eines multimodalen Programmes zur Aktivierung und beruflichen Umorientierung]. *Zeitschrift für Orthopädie und ihre Grenzgebiete* 2006;**144**(3):255–66.

Grossi G, Santell B. Quasi-experimental evaluation of a stress management programme for female county and municipal employees on long-term sick leave due to work-related psychological complaints. Journal of rehabilitation medicine. 2009 Jul 5;41(8):632-8.

Hagen EM, Ødelien KH, Lie SA, Eriksen HR. Adding a physical exercise programme to brief intervention for low back pain patients did not increase return to work. Scandinavian journal of public health. 2010 Nov;38(7):731-8.

Hagen SE, Wright DW, Finch R, Talamonti WJ, Edington DW. Impact of compliance to oral hypoglycemic agents on short-term disability costs in an employer population. *Population health management*. 2014 Feb 1;**17**(1):35-41.

Haldorsen EM, Grasdal AL, Skouen JS, Risa AE, Kronholm K, Ursin H. Is there a right treatment for a particular patient group? Comparison of ordinary treatment, light multidisciplinary treatment, and extensive multidisciplinary treatment for long-term sick-listed employees with musculoskeletal pain. *Pain* 2002;**95**(1-2):49–63.

Haldorsen EM, Kronholm K, Skouen JS, Ursin H. Multimodal cognitive behavioral treatment of patients sicklisted for musculoskeletal pain: a randomized controlled study. *Scandinavian Journal of Rheumatology* 1998;**27**(1): 16–25.

Hamersma S. The effects of an employer subsidy on employment outcomes: A study of the work opportunity and welfare-to-work tax credits. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*. 2008 Jun;27(3):498-520.

Harbord, Roger M. and Julian P. T. Higgins. 2008. "Meta-regression in Stata." *Stata Journal* 8(4): 493–519.

Hees HL, de Vries G, Koeter MW, Schene AH. Adjuvant occupational therapy improves long-term depression recovery and return-to-work in good health in sick-listed employees with major depression: results of a randomised controlled trial. *Occupational and Environmental Medicine* 2013;**70**(4):252–60.

Henchoz Y, de Goumoens P, So AK, Paillex R. Functional multidisciplinary rehabilitation versus outpatient physiotherapy for non specific low back pain: randomized controlled trial. *Swiss Medical Weekly* 2010;**140**:w13133. [DOI: 10.4414/smw.2010.13133

Heslin M, Howard L, Leese M, McCrone PA, Rice C, Jarrett M, Spokes T, Huxley P, Thornicroft G. Randomized controlled trial of supported employment in England: 2 year follow-up of the Supported Work and Needs (SWAN) study. World Psychiatry. 2011 Jun;10(2):132.

Hill CJ, Bloom HS, Black AR, Lipsey MW. Empirical Benchmarks for Interpreting Effect Sizes in Research. MDRC Working Papers on Research Methodology, July 2007. New York: MDRC. [https://www.mdrc.org/sites/default/files/full_84.pdf]

Hlobil H, Uegaki K, Staal JB, de-Bruyne MC, Smid T, van MW. Substantial sick leave costs savings due to a graded activity intervention for workers with non-specific sub-acute low back pain. *European Spine Journal* 2007;**16**(7):919–24.

Hoefsmit N, Houkes I, Nijhuis FJ. Intervention characteristics that facilitate return to work after sickness absence: a systematic literature review. *Journal of occupational rehabilitation*. 2012 Dec 1;**22**(4):462-77.

Hoffmann H, Jäckel D, Glauser S, Mueser KT, Kupper Z. Long-term effectiveness of supported employment: 5-year follow-up of a randomized controlled trial. American Journal of Psychiatry. 2014 Nov 1;171(11):1183-90.

Hoverstad T, Koefoed G, Gudding IH. Intervention against sick leave in an industrial company. *Tidsskrift for den Norske lægeforening* 1994;**114**(11):1317–20.

Howard KJ, Mayer TG, Gatchel RJ. Effects of presenteeism in chronic occupational musculoskeletal disorders: stay at work is validated. Journal of Occupational and Environmental Medicine. 2009 Jun 1;51(6):724-31.

Hubbard G, Gray NM, Ayansina D, Evans JM, Kyle RG. Case management vocational rehabilitation for women with breast cancer after surgery: a feasibility study incorporating a pilot randomised controlled trial. *Trials* 2013;**14**:175. [DOI: 10.1186/1745-6215-14-175

Huber MJ, Oswald G, Chan F, Shaw LR, Wilson J. A call for a national evidence-based programs and practices registry in vocational rehabilitation. *Journal of Vocational Rehabilitation*. 2017 Jan 1;**46**(1):11-8.

Hullegie P, Koning P. Employee Health and Employer Incentives. Ministry of Social Affairs and Employment, University of Amsterdam, Tinbergen Institute 2014.

Huppe A, Glaser-Moller N, Raspe H. Offering multidisciplinary medical rehabilitation to workers with work disability due to musculoskeletal disorders: results of randomized controlled trial [Trägerübergreifendes Projekt zur Früherkennung von Rehabilitationsbedarf bei Versicherten mit muskuloskelettalen Beschwerden durch Auswertung von Arbeitsunfähigkeitsdaten: Ergebnisse einer randomisierten, kontrollierten Evaluationsstudie]. *Gesundheitswesen* 2006;**68**(6):347–56.

Indahl A, Haldorsen EH, Holm S, Reikeras O, Ursin H. Five-year follow-up study of a controlled clinical trial using light mobilization and an informative approach to low back pain. *Spine* 1998;**23**(23):2625–30.

ISRCTN31632033. The efficacy and predicting variables of a multidisciplinary disability resolution (MDR) program for CFS patients receiving long term disability benefits from income protection insurers. www.isrctn.com/ ISRCTN31632033 (first received 4 April 2006).

Jensen AG. A two-year follow-up on a program theory of return to work intervention. *Work (Reading, MA)* 2013;**44** (2):165–75.

Jensen C, Jensen OK, Nielsen CV. Sustainability of return to work in sick-listed employees with low-back pain. Two-year follow-up in a randomized clinical trial comparing multidisciplinary and brief intervention. *BMC Musculoskeletal Disorders* 2012;**13**:156. [DOI: 10.1186/ 1471-2474-13-156

Jensen I, Nygren A, Gamberale F, Goldie I, Westerholm P, Jonsson E. The role of the psychologist in multidisciplinary treatments for chronic neck and shoulder pain: a controlled cost-effectiveness study. *Scandinavian Journal of Rehabilitation Medicine* 1995;**27**(1):19–26.

Jinnett K, Parry T. Valuing lost work time: connecting medication adherence and short-term disability. *Am J Pharm Benefits*. 2012;**4**:e56-64.

Johansson C, Dahl J, Jannert M, Melin L, Andersson G. Effects of a cognitive-behavioral pain-management program. *Behaviour Research and Therapy* 1998;**36**(10): 915–30.

Johnston MV, Sherer M, Whyte J. Applying evidence standards to rehabilitation research. *American journal of physical medicine & rehabilitation*. 2006 Apr 1;**85**(4):292-309.

Johnston MV, Vanderheiden GC, Farkas MD, Rogers ES, Summers JA, Westbrook JD. The Challenge of Evidence in Disability and Rehabilitation Research and Practice: A Position Paper. SEDL. 2009.

Jousset N, Fanello S, Bontoux L, Dubus V, Billabert C, Vielle B, et al. Effects of functional restoration versus 3 hours per week physical therapy: a randomized controlled study. *Spine* 2004;**29**(5):487–93.

Juras, Randall, Amy Minzner, Jacob Klerman, Peter Honnef, and Lauren Dunton. 2016. “The Effect of Marketing on Demand for OSHA’s On-site Consultation Program.” Washington, DC: DOL. [https://www.dol.gov/asp/evaluation/completed-studies/OSC_Marketing_Report.pdf]

Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJ, Ostelo RW, Guzman J, Van Tulder MW. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: Cochrane systematic review and meta-analysis. Bmj. 2015 Feb 18;350:h444.

Karlson B, Jonsson P, Palsson B, Abjornsson G, Malmberg B, Larsson B, et al. Return to work after a workplaceoriented intervention for patients on sick-leave for burnout—a prospective controlled study. *BMC Public Health* 2010; **10**:301. [DOI: 10.1186/1471-2458-10-301

Karrholm J, Ekholm K, Ekholm J, Bergroth A, Ekholm KS. Systematic co-operation between employer, occupational health service and social insurance office: a 6-year followup of vocational rehabilitation for people on sick-leave, including economic benefits. *Journal of Rehabilitation Medicine* 2008;**40**(8):628–36.

Karrholm J, Ekholm K, Jakobsson B, Ekholm J, Bergroth A, Schuldt K. Effects on work resumption of a co-operation project in vocational rehabilitation. Systematic, multiprofessional, client-centred and solution-oriented cooperation. *Disability and Rehabilitation* 2006;**28**(7): 457–67.

Killackey E, Jackson HJ, McGorry PD. Vocational intervention in first-episode psychosis: individual placement and support v. treatment as usual. *The British journal of psychiatry*. 2008 Aug;193(2):114-20.

Kimbrough E, Lao L, Berman B, Pelletier KR, Talamonti WJ. An integrative medicine intervention in a Ford Motor Company assembly plant. *Journal of Occupational and Environmental Medicine* 2010;**52**(3):256–7.

Kirby S, Riley R. Compulsory work-focused interviews for inactive benefit claimants: an evaluation of the British ONE pilots. *Labour Economics* 2004;**11**(4):415–29.

Kittel J, Karoff M. Improvement of worklife participation through vocationally oriented cardiac rehabilitation? Findings of a randomized control group study. *Rehabilitation* 2008;**47**(1):14–22.

Klaber Moffett JA, Carr J, Howarth E. High fear-avoiders of physical activity benefit from an exercise program for patients with back pain. *Spine* 2004;**29**(11):1167-72; discussion 73.

Knapp, G., and J. Hartung. 2003. Improved tests for a random effects meta-regression with a single covariate. Statistics in Medicine 22: 2693-2710.

Knapp S, Briest J, Bethge M. Work-related rehabilitation aftercare for patients with musculoskeletal disorders: results of a randomized-controlled multicenter trial. International Journal of Rehabilitation Research 2015; Vol. 38, issue 3: 226–32.

Knight KK, Goetzel RZ, Fielding JE, Eisen M, Jackson GW, Kahr TY, et al. An evaluation of Duke university's LIVE FOR LIFE health promotion program on changes in worker absenteeism. *Journal of Occupational and Environmental Medicine* 1994;**36**(5):533–6.

Kornfeld R, Rupp K. The net effects of the Project NetWork return-to-work case management experiment on participant earnings, benefit receipt, and other outcomes. *Social Security Bulletin* 2000;**63**(1):12–33.

Krough, J., Saltin, B., Gludd, C., and Nordentoft, M., (2009). The DEMO Trial: A Randomized, Parallel Group, Observer-Blinded Clinical Trial of Strength Versus Aerobic, Versus Relaxation Training for Patients with Mild to Moderate Depression. Journal of Clinical Psychiatry 70:6, 790-800.

Kunda, Ziva. *Social Cognition: Making Sense of People.* MIT Press. 1999.

Kuoppala J, Lamminpää A. Rehabilitation and work ability: a systematic literature review. *Journal of Rehabilitation Medicine*. 2008 Nov 5;**40**(10):796-804.

Kyes KB, Wickizer TM, Franklin G, Cain K, Cheadle A, Madden C, et al. Evaluation of the Washington State Workers Compensation Managed Care Pilot Project I: medical outcomes and patient satisfaction. *Medical Care* 1999;**37**(10):972–81.

Lai HS, Chan CC. Implementing a pilot work injury management program in Hong Kong. *Journal of Occupational Rehabilitation* 2007;**17**(4):712–26.

Lamb SE, Gates S, Underwood MR, Cooke MW, Ashby D, Szczepura A, et al. Managing Injuries of the Neck Trial (MINT): design of a randomised controlled trial of treatments for whiplash associated disorders. *BMC Musculoskeletal Disorders* 2007;**8**:7.

Lambeek LC, van Mechelen W, Knol DL, Loisel P, Anema JR. Randomised controlled trial of integrated care to reduce disability from chronic low back pain in working and private life. *BMJ* 2010;**340**:c1035. [DOI: 10.1136/bmj.c1035

Lander F, Friche C, Tornemand H, Andersen JH, Kirkeskov L. Can we enhance the ability to return to work among workers with stress-related disorders?. BMC public health. 2009 Dec;9(1):372.

Langi FF, Oberoi A, Balcazar FE, Awsumb J. Vocational rehabilitation of transition-age youth with disabilities: A propensity-score matched study. Journal of occupational rehabilitation. 2017 Mar 1;27(1):15-23.

Lerner D, Adler DA, Rogers WH, Chang H, Greenhill A, Cymerman E, et al. A randomized clinical trial of a telephone depression intervention to reduce employee presenteeism and absenteeism. *Psychiatric Services* 2015;**66** (6):570–7. **Li 2006**

Li EJ, Li-Tsang CW, Lam CS, Hui KY, Chan CC. The effect of a "training on work readiness" program for workers with musculoskeletal injuries: a randomized control trial (RCT) study. *Journal of Occupational Rehabilitation* 2006; **16**(4):529–41.

Linden M, Muschalla B, Hansmeier T, Sandner G. Reduction of sickness absence by an occupational health care management program focusing on self-efficacy and selfmanagement. *Work (Reading, MA)* 2014;**47**(4):485–9.

Lindh M, Lurie M, Sann H. A randomized prospective study of vocational outcome in rehabilitation of patients with non-specific musculoskeletal pain: A multidisciplinary approach to patients identified after 90 days of sick-leave. *Scandinavian Journal of Rehabilitation Medicine* 1997;**29**(2): 103–12.

Lindström I, Ohlund C, Eek C, Wallin L, Peterson LE, Fordyce WE, et al. The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study with an operant-conditioning behavioral approach. *Physical Therapy* 1992;**72**(4):279–90.

Lindström I, Ohlund C, Nachemson A. Physical performance, pain, pain behavior and subjective disability in patients with subacute low back pain. *Scandinavian Journal of Rehabilitation Medicine* 1995;**27**(3):153–60.

Linton SJ, Boersma K, Traczyk M, Shaw W, Nicholas M. Early workplace communication and problem solving to prevent back disability: results of a randomized controlled trial among high-risk workers and their supervisors. *Journal of occupational rehabilitation*. 2016 Jun 1;**26**(2):150-9.

Linton SJ, Bradley LA, Jensen I, Spangfort E, Sundell L. The secondary prevention of low back pain: a controlled study with follow-up. *Pain* 1989;**36**(2):197–207.

Li-Tsang CW, Li EJ, Lam CS, Hui KY, Chan CC. The effect of a job placement and support program for workers with musculoskeletal injuries: a randomized control trial (RCT) study. *Journal of Occupational Rehabilitation* 2008; **18**(3):299–306.

Livermore G, Mamun A, Schimmel J, Prenovitz S. Executive Summary of the Seventh Ticket to Work Evaluation Report. Mathematica Policy Research; 2013 Jul 30.

Livermore G, Prenovitz S, Schimmel J. Employment-Related Outcomes of a Recent Cohort of Work Incentives Planning and Assistance (WIPA) Program Enrollees. Mathematica Policy Research. Washington DC. 2011 Sep 19.

Loeppke R, Haufle V, Jinnett K, Parry T, Zhu J, Hymel P, Konicki D. Medication adherence, comorbidities, and health risk impacts on workforce absence and job performance. *Journal of occupational and environmental medicine*. 2011 Jun 1;**53**(6):595-604.

Loisel P, Lemaire J, Poitras S, Durand MJ, Champagne F, Stock S, et al. Cost-benefit and cost-effectiveness analysis of a disability prevention model for back pain management: a six year follow up study. *Occupational and Environmental Medicine* 2002;**59**(12):807–15.

Magnussen L, Strand LI, Skouen JS, Eriksen HR. Motivating disability pensioners with back pain to return to work—a randomized controlled trial. *Journal of Rehabilitation Medicine* 2007;**39**(1):81–7.

Mallon TM, Cloeren M, Firestone LM, Burch HC. Contract case managers prove cost effective in federal workers compensation programs. *Military Medicine* 2008; **173**(3):253–8.

Marhold C, Linton SJ, Melin L. A cognitive-behavioral return-to-work program: effects on pain patients with a history of long-term versus short-term sick leave. *Pain* 2001;**91**(1-2):155–63.

Marnetoft SU, Selander J, Bergroth A, Ekholm J. Vocational rehabilitation—early versus delayed. The effect of early vocational rehabilitation compared to delayed vocational rehabilitation among employed and unemployed, long-term sick-listed people. *International Journal of Rehabilitation Research* 1999;**22**(3):161–170.

Marnetoft SU, Selander J. Multidisciplinary vocational rehabilitation focusing on work training and case management for unemployed sick-listed people. *International journal of rehabilitation research* 2000;**23**(4): 271–9.

Martin MH, Nielsen MB, Madsen IE, Petersen SM, Lange T, Rugulies R. Effectiveness of a coordinated and tailored return-to-work intervention for sickness absence beneficiaries with mental health problems. *Journal of Occupational Rehabilitation* 2013;**23**(4):621–30.

Matchar DB, Harpole L, Samsa GP, Jurgelski A, Lipton RB, Silberstein SD, et al. The headache management trial: a randomized study of coordinated care. *Headache* 2008;**48** (9):1294–310.

McCluskey S, Burton AK, Main CJ. The implementation of occupational health guidelines principles for reducing sickness absence due to musculoskeletal disorders. *Occupational Medicine* 2006;**56**(4):237–42.

Meijer EM, Sluiter JK, Heyma A, Sadiraj K, Frings-Dresen MH. Cost-effectiveness of multidisciplinary treatment in sick-listed patients with upper extremity musculoskeletal disorders: a randomized, controlled trial with one-year follow-up. *International Archives of Occupational and Environmental Health* 2006;**79**(8):654–64.

Michalopoulos C, Wittenburg D, Israel DA, Schore J, Warren A, Zutshi A, Freedman S, Schwartz L. Impacts on health and employment at twelve months. Technical report, MDRC; 2011 Feb.

Michon H, van Busschbach JT, Stant AD, van Vugt MD, van Weeghel J, Kroon H. Effectiveness of individual placement and support for people with severe mental illness in the Netherlands: A 30-month randomized controlled trial. Psychiatric rehabilitation journal. 2014 Jun;37(2):129.

Mitchell RI, Carmen GM. The functional restoration approach to the treatment of chronic pain in patients with soft tissue and back injuries. *Spine* 1994;**19**(6):633–42.

Mortelmans AK, Donceel P, Lahaye D, Bulterys S. Does enhanced information exchange between social insurance physicians and occupational physicians improve patient work resumption? A controlled intervention study. *Occupational and Environmental Medicine* 2006;**63**(7): 495–502.

Myhre K, Marchand GH, Leivseth G, Keller A, Bautz-Holter E, Sandvik L, et al. The effect of work-focused rehabilitation among patients with neck and back pain: a randomized controlled trial. *Spine* 2014;**39**(24): 1999–2006.

Netterstrom B, Friebel L, Ladegaard Y. Effects of a multidisciplinary stress treatment programme on patient return to work rate and symptom reduction: results from a randomised, wait-list controlled trial. *Psychotherapy and Psychosomatics* 2013;**82**(3):177–86.

Nieuwenhuijsen K, Faber B, Verbeek JH, Neumeyer-Gromen A, Hees HL, Verhoeven AC, van der Feltz-Cornelis CM, Bültmann U. Interventions to improve return to work in depressed people. Cochrane Database of Systematic Reviews. 2014(12).

Nilsson I, Von BL. An attempt to work rehabilitation after long sick-leave. *Work (Reading, MA)* 1996;**7**(3):183–9.

Noordik E, van der Klink JJ, Geskus RB, de Boer MR, van Dijk FJ, Nieuwenhuijsen K. Effectiveness of an exposure based return-to-work program for workers on sick leave due to common mental disorders: a cluster-randomized controlled trial. *Scandinavian Journal of Work, Environment & Health* 2013;**39**(2):144–54.

Norrefalk JR, Linder J, Ekholm J, Borg K. A 6-year followup study of 122 patients attending a multiprofessional rehabilitation programme for persistent musculoskeletal related pain. *International Journal of Rehabilitation Research* 2007;**30**(1):9–18.

Ntsiea MV, Van Aswegen H, Lord S, Olorunju S S. The effect of a workplace intervention programme on return to work after stroke: a randomised controlled trial. *Clinical Rehabilitation* 2015;**29**(7):663-73.

Nystuen P, Hagen KB. Solution-focused intervention for sick listed employees with psychological problems or muscle skeletal pain: a randomised controlled trial [ISRCTN39140363]. *BMC Public Health* 2006;**6**:69.

Odeen M, Ihlebaek C, Indahl A, Wormgoor ME, Lie SA, Eriksen HR. Effect of peer-based low back pain information and reassurance at the workplace on sick leave: a cluster randomized trial. *Journal of Occupational Rehabilitation* 2013;**23**(2):209–19.

Okpaku S, Anderson KH, Sibulkin AE, Butler JS, Bickman L. The effectiveness of a multidisciplinary case management. *Psychiatric Rehabilitation Journal* 1997;**20**(3):34–41.

Oshima I, Sono T, Bond GR, Nishio M, Ito J. A randomized controlled trial of individual placement and support in Japan. Psychiatric rehabilitation journal. 2014 Jun;37(2):137.

Østerås N, Gulbrandsen P, Benth JS, Hofoss D, Brage S. Implementing structured functional assessments in general practice for persons with long-term sick leave: a cluster randomised controlled trial. *BMC Family Practice* 2009;**10**: 31. [DOI: 10.1186/1471-2296-10-31

Palmer KT, Harris EC, Linaker C, Barker M, Lawrence W, Cooper C, Coggon D. Effectiveness of community-and workplace-based interventions to manage musculoskeletal-related sickness absence and job loss: a systematic review. Rheumatology. 2011 Mar 16;51(2):230-42.

Pedersen P, Søgaard HJ, Labriola M, Nohr EA, Jensen C. Effectiveness of psychoeducation in reducing sickness absence and improving mental health in individuals at risk of having a mental disorder: a randomised controlled trial. *BMC Public Health* 2015;**15**:763. [DOI: 10.1186/ s12889-015-2087-5

Peters K, Carlson J. Worksite stress management with high-risk maintenance workers: a controlled study. *International Journal of Stress Management* 1999;**6**(1):21–44.

Petit A, Roche-Leboucher G, Bontoux L, Dubus V, Ronzi Y, Roquelaure, Y, et al. Effectiveness of three treatment strategies on occupational limitations and quality of life for patients with non-specific chronic low back pain: is a multidisciplinary approach the key feature to success? Study protocol for a randomized controlled trial. *BMC Musculoskeletal Disorders* 2014;**15**(1):131.

Pettersson-Lidbom P, Skogman Thoursie P. Temporary disability insurance and labor supply: evidence from a natural experiment. Available at SSRN 988356. 2006 Mar 31.

Ponzer S, Molin U, Johansson SE, Bergman B, Törnkvist H. Psychosocial support in rehabilitation after orthopedic injuries. *The Journal of Trauma* 2000;**48**(2):273–9.

Poulsen OM, Aust B, Bjorner JB, Rugulies R, Hansen JV, Tverborgvik T, et al. Effect of the Danish return-to-work program on long-term sickness absence: results from a randomized controlled trial in three municipalities. *Scandinavian Journal of Work, Environment & Health* 2014; **40**(1):47–56.

Purdon S, Stratford N, Taylor R, Natarajan L, Bell S, Wittenburg D. Impacts of the job retention and rehabilitation pilot. Leeds: Department for Work and Pensions. 2006. Research Report No 342.

Raftery MN, Murphy AW, O'Shea E, Newell J, McGuire BE. Effectiveness of a cognitive behavioural therapy-based rehabilitation programme (Progressive Goal Attainment Program) for patients who are work-disabled due to back pain: study protocol for a multicentre randomised controlled trial. *Trials* 2013;**14**:290. [DOI: 10.1186/ 1745-6215-14-290

Rebergen DS, Bruinvels DJ, Bezemer PD, van der Beek AJ, van MW. Guideline-based care of common mental disorders by occupational physicians (CO-OP study): a randomized controlled trial. *Journal of Occupational and Environmental Medicine/American College of Occupational and Environmental Medicine* 2009;**51**(3):305–12.

Rebergen DS, Bruinvels DJ, van der Beek AJ, van MW. Design of a randomized controlled trial on the effects of counseling of mental health problems by occupational physicians on return to work: the CO-OPstudy. *BMC Public Health* 2007;**7**:183. [DOI: 10.1186/ 1471-2458-7-183

Rebergen DS, Bruinvels DJ, van-Tulder MW, van der Beek AJ, van MW. Cost-effectiveness of guideline-based care for workers with mental health problems. *Journal of Occupational and Environmental Medicine* 2009;**51**(3): 313–22.

Richmond H, Hall AM, Copsey B, Hansen Z, Williamson E, Hoxey-Thomas N, Cooper Z, Lamb SE. The effectiveness of cognitive behavioural treatment for non-specific low back pain: a systematic review and meta-analysis. PloS one. 2015 Aug 5;**10**(8):e0134192.

Roche G, Ponthieux A, Parot SE, Jousset N, Bontoux L, Dubus V, et al. Comparison of a functional restoration program with active individual physical therapy for patients with chronic low back pain: a randomized controlled trial. *Archives of Physical Medicine and Rehabilitation* 2007;**88** (10):1229–35.

Rolving N, Oestergaard LG, Willert MV, Christensen FB, Blumensaat F, Bunger C, Nielsen CV. Description and design considerations of a randomized clinical trial investigating the effect of a multidisciplinary cognitive behavioural intervention for patients undergoing lumbar spinal fusion surgery. *BMC Musculoskeletal Disorders* 2014; **15**(1):62.

Rosenheck, R., Estroff, S., Kyaw, S., Haiqun, L., Mueser, K., Robinson, D., Schooler, N., Marcy, P., Kane, J., (2017), Incomes and Outcomes: Social Security Disability Benefits in First-Episode Psychosis.. American Journal of Psychiatry. 174: 886-894; dci:10.1176/appi.ajp.2017.16111273.

Ross RH, Callas PW, Sargent JQ, Amick BC, Rooney T. Incorporating injured employee outcomes into physical and occupational therapists' practice: a controlled trial of the Worker-Based Outcomes Assessment System. *Journal of Occupational Rehabilitation* 2006;**16**(4):607–29.

Rossignol M, Abenhaim L, Seguin P, Neveu A, Collet JP, Ducruet T, et al. Coordination of primary health care for back pain. A randomized controlled trial. *Spine* 2000;**25** (2):251-8; discussion 258-9.

Rost K, Smith JL, Dickinson M. The effect of improving primary care depression management on employee absenteeism and productivity. A randomized trial. *Medical Care* 2004;**42**(12):1202–10.

Rupp K, Wood M, Bell SH. Targeting people with severe disabilities for return-to-work: the Project NetWork demonstration experience. *Journal of Vocational Rehabilitation* 1996;**7**(1):63–91.

Sacks S, McKendrick K, Sacks JY, Banks S, Harle M. Enhanced outpatient treatment for co-occurring disorders: main outcomes. *Journal of Substance Abuse Treatment* 2008; **34**(1):48–60.

Salazar MK. Evaluation of a case management program. Summary and integration of findings. *American Association of Occupational Health Nurses* 2010;**47**(9):416–23.

Schandelmaier S, Ebrahim S, Burkhardt SC, de Boer WE, Zumbrunn T, Guyatt GH, Busse JW, Kunz R. Return to work coordination programmes for work disability: a meta-analysis of randomised controlled trials. PloS one. 2012 Nov 19;7(11):e49760.

Scheer SJ, Radack KL, O'Brien DR Jr. Randomized controlled trials in industrial low back pain relating to return to work. Part 1. Acute interventions. *Archives of Physical Medicine and Rehabilitation* 1995;**76**(10):966–73.

Schene AH, Koeter MW, Kikkert MJ, Swinkels JA, McCrone P. Adjuvant occupational therapy for work-related major depression works: randomized trial including economic evaluation. *Psychological Medicine* 2007;**37**(3): 351–62.

Schoenbaum M, Unützer J, Sherbourne C, Duan N, Rubenstein LV, Miranda J, Meredith LS, Carney MF, Wells K. Cost-effectiveness of practice-initiated quality improvement for depression: results of a randomized controlled trial. *JAMA*. 2001 Sep 19;286(11):1325-30.

Scholz SM, Andermatt P, Tobler BL, Spinnler D. Work incapacity and treatment costs after severe accidents: standard vs. intensive case management in a 6-year randomized controlled trial. *Journal of Occupational Rehabilitation* 2016;**26**(3):319–31.

Schultz IZ, Crook J, Berkowitz J, Milner R, Meloche GR, Lewis ML. A prospective study of the effectiveness of early intervention with high-risk back-injured workers—a pilot study. *Journal of Occupational Rehabilitation* 2008;**18**(2): 140–51.

Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340(c322) [*https://doi.org/10.1136/bmj.c332*].

Sevak P, Kehn M, Honeycutt T, Livermore G. Minnesota Substantial Gainful Activity (SGA) Project Demonstration: Final Evaluation Report. Report submitted to the US Department of Education, Rehabilitation Services Administration. Washington, DC: Mathematica Policy Research. 2017 Sep.

Sevak P, Martin F, Livermore G, Honeycutt T, Morris E. Kentucky Substantial Gainful Activity (SGA) Project Demonstration: Final Evaluation Report. Report submitted to the US Department of Education, Rehabilitation Services Administration. Washington, DC: Mathematica Policy Research. 2017 Sep.

Shete KM, Suryawanshi P, Gandhi N. Management of low back pain in computer users: a multidisciplinary approach. *Journal of Craniovertebral Junction and Spine* 2012;**3**(1): 7–10.

Shu-Kei C. Job coach model for occupational shoulder soft tissue injuries rehabilitation. *Dissertation Abstracts International*. Vol. **68**, Hong Kong: Chinese University of Hong Kong, 2007.

Sjöström R, Asplund R, Alricsson M. Back to work: evaluation of a multidisciplinary rehabilitation program with emphasis on mental symptoms; A two-year follow up. *Journal of multidisciplinary healthcare* 2012;**5**:145–51. [DOI: 10.2147/JMDH.S32372

Skouen JS, Grasdal A, Haldorsen EM. Return to work after comparing outpatient multidisciplinary treatment programs versus treatment in general practice for patients with chronic widespread pain. *European Journal of Pain* 2006;**10**(2):145–52.

Skouen JS, Grasdal AL, Haldorsen EM, Ursin H. Relative cost-effectiveness of extensive and light multidisciplinary treatment programs versus treatment as usual for patients with chronic low back pain on long-term sick leave: randomized controlled study. *Spine* 2002;**27**(9):901–9.

Smith JL, Rost KM, Nutting PA, Libby AM, Elliott CE, Pyne JM. Impact of primary care depression intervention on employment and workplace conflict outcomes: is value added?. Journal of Mental Health Policy and Economics. 2002 Mar;5(1):43-50.

Søgaard HJ, Bech P. The effect on length of sickness absence by recognition of undetected psychiatric disorder in long-term sickness absence. A randomized controlled trial. Scandinavian journal of public health. 2009 Nov;37(8):864-71.

Stapelfeldt CM, Christiansen DH, Jensen OK, Nielsen CV, Petersen KD, Jensen C. Subgroup analyses on return to work in sick-listed employees with low back pain in a randomised trial comparing brief and multidisciplinary intervention. *BMC Musculoskeletal Disorders* 2011;**12**:112. [DOI: 10.1186/1471-2474-12-112

Stapelfeldt CM, Labriola M2 Jensen AB, Andersen NT, Momsen AM, Nielsen CV. Municipal return to work management in cancer survivors undergoing cancer treatment: a protocol on a controlled intervention study. *BMC Public Health* 2015;**15**:720. [DOI: 10.1186/ s12889-015-2062-1

Stapleton, David, Robert Anfield, Robert Burns, Winthrop Cashdollar, Benjamin Doornink, Brian Gifford, Mary Harris, and Kevin Ufier. 2015. "Targeting Early Intervention to Workers Who Need Help to Stay in the Labor Force."

Steenstra IA, Anema JR, Bongers PM, de-Vet HC, van MW. Cost effectiveness of a multi-stage return to work program for workers on sick leave due to low back pain, design of a population based controlled trial [ISRCTN60233560]. *BMC Musculoskeletal Disorders* 2003;**4**:26.

Steenstra IA, Anema JR, Bongers PM, Vet-de HC, Knol DL, Loisel P, et al. Multidisciplinary rehabilitation for subacute low back pain: graded activity or workplace intervention or both? A randomized controlled trial. *Spine* 2007;**32**(3):299–300.

Steenstra IA, Anema JR, van-Tulder MW, Bongers PM, de-Vet HC, van MW. Economic evaluation of a multi-stage return to work program for workers on sick-leave due to low back pain. *Journal of Occupational Rehabilitation* 2006;**16** (4):557–78.

Steenstra IA, Knol DL, Bongers PM, Anema JR, van Mechelen W, de Vet HC. What works best for whom? An exploratory, subgroup analysis in a randomized, controlled trial on the effectiveness of a workplace intervention in low back pain patients on return to work. *Spine* 2009;**34**(12): 1243–9.

Streibelt M, Bethge M. Effects of intensified work-related multidisciplinary rehabilitation on occupational participation: a randomized-controlled trial in patients with chronic musculoskeletal disorders. *International Journal of Rehabilitation Research* 2014;**37**(1):61–6.

Streibelt M, Blume C, Thren K, Müller FW. Economic evaluation of medically occupationally orientated rehabilitation in patients with musculoskeletal disorders— a cost-benefit analysis from the perspective of the German statutory pension insurance scheme [Ökonomische Evaluation einer medizinisch–beruflich orientierten Maßnahme bei Patienten mit muskuloskeletalen Erkrankungen—Eine Kosten–Nutzen–Analyse aus Rentenversicherungsperspektive]. *Die Rehabilitation* 2008; **47**(3):150–7.

Suijkerbuijk YB, Schaafsma FG, van Mechelen JC, Ojajärvi A, Corbière M, Anema JR. Interventions for obtaining and maintaining employment in adults with severe mental illness, a network meta-analysis. Cochrane Database of Systematic Reviews. 2017(9).

Sundberg T, Petzold M, Wändell P, Rydén A, Falkenberg T. Exploring integrative medicine for back and neck pain—a pragmatic randomised clinical pilot trial. *BMC Complementary and Alternative Medicine* 2009;**9**:33. [DOI: 10.1186/1472-6882-9-33

Svendsen SW, Christiansen DH, Haahr JP, Andrea LC, Frost P. Shoulder function and work disability after decompression surgery for subacromial impingement syndrome: a randomised controlled trial of physiotherapy exercises and occupational medical assistance. *BMC Musculoskeletal Disorders* 2014;**15**:215. [DOI: 10.1186/ 1471-2474-15-215

Tamminga SJ, Hoving JL, Frings-Dresen MH, de Boer AG. Cancer@Work—a nurse-led, stepped-care, e-health intervention to enhance the return to work of patients with cancer: study protocol for a randomized controlled trial. *Trials* 2016;**17**:1.

Tamminga SJ, Verbeek JH, Bos MM, Fons G, Kitzen JJ, Plaisier PW, et al. Effectiveness of a hospital-based work support intervention for female cancer patients—a multi-centre randomised controlled trial. *PLOS ONE* 2013;**8**(5): e63271.

Thunnissen M, Duivenvoorden H, Busschbach J, Hakkaartvan Roijen L, van Tilburg W, Verheul R, et al. A randomized clinical trial on the effectiveness of a reintegration training program versus booster sessions after short-term inpatient psychotherapy. *Journal of Personality Disorders* 2008;**22**(5): 483–95.

Tsang HW, Chan A, Wong A, Liberman RP. Vocational outcomes of an integrated supported employment program for individuals with persistent and severe mental illness. Journal of behavior therapy and experimental psychiatry. 2009 Jun 1;40(2):292-305.

Twamley EW, Vella L, Burton CZ, Becker DR, Bell MD, Jeste DV. The efficacy of supported employment for middle-aged and older people with schizophrenia. Schizophrenia research. 2012 Mar 1;135(1-3):100-4.

U.S. General Accounting Office (GAO). (2001). Work Opportunity Tax Credit: Employers do not appear to dismiss employees to increase tax credits. Report to the Chairman, Subcommittee on Oversight, Committee on Ways and Means, House of Representatives. (GAO-01-329). Washington. DC: U.S. General Accounting Office. Retrieved April 10, 2018. From https://www.gao.gov/new.items/d01329.pdf.

U.S. General Accounting Office (GAO). (2002). Business tax incentives: Incentives to employ workers with disabilities receive limited use and have an uncertain impact. Report to Congressional Committees. {GAO-03-39). Washington, DC: U.S. General Accounting Office. Retrieved April 10, 2018, from https://www.gao.gov/new.items/d0339.pdf.

U.S. Department of Labor (DOL). (ND). Temporary Disability Insurance. Retrieved April 10, 2018 from https://workforcesecurity.doleta.gov/unemploy/pdf/temporary.pdf

Van Beurden KM, van der Klink JJ, Brouwers EP, Joosen MC,Mathijssen JJ, Terluin B, et al. Effect of an intervention to enhance guideline adherence of occupational physicians on return-to-work self-efficacy in workers sick-listed with common mental disorders. *BMC Public Health* 2015;**15**: 796. [DOI: 10.1186/s12889-015-2125-3

Van Beurden KM, Vermeulen SJ, Anema JR, van der Beek AJ. A participatory return-to-work program for temporary agency workers and unemployed workers sick-listed due to musculoskeletal disorders: a process evaluation alongside a randomized controlled trial. *Journal of Occupational Rehabilitation* 2012;**22**(1):127–40.

Van der Feltz-Cornelis CM, Hoedeman R, de Jong FJ, Meeuwissen JA, Drewes HW, van der Laan NC, et al. Faster return to work after psychiatric consultation for sicklisted employees with common mental disorders compared to care as usual. A randomized clinical trial. *Journal of Neuropsychiatric Disease and Treatment* 2010;**6**:375–85.

Van der Klink JJ, Blonk RW, Schene AH, van-Dijk FJ. Reducing long term sickness absence by an activating intervention in adjustment disorders: a cluster randomised controlled design. *Occupational and Environmental Medicine* 2003;**60**(6):429–37.

Van Oostrom SH, Anema JR, Terluin B, de Vet HC, Knol DL, van MW. Cost-effectiveness of a workplace intervention for sick-listed employees with common mental disorders: design of a randomized controlled trial. *BMC Public Health* 2008;**8**:12. [DOI: 10.1186/1471-2458-8-12

Van Oostrom SH, Driessen MT, de Vet HC, Franche RL, Schonstein E, Loisel P, van Mechelen W, Anema JR. Workplace interventions for preventing work disability. *Cochrane database of systematic reviews*. 2009(2).

Van Oostrom SH, van Mechelen W, Terluin B, de Vet HC, Knol DL, Anema JR. A workplace intervention for sicklisted employees with distress: results of a randomised controlled trial. *Occupational and Environmental Medicine* 2010;**67**(9):596–602.

Van Vilsteren M, van Oostrom SH, de Vet HC, Franche RL, Boot CR, Anema JR. Workplace interventions to prevent work disability in workers on sick leave. Cochrane Database of Systematic Reviews. 2015(10).

Van Vilsteren M, Boot CR, Twisk JW, Steenbeek R, Voskuyl AE, van Schaardenburg D, Anema JR. One year effects of a workplace integrated care intervention for workers with rheumatoid arthritis: results of a randomized controlled trial. Journal of occupational rehabilitation. 2017 Mar 1;27(1):128-36.

Varekamp I, Verbeek JH, de Boer A, van Dijk FJ. Effect of job maintenance training program for employees with chronic disease—a randomized controlled trial on self efficacy, job satisfaction, and fatigue. *Scandinavian Journal of Work, Environment & Health* 2011;**37**(4):288–97.

Vermeulen SJ, Anema JR, Schellart AJ, Knol DL, van Mechelen W, van der Beek AJ. A participatory return-to-work intervention for temporary agency workers and unemployed workers sick-listed due to musculoskeletal disorders: results of a randomized controlled trial. *Journal of Occupational Rehabilitation* 2011;**21**(3):313–24.

Vermeulen SJ, Anema JR, Schellart AJ, van MW, van der Beek AJ. Cost-effectiveness of a participatory return-to-work intervention for temporary agency workers and unemployed workers sick-listed due to musculoskeletal disorders: design of a randomised controlled trial [Protocol]. *BMC Musculoskeletal Disorders* 2010;**11**:60. [DOI: 10.1186/ 1471-2474-11-60

Vidor C, Leroyer A, Christophe V, Seillier M, Foncel J, Van de Maele J, et al. Decrease social inequalities return-to-work: development and design of a randomised controlled trial among women with breast cancer. *BMC Cancer* 2014; **14**:267. [DOI: 10.1186/1471-2407-14-267

Vlasveld MC, van der Feltz-Cornelis CM, Ader HJ, Anema JR, Hoedeman R, van Mechelen W, et al. Collaborative care for sick-listed workers with major depressive disorder: a randomised controlled trial from the Netherlands Depression Initiative aimed at return to work and depressive symptoms. *Occupational and Environmental Medicine* 2013; **70**(4):223–30.

Vogel N, Schandelmaier S, Zumbrunn T, Ebrahim S, de Boer WEL, Busse JW, Kunz R. Return-to-work coordination programmes for improving return to work in workers on sick leave. Cochrane Database of Systematic Reviews 2017, Issue 3. Art. No.: CD011618.

Volker D, Zijlstra-Vlasveld MC, Anema JR, Beekman AT, Brouwers EP, Emons WH, et al. Effectiveness of a blended web-based intervention on return to work for sick-listed employees with common mental disorders: results of a cluster randomized controlled trial. *Journal of Medical Internet Research* 2015;**17**(5):e116.

Vonk Noordegraaf A, Anema JR, van Mechelen W, Knol DL, van Baal WM, van Kesteren PJ, et al. A personalised eHealth programme reduces the duration until return to work after gynaecological surgery: results of a multicentre randomised trial. *BJOG: An International Journal of Obstetrics and Gynaecology* 2014;**121**(9):1127-35; discussion 36.

Waddell G, Burton AK, Kendall Nicholas AS. Vocational rehabilitation-what works, for whom, and when?. Report for the Vocational Rehabilitation Task Group HMSO, London ISBN 97801170386150. 2008.

Waghorn G, Dias S, Gladman B, Harris M, Saha S. A multi-site randomised controlled trial of evidence-based supported employment for adults with severe and persistent mental illness. Australian Occupational Therapy Journal. 2014 Dec;61(6):424-36.

Wagner S, Lau H, Frech-Tamas F, Gupta S. Impact of medication adherence on work productivity in hypertension. *Am J Pharm Benefits*. 2012;**4**(4):e88-96.

Wang PS, Simon GE, Avorn J, Azocar F, Ludman EJ, McCulloch J, et al. Telephone screening, outreach, and care management for depressed workers and impact on clinical and work productivity outcomes: a randomized controlled trial. *JAMA* 2007;**298**(12):1401–11.

Wason, Peter C. On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology,* 1960; 12(3): 129–40.

Westman A, Linton SJ, Ohrvik J, Wahlén P, Theorell T, Leppert J. Controlled 3-year follow-up of a multidisciplinary pain rehabilitation program in primary health care. *Disability and Rehabilitation* 2010;**32**(4):307–16.

Whalen D, Gimm G, Ireys H, Gilman B, Croake S. Demonstration to maintain independence and employment (DMIE). Final report submitted to the Centers for Medicare & Medicaid Services. Washington, DC: Mathematica Policy Research. 2012 Feb 13.

Whitehurst DG, Lewis M, Yao GL, Bryan S, Raftery JP, Mullis R, et al. A brief pain management program compared with physical therapy for low back pain: results from an economic analysis alongside a randomized clinical trial. *Arthritis & Rheumatology* 2007;**57**(3):466–73.

Wickizer TM, Franklin G, Fulton-Kehoe D, Gluck J, Mootz R, Smith-Weller T, Plaeger-Brockway R. Improving quality, preventing disability and reducing costs in workers' compensation healthcare: a population-based intervention study. *Medical care*. 2011 Dec 1:1105-11.

Willert MV, Thulstrup AM, Bonde JP. Effects of a stress management intervention on absenteeism and return to work-results from a randomized wait-list controlled trial. *Scandinavian Journal of Work, Environment & Health* 2011; **37**(3):186–95.

Williams A, Eccleston C, Morley S. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 11 2012.

Williams, R. M., Westmorland, M. G., Lin, C. A., Schmuck, G., & Creen, M. Effectiveness of workplace rehabilitation interventions in the treatment of work-related low back pain: A systematic review. Disability and Rehabilitation 2007, 29 (8), 607–624.

Wong K, Chiu R, Tang B, Mak D, Liu J, Chiu SN. A randomized controlled trial of a supported employment program for persons with long-term mental illness in Hong Kong. Psychiatric services. 2008 Jan;59(1):84-90.

Yassi A, Tate R, Cooper JE, Snow C, Vallentyne S, Khokhar JB. Early intervention for back-injured nurses at a large Canadian tertiary care hospital: an evaluation of the effectiveness and cost benefits of a two-year pilot project. Occupational Medicine 1995;45(4):209–14.

Zaman AG, Tytgat KM, Klinkenbijl JH, Frings-Dresen MH, de Boer AG. Design of a multicentre randomized controlled trial to evaluate the effectiveness of a tailored clinical support intervention to enhance return to work for gastrointestinal cancer patients. *BMC Cancer* 2016;**16**:303. [DOI: 10.1186/s12885-016-2334

Zhang M, Yan H, Phillips MR. Community-based psychiatric rehabilitation in Shanghai. Facilities, services, outcome, and culture-specific characteristics. *The British Journal of Psychiatry. Supplement* 1994;**165**(24):70–9.