

Work programmes for welfare recipients

Report from the Norwegian Knowledge Centre for the Health Services
(Nasjonalt kunnskapssenter for helsetjenesten) No 20-2006
A Campbell review



Background: Welfare-to-work programmes have replaced passive welfare reciprocity as a means of fighting poverty in many developed countries during the latest decades. There is a belief that placing welfare recipients into subsidised jobs and/or strengthening their skills and knowledge will help them acquire steady jobs. There has, however, been no systematic review of the effects of such programmes on employment, earnings and welfare payments searching systematically for studies from all parts of the world. **Objectives:** To estimate the effects of work programmes, including elements such as job search assistance, job search training, subsidised employment, job clubs, vocational training, etc. on welfare recipients. employment and economic self-sufficiency. **Selection criteria:** Randomised controlled trials, quasi-randomised trials, or cluster randomised trials of welfare-to-work programmes. **Data collection and analysis:** Studies were evaluated independently by two reviewers according to a data extraction form. The GRADE system was used for quality assessment. Outcomes on employment, earnings, welfare payments, and proportion on welfare were

(fortsetter på baksiden)

Review Team:

Geir Smedslund, Norwegian Knowledge Centre for the Health Services

Kåre Birger Hagen, National Resource Centre for Rehabilitation in Rheumatology,

Diakonhjemmet Hospital

Asbjørn Steiro, Norwegian Knowledge Centre for the Health Services

Torill Johme, Hans Nielsen Hauges G 20, 0481 Oslo, Norway

Therese Kristine Dalsbø, Almeveien 36, 0855 Oslo, Norway

Mons Georg Rud, Bleikerhaugen 17 C, 1387 Asker, Norway

Nasjonalt kunnskapssenter for helsetjenesten

Postboks 7004, St. Olavs plass

N-0130 Oslo

(+47) 23 25 50 00

www.kunnskapssenteret.no

Rapport: ISBN 82-8121-120-2 ISSN 1503-9544

20-2006

Nasjonalt kunnskapssenter for helsetjenesten



(fortsettelsen fra forsiden) included in meta-analyses. **Main results:** Randomised controlled impact evaluations of welfare-to-work programmes came almost exclusively from the United States. A total of 46 programmes with more than 412.000 participants were included in this review. Overall, 60.9 percent of intervention participants were employed at the follow-ups. But 57.9 percent of control participants were also employed. For all four outcomes, there was significant heterogeneity which could not be sufficiently explained by moderator analysis. The GRADE quality assessment showed that for all four outcomes, the quality of evidence was very low. **Authors' conclusions:** Welfare-to-work programmes in the USA have shown small, but consistent effects in moving welfare recipients into work, increasing earnings, and lowering welfare payments. The results are not clear for reducing the proportion of recipients receiving welfare. Little is known about the impacts of welfare-to-work programmes outside of the USA.

Title	Work programmes for welfare recipients
Institution	Norwegian Knowledge Centre for the Health Services
Head of centre	John-Arne Røttingen, MD, PhD, Director
Authors	Geir Smedslund, Kåre Birger Hagen, Asbjørn Steiro, Torill Johme, Therese Kristine Dalsbø, Mons Georg Rud
ISBN	82-8121-120-2
ISSN	1503-9544
Report number	No 20 – 2006
Type of report	Campbell systematic review
Project number	262
Number of pages	121
Financial sources	The Norwegian Directorate for Health and Social Affairs

Norwegian Knowledge Centre for the Health Services is a governmental centre, with a mission to support improvement of health services in Norway. The centre's mission is achieved through supporting decisions about health services by providing expert information and advice founded on knowledge-based summaries, research and development and teaching and presentation in the field of health services.

Norwegian Knowledge Centre for the Health Services

Oslo, September 2006

Summary

Background: Welfare-to-work programmes have replaced passive welfare reciprocity as a means of fighting poverty in many developed countries during the latest decades. There is a belief that placing welfare recipients into subsidised jobs and/or strengthening their skills and knowledge will help them acquire steady jobs. There has, however, been no systematic review of the effects of such programmes on employment, earnings and welfare payments searching systematically for studies from all parts of the world.

Objectives: To estimate the effects of work programmes, including elements such as job search assistance, job search training, subsidised employment, job clubs, vocational training, etc. on welfare recipients' employment and economic self-sufficiency.

Search strategy: We searched the following electronic databases: C2-SPECTR, Cochrane Library, MEDLINE, EMBASE, PsycINFO, Sociological Abstracts, Cinahl, Caredata, Eric, BIBSYS, SIGLE, IBSS, PAIS, and Social Science Citation Index. References from included primary reports and relevant reviews were scanned and content experts were contacted. A number of possibly relevant websites were searched.

Selection criteria: Randomised controlled trials, quasi-randomised trials, or cluster-randomised trials of welfare-to-work programmes.

Data collection and analysis: Studies were evaluated independently by two reviewers according to a data extraction form. The GRADE system was used for quality assessment. Outcomes on employment, earnings, welfare payments, and proportion on welfare were included in meta-analyses.

Main results: Randomised controlled impact evaluations of welfare-to-work programmes came almost exclusively from the United States. A total of 46 programmes with more than 412 thousand participants were included in this review. Participants were randomised to intervention or control group, and we report follow-up outcomes from the end of the intervention and up to six years.

Overall, 60.9 percent of intervention participants were employed at the follow-ups. But 57.9 percent of control participants were also employed. The random effects risk ratio (RR) for employment was 1.097 at the one-year follow-up with 95 percent confidence interval (CI) 1.006-1.196. At two years the random effects RR was 1.092 (95% CI: 1.032-1.157), and at five years the random-effects RR was 1.037 (1.004-1.071). We estimated the overall *number needed to treat* to be 33 (95% confidence interval: 30-37). In other words, an average of 33 welfare recipients had to receive one of the work programmes in this review in order to predict that one more of them would become employed.

The effect on earnings was small. At one year follow-up, the random effects Hedges' *g* was 0.043 (95% CI: 0.011-0.076). At two years the random effects *g* was 0.044 (0.022-0.066). At five years the random-effects *g* was 0.011 (-0.029-0.050). The mean earnings (weighted by sample size) across all the intervention outcomes (in year 2005 US dollars) was \$ 11,021 compared to \$ 8,843 in the control groups. Using the *Binominal Effect Size Display* (BESD) this roughly corresponds to a positive impact for 51.1 percent in the intervention group and for 48.9 percent in the control group.

The effect on welfare payments at one year was also small (random-effects Hedges' $g = 0.038$, 95% CI: -0.022-0.098). At two years, the random-effects g was 0.053 (-0.005-0.111), and at five years the fixed-effects g was 0.044 (0.028-0.060). The programmes reduced the welfare payments from \$ 21,719 to \$ 18,777 when averaging across all studies. Here the BESD indicated an improvement for 51.2 % of the participants in the intervention groups and for 48.8 % in the control groups.

Finally, the effect on the proportion of participants on welfare after one year showed a random effects risk ratio of 0.967 (95% CI: 0.926-1.009). After two years it was of similar magnitude (random-effects RR: 0.946, 95% CI: 0.886-1.010). Finally, after five years, the fixed-effects RR was 1.003 (0.984-1.023). After taking part in a programme, 68 percent (weighted by sample size) were on welfare, compared to 72 percent in the control groups. The overall number needed to treat indicates that, on average, 27 welfare recipients (95% CI: 24-30) had to take part in a programme in order to get an additional person off welfare (overall risk ratio for all outcomes = 0.963, 95% CI: 0.948-0.978).

For all four outcomes, there was significant heterogeneity which could not be sufficiently explained by moderator analysis.

The GRADE quality assessment showed that for all four outcomes, the quality of evidence was very low.

Authors' conclusions: Welfare-to-work programmes in the USA have shown small, but consistent effects in moving welfare recipients into work, increasing earnings, and lowering welfare payments. The results are not clear for reducing the proportion of recipients receiving welfare. Little is known about the impacts of welfare-to-work programmes outside of the USA.

Contributions and acknowledgements

	Planning and designing protocol and review	Literature searches	Screening and data extraction	Statistical analyses	Writing of protocol and review	Reading and commenting on drafts
Smedslund	x		x	X	x	
Hagen	x		x		x	x
Steiro	x		x			x
Johme	x	x	x			x
Dalsbø	x		x			x
Rud	x		x			x

Thanks to Espen Dahl and Ivar Lødemel for giving advice as content experts and commenting on parts of drafts. Thanks to Sari Susanna Ormstad for running an updated literature search in November 2005. Thanks to Gunn Elisabeth Vist for help with grading the outcomes.

Table of Contents

SUMMARY	2
CONTRIBUTIONS AND ACKNOWLEDGEMENTS	4
TABLE OF CONTENTS	5
BACKGROUND	7
RATIONALE FOR THIS REVIEW.....	7
OBJECTIVES OF THIS REVIEW.....	8
METHODS	8
INCLUDED DESIGNS.....	8
EXCLUDED DESIGNS.....	8
PARTICIPANTS.....	9
INTERVENTIONS.....	9
OUTCOME MEASURES.....	9
SEARCH STRATEGY FOR IDENTIFICATION OF RELEVANT STUDIES.....	10
CRITERIA FOR EVALUATING ELIGIBILITY OF RETRIEVED STUDIES.....	12
QUALITY ASSESSMENTS.....	13
DATA MANAGEMENT AND EXTRACTION.....	14
DETAILS OF STUDY CODING CATEGORIES.....	15
CRITERIA FOR DETERMINATION OF INDEPENDENT FINDINGS.....	17
STATISTICAL PROCEDURES AND CONVENTIONS.....	18
CHANGES IN THIS REVIEW FROM THE PUBLISHED PROTOCOL	18
RESULTS	18
THE SCREENING PROCESS.....	18
ONGOING STUDIES.....	19
STUDIES AWAITING ASSESSMENT.....	19
RAW DATA.....	19
METHODOLOGICAL QUALITY OF INCLUDED STUDIES.....	20
<i>Inter-rater reliability</i>	20
TYPE OF DESIGN.....	21
PARTICIPANTS.....	22
INTERVENTION.....	22
REPORTING OF OUTCOMES.....	23
IMPACT ON EMPLOYMENT, EARNINGS, WELFARE PAYMENT, AND PROPORTION ON WELFARE.....	24
<i>Impact on Employment</i>	24
<i>Impact on Earnings</i>	25
<i>Impact on Welfare Payments</i>	26
<i>Impact on Welfare Proportion</i>	26
THE DIFFERENTIAL EFFECT OF EMPLOYMENT FOCUS AND EDUCATION FOCUS.....	26
INTEGRATED OR TRADITIONAL CASE-MANAGEMENT.....	27
PUBLICATION BIAS.....	27
EXPLORATORY MODERATOR ANALYSES.....	27
<i>Continuous moderators</i>	28
<i>Categorical moderators</i>	29
DISCUSSION	30
REVIEWERS' CONCLUSIONS	31
IMPLICATIONS FOR PRACTICE	31
IMPLICATIONS FOR RESEARCH	32
PLANS FOR UPDATING THE REVIEW	32
POTENTIAL CONFLICTS OF INTEREST	32
TABLES AND FIGURES	46

APPENDIX A: DETAILS OF THE SEARCH STRATEGY 111
APPENDIX B: INCLUSION FORM AND DATA EXTRACTION FORM..... 114

Background

The problem: Long-term welfare reciprocity is a source of vast costs for society, and welfare recipients are over-represented regarding economic, mental and social problems. Data from 2002 show that in the UK 21 percent of household income came from social benefits (Office for National Statistics 2005). In 1998 the proportion of households receiving public assistance varied between approximately 20 percent in Spain and the United States and 57 percent in Belgium (OECD 2005). Welfare reciprocity therefore is a subject of great interest to policy makers and politicians (Martin 1998) . In most developed countries, there is some kind of unemployment benefit, but this is only accessible to people who have had some previous employment, and even then these benefits are usually time limited. Most developed countries have some kind of welfare benefit or social assistance for unemployed people who are not entitled to unemployment benefits.

Rationale for intervention: In order to reduce the extent of long-term welfare reciprocity, there is increasing consensus among policy- and decision makers in developed countries that people should, as a rule, not passively receive benefits if they have some ability to work. It is believed that some kind of work-related activation will help welfare recipients develop the skills needed to enter the regular job market. Therefore, many countries demand that (some) receivers of welfare benefits enroll in some kind of work-related programme. These programmes are often labelled “welfare-to-work” or “workfare”. Other aims include increasing quality of life, self-confidence, skills and work morale. Thus, for some persons the goal is to obtain work, but for others the goals are more modest like social inclusion, quality of life, etc.

Levitas (1998) distinguished among several discourses about persons outside the labour market. The *moral underclass discourse* implies a view of welfare recipients as lazy. In this view, if welfare benefits are too generous, recipients will have low motivation to seek employment. This discourse is closely tied to a *paternalistic discourse* in which the welfare bureaucracy must impose strict work requirements (Dahl 2003). It also blames welfare reciprocity on the individual client. There can also be structural causes of welfare reciprocity, such as lack of jobs. If the state has an obligation to provide jobs, and the clients are expected to take the available jobs, the relationship between the individual and the state is one of reciprocity. Levitas labelled this the *social integrationist discourse*.

It is difficult to compare effects of interventions across different countries. Generally, effects vary not only with the nature of the intervention, but also with characteristics of participants, the conditions of the local labour market, and how other welfare institutions are shaped and organised.

Aspects of programmes: The programmes vary on several dimensions: Individuals are either directly placed in jobs, or the placement is preceded by some kind of training period. The main aims are to obtain competitive work or to enhance quality of life and social integration. The programmes are either mandatory or voluntary. Voluntary programmes might have larger impacts because people who volunteer to participate in a welfare-to-work programme may be more motivated, on average, than individuals who are required to participate.

Rationale for this review

We are not aware of any systematic reviews of the effects of work programmes for welfare recipients worldwide. Several overviews of controlled trials have been performed (mostly

from the USA; Fischer 1995; Michalopoulos, Schwartz, & Adams-Ciardullo 2000; Cebulla et al. 2005; Grogger, et. al. 2002; Bloom, Hill, & Riccio 2003). In Europe, a recent collaborative overview (Cornwell et al. 2002) was performed in six European countries (Denmark, France, Germany, the Netherlands, Norway, and the UK). The results were not directly comparable because the countries have different ways of organising their social services, and they had somewhat different methodological approaches.

The group of individuals on welfare differs among countries. In the USA, welfare is mostly provided to poor families with children. In 1996, the Temporary Assistance for Needy Families (TANF) programme replaced the Aid to Families with Dependent Children (AFDC) as the nation's main welfare programme. Most TANF recipients are single mothers, but there are also some families with two unemployed parents. The *Food Stamp Program* is also an important transfer programme for the poor in the U. S. In addition there are EITC (Earned Income Tax Credit) and Medicaid. Some countries do not have a sharp distinction between unemployment benefits and welfare benefits, while other countries (for instance Norway and Denmark) distinguish between individuals who are entitled to unemployment benefit if they have been employed in the past, and individuals who are only entitled to welfare. In these countries, all citizens have the right to receive welfare benefit if they have no other source of income.

Objectives of this review

- To estimate the effects of work programmes for welfare recipients on work status, earnings, and welfare receipt.
- To explore sources of heterogeneity. The review tries to identify approaches that appear to be more or less effective overall and, to the extent that adequate data are available, to identify subgroups that are more or less likely to benefit than others.
- To identify any adverse effects of such programmes (e.g. displacement or negative effect in income or work status).

Methods

Included designs

We chose to limit the review to randomised controlled trials because this is the most rigorous design for studies aimed at establishing causal inference.

Criteria for inclusion and exclusion of studies for this review:

Included designs:

- Randomised controlled trials (RCTs)
- Cluster randomised trials
- Quasi randomised trials (allocation not truly random, e. g. using case record numbers, alternation, date of admission, date of birth.)

Excluded designs

- Studies using the following designs were collected, listed, and described:
 - Non-randomised concurrently controlled studies (two groups defined by the investigators and dependent variables measured on both groups before and after the intervention)

- Cohort studies (two groups defined by the interventions and dependent variables measured on both groups before and after the intervention).

There were no restrictions regarding country of publication. There were no language restrictions.

Participants

Welfare recipients. This includes both persons who already receive welfare and people who applied for welfare. People who receive food stamps were also included. We excluded persons entitled to unemployment benefit and persons with pensions of any kind.

Interventions

Interventions intended to help welfare recipients become self-sufficient typically come as “packages” involving several components. Therefore, it is usually not possible to measure the effect of one component apart from the effect of the other components of the programme. We included interventions that offered one or more of the following: *time limited work experience, job search assistance, remedial education, job clubs, financial incentives, time limits on financial support, or vocational training*. The interventions could be either mandatory or voluntary. In a voluntary programme random assignment takes place after individuals volunteer, and in a mandatory programme volunteering does not take place at all; individuals who meet certain criteria are simply randomly assigned.

The term “time limited work experience” needs some clarification. The work experience is not a permanent job. The duration of the work must be explicitly stated. This duration was recorded for each study.

Control/comparison conditions

The main control or comparison condition was ordinary (passive) social economic assistance or the usual services available to welfare recipients.

Outcome measures

All outcomes involved a comparison between different groups. The comparisons were post intervention measures. Accepted sources of outcome data were state and county administrative records, recordings by programme personnel or employers and self-reports. The type of data source was recorded for each study.

Primary Outcomes:

(1) Work status

- a) proportion of attendants who obtained competitive work (= work with standard wages and which anyone can apply for).
- b) duration of employment during a given time interval.
- c) elapsed time until obtained work
- d) duration on welfare benefit during a given time interval (as above)
- e) proportion on welfare

(2) Earnings

- a) mean or median earnings at different time points

- b) total individual earnings (e.g., income + social insurance + social assistance + welfare benefit)
- c) total household earnings

(3) Welfare Payments

Secondary Outcomes:

(1) Skills and satisfaction

- a) quality of life
- b) social skills
- c) self confidence
- d) social activity, participation and integration

(2) Adverse effects

- a) crime
- b) drug and alcohol use
- c) displacement (when programme participants get jobs, this worsens the chances of non-participants to get jobs)
- d) decrease in combined income
- e) other

Follow-up durations

Because duration of follow-up was expected to vary across studies, the exact duration of follow-up was recorded for each study. After data collection was finished, duration times were analysed based on the available data. Follow-up times were grouped into “one year”, “two years”, “five years”, “1-3 years” etc.

Search strategy for identification of relevant studies

Strategy for electronic database searches

We searched the following databases:

- C2-SPECTR (11 December 2003),
- Cochrane Library (including DARE, CCTR, HTA database (2003 Issue 2),
- MEDLINE (1966 to November week 3 2005),
- EMBASE (1980 to 2005 week 47),
- PsycINFO (1872 to November week 2 2005),
- Sociological Abstracts (1963 to November 2003),
- Cinahl (1982 to November week 3 2005),
- Caredata (15 September 2003),
- Eric (1966 to November 2003),
- SIGLE (for grey literature, 12 March 2003),
- IBSS (10 March 2004),
- PAIS (23 March 2004),
- Dissertation Abstracts International (2003¹),
- BIBSYS,
- and Social Science Citation Index (2003).

The African Trials Registry (<http://www.mrc.ac.za/ATR/>) was searched on 18 January, 2006 for the years 1955-2002.

Search words in title and keywords were ‘work’, ‘employment’, ‘unemployment’, ‘welfare’, and ‘social assistance’.

We used the same text words across all databases (unless specifically stated) and used the specialised controlled vocabularies for each database. For details of the search strategy, see Appendix A.

Other sources of information

In addition, references from included primary reports were scanned. We also scanned the reference lists from reviews that we found through our search strategy. The following authors of included studies and other potential experts in the field were contacted: Espen Dahl, Ivar Lødemel, Mathematica Policy Research and Abt Associates were contacted by email. Specifically, Gayle Hamilton at MDRC provided comments on a draft version in March 2005. The following centres of the Cochrane Collaboration were contacted by e-mail on 12 January 2006 with a request for trials: Australasian, Brazilian, Chinese, Dutch, German, Italian, Russian (under Nordic branch), South African, and the one in Bahrain. We searched the book “The Digest of Social Experiments” (Greenberg & Shroder 2004). This book documents 240 completed and 21 ongoing social experiments updated until April 2003. Another source of study data is a database of US welfare-to-work programs constructed and maintained by Andreas Cebulla, Stacey Bouchet and David Greenberg (Cebulla et. al. 2005). This database is open to the public and is said to include all known random assignment evaluations of U. S. welfare-to-work programs through 2003 that were targeted specifically at AFDC/TANF recipients.

The following websites were searched:

¹ Unfortunately, the exact dates of search in this database, BIBSYS, and Social Science Citation Index were lost.

Abt Associates (USA) (Searched 29 November, 2005)
(<http://www.abtassociates.com>)

Department for Work and Pensions Social Research Branch (UK) (Searched 29 November, 2005)
(<http://www.dwp.gov.uk/asd/asd5/>)

Joseph Rowntree Foundation (UK) (Searched 29 November, 2005)
(<http://www.jrf.org.uk/knowledge/findings/>)

MDRC (USA) (Searched 5 April, 2006)
(<http://www.mdrc.org>)

Mathematica Policy Research (USA) (Searched 30 November, 2005)
(<http://www.mathematica-mpr.com/>)

National Centre for Social Research (UK) (Searched 30 November, 2005)
(http://www.natcen.ac.uk/natcen/pages/op_employment.htm)

RAND Corporation, Social Welfare (USA) (Searched 30 November, 2005)
(http://www.rand.org/research_areas/population/)

Regard (UK) (Searched 30 November, 2005)
(<http://www.regard.ac.uk/ESRCInfoCentre/index.aspx>)

Research Forum at the National Center for Children in Poverty (NCCP) (Searched 28 March 2006)
<http://www.researchforum.org/>

Social Work Research Centre (UK) (Searched 30 November, 2005)
(<http://www.dass.stir.ac.uk/sections/sw-research/index.php>)

The United States General Accounting Office (USA) (Searched 30 November, 2005)
(<http://www.gao.gov/>)

Urban Institute (USA) (Searched 30 November, 2005)
(<http://www.urban.org/>)

U.S. Department for Health & Human Services. Administration for Children and Families
(http://www.acf.hhs.gov/acf_policy_planning.html#research) (Searched 21 March 2006)

Criteria for evaluating eligibility of retrieved studies

Two reviewers independently scanned the abstracts and titles of identified reports for eligibility, according to the inclusion criteria above. Full copies of all those deemed eligible by one of the reviewers were retrieved for closer examination. Consensus was reached by discussion and consultation with a third reviewer, if necessary. All studies which initially appeared to meet the inclusion criteria but, based on the full text reports, did not meet the inclusion criteria were detailed in a table of excluded studies.

Quality assessments

We assessed components that may contribute to the measured effectiveness of the included interventions. Two reviewers independently assigned each selected study to quality categories described below. Uncertainty or disagreement was solved by discussion with a third reviewer.

Generation of allocation sequence

MET = Resulting sequences are unpredictable (explicitly stated use of either computer-generated random numbers, table of random numbers, drawing lots or envelopes, coin tossing, shuffling cards, or throwing dice).

UNCLEAR = Vague statement that the study was randomised but not describing the generation of the allocation sequence or statement(s) indicating that random allocation was used in some but not all cases.

NOT MET = Explicit description of inadequate generation of sequence, (e.g., using case record numbers, alternation, date of admission, date of birth).

Concealment of allocation sequence

MET = Participants and investigators cannot foresee assignment, e.g. central randomisation performed at a site remote from trial location; or use of sequentially numbered, sealed, opaque envelopes).

UNCLEAR = Vague statement that the study was randomised but not describing the concealment of the allocation sequence.

NOT MET = Explicit statement that allocation was not concealed OR statement indicating that participants or investigators can foresee upcoming assignment (e.g., open allocation schedule, unsealed or non-opaque envelopes).

Control of initial difference in prognostic factors between groups

In a properly randomised study, all initial differences between groups will be caused by chance. But these initial differences may be important and can affect outcomes, especially if the study is not large (as N increases, the probability of a meaningful difference between randomised groups decreases). This applies to all prognostic variables, both known and unknown. But in improperly randomised designs and RCTs with differential attrition, there may be important differences between groups. These differences can be systematic, and they can appear in unmeasured variables as well as in the measured ones. It is generally possible to control for the latter but not the former. Matching can be used before the intervention to make groups more similar, and regression methods can be used after the intervention to control for initial differences, but all these methods may introduce bias in the results (Deeks et al. 2003).

Studies, in which both generation and concealment of allocation sequence are MET, were coded as MET below.

MET = Control for one or more prognostic factors. Also score MET when there is no control for prognostic factors because there was no imbalance (statistically significant difference between groups) in measured variables.

UNCLEAR = Sufficient information could not be obtained.

NOT MET = Imbalance in prognostic factors and failure to control for this imbalance.

Prevention of Performance Bias

MET = other interventions avoided or used similarly across comparison groups.

UNCLEAR = Use of other interventions not reported and could not be verified by contacting the investigators.

NOT MET = Dissimilar use of other interventions across comparison groups, i. e. differences in the care provided to the participants in the comparison groups other than the intervention under investigation.

Prevention of Detection Bias

MET = Assessor unaware of the assigned treatment when collecting outcome measures. This was also scored as met if the outcome was questionnaire data or data from administrative records.

UNCLEAR = Blinding of assessor not reported and could not be verified by contacting the investigators.

NOT MET = Assessor aware of the assigned treatment when collecting outcome measures.

Prevention of Attrition Bias

MET = Losses to follow up less than or equal to 20% and equally distributed between comparison groups (proportion of total loss to follow-up equal to or less than 20% in group with the highest loss to follow-up).

UNCLEAR = Losses to follow up not reported.

NOT MET = Losses to follow up greater than 20% or not equally distributed between comparison groups.

Intention-to-treat analysis

MET = Intention to treat analysis performed or possible with data provided.

UNCLEAR = Intention to treat not reported, and could not be undertaken by contacting the investigators.

NOT MET = Intention to treat analyses not done and not possible for reviewers to calculate independently.

This review has been graded using GRADE (GRADE Working Group, 2004). GRADE stands for Grading of Recommendations Assessment, Development and Evaluation. The evidence is graded separately for each outcome. Evidence from effect studies is downgraded if the *design* is not a randomised controlled trial. *Limitations* might be lack of allocation concealment, lack of blinding, or a large attrition. *Consistency* refers to the similarity of estimates of effect across studies. Finally, *directness* refers to the extent to which the people, interventions and outcome measures are similar to those of interest. The overall quality of the evidence is categorised as high, moderate, low, or very low. The reasons for downgrading the evidence are always made explicit in GRADE.

Data management and extraction

Two reviewers independently extracted information from the full text report on study characteristics using a specially designed, pretested data extraction form (Appendix B). For cases in which outcome information was missing from the original reports, attempts were made to retrieve the necessary data for the analysis from the original investigators or from the Internet. For example, missing unemployment rates were collected by going to the website of the U. S. Department of Labor (<http://data.bls.gov/PDQ/outside.jsp?survey=la>). Inter-rater agreement (i. e. coding reliability) was assessed and reported as percent agreement. Disagreements were resolved by meeting and discussing coded items. Data were entered into Comprehensive Meta-Analysis version 2 (Borenstein et al. 2004)

Details of study coding categories

Study characteristics: Country (or countries) of origin, year of publication, publication type, trial quality (see above), employment rate in the area.

Design: Randomised controlled trial, cluster-randomised controlled trial, or quasi-randomised controlled trial.

Participants: age, ethnicity, gender, education level, number of children, age of youngest child, previous work experience, welfare history.

Intervention: employment-focused or education-focused, job search first or varied first activity, mandatory or voluntary, duration of intervention, extent of work (e.g. working hours per week), funding agent and implementing agent.

Data synthesis

Continuous measures (e.g., earnings) were calculated as standardised mean differences (Hedges' g). We reported the 95% confidence intervals for all effect estimates. We also reported the post-intervention earnings and welfare payments in US dollars converted to year 2005 equivalents using the following website:

http://oregonstate.edu/Dept/pol_sci/fac/sahr/sahr.htm# Conversion factor tables are available

For studies with other currencies, we first converted the amounts into US dollars and then converted to their 2005 equivalents.

Discrete outcomes

We expressed binary outcome measures (e.g., employed/unemployed) as risk ratios (relative risks). The odds ratio is not the correct outcome measure in prospective studies². We also reported the number needed to treat (NNT, with 95% confidence intervals) for summary findings.

When computing summary statistics, the primary studies were weighted by the inverse of their variance.

Identifying heterogeneity

We used the advice described in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2005) for identifying statistical heterogeneity. If confidence intervals for the results of individual studies (generally depicted graphically using horizontal lines) have poor overlap, this generally indicates the presence of statistical heterogeneity. More specifically we first performed a Chi-square test for heterogeneity. If this approached statistical significance ($p < 0.10$), we looked at the results of an I-Squared test. If this was greater than 25%, we concluded that there was heterogeneity in the results.

Addressing heterogeneity

² Altman, Deeks, and Sackett (1998) wrote, "The odds ratio is valuable in case-control studies where events are usually rare and the relative risk cannot validly be estimated directly. In prospective studies interpretation of the odds ratio as an approximation to the relative risk becomes unreliable when events are common, and thus its use for prospective studies, especially randomised trials and systematic reviews, has been criticised. The distortion is especially large when the event rate is high in only one group. The odds ratio should not be interpreted as an approximate relative risk unless the events are rare in both groups (say, less than 20-30%). The odds ratio remains especially useful when researchers need to adjust for other variables, for which logistic regression is the usual approach. While such analyses are valid, when the objective is to communicate study results to an audience unfamiliar with the relation between odds ratios and relative risks, surely it makes no sense also to report the relative risk when this differs markedly from the odds ratio."

When heterogeneity was judged to be large, we examined potential sources using meta-regression, and stratified analyses using categorical moderators.

We performed meta-regressions with the following intervention and contextual moderators:

- duration of intervention
- percent sanctioned
- unemployment in the area during the evaluation years (using, if not found in the report, data from the U.S. Department of Labor, Bureau of Labor Statistics at <http://data.bls.gov/PDQ/outside.jsp?survey=la>)
- year when data collection started

The effects of participant characteristics are best measured by looking at separate subgroup results for each primary study (e.g. combining results only for Whites), but this was not feasible because subgroup results were only reported in some of the reports. Instead we entered aggregate data from the primary studies as predictors in meta-regressions:

- mean age
- percent of males
- percent Whites
- percent Blacks
- percent Hispanics
- percent other ethnic group
- percent with GED (General Educational Diploma) or high school diploma

One moderator was entered at a time, so we could not measure the effect of the moderators controlling for the others (this was not a multiple regression). We report zero-order correlations among the continuous moderators.

We performed separate analyses across the following intervention and setting related factors:

- labour force attachment (employment-focused) or human capital development (education-focused) strategy
- job search first or varied first activity
- mandatory versus voluntary programmes
- time limits or not
- financial incentives or not
- job search /job clubs or not
- educational intervention or not
- child care support or not
- work experience or not
- skills training or not
- the current U.S president
- whether the intervention was conducted before the 1988 Family Support Act (first era), during 1988 through 1995 (second era), or after 1996³ (third era).

³ The Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) marked the transition from the older AFDC system to the newer TANF). AFDC stands for Aid to Families with Dependent Children, and was operated between 1937 and 1995. TANF stands for Temporary Assistance for Needy Families. Key elements of TANF include a lifetime limit of five years (60 months) on the amount of time a family with an adult can receive assistance funded with federal funds, increasing work participation rate requirements which states must meet, and broad state flexibility on program design.

- The evaluator of the programme
- Number of intervention elements (time limits, financial incentives, job search, education, child care, work experience, skills training)

We performed separate analyses across different levels of the following design quality factors:

- proper generation of the randomisation sequence
- proper concealment of the randomisation sequence
- prevention of performance bias
- prevention of detection bias
- risk of attrition bias
- intent-to-treat analysis

Publication bias

To explore possible publication bias, we report a funnel plot for each main outcome showing both observed studies and studies imputed to produce symmetric plots (how it would have looked with no evidence of publication bias).

Sensitivity Analyses

Sensitivity analyses were used to evaluate whether the pooled effect sizes were robust across different ways of calculating standard errors based on significance levels. Reporting of nonsignificant results was not only analysed using $p = 0.1$, but also using $p = 0.55$ and $p = 0.99$.

Sensitivity analyses were also used for exploring the effects of continuous moderators. For the main analyses, we excluded studies with missing data in the meta-regressions. In another scenario, we imputed the variable means for studies with missing values. We compared the patterns of significant slopes over the two ways of dealing with missing values.

Fixed vs. Random Effects Models

We used fixed effects models when heterogeneity was small (as defined above). We used random effects models when heterogeneity was large, and could not be explained by intervention/setting factors or meta-regression.

Criteria for determination of independent findings

In many instances, several different outcome data are measured on the same subjects in the primary studies (e.g. employment status and earnings). Sometimes the same outcome is measured at multiple points in time. Because these data are from the same sample of participants, and, therefore, are not independent estimates of treatment effect, we analysed the data in such a way that any one analysis contained a single outcome from a particular time period after random assignment. When two intervention groups shared the same control group, we avoided including both intervention groups in the same analysis. Specifically, we used four separate data files for employment, earnings, welfare payments, and proportion on welfare. Within each data file, we grouped the outcomes according to follow-up time. Intervention groups sharing the same control group were not analysed together. The only exceptions were the direct comparisons in Atlanta, Columbus, Grand Rapids, and Riverside, in which we compared two intervention groups sharing a common control group (not a meta-analysis). For the exploratory analyses (meta-regression and the meta-analytic analogue of

ANOVA) we used the mean effect magnitude (an option in Comprehensive Meta-Analysis) for studies in which the same outcome was measured at different follow-up times.

Statistical procedures and conventions

Results were analysed using Comprehensive Meta-Analysis software version 2 (Borenstein, et. al., 2004). Prognostic variables were used in analyses if there were less than or equal to 20% missing data. Graphical presentations of effects (forest plots) were produced using Comprehensive Meta-Analysis.

Changes in this review from the published protocol

The main change is that we have extended the inclusion criteria to all programmes that the field perceives as “welfare-to-work” programmes. We previously excluded studies where the programmes did not offer time-limited work experience. But the external reviewers argued that work experience was only a main part of welfare-to-work programmes in the 1960s and 1970s. This component has become less and less important. Wider inclusion criteria are more meaningful and useful for policy purposes.

We have also assessed the quality of the evidence using GRADE (GRADE Working Group, 2004). For each of the four outcomes we graded the evidence for the 2 year follow-up time.

Results

The screening process

Figure 1 is a flowchart of the inclusion process. After having excluded 191 citations from reading the fulltext reports, we were left with 113 citations. Another 45 citations (mainly from Europe) were excluded because they were either nonrandomised controlled studies or they did not include work experience. We included 68 citations when using the limited protocol inclusion criteria. After incorporating the extended inclusion criteria, we included another 34 citations, making the total 102 included citations.

Citations can be reports, journal articles, books, book chapters, etc. Each programme can be reported in several citations and/or several programmes can be reported in one citation. Finally, a programme can be implemented in several sites. Table 1 shows the 46 included programmes described in 102 citations. It was difficult to arrive at the number of programmes because many had several different names. For instance, the programme “Welfare Reform Indiana” is also labelled “Indiana Manpower Placement and Comprehensive Training (IMPACT) Program”. The programme “EWEP (Experimental Work Experience Program)” is also called “San Diego Job Search and Work Experience Demonstration”. And “EMPOWER” also goes under the name “Arizona Works”. A further complication is that even if the same programme was conducted in different states, the programme operators were free to include or exclude programme components, and they had varying degrees of monitoring and sanctioning.

Multisite programmes with different intervention conditions: Many of the programmes were conducted in several sites. The highest number of sites was in the Food Stamp Education and Training Program (53 sites). Sometimes separate results for each site were reported for a programme, but at other times results were summed up for all sites. A few sites reported more than one intervention condition (Atlanta, Columbus, Grand Rapids, and Riverside). Atlanta,

for instance, had one group randomised to receive a labour force attachment strategy and another group receiving a human capital development strategy.

Intervention conditions: Table 1 describes 46 programmes but it has 58 rows. This is because the GAIN evaluation had 6 conditions and the NEWS evaluation had 8 conditions.

Different endpoints: For some of the 58 conditions there were more than one follow-up time. Because of this, there were e. g. 73 different employment endpoints.

Excluded studies

Table 2 (list of excluded studies) consists of 24 citations from 16 studies which were not randomised controlled trials (e. g. cohort studies or nonrandomised studies). Most of these studies were from the United States, but some were from Norway, Belgium, Canada, and the United Kingdom.

Ongoing studies

We identified two ongoing studies which possibly conform to our inclusion criteria:

The first one is a Dutch study (Blonk & Brenninkmeijer 2004) which investigated effects of the JOBS programme in Holland.

The other is the ERA Programme (The Employment Retention and Advancement Program; Anderson & Martinson 2003, 2005). It is conducted in several U. S. sites (Los Angeles County, Riverside County (2), Duval County, Leon County, Cook County, St. Clair County, Medford, Eugene, Hennepin County, New York City (2), and Portland). The programme is also being evaluated in the UK (Walker, et. al., 2006). Over the course of about 15 months, Jobcentre Plus randomly assigned over 16,000 people, making the ERA evaluation the largest random assignment test of a social policy's effectiveness in the UK to date.

Studies awaiting assessment

According to Vasily V. Vlassov, director of the Russian branch of the Nordic Cochrane Centre (personal communication, 24 January, 2006), there are a number of Russian databases with free internet access where trials may be described. (A list of databases is at http://www.inion.ru/product/db_2.htm.) We have not been able to access them because of the language barrier.

Raw data

We would like to draw attention to the four included Microsoft Excel files labelled "employment.xls", "earnings.xls", "welfare payments.xls", and "welfare proportion.xls". The files contain all the raw data collected for this review. There are two reasons for including these files, both having to do with transparency. Firstly, interested researchers will be able to transfer the data into their statistical software package of their likings and check our results, or they may do additional analyses, like multiple meta-regressions. Secondly, anyone having first-hand knowledge of the included primary studies might point out possible errors in the data files or add missing values. Appendix B (data extraction form) provides explanations of the variables in the attached Excel files.

Description of included studies

A main finding is that there are many randomised trials from the United States but few such trials from outside North America (two Canadian programmes were included; see Table 1).

Year when data collection started: Figure 2 shows that the first programme of this type started in the late 1960s. The number of studies was low in the 1970s. It increased sharply around 1980, and increased further in the late 1980s. After peaking in the period 1990-1994, it decreased somewhat in the late 1990s. The numbers refer to intervention sites (n= 59) with separately reported results.

Policy context and evaluators: In 1988, US Congress passed the Family Support Act, which offered states \$1 billion a year in new federal money for welfare-to-work programmes, but only if they put up matching state resources. Most of the studies reported here were conducted by the MDRC (previously known as the Manpower Demonstration Research Corporation). In their studies, many background variables have been recorded in the same way across different sites and different programmes. Some other studies are evaluated by Mathematica Policy Research or Abt Associates. Finally, a few programmes have been evaluated by universities and federal agencies.

Program goals and components: Goals like preventing social exclusion and increasing quality of life were not emphasised in the included reports. No study had the aim of drug rehabilitation. Almost all programmes had multiple components. This could be, for instance, job search assistance, job clubs, educational classes, case management, time limits on welfare receipt, financial incentives, or child care or transportation services.

Data sources: Most studies had data on earnings and employment from state and county administrative records (register data).

Publication Type: Of the 102 included citations, only 11 were journal articles. The rest were reports, books, and book chapters.

Methodological quality of included studies

Inter-rater reliability

Six raters worked independently and then compared results in pairs. The results reported here were recorded after a pilot test period of the first 12 programmes during which we had meetings and negotiated how to code the studies consistently according to a version of the data extraction form (Appendix B is the latest version). We then coded agreement for 11 programmes⁴. In total, we coded agreement for 55 variables 11 times, summing up to 605 possible disagreements (Table 3). Note that agreement means different things depending on the type of question. For variables coded “Met”, “Unclear” and “Not Met”, there are six possible ways of disagreeing. For variables such as mean age, there are in principle an infinite number of ways of disagreeing. In such cases, raters had to report exactly the same numerical value in order to get an agreement. The data extraction form (Appendix B) was edited several times during the pilot period, but ultimately all 46 programmes were coded in a consistent way.

⁴ In November 2005 we included two additional programmes (WRP in Vermont and CWEP in Washington). These programmes are not included in the inter-rater agreement reporting in Table 3. The 21 programmes that were included in March 2006 were also not included in the inter-rater data.

For all pairs of raters and all variables we achieved the proportional agreement $P_o = .82$. Table 3 shows in detail how the disagreements were distributed over the 41 variables with disagreement. For the remaining 14 variables⁵ there was perfect agreement.

Type of design

All included programmes at all sites were reported as randomised controlled trials (although the exact randomisation procedure was not always explicitly reported). Results are for individual sites where these have been reported in such a way that we could calculate an effect size for each site. (For some programmes, e. g. the Food Stamp Employment and Training Programme [53 sites], we could not obtain separate results for each site).

Columns 8-13 in Table 1 show how we judged the reporting of the methodological quality in each programme. In only three instances (EWEP, WDP, Work Program) were all methodological indicators coded as met (truly random allocation, acceptable allocation concealment, prevention of performance bias, prevention of detection bias, prevention of attrition bias, intention-to-treat analysis performed). In five cases all indicators were coded as unclear. There were few cases where the indicators were explicitly not met. (The MASSWEP programme had high attrition. The FIP programme had performance bias, etc.).

In cases where different citations gave conflicting information on the same programme, we gave the authors the benefit of the doubt according to the following rules: If at least one citation was coded as “met”, the whole programme was coded “met” on that variable. But if one citation was “not met” and the rest were “unclear”, the programme was coded as “not met” on that variable.

Prevention of detection bias was met on 48 out of the 58 intervention conditions (Table 4). This was mainly because most studies used state and county records which were defined as being relatively free of this bias.

Prevention of attrition bias was met on 33 conditions because there are systems in place for registering welfare payments, earned income, etc.

Intention-to-treat was explicitly mentioned for 28 of the intervention conditions.

Concealment of the randomisation was met for only 15 conditions, and the proper generation of the randomisation allocation sequence was only met for 10 intervention conditions. It should be stressed that this quality coding refers to the reporting of methodological quality. The generation of the randomisation allocation sequence could have been completely satisfactory in many cases, while the reporting was insufficient for us to code it as “met”.

Finally, prevention of performance bias was met for 21 conditions. This was because in most programmes, the control group participants were free to take part in other programmes while the programme was evaluated. This is only a problem if the purpose of the study is to find the theoretical, counterfactual impact of the intervention. In practice, however, one is often more interested in whether the intervention received by the programme group “adds value” relative

⁵ Main reference, publication type, publication status, publication year, design type, control variables, proportion of Hispanics, mean years of education, broadness of enrolment, degree of monitoring, table/figure where outcome was reported, sample size in intervention group number 3, sample size in intervention group number 4, sample size in control group number 3.

to whatever assistance this group would normally have received. Table 4 shows the number and percent of conditions with reports coded as “met”, “unclear” and “not met” for a number of quality indicators.

Quality of evidence

The GRADE evidence profile (Table 14) concluded that the quality of evidence is very low for all four outcomes. The reasons for this low grading are (1) there are serious limitations in the reporting of the generation of the randomisation sequence and the concealment of this sequence, (2) there is important unexplained inconsistency (heterogeneity) in the results, and (3) there is major uncertainty in the directness of the results because almost all studies are from the USA.

Participants

Our best estimate of the sample size in this systematic review is 412,045 participants, of which 245,509 were randomised to an intervention group and 166,536 were randomised to a control group. Almost 9 out of 10 were women (Table 7). This is because lone parents are usually women. The few males are either unemployed men in two-parent families with small children or heads of single-parent families.

In 1990, the U.S. Bureau of the Census (U.S. Census Bureau 1990) classified race into four major groups: (a) White; (b) Black; (c) American Indian, Eskimo, and Aleut; and (d) Asian and Pacific Islander. According to the 1990 Census, the White population accounted for 82.5 % and the Black population for 12.4 %. The remaining 5.1 percent were in categories (c) and (d). According to Table 7, Black citizens are over-represented and Whites are under-represented in the studies in this review.

Also according to the Census, the percent of people 25 years and over who have completed four years of high school or more rose steadily from 69 percent in 1980 to 84 percent in 2003. Among the participants in this review (some of whom were under age 25), 43 percent had a general educational diploma (GED) and 46 percent had a high school diploma. A GED is received by high school dropouts who pass a qualifying test.

Study participants can also be classified either as recent *applicants* of welfare when they enter a study, or people who already receive welfare (*recipients*). Also, they can be divided into single-parent families or two-parent families. We have not taken these distinctions into account in this version of the review, but we might consider doing this in future updates.

Intervention

Table 5 shows that of 73 endpoints⁶ we classified 46 as employment-focused. Another name for employment-focused is labour force attachment. This means that we judged that the aim was to try and get participants as quickly as possible into jobs, even if that meant accepting the lowest paying jobs. The rationale behind such a focus is that the best place to acquire job-relevant skills and build competence is a real workplace. The education-focused approach (also called the human capital development approach) reasons that if you start by educating the participants, they will eventually be qualified for better paying jobs, and not the low paying ones which can not sustain a family economically. For 16 endpoints, the intervention focus was classified as education-focused.

⁶ “Endpoints” refer to sites with separate results. Studies with one site report only one endpoint, whereas a study with e. g. seven sites is listed here with seven endpoints.

Nine endpoints (labelled “other”) had a detailed screening process at the outset. The results of this screening influenced whether the participant started job search immediately or began with some kind of educational activity. Finally, in two instances we could not decide what type of intervention focus was in place.

The first activity is closely related to the intervention focus in that employment-focused interventions tend to start with job search assistance, job clubs or work placement, while the education-focused ones usually have more varied first activity, often depending on the job-readiness of the individual client. For 24 endpoints, we coded the intervention as job-search first. For 20 endpoints, we coded varied first activity, and for 26 we coded it as other (Table 5).

Sanctioning: If participants did not comply with their assignments in the mandatory programmes, programme managers could sanction them. This could mean denying them their welfare payments. Across all the 34 mandatory programmes (74 percent of the programmes were classified as mandatory), we estimated that about 17 percent of participants were sanctioned.

Reporting of outcomes

As shown in Table 6, most of the 73 endpoints (N = 46, 63%) had data on earnings and employment from state and county administrative records (register data).

Meta-analysis

Figures 3 through 6 show meta-analytic summaries of effects grouped according to follow-up time. For each follow-up time, there are two rows that show the fixed-effects and random-effects estimates. A few studies have results for more than one follow-up. These are included in the rows labelled “Combined”. The effect magnitudes with lower and upper bounds of the confidence intervals are shown in the middle part. The schematic on the right is a summary forest plot. The middle of the (orange-coloured) diamonds is located at the estimated effect magnitude, and their widths indicate the length of the 95 percent confidence interval around the estimate. At the bottom, there are two summary rows which give an impression of the overall estimates (with green diamonds). Note that these summaries are meaningful only if the effects do not vary systematically with follow-up time. For proportional outcomes (proportion employed and proportion on welfare) we use risk ratios. The risk ratio is e.g. the relative “risk” of becoming employed. If, for instance, 55 out of 100 in the intervention group become employed and 50 out of 100 in the control group become employed, the risk ratio is $0.55/0.50 = 1.1$. A risk ratio of 1.00 indicates no effect. A number greater than 1 indicates that the intervention group is becoming employed to a greater degree, and a number smaller than 1 indicates that the control group is becoming employed to a greater degree.

For continuous outcomes (such as earnings) we use Hedges’ *g*. A *g* of zero means no difference between the intervention group and the control group. Large samples usually have more precise estimates than small samples. This precision is measured by the standard error of the estimate.

Each study is weighted by its precision. This means that large and precise studies are given more weight in the analyses. While the confidence intervals for an individual study can be

wide, the meta-analysis confidence intervals are much narrower, signalling a more precise estimate.

Heterogeneity is a measure of the spread of results. Statistical tests of homogeneity (i.e. lack of heterogeneity) assess whether the individual study results are likely to reflect a single underlying effect, as opposed to a distribution of effects. If this test fails to detect heterogeneity among results, then it is assumed that the differences observed between individual studies are a consequence of sampling variation and simply due to chance.

Impact on employment, earnings, welfare payment, and proportion on welfare

Impact on Employment

Figure 3 shows the impact on employment graphically for all follow-up lengths. Each follow-up time has two separate meta-analyses (fixed-effects and random-effects). Thus, there are two separate meta-analyses for ‘employment 2 years’ and two others for ‘employment 1-3 years’. Figure 3 shows that the employment at 1 year random-effects risk ratio is 1.097 (95% CI: 1.006-1.196). The employment at 2 years random-effects risk ratio is 1.092 (95% CI: 1.032-1.157), and the five year random-effects risk ratio is 1.037 (1.004-1.071). Homogeneity tests show that, in addition, the following follow-up times should use random-effects summaries: 0.25 years and 1-3 years. The remaining follow-ups were homogeneous and should use fixed effects.

There are three overall impressions: (1) Effects are more and less consistently in favour of the intervention group; (2) effects vary, and (3) effects are small. The intervention group does better than the control group, but the effect is small. The overall effects are heterogeneous ($Q_{\text{all studies}} = 414.967$, $df = 59$, $p = .000$, $I^2 = 85.8$).

Overall, 60.9 percent of intervention participants (weighted by sample size) were employed at the follow-ups. But 57.9 percent of control participants were also employed. Another way to look at these results is to compute the number needed to treat (NNT). The absolute risk reduction (ARR) was $0.609 - 0.579 = 0.03$. The number needed to treat was $1/ARR \approx 33$ (95% confidence interval: 30-37). This was estimated by first computing the confidence limits for the ARR⁷ and computing the confidence limits for the NNT as the reciprocals of the ARR confidence limits. In other words, an average of 33 welfare recipients had to receive one of the work programmes in this review in order to predict employment of one more recipient.

When 54 employment outcomes were ranked according to effect size, the GAIN programme in Riverside ranked in 6th place, while the same intervention in Alameda, Los Angeles, San Diego, Tulare, and Butte ranked 9, 13, 21, 28, and 47, respectively. This is an indication that the local context in which a programme is executed has a more powerful effect on the outcome than the programme itself.

⁷ Standard deviation for **ARR** = $ARR = \sqrt{p(1-p)\left(\frac{1}{nG_1} + \frac{1}{nG_2}\right)}$ where p = the weighted (pooled)

proportion for both groups $p = \frac{(nG_1 pG_1 + nG_2 pG_2)}{(nG_1 + nG_2)}$. nG_1 is the number in group 1, nG_2 is the number in group 2, pG_1 is the proportion in group 1 and pG_2 is the proportion in group 2.

Impact on Earnings

Because few studies reported the standard errors necessary for direct computation of Hedges' g , we had to use p -values to estimate most of the standard errors. This was the only way to estimating effect sizes. Only a few studies reported exact p -values. If a standard error was reported, we used it. If not, we used p -values. If an exact p -value was reported, we used this value. Otherwise, we used the following conversion:

“Not significant” → $p = 0.1$

The included studies mostly used $p = 0.1$ as the critical value for choosing between significant/not significant. We decided that any other value of p would be arbitrary. This probably caused some comparisons reported as nonsignificant to have caused an upward bias to effect-magnitudes. Therefore we ran two sensitivity analyses. In that way, we established a range of probable results. In the first scenario we used $p = 0.99$ for nonsignificant findings. In the second scenario, we used $p = 0.55$ (midpoint between 0.1 and 1.0) for nonsignificant findings.

$P < 0.1$ → $p = 0.075$

The rationale is simply that this value is half way between 0.1 and 0.05.

$P < 0.05$ → $p = 0.03$

This value is half the way between 0.05 and 0.01

$P < 0.01$ → $p = 0.01$

Every smaller value would be arbitrary. This causes some comparisons with very low p -values to be underestimated. However, some exploratory analyses with extremely small p -values (not shown) did not change the estimated effect size to any large degree.

Figure 4 shows the impact on earnings. Overall, the intervention groups did slightly better than the control group, but the difference was small. The results are heterogeneous ($Q = 180.96$, $df = 65$, $p = .000$, $I^2 = 64.1$). At the one-year follow-up, the random-effects Hedges' g was 0.043 (95% CI: 0.011-0.076). At two years the random-effects g was 0.044 (0.022-0.066), and at five years the random-effects g was 0.011 (-0.029-0.050). Apart from these, the 1-3 years analysis should also use random effects.

Sensitivity analyses showed that using $p = 0.99$ for nonsignificant results lowered the overall g from 0.047 to 0.032 (0.022-0.042), and that using $p=0.55$ produced a g of 0.035 (0.024-0.045). The three confidence intervals overlap, and, hence, we used $p=0.1$ for all nonsignificant analyses.

The mean earnings (in year 2005 US dollars) across all the intervention outcomes were \$ 11,021 compared to \$ 8,843 in the control groups. Two studies were conducted in Canada. We converted the Canadian dollars into US dollars using the exchange rate of 1CAD = 0.873 USD (17 April, 2006).

Using the *Binominal Effect Size Display* (BESD⁸) the results roughly correspond to a positive effect for 51.1 % in the intervention groups and for 48.9 percent in the control group.

⁸ BESD shows the proportion in the intervention group with a positive result as $0.5 + r/2$ and the proportion in the control group with positive result as $0.5 - r/2$. First we had to convert the standardised mean difference into r .

On a ranked list of 64 effect magnitudes, the different sites at which the GAIN programme were implemented ranks 5, 18, 29, 36, 43 and 44. Similarly, the NEWWS evaluation sites rank 12, 13, 14, 24, 25, 27, 28, 33, 38, 41, and 57. This is again an indication that aspects of local implementation may matter more than the programme model itself (note that these comparisons are inherently quasi-experimental).

Impact on Welfare Payments

Figure 5 shows the impact on welfare payments. Once again, the intervention group mostly does better than the control group (the intervention group receives less welfare payments), but the effect is small (overall random effects Hedges' $g = 0.043$, 95% CI: 0.034-0.053). The results are heterogeneous ($Q = 105.7$, $df = 42$, $p = .000$, $I^2 = 60.3$). At one year follow up the random-effects g was 0.038 (95% CI: -0.022-0.098). At two years, the random-effects g was 0.053 (-0.005-0.111), and at 5 years the fixed-effects g was 0.044 (0.028-0.060). The 1-3 years follow-up was heterogeneous and should use random effects.

The BESD indicates that 51.2 percent in the intervention groups and 48.8 percent in the control groups would receive less welfare payments as a result of taking part in a programme.

Using $p = 0.99$ for nonsignificant results reduces g from 0.043 to 0.040 (0.026-0.053), and using $p = 0.55$ estimates g to be 0.042 (0.029-0.055). Again, the CI's overlap, and we used $p = 0.1$ for all nonsignificant analyses.

On a ranked list of 43 effect magnitudes, the sites of the GAIN evaluation rank 7, 13, 20, 22, 29, and 39. The California Work Pays evaluation sites rank 2, 14, and 25.

Impact on Welfare Proportion

Figure 6 shows the impact of the programmes on the proportion of people who were still receiving welfare at the follow-ups. Note that "favours treatment" is on the left side. Overall, the intervention groups do slightly better than the control groups (the intervention group participants are less likely to be on welfare at the follow-up), but the effect is small. At the one-year follow-up, the random-effects risk ratio is 0.967 (95% CI: 0.926-1.009). At two years, the random-effects risk ratio is 0.946 (0.886-1.010), and at five years, the fixed-effects risk ratio is 1.003 (0.984-1.023). The results for the 1-2 years and the 3 years follow-ups were also heterogeneous. The overall results are heterogeneous ($Q = 529.292$, $df = 38$, $p = .000$, $I^2 = 92.8$). Overall, the proportion of intervention participants living on welfare at the follow-ups was 68.4% compared to 72.1% in the control groups. This gives an absolute risk reduction of 0.037 and a NNT of 27 (95% confidence interval 24-30).

On a ranked list of 40 effect magnitudes, the California Work Pays evaluation sites rank 10, 18, and 37.

The differential effect of employment focus and education focus

As part of the JOBS (NEWWS) evaluation, in three sites (Atlanta, Grand Rapids and Riverside), subjects were randomised into either a HCD (human capital development) condition, a LFA (labour force attachment) condition, or a control condition.

LFA versus HCD

Employment: A random-effects meta-analysis at the three sites showed that the LFA approach had a slightly better effect on employment (risk ratio: 1.094, 95% CI: 1.056-1.133) than the HCD approach (risk ratio: 1.049, 95% CI: 1.003-1.098), but the confidence intervals have a large overlap. Overall, the results were heterogeneous ($Q = 25.7$, $df = 5$, $p = .000$, $I^2 = 80.5$). The subgroups were also heterogeneous (HCD: $Q = 8.7$, $df = 2$, $p = .013$, $I^2 = 76.9$ / LFA: $Q = 4.8$, $df = 2$, $p = .087$, $I^2 = 59.0$)

No direct comparisons on the effects on earnings, welfare payments or welfare proportion were available for the LFA and HCD approaches.

Integrated or traditional case-management

At Columbus, Ohio the participants were randomised into either a traditional case management approach or an integrated approach. In the traditional case management model, welfare eligibility and employment programme functions were performed by separate staff members. In the integrated case management model, these two functions were performed by the same staff member. At two years the two approaches generated almost exactly the same earnings that were also close to the control group (integrated: Hedges' $g = 0.076$, 95% CI: 0.018-0.133, traditional: Hedges' $g = 0.075$, 95% CI: 0.018-0.132. At five years, the approaches did not have a statistically significant effect on employment, and neither did they differ (integrated: risk ratio = 1.022, 95% CI: 0.999-1.045, traditional: risk ratio = 1.012, 95% CI: 0.989-1.035).

Publication bias

Figures 7-10 are funnel plots showing for each of the four outcomes the relation between each study's effect size and its standard error. In the absence of publication bias the studies will be distributed symmetrically about the combined effect size. By contrast, in the presence of bias, the bottom of the plot would tend to show a higher concentration of studies on one side of the mean than the other. This would reflect the fact that smaller studies (which appear toward the bottom) are more likely to be published if they have larger than average effects, which makes them more likely to meet the criterion for statistical significance. We have used a "trim and fill" function which builds on the key idea behind the funnel plot; that in the absence of bias the plot would be symmetric about the summary effect. If there are more small studies on the right than on the left, the concern is that studies may be missing from the left. The trim and fill procedure imputes these missing studies, adds them to the analysis, and then re-computes the summary effect size. Figures 7-10 indicates that there is some asymmetry in the plots for employment, earnings, and welfare payments, but not for welfare proportion.

Exploratory moderator analyses

Because the results were heterogeneous, we attempted to explore this heterogeneity using moderator analyses. Note that by doing this, randomisation is lost, and, hence, the following are quasi-experimental results. Also note that they are exploratory hypothesis-generating analyses. They are not theory-driven, hypothesis-testing analyses.

Data on a number of variables were collected, but not used as possible moderators. Country where the study was conducted was e.g. not used because all but two studies came from the USA. For the same reason, publication type was not used because the majority of citations were reports. Publication year was not used because we used the year when the data collection

started. Design was not used, since all studies used RCT. Some potentially interesting variables were not used because of a large number of missing values (number of children, age of youngest child, previous work experience, previous welfare history, extent of work experience). Finally we did not use data on funding and implementing agents. Studies were mostly funded and implemented by federal and state authorities.

Continuous moderators

Table 7 provides descriptive data on continuous moderators. Table 8 shows the zero-order, non-parametric Spearman correlations among moderators.

Tables 9 through 12 show the results of meta-regressions using the continuous moderators. As indicated above, one moderator was entered into these regressions at a time; thus, these are bivariate (not multivariate) regressions that do not control for influences of other moderators.

Table 9 shows for instance that if, for example, the mean age in a study is raised by one year, the proportion becoming employed is expected to be 0.8 percentage points higher (slope: 0.00829). In the following we list the regressions where the 95 percent confidence interval for the regression slope does not cross the point of no effect. An asterisk is also listed at the end of these variable names in the tables. All in all, there were 48 such comparisons. We also note the deviations from this pattern in the sensitivity analyses using imputation of the mean for missing data.

Employment (Table 9)

- Positive associations: mean age, percent other ethnic group, duration of intervention, and regional unemployment.
- Negative associations: proportion of Whites, proportion of Blacks, and percent sanctioned
- In the sensitivity analysis, proportion of Hispanics also showed a positive association.

Earnings (Table 10)

- Positive association: proportion of Blacks
- Negative associations: year when data collection started, proportion of Whites, proportion of Hispanics, and proportion with a general educational diploma or high-school degree.
- In the sensitivity analysis, proportion sanctioned showed a negative association. We also failed to reproduce the negative association with GED.

Welfare payments (Table 11)

- Negative associations: proportion of males, proportion of Whites
- In the sensitivity analysis, duration of intervention showed a positive association.

Welfare proportion (Table 12)

- Negative associations: mean age, proportion of males, proportion of Hispanics, proportion of other ethnic group, duration of intervention, proportion sanctioned, regional unemployment, and number of intervention elements.
- In the sensitivity analysis, we also found a positive association with proportion of Blacks, and a negative association with proportion of Whites.

The proportion of males, proportion of Whites, and percent sanctioned have negative slopes for all outcomes, but these slopes are not always statistically significant. These findings indicate (but only very weakly) that the programmes might have been more effective for women, and for non-Whites, and that programmes with less sanctioning have been more effective. For the other variables, the direction of the slope varies. The sensitivity analyses found five additional significant slopes (11%), while one significant slope (2%) in the main analysis was not significant in the sensitivity analysis.

For all regressions, the residual Q was significant at $p = .00000$. This indicates that any single moderator explains only a very small part of the variability in effect magnitudes.

Categorical moderators

Tables 13 through 20 show the effects of categorical moderators on employment, earnings, welfare payments, and welfare proportion. An asterisk is listed at the end of the variable name if the between-group heterogeneity p-value is below 0.05. These are interpreted as statistically significant differences among the categories of moderators.

Design qualities:

Table 13 shows the separate risk ratios for each level of the design quality variables regarding effect on employment. The number (k) of outcomes contributing to each level is shown first. For example: we coded 12 outcomes as “Met” on random allocation. The separate meta-analysis for these outcomes showed a risk ratio of 1.064 with a 95 percent confidence interval from 1.016 to 1.115. This risk ratio was significantly different from zero ($p = .008$). The $Q_{\text{within groups}}$ was 53.9 and the I-squared was 79.6. The $Q_{\text{between groups}}$ was 26, which shows that this meta-analytic analogue of ANOVA is significant. In other words, the effect magnitude varies over the different quality ratings of random allocation. Specifically, studies coded as “Not Met” have smaller risk ratios than studies coded as “Met” and “Unclear”.

Tables 15, 17, and 19 do the same as Table 13 for the effects of design characteristics on earnings, welfare payments and welfare proportion.

Overall, these analyses show that “Not Met” and “Unclear” are associated with very high or very low effect magnitudes. High-quality reporting (Met) is associated with larger effect magnitudes in 54% (13/24) of the comparisons. The “Unclear” category had strongest effects in 21% of the comparisons (5/24), and the “Not Met” category had strongest effect sizes in 25% (6/24) of the comparisons.

A coding of “Met” on allocation concealment is always associated with the largest effect magnitude (although the difference was not always statistically significant).

Other categorical moderators:

Tables 14, 16, 18, and 20 show estimated effects of the other categorical variables on employment, earnings, welfare payments and welfare proportion.

Other consistent findings:

- Voluntary programmes show larger effects than mandatory programmes, but differences are not statistically significant.
- Programmes with work experience show larger effects than the ones without work experience (again differences are not always significant).

- Programmes with job search show larger effects than programmes without job search.²⁰

We note once more that these associations are not of a causal nature and should be taken only as exploratory analyses.

Discussion

A major finding of this review is that randomised effect evaluations of welfare-to-work programmes come almost exclusively from the USA. Why is this so? In the USA, having a randomly allocated control group has sometimes been a condition for receiving money to evaluate a programme. But randomisation has also sometimes been a condition for receiving federal waivers that allow proposed programmes to go forward. The latter has probably been more important in stimulating the use of random assignment in the United States than the former.

In Europe, randomisation is often regarded as unethical, because the control group is denied a possibly beneficial intervention. In the United States this is not looked upon as a problem. In fact, they argue, whenever only some of the welfare recipients can receive a programme, random allocation is the fairest way of sharing these scarce resources. It has also been claimed that not using the best design is unethical because if not, one will never learn whether the intervention works, is ineffective or harmful. “Possibly beneficial” interventions are not necessarily beneficial and may, in fact, be harmful; and we will not know whether something is beneficial or harmful until we do randomised controlled trials.

How reliable are the outcome data? The programmes mainly used administrative data from state and county registers. A few also used client surveys. Each data source has strengths and limits. Whereas administrative data are relatively free from recall and expectancy biases, they will miss information on some events, e. g. employment in jobs that are not reported to the authorities. We can see no reason why this would affect intervention and control groups differently, but there may be some bias due to differential employment rates and differential attrition.

How reliable are the study quality ratings? Study quality is not the same as the quality of reporting. For instance, we have reason to believe that all the evaluations conducted by MDRC used a proper method for generating the random allocation sequence. But since this was often not explicitly reported, we had to code the studies as “Unclear”.

How strong and robust are the effects? The evaluations have clearly shown that the programmes can increase the probability of becoming employed, increase earnings, and reduce welfare payments. The results are not so clear about a reduction of the proportion of participants on welfare. However, no meta-analysis significantly favoured the control group. This was true even for welfare proportion. For every follow-up time and all of the four main outcomes, either the intervention group did significantly better than the control group, or there was no statistically significant difference between the groups (Figures 3 through 6).

For welfare recipients, the programmes only have a net effect if it can raise their total income. This occurs only if the increase in earnings is greater than the reductions in welfare payment. Cost-benefit analysis is not part of the present review though.

Some readers of the welfare-to-work literature may be unfamiliar with the reporting of results as effect sizes. They are used to having results presented in their original metrics (such as

dollars for earnings and percent for employment). We have attached all the collected data (as Microsoft Excel files) in Appendix C. For each study and outcome the proportion employed or on welfare is listed, and the dollar amounts in 2005 equivalents are used for continuous measures. We invite readers to examine these data.

Can these results be generalised to the rest of the world (e. g. Europe)? The results cannot be generalised because European countries have welfare systems that are very different from the American system. In many European countries, everyone typically have access to welfare if they are outside the labour market. In Norway, the workfare programme has lead to about 23 percent becoming employed (Lødemel et al. 2004) compared to about 60 percent in the US studies reviewed here. The only way to measure the effect of workfare under the European welfare model is to do randomised controlled evaluations.

The two Canadian programmes (SSP and SSP+) seem to perform somewhat better than the US programmes according to a ranking of effect sizes. On a list of 54 employment outcomes, SSP+ places second and SSP places fourth. On a list of 44 welfare payment outcomes, SSP+ is the “winner”, while SSP is at place 19. Finally, on a list of 36 welfare proportion outcomes, SSP places second. Again, these are only quasi-experimental comparisons.

Are mandatory programmes more effective than voluntary ones? To our knowledge, no welfare-to-work programme has randomised welfare recipients to either a mandatory condition or a voluntary one (and this is may not be feasible), so this question is still largely unanswered. Also, the operational differences between voluntary and mandatory approaches are often cloudy, rather than representing a clear dichotomy (Gueron, Pauly, & Lougy 1991). However, we have classified each programme as either mandatory or voluntary and performed separate analyses for the two subgroups. The exploratory analyses showed that the voluntary programmes have larger effect sizes, but these differences are not statistically significant and this is only indirect evidence that they are more effective than the mandatory ones.

Missing outcomes: We were interested in whether the programmes would affect quality of life, social skills and self-confidence of the participants, and whether the programmes would have any adverse effects on the participants. These outcomes were lacking in most of the included studies.

Reviewers' conclusions

Work programmes for welfare recipients in the USA have had a small but consistently positive effect on employment, earnings, and welfare payments during the first six years after the intervention. The effects on welfare proportion are close to zero. Because evaluations from outside North America lack randomised controls, no conclusions can be drawn about how such interventions might work under other kinds of welfare systems and in other countries.

Implications for practice

US policy makers should weigh the cost of putting 33 welfare recipients through a work programme against the anticipated benefit of employment for one more person (i.e., employment of 20/33 people instead of 19/33). They should also consider the benefit of putting 27 welfare recipients through a programme against the anticipated benefit of reducing the percent on welfare from 19/27 to 18/27. (The cost of enrolling a welfare

recipient in a programme should be weighed against a predicted raise in earnings of approximately 25 % and a predicted reduction in welfare payments of approximately 14 %.

Only two studies have been evaluated in Canada, but the results are somewhat more promising in this country than in the USA.

We do not think this review should have any implications for practice in other countries.

Implications for research

Policy makers and research funders should allocate resources to conducting randomised controlled trials in countries outside of the USA. When reporting results from randomised trials, authors should report both mean values and standard deviations for continuous outcomes. At the very least, exact p-values should be reported. We welcome studies in which randomised trials are accompanied by qualitative evaluations. Without the latter, systematic reviews can only inform society about what works, not why or how.

Plans for updating the review

Searches will be performed every two years after publication of the full review. Results of searches will be published and classified either as (1) search performed [date] but no new studies found, (2) minor update: new studies found, but conclusions are unchanged, or (3) major update: new studies found and consumers who read the previous version are advised to read the whole update. Future versions of the review will, hopefully, be able to synthesise results from randomised trials and qualitative evaluations.

Potential conflicts of interest

None known.

INCLUDED CITATIONS

1. Aber, J. L., Brooks, G. J., & Maynard, R. A. 1995, "Effects of welfare reform on teenage parents and their children. [Review] [60 refs]", *Future Choices.*, vol. 5, pp. 53-71.
2. Auspos, P, et al. 1988, *Maine. The Demonstration of State Work/Welfare Initiatives. Final Report on the Training Opportunities in the Private Sector Program*, MDRC, Manpower Demonstration Research Corporation, New York.
3. Beccera, R. M., Lew, V., Mitchell, M. N., & Ono, H. 1998, *California Work Pays Demonstration Project: Report of first forty-two months. January 1993 through June 1996*, School of Public Policy and Social Research, University of California, Los Angeles, California, Los Angeles.
4. Beecroft, E., Lee, W., & Long, D. 2003, *The Indiana Welfare Reform Evaluation: Five-Year Impacts, Implementation, Costs and Benefits*, Abt Associates Inc., Cambridge, MA.
5. Bell, S. H., Burstein, N. R., & Orr, L. L. 1987, *Overview of evaluation results: Evaluation of the AFDC Homemaker-Home Health Aide Demonstrations*, Abt. Associates Inc., Cambridge, Massachusetts.
6. Bell, S. H. & Orr, L. L. 1994, "Is Subsidized Employment Cost-Effective for Welfare Recipients? Experimental Evidence from Seven State Demonstrations", *The Journal of Human Resources*, vol. Winter 1994.
7. Bloom, D. & AND, O. T. H. E. 1997, *The Family Transition Program: Implementation and Early Impacts of Florida's Initial Time-Limited Welfare Program*, Manpower Demonstration Research Corp, New York.
8. Bloom, D., Hendra, R., Michalopoulos, C., Redcross, C., Scrivener, S., & Walter, J. 2002a, *WRP: Final Report on Vermont's Welfare Restructuring Project*, MDRC (Manpower Demonstration Research Corporation), Washington.
9. Bloom, D., Kempl, J., Morris Pamela A, Scrivener, S., & Walter, J. 2000a, *The Family Transition Program: Final Report on Florida's Initial Time-Limited Welfare Program*, Manpower Demonstration Research Corporation, Washington.
10. Bloom, D., Melton, L., Michalopoulos, C., Scrivener, S., & Walter, J. 2000b, *Jobs First: Implementation and Early Impacts of Connecticut's Welfare Reform Initiative*, Manpower Demonstration Research Corporation.
11. Bloom, D., Scrivener, S., Michalopoulos, C., Morris, P., Hendra, R., Adams-Ciardullo, D., & Walter, J. 2002b, *Jobs First: Final Report on Connecticut's Welfare Reform Initiative*.
12. Bloom, H., Orr, L. L., Cave, G., Bell, S. H., Doolittle, F., & Lin, W. 1997, "The Benefits and Costs of JTPA Programs: Key Findings from the national JTPA Study", *The Journal of Human Resources*, vol. 32, no. 3 (summer), pp. 549-76.

13. Bloom, H., Orr, L., & Gave, G. 1992, *The National JTPA Study: Title IIA impacts on earnings and employment at 18 months*. Abt Associates, Bethesda, MD.
14. Bos, J. M., Huston, A. C., Granger, R. C., Duncan, G. J., Brock, T. W., & McLoyd, V. C. 1999, *New Hope for people with low incomes: Two-year results of a program to reduce poverty and reform welfare*, MDRC.
15. Brock, T., Doolittle, F., Fellerath, V., Greenberg, D., Hollister, R., Jr., & Wiseman, M. 1997, *Creating New Hope: Implementation of a Program To Reduce Poverty and Reform Welfare*, Manpower Demonstration Research Corporation.
16. Burghardt, J., Rangarajan, A., Gordon, A., & Kisker, E. 1992, *Evaluation of the Minority Female Single Parent Demonstration: Summary report*, Mathematica Policy Research.
17. Cancian, M., Kaplan, T., & Rothe, I. 2000, *Wisconsin's Self-Sufficiency /Pay for Performance: Results and lessons from a social experiment*, Institute for Research on Poverty, University of Wisconsin-Madison.
18. Carcagno, G, et al. 1982, "Using Private Employment Agencies to Place Public Assistance Clients in Jobs", *Journal of Human Resources*.v17 n1 p132-43 Win.
19. Couch, K. 1992, "New evidence on the long-term effects of employment training programs", *Journal of Labor Economics*, vol. 10, pp. 380-388.
20. Farrel, M. 2000, *Implementation, Participation Patterns, Costs, and Two-Year Impacts of the Detroit Welfare-to-work Program*, Manpower Demonstration Research Corporation.
21. Fein, D. J., Beecroft, E., & Blomquist, J. D. 1994, *The Ohio Transitions To Independence Demonstration - Final Impacts For Jobs And Work Choice*, Abt. Associate Inc., Cambridge, MA.
22. Fein, D. J., Beecroft, E., Karweit, J., Holcomb, P. A., Clark, S. J., O'Brien, C. T., & Ratcliffe, C. E. 1997, *The Indiana Welfare Reform Evaluation: Assessing Program Implementation and Early Impacts on Cash Assistance*, Abt. Associates Inc., Cambridge, MA.
23. Fein, D. 1994, *The Ohio Transitions to Independence Demonstration. JOBS Assignments in Ohio: Patterns and Impacts. Executive Summary*.
24. Fein, D., Beecroft, E., Hamilton, W., Lee, W., Holcomb, P. A., Thompson, T., & Ratcliffe, C. 1998, *The Indiana welfare reform evaluation: Program implementation and economic impacts after two years*. Abt Associates Inc., Bethesda, MD.
25. Fein, D. & Karweit, J. A. 1997, *The ABC Evaluation: The Early Economic Impacts of Delaware's A Better Chance Welfare Reform Program*, Abt. Associates, Cambridge, Mass.
26. Fein, D. & Long, D. A. 1999, *The ABC evaluation: Delaware's a better chance welfare reform program at four years* Abt Associates Inc, Bethesda, MD.

27. Fraker, T. M., Ross, C. M., Stapulonis, R. A., Olsen, R. B., Kova, M. D., Dion, M. R., & Rangarajan, A. 2002, *The Evaluation of Welfare Reform in Iowa*, Mathematica Policy Research, Inc., Washington, DEA-0699.
28. Fraker, T., Nixon, L., Jacobson, J., Gordon, A., & Martin, T. 1998, *Iowa's Family investment program: two-year impacts*, Mathematica Policy Research, Washington, DC.
29. Freedman, S. 1995, *The JOBS evaluation: Early findings on program impacts in three sites*, U.S. Department of Health and Human Services, Administration for Children and Families, Washington, DC.
30. Freedman, S. 2001, *Four-Year Impacts of Ten Programs on Employment Stability and Earnings Growth. The National Evaluation of Welfare-to-Work Strategies*.
31. Freedman, S. 2003, *Pursuing Economic Security for Young Adults: Five-Year Impacts of Pre-Employment Services in the National Evaluation of Welfare-to-Work Strategies*, Manpower Demonstration Research Corporation.
32. Freedman, S., Bryant, J., Cave, G., Bangser, M., Friedlander, D., Goldman, B., & Long, D. 1988, *New Jersey. The Demonstration of State Work/Welfare Initiatives. Final Report on the Grant Diversion Project*, MDRC, Manpower Research and Demonstration Project, New York.
33. Freedman, S., Friedlander, D., Hamilton, G., Rock, J., Mitchell, M., Nudelman, J., Schweder, A., & Storto, L. 2001a, *Evaluating Alternative Welfare-to-Work Approaches: Two-Year Impacts for Eleven Programs. National Evaluation of Welfare-to-Work Strategies*.
34. Freedman, S., Knab, J., Gennetian, L., & Navarro, D. 2001b, *The Los Angeles Jobs-First GAIN Evaluation: Final Report on a Work First Program in a Major Urban Center*.
35. Freedman, S., Mitchell, M., & Navarro, D. 1997, *The Los Angeles Jobs-First GAIN Evaluation. First-Year Findings on Participation Patterns and Impacts*, MDRC, Manpower Demonstration Research Corporation, New York.
36. Friedlander, D. 1987, *Maryland: Supplemental Report on the Baltimore Options Program*, Manpower Demonstration Research Corporation, MDRC, New York.
37. Friedlander, D. 1988, *Subgroup Impacts and Performance Indicators for Selected Welfare Employment Programs*.
38. Friedlander, D, et al. 1987, *West Virginia. The Demonstration of State Work/Welfare Initiatives. Final Report on the Community Work Experience Demonstrations*, MDRC, Manpower Demonstration Research Corporation, New York.
39. Friedlander, D. & Burtless, G. 1994, *Five years after: The long-term effects of welfare-to-work programs*. Russel Sage Foundation, New York.

40. Friedlander, D. & et.al 1985, *Arkansas: Final Report on the WORK Program in Two Counties*, MDRC, Manpower Demonstration Research Corporation, New York.
41. Friedlander, D., Freedman, S., Hamilton, G., & Quint, J. 1987, *Illinois: The Demonstration of State Work/Welfare Initiatives: Final Report on Job Search and Work Experience in Cook County*, MDRC, Manpower Demonstration Research Corporation, New York.
42. Friedlander, D. & Goldman, B. 1988, *Employment and Welfare Impacts of the Arkansas Work Program: A three-year follow-up study in two counties*, Manpower Demonstration Research Corporation, New York.
43. Friedlander, D. & Hamilton, G. 1993, *Saturation Work Initiative Model in San Diego: A Five-Year Follow-up Study*.
44. Friedlander, D. & Hamilton, G. 1996, "The Impact of a Continuous Participation Obligation in a Welfare Employment Program", *Journal-of-Human-Resources*, vol. 31, no. 4, pp. 734-756.
45. Friedlander, D., Hoerz, G., Long, D., Quint, J., Goldman, B., & Gueron, J. M. 1985, *Maryland: Final Report on the Employment Initiatives Evaluation*, MDRC, Manpower Demonstration Research Corporation, New York.
46. Friedlander, D. & Long, D. 1987, *A Study of Performance Measures and Subgroup Impacts in Three Welfare Employment Programs.*, MDRC, Manpower Demonstration Research Corporation, New York, RR-87-28.
47. Friedlander, D., Riccio, J., & Freedman, S. 1993, *GAIN: Two-Year Impacts in Six Counties. California's Greater Avenues for Independence Program.*
48. Friedman, B., Davenport, B., Evans, R., Hahn, A., Hausman, L., & Papirno, C. 1980, *An evaluation of the Massachusetts Work Experience Program*, Brandeis University, Florence Heller Graduate School.
49. Gennetian, L., et al. 2000, *The effects of welfare reform in rural Minnesota: experimental findings from the Minnesota Family Investment Program.*
50. Gennetian, L. A. & Miller, C. 2002, "Children and welfare reform: A view from an experimental welfare program in Minnesota", *Child Development*, vol. 73, no. 2, pp. 601-620.
51. Goldman, B., Friedlander, D., & Long, D. 1986, *Final report on the San Diego Job Search and Work Experience Demonstration*, MDRC, Manpower Research Demonstration Corporation, New York.
52. Gordon, A. & James-Burdumy, S. 2002, *Impacts of the Virginia Initiative for Employment Not Welfare. Final Report*, Mathematica Policy Research, Inc., Princeton, New Jersey, 8477-120.

53. Gordon, A. & Agodini, R. 1999, *Early impacts of the Virginia Independence Program. Final report submitted to the Virginia Department of Social Services, Mathematica Policy Research, Inc., Princeton.*
54. Hamilton, G., Brock, T., Farrel, M., Friedlander, D., Hunter-Manns, J., Walter, J., & Weisman Joanna 1997, *National Evaluation of Welfare-to-Work Strategies: Evaluating two Welfare-to-Work Program Approaches: two-year findings on the on the labor force attachment and human capital development program in three sites, Manpower Demonstration Research Corporation.*
55. Hamilton, G., Freedman, S., Gennetian, L. A., Michalopoulos, C., Walter, J., Adams-Ciardullo, D., & Gassman-Pines, A. 2001, *How Effective Are Different Welfare-to-Work Approaches? Five-year Adult and Child Impacts for Eleven Programs, Manpower Demonstration Research Corporation.*
56. Hamilton, G. & Friedlander, D. 1989, "Saturation Work Initiative Model in San Diego. Final Report", *Manpower Demonstration Research Corp., New York, N.Y.* p. 213.
57. Hamilton, W. L., Burstein, N. R., Baker, A. J., Earle, A., Gluckman, S., Peck, L., & White, A. 1996, *The New York State Child Assistance Program: Five-year impacts, costs, and benefits, Abt Associates.*
58. Hamilton, W. L., Burstein, N. R., Hargreaves, M., Moss, D. A., & Walker, M. 1993, *The New York State Child Assistance Program: Program impacts, costs, and benefits, Abt Associates.*
59. Hendra, R., Michalopoulos, C., & Bloom, D. 2002, *Three-Year Impacts of Connecticut's Jobs First Welfare Reform Initiative.*
60. Hogarth, S. & Martin, R. 1988, *Pennsylvania Saturation Work Program: Process evaluation, Pennsylvania Department of Public Welfare.*
61. Hogarth, S., Martin, R., & Nazar, K. 1989, *Pennsylvania Saturation Work Program: Impact evaluation, Pennsylvania Department of Public Welfare.*
62. Hollister, R., Kemper, P., & Maynard, R. 1984, *The national supported work demonstration University of Wisconsin Press, Madison.*
63. Horwitz, S. M. & Kerker, B. D. 2001, "Impediments to employment under welfare reform: the importance of physical health and psychosocial characteristics", *Women & Health.*, vol. 32, no. 1-2, pp. 101-117.
64. Kemper, P., Long, D. A., & Thornton, C. 1981, *The Supported Work Evaluation: Final Benefit-Cost Analysis, Manpower Demonstration Research Corporation, New York, Vol 5 of the Final Report.*
65. Kemple, J. J., Friedlander, D., & Fellerath, V. 1995, *Florida's Project Independence. Benefits, Costs, and Two-Year Impacts of Florida's JOBS Program, MDRC, Manpower Demonstration Research Corporation, New York.*

66. Kemple, J. & Haimson, J. 1994, *Florida's Project Independence. Program Implementation, Participation Patterns, and First-Year Impacts*, MDRC, Manpower Demonstration Research Corporation, New York.
67. Kisker, E. E., Rangarajan, A., & Boller, K. 1998, *Moving into adulthood: Were the impacts of mandatory programs for welfare-dependent teenage parents sustained after the programs ended?*, Mathematica Policy Research.
68. Kornfeld, R., Peck, L., Porcari, D., Straubinger, J., Johnson, Z., & Cabral, C. 1999, *Evaluation of the Arizona EMPOWER Welfare Reform Demonstration - Impact Study Interim Report*, Abt Associates Inc., Cambridge, MA.
69. Lei, Y. & Michalopoulos, C. 2001, *SSP Plus at 36 Months: Effects of Adding Employment Services to Financial Work Incentives*.
70. Lerman, R., Friedman, B., Ajemian, S., Fairchild, C. K., Jastrzab, J., Kulik, J., Logan, C., Papirno, C., & Seitchik, A. 1986, *Food Stamp Work Registration and Job Search Demonstration. Final Report*, Brandeis University, Center for Human Resources, Abt Associates.
71. Lin, W., Robins, P. K., Card, D., Harknett, K., & Lui-Gurr, S. 1998, *When Financial Incentives Encourage Work. Complete 18-Month Findings from the Self-Sufficiency Project*, Social Research and Demonstration Corporation, Canada, R-99-2E.
72. Masters, S. 1981, "The Effects of Supported Work on the Earnings and Transfer Payments of Its AFDC Target Group", *Journal of Human Resources*.v16 n4 p600-36 Fall.
73. Masters, S. & Maynard, R. 1981, *The Impact of Supported Work on Long-Term Recipients of AFDC Benefits*, MDRC, Manpower Demonstration Research Corporation, New York, Vol 3 of the Final Report on The Supported Work Evaluation.
74. Maynard, R. A., Nicholson, W., & Rangarajan, A. 1993, *Breaking the cycle of poverty: The effectiveness of mandatory services for welfare-dependent teenage parents*. Mathematica Policy Research, Princeton, NJ.
75. Maynard, R., et al. 1981, *Supported Work Demonstration: Effects During the First 18 Months after Enrollment*.
76. Maynard, R., Brown, R., Shore, J., Jackson, R., Leach, V., & Wooldridge, J. 1979, *The National Supported Work Demonstration: Effects during the first 18 months after enrollment*, Mathematica Policy Research, Inc., Princeton, New Jersey.
77. Michalopoulos, C., Card, D., Gennetian, L. A., Harknett, K., & Robins, P. K. 2000, *The Self-Sufficiency Project at 36 Months*, Social Research and Demonstration Corporation, Ottawa.

78. Michalopoulos, C. & Schwartz, C. 2001, *What Works Best for Whom: Impacts of 20 Welfare-to-Work Programs by Subgroup. Executive Summary. National Evaluation of Welfare-to-Work Strategies.*
79. Miller, C., Knox, V., Auspos, P., Hunter-Manns, J., & Orenstein, A. 1997, *Making Welfare Work and Work Pay: Implementation and 18-Month Impacts of the Minnesota Family Investment Program.*
80. Miller, C., Knox, V., Gennetian, L., Dodoo, M., Hunter, J., & Redcross, C. 2002, *Reforming Welfare and Rewarding Work: Final Report on the Minnesota Family Investment Program. Volume 1: Effects on Adults [and] Volume 2: Effects on Children.*
81. Nelson, H. 1984, *Evaluation of the Community Work Experience Program*, Washington State Department of Social and Health Services, Division of Administration and Personnel, Washington.
82. Olson, J., Schexnayder, D., & O'Shea, D. 1998, *Participation Patterns and Program Impacts of Hawaii's JOBS WORKS! Demonstration Project.*
83. Orr, L. L. & AND, O. T. H. E. 1996, *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study.*
84. Pawasarat, J. & Quinn, L. 1993, *Wisconsin Welfare Employment Experiments: An Evaluation of the WEJT and CWEP Programs*, University of Wisconsin, Wisconsin.
85. Puma, M. J. & Burstein, N. R. 1994, "The national evaluation of the Food Stamp Employment and Training program", *Journal of Policy Analysis and Management*, vol. 13, no. 2, pp. 311-330.
86. Puma, M. J., Burstein, N. R., Merrell, K., & Silverstein, G. 1990, *Evaluation of the Food Stamp Employment And Training Program. Final Report. Volume I.*
87. Quets, G., Robins, P., Pan, E., Michalopoulos, C., & Card, D. 2001, *Does SSP Plus Increase Employment? The Effect of Adding Services to the Self-Sufficiency Project's Financial Incentives.*
88. Quint, J., Ball, J., Goldman, B., Gueron, J., & Hamilton, G. 1984, *Maryland: Interim findings from the Maryland Employment Initiatives Programs*, Manpower Demonstration Research Corporation.
89. Riccio, J., et al. 1986, *Virginia Employment Services Program. Final Report.*
90. Riccio, J., et al. 1994, *GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program. California's Greater Avenues for Independence Program*, MDRC, Manpower Demonstration Research Corporation, New York.
91. Schexnayder, D. T. 2003, *Achieving Change for Texans: Evaluation of the Texas Welfare Reform Waiver: Final Summary Report*, Ray Marshall Center for the Study of Human Resources, The University of Texas at Austin, Texas, Report 1 of 6.

92. Schexnayder, D. T., Schroeder, D., Olson, J., & Kum, H. 2003, *Achieving Change for Texans: Evaluation of the Texas Welfare Reform Waiver: Final Impact Report*, Ray Marshall Center for the Study of Human Resources, The University of Texas at Austin, Texas, Report 4 of 6.
93. Schexnayder, D., Olson, J., Schroeder, D., Betsinger, A., & Sim, S. C. 1998, *Achieving Change for Texans. Evaluation. Net Impacts through December 1997*.
94. Scrivener, S., Hamilton, G., Farrel, M., Freedman, S., Friedlander, D., Mitchell, M., Nudelman, J., & Schwartz, C. 2003, *Implementation, Participation Patterns, Costs, and Two-Year Impacts of the Portland (Oregon) Welfare-to-Work Program*, Manpower Demonstration Research Corporation.
95. Scrivener, S. & Walter, J. 2002, *Evaluating Two Approaches to Case Management: Implementation, Participation Patterns, Costs, and Three-Year Impacts of the Columbus Welfare-to-Work Program. National Evaluation of Welfare-to-Work Strategies*.
96. Storto, L., Hamilton, G., Schwartz, C., & Scrivener, S. 2000, *Oklahoma City's ET & E Program: Two-Year Implementation, Participation, Cost and Impact Findings NEWS*, Manpower Demonstration Research Corporation.
97. United States General Accounting Office 1996, *Job Training Partnership Act. Long-Term Earnings and Employment Outcomes. Report to Congressional Requesters*.
98. Werner, A. & Kornfeld, R. 1997, *The Evaluation of "To Strengthen Michigan Families" - Final Impact Report*, Abt Associates Inc., Cambridge, MA.
99. Werner, A., Mills, G., Walker, M., & Kutty, N. 1994, *The evaluation of the New York State Comprehensive Employment Opportunity Support Centers Program: Impact study*, Abt Associates.
100. Wilkinson, K. P. & Ross, P. J. 1972, "Evaluation of the Mississippi AFDC Experiment", *Social-Service-Review*, vol. 46, no. 3, pp. 363-377.
101. Wilkinson, K. & Ross, P. 1977, *Experiments in Social Welfare: An Empirical Evaluation of the "Mississippi Project"*. Mississippi State University, Social Science Research Center Report 38.
102. Zambrowski, A. & Gordon, A. 1993, *Evaluation of the Minority Female Single Parent Demonstration: Fifth-Year Impacts at CET*, Mathematica Policy Research, Inc., Document No. PR93-58.

EXCLUDED CITATIONS

1. Angelo, L. & March, J. A. 1997, "Do training programs help AFDC recipients leave the welfare rolls? An evaluation of New York City's BEGIN (Begin Employment Gain Independence Now) program", *Journal of Health & Human Services Administration.*, vol. 20, no. 1, pp. 83-104.
2. Bloom, H. & Riccio, J. A. 2002, *Using Place-Based Random Assignment and Comparative Interrupted Time-Series Analysis to Evaluate the Jobs-Plus Employment Program for Public Housing Residents*, Manpower Demonstration Research Corporation.
3. Bourn John 2002, *The New Deal for young people; report by the Comptroller and Auditor General. HC 639 (session 2001-2002: 28 February 2002)*, National Audit Office, Norwich.
4. Dahl, E. 2003, "Does "Workfare" Work? - The Norwegian Experience", *International Journal of Social Welfare*, vol. 12, pp. 274-288.
5. Hales, J., Lessof, C., & Roth, W. 2000, *Evaluation of the New Deal for Lone Parents: Early lessons from the Phase One prototype - synthesis report*, Department of Social Security, London, 108.
6. Hales, J., Roth, W., Barnes, M., Millar, J., Lessof, C., Gloyer, M., & Shaw, A. 2000, *Evaluation of the New Deal for Lone Parents: Early Lessons from the Phase One Prototype - Findings of Surveys*, Department of Social Security, Leeds, Research report No 109.
7. Hales, J., Taylor, R., Mandy, W., & Miller, M. 2003, *Evaluation of Employment Zones: Report on a Cohort Survey of Long-Term Unemployed People in the Zones and a Matched Set of Comparison Areas*, Department for Work and Pensions, Research Management, Research and Analysis Publications, Sheffield, WAE Report 176.
8. Janzen, F., Taylor, M. J., & Weathers, S. 1997, *An evaluation of Utah's Single Parent Employment Demonstration Program*, Social Research Institute.
9. King, C., Schexnayder, D. T., O'Shea, D., Olson, J., Tse-i Pan, D., Roche, R. A., & Trunkey, D. R. 1994, *Texas JOBS Program Evaluation. Final Report*.
10. Leigh, D. E. 1995, "Can a voluntary workfare program change the behavior of welfare recipients? New evidence from Washington state's Family Independence Program (FIP)", *Journal of Policy Analysis & Management*, vol. 14, no. 4, pp. 567-589.
11. Lessof, C., Miller, M., Phillips, M., Pickering, K., Purdon, S., & Hales, J. 2003, *New Deal for Lone Parents Evaluation: Findings From the Quantitative Survey*, Department of Work and Pensions WAE, Sheffield, 147.

12. Long, S. K. & Wissoker, D. A. 1995, "Welfare Reform at Three Years: The Case of Washington State's Family Independence Program", *Journal of Human Resources*, vol. 30, no. 4, pp. 766-90.
13. Nicaise, I. 2000, "Giving fish or teaching to fish? A Cost-Benefit Analysis of Belgian Employment-training projects for minimum income recipients", *Public Finance and Management*, vol. Forthcoming.
14. Nightingale, D. S., Wissoker, D., Burbridge, L., Bawden, D., & Jeffries, N. 1990, *Evaluation of the Massachusetts Employment and Training Program*, The Urban Institute, Washington, DC, Report 94-1.
15. Pedersen, J.-T. 1998, *Er arbeidslinjen egnet til å hjelpe klientene ut av stønadsrollen? : langtidsoppfølging av effekter av arbeidsmarkedstiltak for unge sosialklienter og rusmisbrukere*, Seksjon for trygdemedisin, Institutt for allmenntilleggsmedisin og samfunnsmedisinske fag, Universitetet i Oslo, Oslo, 98:4.
16. Schexnayder, D. & Olson, J. 1995, *Texas JOBS Program Evaluation. Second Year Impacts*.
17. Schexnayder, D. & Olson, J. 1998, *Hawaii Food Stamp Employment and Training/JOBS Conformance Demonstration: Impact Evaluation Final Report*.
18. Schexnayder, D. T. & Olson, J. 1997, *Texas Food Stamp Employment and Training/JOBS Conformance Demonstration: Impact Evaluation Final Report*, Center for the study of Human Resources, LBJ School of Public Affairs, The University of Texas, Austin.
19. Schiller, B., et al. 1976, *The Impact of WIN II: A Longitudinal Evaluation of the Work Incentive Program (WIN)*.
20. Schiller, B. 1978, "Lessons from WIN: A Manpower Evaluation", *Journal of Human Resources*, vol. 13, no. 4, pp. 502-523.
21. Stevens, H. 1997, "The Effectiveness of Community Employment Programs for Social Assistance Recipients: An Evaluation of the City of Winnipeg's Community Services Programs", *The Canadian Journal of Program Evaluation/La Revue Canadienne D'Evaluation de Programme*, vol. 12, no. 1, summer, pp. 71-85.
22. Werner, A., Rodda, D., Pan, E., & Plimpton, L. 1997, *Evaluation of the Alabama Avenues to Self-Sufficiency Through Employment and Training Services (ASSETS) Demonstration: Final Report*: Abt. Associates, Cambridge, Mass.
23. White, M. & Riley, R. 2002, *Findings from the Macro evaluation of the New Deal for Young People*, Department for Work and Pension, Research Report No 168.
24. Wouters, M., Van Meensel, R., & Nicaise, I. 1994, *De TOK-projecten en hun cursisten, drie jaar later. Follow-up onderzoek van de projecten van 1989*, Hoger Instituut voor de Arbeid, Katholieke Universiteit Leuven, Leuven.

ADDITIONAL REFERENCES

1. Altman, D., Deeks, J., & Sackett, D. 1998, "Odds ratios should be avoided when events are common", *British Medical Journal*, vol. 317, p. 131.
2. Anderson J, Martinson K. *Service Delivery and Institutional Linkages Early Implementation Experiences of Employment Retention and Advancement Programs* [MDRC]. [updated 2003]. Available from: <http://www.mdrc.org/publications/356/full.pdf>.
3. Blonk R, Brenninkmeijer V. *Experimenteren in activering. (Experimenting in activation)*. In: Vos Ed, Besseling J, Wevers C, editors. *Verraderlijk effectief (Insideous effective)*. Hoofddorp: TNO Arbeid; 2004.
4. Cebulla, A., Bouchet, S., & Greenberg, D. *User's guide to a database of welfare-to-work programs*. 2005. Loughborough University, UK.
5. *Comprehensive Meta-Analysis. Manual [computer program]. Version 2.0*. Englewood, NJ: BioStat; 2004.
6. Cornwell E, Enjolras B, Fraisse L, Lødemel I, Olsen L, Rosdahl A, et al. *Workfare in six European nations. Findings from evaluations and recommendations for future development*. 2002.
7. Dahl E. Does 'workfare' work? - the Norwegian experience. *International Journal of Social Welfare* 2003; 12:274-88.
8. Deeks J, Dinnes J, D'Amico R, Sowden A, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7(27):1-188.
9. Fischer RL. Job Training as a Means to "Ending Welfare as We Know It": A Meta-Analysis of U.S. Welfare Employment Program Effects. *DAI* 1995;56(July):371-A-372-A.
10. GRADE Working Group 2004, "Grading quality of evidence and strength of recommendations", *British Medical Journal*, vol. 328, pp. 1490-1494.
11. Greenberg D, Shroder M. *The digest of social experiments*. Washington D.C.: Urban Institute Press; 2004.
12. Grogger J, et al. Conflicting benefits: trade-offs in welfare reform. *Rand. Review* 2002; 26(3): 16-23.
13. Gueron JM, Pauly E, Lougy CM. *From welfare to work: a Manpower demonstration research corporate study*. New York: Russel Sage Foundation; 1991.

14. Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions 4.2.5* [updated May 2005]. In: *The Cochrane Library, Issue 3, 2005*. Chichester, UK: John Wiley & Sons, Ltd.

15. Levitas R. *The inclusive society? Social exclusion and New Labour*. Hampshire: Macmillan Press; 1998.
16. Lødemel I, Johannessen A, Buschmann E, Nilsen KL, Velvin G. *Nytter det? Erfaringer fra Tiltaksforsøket 2000-2004. Evaluering av Forsøk med kommunalt ansvar for aktive, arbeidsretta tiltak for langtidsmottakere av sosialhjelp (Does it work? Lessons from the demonstration research project 2000-2004. Evaluation of municipal responsibility for active, work-oriented initiatives for long-term welfare recipients: Gruppe for inkluderende velferd Høgskolen i Oslo; 2004. Delrapport 3.*
17. Martin JP. *What works among active labour market policies: evidence from OECD countries' experiences.*: Directorate for Education, Employment, Labour and Social Affairs, Employment, Labour and Social Affairs Committee, OECD; 1998. 35.
18. Michalopoulos C, Schwartz C, Adams-Ciardullo D. *National evaluation of welfare-to-work strategies. What works best for whom: Impacts of 20 welfare-to-work programs by subgroup.*: U.S.Department of Health and Human Services, Administration for Children and Families, Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Education, Office of the Under Secretary, Office of Vocational and Adult Education; 2000.
19. OECD. *Benefit recipiency ratio for the working age population (15-64), 1980-1998* [OECD]. [updated 2005]. Available from: <http://www.oecd.org/dataoecd/61/28/1875507.xls>.
20. Office for National Statistics. *Composition of household income* [Office for National Statistics]. [updated 2005]. Available from: <http://www.statistics.gov.uk/STATBASE/Expodata/Spreadsheets/D7419.xls>.
21. U S Census Bureau. *General population and housing characteristics: 1990*. [updated 2005]. Available from: http://factfinder.census.gov/servlet/OTTable?_bm=n&_lang=en&_qr_name=DEC_1990_STF1_DP1&_ds_name=DEC_1990_STF1_&_geo_id=01000US.
22. U.S.Census Bureau. *SSI, TANF1, and Food Stamp Recipients: 1980 to 1998* [U S Census Bureau]. [updated 2005]. Available from: www.census.gov/prod/99pubs/99statab/sec12.pdf.
23. Walker,R., Hoggart,L., Hamilton,G., Blank,S. (2006). *Making random assignment happen: Evidence from the UK Employment Retention and Advancement (ERA) demonstration*, Department for Work and Pensions, Research Report No 330.

Tables and Figures

Table 1. Overview of the 46 included programmes

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ⁹
ABC (A Better Chance Welfare Reform Program) (Fein & Karweit 1997; Fein & Long 1999)	Delaware	1995	2138 / 1821	Job search and placements, expanded care and other work supports, time limits.	M	Continued under traditional AFDC rules.	M	M	U	M	M	U	EP-J

⁹ M= MET, U=UNCLEAR, N=NOT MET, EP= employment-focused, ED=education-focused, OT = other than employment-focused or education-focused, J=job search first, V=varied first activity.

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
ACT (Achieving Change for Texans) (Schexnayder 2003; Schexnayder et al. 2003)	Texas	1996	22,559 / 22,293	Time limit experiment: Time limits on TANF benefits. RER (Responsibilities, Employment, and Resources) Non-Choices: Responsibility agreement to comply with e.g. health screenings for children, school attendance, parenting skills training classes RER Choices: Combination of time limits and responsibility agreements.	M	Continued under traditional AFDC rules.	U	U	M	M	U	U	EP-OT
AFDC (Aid to Families with Dependent Children) Experiment (Wilkinson & Ross 1972; Wilkinson & Ross 1977)	Mississippi	1967	31 / 56	Training and income group: a special care worker secured jobs in the community for the clients. Increased financial assistance in the form of payment for attending vocational orientation classes and for participating in on-	V	Clients continued to receive usual AFDC grants and regular case work services.	U	M	M	M	M	U	EP-V

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
AFDC Homemaker-Home Health Aide Demonstrations (Bell & Orr 1994) (Bell, Burstein, & Orr 1987)	Arkansas, Kentucky, New Jersey, New York, Ohio, South Carolina, Texas	1967	3912 / 4625	Job skills training and subsidised employment.	V	AFDC recipients excluded from CWEP (Community Work Experience Program).	U	U	M	M	U	U	EP-V
CPREP (Colorado Personal Responsibility and Employment Program)(Cebulla, Bouchet, & Greenberg 2005)	Colorado	1994	10065 / 10018	Increased the amount of earnings without grant reductions. New case managers and stronger sanctions for non-compliance relative to AFDC. \$500 cash incentive for completion of high school.	M	Existing AFDC services.	U	U	U	U	U	U	U-U
CWEP (Community Work Experience Program) (Nelson 1984)	Pierce and Spokane, Washington	1982	64 / 42	Placement in unpaid jobs provided by contracting non-profit agencies or organisations which provide supervision and training in return for the work done by participants	M	AFDC reciprocity.	M	M	M	M	U	U	EP-V

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
CWEP (Community Work Experience Program) for AFDCs) (Friedlander et al. 1986)	West Virginia	1982	1845 / 1834	Work requirements equal to the welfare benefit divided by the \$ 3.35 minimum wage rate.	M	AFDC recipients excluded from CWEP.	U	U	M	M	M	M	EP-J
EMPOWER (Employing and Moving People Off Welfare and Encouraging Responsibility) (Arizona) (Also called Arizona Works) (Kornfeld et al. 1999)	Arizona	1995	2876 / 2998 (Phoenix: 2411 / 2467 Navajo: 465 / 531)	Time-limited assistance, family benefit cap, restricted eligibility for unved minor parents, mandatory job participation for teen parents, strict job sanctions, extended transitional medical assistance, transitional child care, elimination of the 100-hour rule.	M	Not subject to the EMPOWER provisions.	U	U	M	M	N	U	EP-OT
ESP (Employment Services Program)/WIN (Work Incentive Program) (Riccio et al. 1986)	Virginia	1983	1269 / 1044	Job search followed by 13 weeks of community work experience, education, or training.	M	They could find and participate in services available elsewhere in the community. The embargo ended after 2 years.	U	U	U	M	M	M	EP-J

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
EWEP (Experimental Work Experience Program) (Friedlander & Long 1987;Goldman, Friedlander, & Long 1986) (Also called San Diego Job Search and Work Experience Demonstration)	San Diego	1982	1502 / 873	EPP (Employment Preparation Program), job search vs. (EWEP). Job search followed by requirement to work in public or non-profit agencies.	M	Regular WIN services.	M	M	M	M	M	M	EP-J
FIP (Family Investment Program) (Fraker et al. 2002;Fraker et al. 1998)	Iowa	1993	4439 / 2200	Earnings disregards, participation requirement in PROMISE Jobs (Promoting Independence and Self-Sufficiency through Employment) + family investment agreement.	M	Received pre-reform AFDC services.	U	U	N	M	M	U	EP-V
FTP (Family Transition Program) (Bloom et al. 1997;Bloom et al. 2000)	Florida	1994	1,405 / 1,410	24-month time-limit on cash assistance, financial work incentives, enhanced services and requirements (child care assistance, case management).	M	Pre-existing AFDC policy.	U	M	M	M	U	M	OT-V
Food Stamp E&T (Education & Training) Program (Puma et al. 1990;Puma & Burstein 1994)	53 US sites in 23 states	1988	8467 / 6710	Job search, job search training, workfare and work experience, vocational skills training.	M	Food stamp recipients eligible for participation in the education & training program.	M	M	U	M	U	M	H-U

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
Food Stamp Work Registration and Job Search Demonstration (Lerman et al. 1986; Cebulla, Bouchet, & Greenberg 2005)	21 sites in Arizona, California, Colorado, DC, Florida, Kentucky, Maine, Michigan, New Mexico, New York, Ohio, Texas, Virginia, and Wyoming	1981	9669 / 6757 (Applicant search model: 4,396 / 4,116) (Job club model: 2,333 / 1,633) (Group job search: 870 / 586) (Job club with workfare: 2,070 / 422)	Eight treatments: (1) In-person registration model, (2) job club model, (3) in-person registration/job club, (4) Food stamp agency/job club, (5) applicant search model, (6) job club model, (7) group job search assistance model, (8) job club/workfare model.	M	Not subject to any work requirements.	U	M	U	M	N	U	EP-J
GAIN (Greater Avenues to Independence) (Riccio et al. 1994; Friedman, Riccio, & Freedman 1993)	Alameda, California	1988	698 / 689	Assessed & grouped according to education level. First step: Education, job search, child care. Second step: vocational or work experience.	M	Could receive alternative services in the community on their own initiative. Some participated in unpaid work.	U	U	N	M	M	M	ED-V

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
GAIN (Greater Avenues to Independence) (Riccio et al. 1994; Friedlander, Riccio, & Freedman 1993)	Butte, California	1988	1780 / 473	Assessed & grouped according to education level. First step: Education, job search, child care. Second step: vocational or work experience.	M	Could receive alternative services in the community on their own initiative. Some participated in unpaid work	U	U	N	M	M	M	ED-V
GAIN (Greater Avenues to Independence) (Riccio et al. 1994; Friedlander, Riccio, & Freedman 1993)	Los Angeles, California	1989	3750 / 2143	Assessed & grouped according to education level. First step: Education, job search, child care. Second step: vocational or work experience.	M	Could receive alternative services in the community on their own initiative. Some participated in unpaid work	U	U	N	M	M	M	ED-V
GAIN (Greater Avenues to Independence) (Riccio et al. 1994; Friedlander, Riccio, & Freedman 1993)	Riverside, California	1988	6234 / 1799	Assessed & grouped according to education level. First step: Education, job search, child care. Second step: vocational or work experience.	M	Could receive alternative services in the community on their own initiative. Some participated in unpaid work	U	U	N	M	M	M	EP-V

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
GAIN (Greater Avenues to Independence) (Riccio et al. 1994;Friedlander, Riccio, & Freedman 1993)	San Diego, California	1988	9488 / 2013	Assessed & grouped according to education level. First step: Education, job search, child care. Second step: vocational or work experience.	M	Could receive alternative services in the community on their own initiative. Some participated in unpaid work.	U	U	N	M	M	M	ED-V
GAIN (Greater Avenues to Independence) (Riccio et al. 1994;Friedlander, Riccio, & Freedman 1993)	Tulare, California	1988	2918 / 1239	Assessed & grouped according to education level. First step: Education, job search, child care. Second step: vocational or work experience.	M	Could receive alternative services in the community on their own initiative. Some participated in unpaid work.	U	U	N	M	M	M	ED-V
Grant Diversion Project (Freedman et al. 1988)	New Jersey	1984	508 / 486	On-the-job training, subsidized job search, education, training & 20 weeks part-time unpaid work at governmental agency or non-profit organization.	V	Eligible for JTPA (job training partnership act) services as well as other services pursued on their own initiative such as work experience, job search, etc.	U	U	U	U	U	U	EP-V

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
<p>JOBS (Job Opportunities and Basic Skills Training) (Freedman 1995;Hamilton et al. 1997;Freedman et al. 2001;Freedman 2001;Hamilton et al. 2001;Scrivener & Walter 2002;Freedman 2003;Farrel 2000;Fein 1994;Storto et al. 2000;Fein, Beecroft, & Blomquist 1994;Scrivener et al. 2003) (Also called the National Evaluation of Welfare-to-Work Strategies or NEWWS)</p>	Atlanta, Georgia	1991	3376 / 3443 (HCD: 1935 / 1946) (LFA: 1441 / 1497)	Orientation & assessment. Ss without barriers to employment recruited. Those without high school diploma (HSD) or General Equivalency Diploma (GED) received CWEP/basic education. The others: job readiness/training. After education, Group 1 assigned same intervention as Group 2 (post secondary education/job skills training/job club).	M	Eligible for welfare as usual. In addition, they were eligible for child care assistance similar to that offered to program group members, provided that they were participating in non-programme activities in which they had enrolled on their own.	U	U	U	M	M	M	ED-U
<p>JOBS (NEWWS)</p>	Columbus, Ohio	1992	5083 / 2159 (Integrated: 2513) (Traditional: 2570)	In the traditional case management model, welfare eligibility and employment program functions were performed by separate staff members. In the integrated case management model, these two functions were performed by the same staff member.	M	Eligible for welfare as usual. In addition, they were eligible for child care assistance similar to that offered to program group members, provided that they were participating in non-programme activities in which they had enrolled on their own.	U	U	U	M	M	M	ED-U

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
JOBS (NEWWS)	Detroit, Michigan	1992	2226 / 2233	Job search, basic education, vocational training, college, work experience or on-the-job training.	M	Eligible for welfare as usual. In addition, they were eligible for child care assistance similar to that offered to program group members, provided that they were participating in non-programme activities in which they had enrolled on their own.	U	U	U	M	M	M	ED-V
JOBS (NEWWS)	Grand Rapids, Michigan	1991	3099 / 1455 (HCD: 1542) (LFA: 1557)	Orientation & assessment. Ss without barriers to employment recruited. Those without high school diploma (HSD) or General Equivalency Diploma (GED) received CWEP/basic education. The others: job readiness/training. After education, Group 1 assigned same intervention as Group 2 (post secondary education/job skills training/job club).	M	Eligible for welfare as usual. In addition, they were eligible for child care assistance similar to that offered to program group members, provided that they were participating in non-programme activities in which they had enrolled on their own.	U	U	U	M	M	M	ED-U

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
JOBS (NEWWS)	Work Choice Ohio (Also called Transitions to Independence)	1992	11021 /2307	Job search, basic education, vocational training, college, work experience or on-the-job training.	M	Continued under the old rules, could enrol in other (non-jobs) education and training services.	U	U	U	M	M	M	ED-U
JOBS (NEWWS)	Oklahoma City, Oklahoma	1991	4309 / 4368	Job search, basic education, vocational training, college, work experience or on-the-job training.	M	Eligible for welfare as usual. In addition, they were eligible for child care assistance similar to that offered to program group members, provided that they were participating in non-programme activities in which they had enrolled on their own.	U	U	U	M	M	M	ED-U
JOBS (NEWWS)	Portland, Oregon	1993	3529 / 2018	Job search, basic education, vocational training, college, work experience or on-the-job training.	M	Eligible for welfare as usual. In addition, they were eligible for child care assistance similar to that offered to program group members, provided that they were participating in non-programme activities in which they had enrolled on their own.	U	U	U	M	M	M	EP-V

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
JOBS (NEWWS)	Riverside, California	1991	4980 / 3342 (LFA: 3384) (HCD: 1596)	Job search, basic education, vocational training, college, work experience or on-the-job training.	M	Eligible for welfare as usual. In addition, they were eligible for child care assistance similar to that offered to program group members, provided that they were participating in non-programme activities in which they had enrolled on their own.	U	U	U	M	M	M	ED-U
Jobs First (Bloom et al. 2000; Horwitz & Kerker 2001; Bloom et al. 2002; Hendra, Michalopoulos, & Bloom 2002)	Connecticut	1996	2381 / 2392	Includes unusually generous financial work incentives and requires recipients to participate in employment-related services targeted toward quick job placement	M	No employment requirement. AFDC. Some employment-related services to recipients but less mandatory than Jobs-First.	U	U	M	M	M	M	EP-J
Jobs First Gain (Freedman, Mitchell, & Navarro 1997; Freedman et al. 2001)	Los Angeles	1996	1152 / 4162	Strong work-first message, warning about time-limited welfare, orientation session, job search assistance (job clubs), basic education, vocational training, unpaid work experience.	M	Precluded from receiving Jobs-First Gain from September 11th 1996 until October 1998. They remained eligible to receive welfare and food stamps.	U	U	N	M	M	U	EP-J

PROGRAMME / REFERENCES	JOBS Works! (Olson, Schexnayder, & O'Shea 1998)
SITE(S)	Hawaii
START OF STUDY	1989
INTERVENTION / CONTROL SAMPLE SIZE	1662 / 1632
INTERVENTION	Job search 30 hours/week until employed, or case closed, or was transferred to appropriate jobs unit. 2. job readiness seminar. 3. subsidized child care. 4 transportation reimbursement.
VOLUNTARY (V) OR MANDATORY (M)	M
SERVICES PROVIDED TO CONTROL GROUP	Received same treatment as the island's regular jobs population.
RANDOM GENERATION OF ALLOCATION	M
CONCEALMENT OF ALLOCATION	M
PREVENTION OF PERFORMANCE BIAS	M
PREVENTION OF DETECTION BIAS	M
PREVENTION OF ATTRITION BIAS	U
INTENT-TO-TREAT ANALYSIS	U
TYPE OF INTERVENTION^o	EP-J

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
JTPA (Job Training Partnership ACT)(Bloom et al. 1997;Bloom, Orr, & Gave 1992;Orr & AND 1996;United States General Accounting Office 1996)	16 sites: Butte, Montana; Cedar Rapids, Iowa; Coosa Valley, Georgia; Corpus Christi, Texas; Decatur, Illinois; Fort Wayne, Indiana; Heartland, Florida; Jackson, Mississippi; Jersey City, New Jersey; Larimer County, Colorado; Marion, Ohio; Oakland, California; northwest Minnesota; Omaha, Nebraska; Providence, Rhode Island;	1987	Total sample (treatment + control): 20,602 Classroom training group: 7090 On-the-job training group: 7412 Other activities group: 6100 On average 68% assigned to treatment conditions.	Classroom training, on-the-job training, job search assistance.	V	Not allowed to receive services for 18 months.	M	M	U	M	M	M	ED-V

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ⁹
Massachusetts Work Experience Program (MASSWEP)(Friedman et al. 1980)	Massachusetts	1978	290 / 725	3 days unpaid work experience/week for 13 weeks. Also 30 1/3 rule ¹⁰ and 100 hour rule.	M	WIN unassigned recipients, received no intervention.	U	M	U	M	N	M	EP-O
Minnesota Family Investment Program (MFIP) (Gennetian et al. 2000; Miller et al. 1997; Miller et al. 2002)	Minnesota (Dakota, Anoka, Hennepin,	1994	7208 / 7431 (MFIP: 5,275) (MFIP incentives-only: 1,933) (Control 1: AFDC: 5,634) (Control 2: AFDC, no services: 1,797)	Working families kept more of their benefits. Child care expenses were paid directly to provider. Participation requirement in education and training activities. Simplified public assistance.	M	Were subject to the benefits and services offered by Minnesota's AFDC system.	U	U	M	M	U	U	EP-O

¹⁰ "30+1/3." Is a federally allowable earned income disregard provision which disregard \$30 plus 1/3 of the remaining income and bases the grant on that. The 100-hour rule means that families with an unemployed parent (AFDC-U) were ineligible for AFDC if the father worked 100 hours or more in a month.

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
Minority Female Single Parent Demonstration (Burghardt et al. 1992;Zambrowski & Gordon 1993)	Atlanta, Georgia; San Jose, California; Providence, Rhode Island; Washington, D.C.	1984	1,841 / 1,334	Job-training immediately. Training in reading and mathematics skills. Hands-on job skills training. Supplemental training courses for those seeking a GED certificate and ESL training.	M	Not eligible to receive services at the centres but could seek services elsewhere.	U	U	U	U	N	U	EP - V
New Hope (Wisconsin New Hope Project)(Bos et al. 1999;Brock et al. 1997)	Milwaukee, Wisconsin	1994	678 / 679	30 hour weekly work requirement to receive earning supplement, and a health insurance plan.	M	Not eligible to New Hope benefits.	M	M	M	M	U	U	EP - J
Options Program (Maryland Employment Initiatives) (Friedlander & Burtless 1994;Friedlander et al. 1985;Friedlander 1987;Quint et al. 1984)	Baltimore	1982	1362 / 1395	Choice among activities. Job search, 3 month unpaid work experience, education & training and child care services.	M	WIN. Enhanced unsubsidized employment, direct placement into labour force.	U	U	U	U	U	M	EP - V

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
PEA (Private Employment Agencies) (Michigan)(Carcagno et al. 1982) Also known as "AFDC Job Counsellors"	Michigan	1976	2,593 / 1,691	Private employment agencies helped the participant find employment.	V	Minimal services: counselling with job referral.	U	U	U	M	N	U	EP - J
National Supported Work Demonstration (1984; Couch 1992; Kemper, Long, & Thornton 1981; Masters 1981; Masters & Maynard 1981; Maynard et al. 1981; Maynard et al. 1979)	Several sites	1975	675 / 676	Subsidised work with supportive environment through peer group support, close supervision and graduated stress ¹¹ . 25% of the time could be used for job-search assistance and other ancillary	V	Did not receive Supported Work.	U	U	U	U	U	M	EP - J

- A. ¹¹ "Graduated stress" allows participants to adjust to major changes through transitional processes. This technique promotes success through gradual adaptation and the strategic use of supports. It can be used, for example, to transition participants from lower levels of work activity to full time equivalent hours (a new state requirement). It can also be used to move people from program participation to unsubsidized employment by providing strong post-placement supports for both the employer and the new employee.

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
New York State Child Assistance Program (CAP) (Hamilton et al. 1993; Hamilton et al. 1996; Cebulla, Bouchet, & Greenberg 2005)	Monroe, Niagara, and Suffolk counties in New York	1988	4,287 (50% in intervention group.	Economic incentives (earned disregard), child support, case management.	V	AFDC.	U	U	M	M	M	M	EP - O
New York State Comprehensive Employment Opportunities Support Centers Program (CEOSC)(Cebulla, Bouchet, & Greenberg 2005; Werner et al. 1994)	Albany, New York	1987	261 / 268	Assessment, education, skills training, job search assistance, child care, counselling, case management, and other supportive services.	V	No services from the program.	U	U	U	U	U	U	OT-V
Project Independence (Florida's JOBS Program)(Kemple, Friedlander, & Fellerath 1995; Kemple & Haimson 1994)	Florida	1990	13,513 / 4,724	Job search, job club, assessment, basic education and training.	M	Were given a list of alternative employment and training services in the community.	U	M	M	M	U	U	OT-V
Pennsylvania Saturation Work Program (Cebulla, Bouchet, & Greenberg 2005; Hogarth & Martin 1988; Hogarth, Martin, & Nazar 1989)	Philadelphia	1985	4,024 / 3,318	Individual case management, workshops, training and education, community work experience, job clubs, and assistance with work plan development.	M	Continued to receive existing County Assistance Office (CAO) services.	U	U	U	U	U	U	EP-U

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
Self-Sufficiency Project (SSP) (Lin et al. 1998; Michalopoulos et al. 2000)	Canada (New Brunswick and British Columbia)	1992	2,880 / 2,849	Temporary monthly cash supplement for at least one year if they left income assistance for full-time work (at least 30 hours a week).	V	Not offered the supplement.	U	U	M	M	U	U	EP - OT
Self-Sufficiency Project Plus (SSP+) (Quets et al. 2001)	Canada (New Brunswick)	1994	293 / 303	Employment plan, resumé services, job clubs, job coaching.	V	Not offered any component of the program.	U	U	M	M	U	U	EP - V
SWIM (Saturation Work Initiative Model) (Friedlander & Hamilton 1993; Friedlander & Hamilton 1996; Hamilton & Friedlander 1989)	San Diego	1995	1604 / 1607	Caseload saturation ¹² Fixed sequence: (information and assessment, followed by 2 weeks job search, followed by 3 months unpaid work, education & training).	M	Embargo for 2 years. AFDC recipients neither assigned to special services nor subject to sanctions.	M	M	N	M	M	M	EP-J

¹² "Caseload saturation," i.e., having as many persons participate in activities for as long as possible.

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
Teenage Parent Demonstration (Kisker, Rangarajan, & Boller 1998; Maynard, Nicholson, & Rangarajan 1993)	Camden and Newark, New Jersey; and Chicago, Illinois	1988	2647 / 2650	Job training or employment, case management, child care, parenting support, financial sanctions for failing to comply.	M	Received regular services that non-demonstration teens would have received.	U	U	U	M	M	U	OT - V
TOPS (Training Opportunities in the Private Sector program) (Auspos et al. 1988)	Maine	1983	297 / 147	Single program within the WIN demonstration system. Prescribed sequence of prevocational training, unpaid work experience and on-the-job training.	V	Could receive some services, including prevocational training, unpaid work experience or on-the-job training, but not the entire TOPS sequence of activities.	U	U	M	M	M	U	EP-V
To Strengthen Michigan Families (Werner & Kornfeld 1997)	Michigan	1992	9,883 / 9,621	Social contracts, financial incentives, broadened AFDC eligibility.	M	No social contract.	N	U	M	M	N	N	EP-J

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
VIEW (Virginia Initiative for Employment Not Welfare) (Gordon & James-Burdumy 2002; Gordon & Agodini 1999)	Virginia	1983	1237 / 1207	Job search 90 days, mandatory work (regular work or CWE/P). Sanctions. 24-months time limit on TANF benefits. Earned income disregards. Supportive services.	M	Remained subject to the old AFDC policies and received employment-related services only under the old JOBS program.	M	M	U	M	M	M	EP-J
Welfare Reform (also called Indiana Manpower Placement and Comprehensive Training (IMPACT) Program (Beecroft, Lee, & Long 2003; Fein et al. 1997; Fein et al. 1998)	Indiana	1995	14836 / 2841	Job-ready subjects participated in job search and at least 20 hours work activity weekly. The not job-ready participated in education & training with minimal job search.	M	Faced the policies and services of the pre-reform AFDC program in Indiana.	U	U	U	U	M	U	O - O

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
Welfare Restructuring Project (Bloom, Hendra, & Michalopoulos 2000; Bloom et al. 1998; Bloom et al. 2002; Hendra & Michalopoulos 1999; Sperber & Bloom 2002)	Vermont	1994	3271 / 1110	Wage-paying job after cash assistance for 30 months. Minimum-wage community service jobs to those who could not find regular, unsubsidized jobs. If non-compliance with work requirement, the state took control of grant, used the money to pay bills, and required to attend frequent meetings at the welfare office. Programme also included modest financial work incentives to encourage and reward work.	M	1 year Medicaid. First \$120 + 33% of remaining earnings disregarded for first 4 months of employment; first \$120 disregarded in Months 5 to 12; first \$90 disregarded thereafter.	U	U	U	M	U	U	EP-J
Win Demonstration Project (WDP) (Friedlander et al. 1987; Friedlander 1988)	Cook County, Illinois	1985	8107 / 3805	Group 1: All WDP services: Orientation and independent job search, and after that IWEF (Illinois Work Experience Project), pre-employment or modified job search. Group 2: same sequence but excluded from IWEF	M	Required to attend orientation and could be sanctioned if they failed to do so. After that they were eligible for only minimal wdp services.	M	M	M	M	M	M	EP-J

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
Wisconsin Self-Sufficiency First (SSF) / Pay for Performance (PFP) Program (Cancian, Kaplan, & Rothe 2000; Cebulla, Bouchet, & Greenberg 2005)	Dane, Dodge, Jefferson, and Waukesha counties in Wisconsin	1996	SSF experimentals: 1,560 SSF controls: 1,413 PFP experimentals: 1,139 PFP controls: 1,132	SSF: Job search assistance for a minimum of 60 hours, including 30 hours of direct employer contact PFP: participation in JOBS program between 20 and 40 hours weekly.	M	Exempted from the SSF and PFP requirements.	N	U	N	M	M	U	EP-J
Wisconsin Welfare Employment Experiment (Pawarat & Quinn 1993; Cebulla, Bouchet, & Greenberg 2005)	Rock County, Wisconsin	1987	538 / 487	The Work Experience and Job Training Program (WEJT) featured remedial education, job search, subsidised employment, job training, and day care. If not employed: requirement to mandatory Community Work Experience Program.	M	Existing services (mainly job search activities).	U	U	U	U	U	U	OT - V

PROGRAMME / REFERENCES	SITE(S)	START OF STUDY	INTERVENTION / CONTROL SAMPLE SIZE	INTERVENTION	VOLUNTARY (V) OR MANDATORY (M)	SERVICES PROVIDED TO CONTROL GROUP	RANDOM GENERATION OF ALLOCATION	CONCEALMENT OF ALLOCATION	PREVENTION OF PERFORMANCE BIAS	PREVENTION OF DETECTION BIAS	PREVENTION OF ATTRITION BIAS	INTENT-TO-TREAT ANALYSIS	TYPE OF INTERVENTION ^o
California Work Pays Demonstration Program (CWPPD) (Beccera et al. 1996; Cebulla, Bouchet, & Greenberg 2005)	Alameda, Los Angeles, and San Bernardino counties, California	1993	One-parent (AFDC-FG) - experimentals: 6,569 -controls: 3,317 Two-parent (AFDC-U) -experimentals: 3,056 -controls: 1,518	Experimentals received a waiver of the \$ 30 and one-third income disregard, and a waiver of the 100-hour rule.	V	General AFDC rules, with expiring disregards, ineligibility after 100 hours, and higher benefits.	U	U	M	M	M	U	EP - O
Work Program (Friedlander & et.al 1985; Friedlander & Goldman 1988)	Arkansas	1983	560 / 567	Mandatory with sanctions. Fixed sequence with job search first. 3 months unpaid work. Child care and transportation services.	M	AFDC. Did not receive any special services.	M	M	M	M	M	M	EP-J

Table 2. Excluded Non-randomised programmes/references with descriptions (16 programmes)

Name of programme/references	Design	Participants	Intervention
Arbeidslinjen (Norway) (Dahl 2003; Pedersen 1998)	PCS	Young substance abusers and young recipients of social assistance.	Employment training for 6-20 months, supported work or subsidised work. Compulsory participation in welfare.
ASSETS (Avenues to Self-Sufficiency through Employment and Training) (Werner et al. 1997)	PCS	AFDC and food stamp recipients.	Broadening requirements to participate in education & training activities and cooperate with child support enforcement and simplifying and rationalizing welfare administration.
BEGIN (Begin Employment Gain Independence Now) (Angelo & March 1997)	PCS	Parents on welfare who needed to improve their reading, writing or math skills or who wanted to learn to speak English.	Language classes, worksite assignments, employment assistance.

Name of programme/references	Design	Participants	Intervention
Belgian Employment Training Project (Nicaise 2000; Wouters, Van Meensel, & Nicaise 1994)	PCS	Welfare benefit recipients.	Temporary employment, specific work experience projects.
BOND (Better Opportunities for New Directions) Demonstration (Texas) (Schexnayder & Olson 1997)	PCS	AFDC and food stamp recipients.	Directed job search, job readiness, vocational training, non-vocational education, work experience, sanction policy, surveys to volunteers, case management, increased access to supportive services.
Community Home Service Program/Community Service Worker Program (Canada) (Stevens 1997)	PCS	Social assistance recipients in the municipal social assistance program.	Short-term employment in the city of Winnipeg. Salary reduces social assistance benefits dollar for dollar.
Employment and Training (ET) Program (Massachusetts) (Nightingale et al. 1990)	PCS	AFDC recipients.	Direct job development, occupational training, post-secondary education, basic/remedial education, intensive assessment, supported work experience, miscellaneous work experience.

Name of programme/references	Design	Participants	Intervention
EZ (Employment Zones), (UK) (Hales et al. 2003)	PCS	Long-termed unemployed.	Personal job accounts, job search and early entry into mainstream employment or self-employment.
FIP (Family Investment Program) Washington (Leigh 1995; Long & Wissoker 1995)	NRS	Female welfare clients.	Financial incentives and increased support services to encourage education and training investment for welfare clients. For welfare staff, FIP was designed to lead to a more client-oriented environment.
FSE (Food Stamp Employment) & PRIDE (Positive Response in Developing Employment) (Hawaii) (Schexnayder & Olson 1998)	NRS	Food stamp recipients.	Basic education, individual job search, job search skills, vocational training, work experience, community education and training.
JOBS program (Texas) (King et al. 1994; Schexnayder & Olson 1995)	NRS	AFDC recipients.	Education, job skills training and support services.

Name of programme/references	Design	Participants	Intervention
Jobs Plus (Bloom & Riccio 2002)	ITS	Residents of public housing developments.	Employment-related services, rent-based work incentives that allowed residents to keep more of their earnings, and activities to promote neighbourhood-to-neighbour support for work.
NDLP (New Deal for Lone Parents) (UK) (Hales, Lessof, & Roth 2000; Hales et al. 2000; Lessof et al. 2003)	NRS	Lone parents on income support.	Personal advisers provided an individually tailored package of advice and support.
New Deal for Young People (Bourn John 2002; White & Riley 2002)	Register data	Long-term unemployed young people aged 18-24.	Personal advice, action plan, job search, work experience, job training, skills training.

Name of programme/references	Design	Participants	Intervention
Utah's Single-Parent Employment Demo Program ¹³ (Cebulla, Bouchet, & Greenberg 2005; Janzen, Taylor, & Weathers 1997)	Started as a RCT, but ended up as a non-experimental study	Single-parent AFDC families or two-parent AFDC families with one disabled parent.	Self-sufficiency plan with counsellor. Either one-time payment or mandated participation in self-sufficiency activities.
Work Incentive Program (Schiller et al. 1976; Schiller 1978)	PCS	Welfare (AFDC) recipients.	Rapid job placements. Mandatory participation. Child-care assistance, basic education, vocational training, subsidized on-the-job training, counselling, and transportation assistance.

PCS: Prospective cohort study. NRS: Non-randomised study. ITS: Interrupted time series.

¹³ PROBABLY SEVERELY FLAWED! Only earnings data from the welfare recipients themselves. As a result, the evaluators were limited to non-experimental comparisons of only those cases that were active on welfare.

Table 3. Variables (n = 41) with disagreement

(There were 605 possible disagreements. The remaining 14 variables had perfect agreement)

Variable	Number (percent)
Aims of the program	6 (11)
Proportion of participants actually having work	6
Random generation	5 (9)
Allocation concealment	5
Attrition bias	5
Mean age	5
Sample size in intervention group 1	4
Sample size in control group	4
Year when data collection ended	3 (5)
Description of education	3
Employment- or education-focused intervention	3
Unemployment rate in the area	3
Previous welfare history	3
Number of children	3
Proportion sanctioned	3
Participation in work experience for control group	3
Did they report control of initial differences	3
Performance bias	3
Detection bias	2 (4)
Intention to treat	2
Age of youngest child	2
Year when data collection started	2
Control for gender?	2
Types of outcome	2
Control for education?	2
Proportion “other ethnicity”	2
Sample size in intervention group number 2	2
Private or public implementing agent	2
Proportion of Caucasians	2
Control for age?	1 (2)
Description of age	1
Proportion of males	1
Proportion of African Americans	1
Previous work experience	1
Job-search first or varied first activity	1
Duration of intervention	1
Extent of work experience (e.g. number of hours per	1
Private or public funding agent	1
Sample size in control group number 2	1
Adverse effects	1
Total number (percent) of disagreements	107 (18)

Table 4. Quality coding of the included programmes (58 sites)

Quality Indicators ¹⁴	Met		Unclear		Not met	
	%	N	%	N	%	N
Random generation of allocation	19	11	78	45	3	2
Allocation concealment	28	16	72	42	0	0
Prevention of performance bias	36	21	47	27	17	10
Prevention of detection bias	84	49	16	9	0	0
Prevention of attrition bias	59	34	31	18	10	6
Intention-to-treat	50	29	48	28	2	1

¹⁴ The quality indicators are explained in the protocol.

Table 5. Description of interventions by endpoint (n = 73)

Variable				
Intervention focus	Employment-focused n= 46 (63%)	Education-focused n =16 (22%)	Other n = 9 (12%)	Unclear n= 2 (3%)
First activity	Job search first n =24 (33%)	Varied first activity n = 20 (27%)	Other n = 26 (36%)	Unclear n =3 (4%)
Proportion sanctioned	0.169 (SD: 0.113)			

Table 6. Type and reporting of outcomes by endpoint (n = 73)

Variable						
Type of outcome	Data from state and county administrative records n = 46 (63%)	Administrative records and self-report n = 22 (30%)	Report by social worker and self-report n = 9 (12%)	Self-report n = 19 (26%)	Administrative records + report by social worker + self-report n = 8 (11%)	Unclear n = 2 (3%)
Reporting of adverse outcomes	38 not reporting (95%)	2 reporting (5%)				

Table 7. Descriptive data on all recorded continuous moderators

Moderator	Mean	Sd	Range
Mean age	31.3	2.97	18-39
Year when data collection started	1989	5.99	1967 - 1996
Percent males	11.3	16.0	0 – 100
Percent Whites	39.2	23.9	0 – 98
Percent Blacks	42.3	26.3	0 - 96
Percent Hispanics	17.4	18.9	0 – 77.6
Percent other group	7.1	12.2	0 – 88
Percent regional unemployment	6.7	1.89	2.8 – 13.3
Percent sanctioned	16.9	11.3	0.7-39.1
Proportion with GED	43.3	19.1	5.7 – 84
Duration of intervention (months)	25.8	11.4	3- 60

Table 8. Zero-order correlations¹⁵ among the continuous moderators

Start year of data collection	Age	Males	Whites	Blacks	Hispanics	Other	Unemployment	GED	Duration of intervention	Sanctions
	-.34	.03	.10	-.29	.16	.23	-.30	.11	.32	.36
		.32	-.29	.09	.27	.15	.22	-.19	-.04	-.20
			.27	-.30	.33	.33	.06	-.00	-.11	-.07
				-.63	-.13	.21	-.01	.03	-.09	-.06
					-.57	-.62	-.24	.10	-.32	.04
						.59	.23	-.07	.44	.29
							.16	.01	.38	-.04
								-.00	-.01	-.21
									-.13	.37
										.31

¹⁵ Non-parametric correlations were used because the variables were not normally distributed.

Table 9: Effects of Continuous Moderators on Employment (meta-regression)

Moderator	Slope	S.e. slope	Lower limit	Upper limit	Residual Q	df	p-value
Collection start	-0.00135	0.00095	-0.00321	0.00051	377.5	52	.00000
Mean age	0.00829*	0.00156	0.00525	0.01134	302.4	52	.00000
Percent males	-0.00025	0.00022	-0.00068	0.00019	428.9	59	.00000
Percent Whites	-0.00033*	0.00016	-0.00063	-0.00003	370.4	58	.00000
Percent Blacks	-0.00029*	0.00014	-0.00056	-0.00002	385.8	59	.00000
Percent Hispanics	0.00026	0.00017	-0.00008	0.00059	320.9	48	.00000
Percent other ethnic group	0.00126*	0.00035	0.00058	0.00195	325.9	51	.00000
Percent with GED or high school diploma	-0.00008	0.00017	-0.00042	0.00025	350.6	56	.00000
Duration of intervention	0.00099	0.00036	0.00029	0.00170	407.3	58	.00000
Percent sanctioned	-0.00203*	0.00038	-0.00277	-0.00129	203.9	34	.00000
Regional unemployment	0.00915*	0.00160	0.00601	0.01229	358.1	60	.00000
Number of intervention elements	0.00625	0.00319	0.00000	0.01250	428.4	61	.00000

* means that the 95% confidence interval for the regression slope does not cross zero; the moderator has a statistically significant effect for that outcome.

Table 10. Effects of continuous moderators on earnings

Moderator	Slope	S.e. slope	Lower limit	Upper limit	Residual Q	df	p-value
Collection start	-0.00265*	0.00076	-0.00415	-0.00115	158.6	61	.00000
Mean age	0.00188	0.00132	-0.00070	0.00447	122.7	47	.00000
Percent males	-0.00037	0.00021	-0.00077	0.00004	175.0	61	.00000
Percent Whites	-0.00055*	0.00018	-0.00091	-0.00019	146.8	59	.00000
Percent Blacks	0.00062*	0.00015	0.00032	0.00091	160.6	62	.00000
Percent Hispanics	-0.00064*	0.00017	-0.00097	-0.00030	98.6	41	.00000
Percent other ethnic group	0.00003	0.00034	-0.00073	0.00060	120.5	42	.00000
Percent with GED or high school diploma	-0.00037*	0.00017	-0.00070	-0.00004	132.0	57	.00000
Duration of intervention	0.00005	0.00034	-0.00061	0.00072	163.2	57	.00000
Percent sanctioned	-0.00053	0.00036	-0.00124	0.00018	92.2	38	.00000
Regional unemployment	0.00308	0.00194	-0.00072	0.00689	176.6	63	.00000
Number of intervention elements	0.00242	0.00294	-0.00334	0.00819	170.0	61	.00000

Table 11. Effects of continuous moderators on welfare payments

Moderator	Slope	S.e. slope	Lower limit	Upper limit	Residual Q	df	p-value
Collection start	0.00045	0.00089	-0.00131	0.00220	117.4	42	.00000
Mean age	0.00013	0.00142	-0.00265	0.00291	90.2	34	.00000
Percent males	-0.00053*	0.00020	-0.00092	-0.00014	110.4	40	.00000
Percent Whites	-0.00065*	0.00027	-0.00119	-0.00011	108.8	38	.00000
Percent Blacks	0.00012	0.00022	-0.00030	0.00055	115.8	41	.00000
Percent Hispanics	0.00014	0.00018	-0.00021	0.00049	105.4	36	.00000
Percent other ethnic group	0.00067	0.00036	-0.00003	0.00138	103.2	35	.00000
Percent with GED or high school diploma	0.00021	0.00016	-0.00011	0.00053	101.5	37	.00000
Duration of intervention	0.00066	0.00041	-0.00014	0.00146	95.3	37	.00000
Percent sanctioned	-0.00020	0.00042	-0.00102	0.00063	39.5	23	.01726
Regional unemployment	0.00009	0.00229	-0.00440	0.00457	116.1	41	.00000
Number of intervention elements	-0.00450	0.00323	-0.01084	0.00184	115.7	41	.00000

Table 12. Effects of continuous moderators on welfare proportion

Moderator	Slope	S.e. slope	Lower limit	Upper limit	Residual Q	df	p-value
Collection start	-0.00023	0.00034	-0.00089	0.00042	581.8	38	.00000
Mean age	-0.00619*	0.00091	-0.00798	-0.00440	369.5	33	.00000
Percent males	-0.00239*	0.00018	-0.00274	-0.00203	370.0	35	.00000
Percent Whites	-0.00004	0.00007	-0.00018	0.00010	216.2	34	.00000
Percent Blacks	0.00009	0.00006	-0.00002	0.00020	213.4	34	.00000
Percent Hispanics	-0.00037*	0.00011	-0.00058	-0.00016	424.8	30	.00000
Percent other ethnic group	-0.00074*	0.00038	-0.00147	-0.00000	433.5	29	.00000
Percent with GED or high school diploma	-0.00013	0.00012	-0.00037	0.00012	440.9	35	.00000
Duration of intervention	-0.00114*	0.00027	-0.00167	-0.00061	271.9	33	.00000
Percent sanctioned	-0.00108*	0.00026	-0.00159	-0.00057	287.8	19	.00000
Regional unemployment	-0.01061*	0.00151	-0.01356	-0.00766	432.3	37	.00000
Number of intervention elements*	-0.00395	0.00161	-0.00710	-0.00079	523.3	37	.00000

Table 13: Effects of design quality moderators on Employment

Moderator	k	Risk ratio	Lower limit	Upper limit	p	Q _w (df)	I ²	Q _B (df)	p _B
Random allocation*									
Met	12	1.064	1.016	1.115	.008	53.9 (11)	79.6	26.0 (3)	.000
Unclear	48	1.070	1.049	1.092	.000	336.4 (47)	86.0		
Not met	2	1.031	0.890	1.194	.686	16.0 (1)	93.8		
Allocation concealment									
Met	15	1.073	1.031	1.118	.001	82.0 (14)	82.9	2.798 (2)	.247
Unclear	47	1.066	1.045	1.089	.000	347.4 (46)	86.8		
Performance bias*									
Met	23	1.055	1.018	1.094	.003	144.9 (22)	84.8	7.93 (3)	.047
Unclear	30	1.057	1.035	1.080	.000	168.0 (29)	82.7		
Not met	9	1.119	1.056	1.187	.000	73.3 (8)	89.1		
Detection bias									
Met	57	1.067	1.047	1.088	.000	421.2 (56)	86.7	0.559 (1)	.454
Unclear	6	1.049	1.006	1.093	.025	8.8 (5)	43.5		
Attrition bias									
Met	38	1.072	1.050	1.095	.000	253.3 (37)	85.4	0.817 (2)	.665
Unclear	16	1.064	1.024	1.106	.001	117.1 (15)	87.2		
Not met	9	1.041	0.979	1.108	.202	37.5 (8)	78.7		
Intention-to-treat									
Met	30	1.077	1.052	1.104	.000	207.7 (29)	86.0	1.955 (2)	.376
Unclear	31	1.055	1.028	1.082	.000	169.8 (30)	82.3		
Not met	2	1.493	0.606	3.677	.384	25.3 (1)	96.0		

* means that differences are statistically significant.

Table 14. Effects of categorical moderators on employment

Moderator	k	Risk ratio	Lower limit	Upper limit	P	Q _w (df)	I ²	Q _B (df)	p _B
Focus of intervention*									
Employment	41	1.087	1.056	1.119	.000	311.7 (40)	87.2	8.630 (3)	.035
Education	12	1.048	1.024	1.073	.000	40.5 (11)	72.9		
Other	8	1.026	0.990	1.062	.156	45.7 (7)	84.7		
Unclear	2	1.032	1.001	1.063	.042	0.5 (1)	0.0		
First activity*									
Job-search-first	20	1.091	1.058	1.125	.000	109.7 (19)	82.7	13.1 (3)	.004
Varied first activity	19	1.094	1.054	1.136	.000	115.1 (18)	84.4		
Other	21	1.032	1.008	1.057	.008	102.6 (20)	80.5		
Unclear	3	0.997	0.922	1.079	.942	14.7 (2)	86.5		
Enforcement									
Mandatory	48	1.061	1.042	1.080	.000	324.1 (47)	85.5	1.816 (1)	.178
Voluntary	15	1.116	1.039	1.199	.003	101.3 (14)	86.2		
Work experience									
Yes	50	1.072	1.052	1.093	.000	333.1 (49)	85.3	2.69 (2)	.260
No	12	1.047	0.992	1.105	.094	91.9 (11)	88.0		
Education									
Yes	29	1.067	1.039	1.095	.000	253.0 (28)	88.9	0.0 (1)	1.000
No	34	1.067	1.041	1.093	.000	174.6 (33)	81.1		
Time limits									
Yes	11	1.074	1.013	1.138	.017	110.7 (10)	91.0	0.066 (1)	.798
No	52	1.065	1.046	1.085	.000	316.8 (51)	83.9		
Financial incentives									
Yes	18	1.072	1.026	1.121	.002	140.4 (17)	87.9	0.080 (1)	.777
No	45	1.065	1.044	1.086	.000	290.7 (44)	84.9		
Job search*									

Moderator	k	Risk ratio	Lower limit	Upper limit	P	Q _w (df)	I ²	Q _B (df)	P _B
Yes	38	1.087	1.064	1.111	.000	231.9 (37)	84.0	8.628 (1)	.003
No	25	1.031	1.003	1.060	.028	147.4 (24)	83.7		
Skills training									
Yes	24	1.056	1.029	1.083	.000	148.2 (23)	84.5	0.678 (1)	.410
No	39	1.071	1.046	1.098	.000	266.1 (38)	85.7		
Child care									
yes	17	1.087	1.047	1.129	.000	83.3 (16)	80.8	1.471 (1)	.225
No	46	1.059	1.038	1.080	.000	325.2 (45)	86.2		
Era									
Before 1988	18	1.075	1.031	1.120	.001	55.3 (17)	69.3	0.291 (2)	.865
1988-1996	41	1.063	1.042	1.085	.000	316.9 (40)	87.4		
After 1996	4	1.082	0.970	1.208	.159	59.1 (3)	94.9		
Administration*									
Johnson second period	1	2.408	1.685	3.442	.000	---	---	69.61 (7)	.000
Ford	3	1.203	0.965	1.499	.101	8.39 (82)	76.2		
Reagan first period	10	1.068	1.039	1.099	.000	9.81 (9)	8.3		
Reagan second period	14	1.080	1.026	1.138	.004	141.4 (13)	90.8		
Clinton	15	1.065	1.031	1.100	.000	92.1 (14)	84.8		
George Bush	17	1.033	1.010	1.056	.005	94.5 (16)	83.4		
Evaluator*									
MDRC	37	1.092	1.070	1.116	.000	228.9 (36)	84.3	14.786 (4)	.005
Mathematica	6	1.033	0.979	1.091	.024	17.0 (5)	70.6		
Abt Associates	7	1.018	0.976	1.061	.405	21.6 (6)	72.2		
University	9	1.021	0.969	1.076	.428	59.1 (8)	86.5		
Other	4	1.013	0.923	1.112	.783	9.5 (3)	68.3		

Table 15: Effects of design quality Moderators on Earnings

Moderator	k	Hedges' g	Lower limit	Upper limit	p	Q _w (df)	I ²	Q _B (df)	p _B
Random allocation									
Met	13	0.066	0.035	0.096	.000	34.1 (12)	64.8	3.01 (2)	.222
Unclear	47	0.044	0.031	0.057	.000	110.9 (46)	58.5		
Not met	3	-0.001	-0.080	0.078	.988	11.0 (2)	81.9		
Allocation concealment									
Met	15	0.056	0.033	0.079	.000	34.5 (14)	59.4	0.948 (1)	.330
Unclear	48	0.043	0.028	0.057	.000	134.6 (47)	65.1		
Performance bias									
Met	23	0.038	0.016	0.061	.001	73.0 (22)	69.9	0.941 (2)	.625
Unclear	30	0.049	0.033	0.066	.000	72.3 (29)	59.9		
Not met	10	0.054	0.030	0.078	.000	12.0 (9)	24.8		
Detection bias									
Met	56	0.047	0.034	0.060	.000	160.4 (55)	65.7	0.119 (1)	.730
Unclear	7	0.040	0.002	0.077	.038	10.0 (6)	40.0		
Attrition bias									
Met	40	0.056	0.043	0.068	.000	63.4 (9)	38.5	5.625 (2)	.060
Unclear	17	0.034	0.010	0.057	.006	53.2 (16)	69.9		
Not met	6	0.008	-0.039	0.054	.741	22.0 (5)	77.3		
Intention-to-treat									
Met	30	0.058	0.042	0.074	.000	63.7 (29)	54.4	5.321 (2)	.070
Unclear	31	0.036	0.018	0.054	.000	82.9 (30)	63.8		
Not met	2	0.004	-0.054	0.062	.897	5.192 (1)	80.7		

Table 16. Effects of categorical moderators on earnings

Moderator	k	Hedges' g	Lower limit	Upper limit	P	Q _w (df)	I ²	Q _B (df)	p _B
Focus of intervention									
Employment	40	0.051	0.034	0.067	.000	120.3 (39)	67.6	4.507 (3)	.212
Education	13	0.053	0.032	0.075	.000	21.5 (12)	44.1		
Other	8	0.016	-0.015	0.046	.317	18.8 (7)	62.6		
Unclear	2	0.047	0.013	0.080	.006	0.254 (1)	0.0		
First activity									
Job-search-first	20	0.063	0.040	0.086	.000	52.9 (19)	64.1	5.927 (3)	.115
Varied first activity	20	0.048	0.035	0.061	.000	20.1 (19)	5.3		
Other	20	0.031	0.008	0.053	.008	64.3 (19)	70.4		
Unclear	3	-0.011	-0.090	0.069	.792	9.9 (2)	79.8		
Enforcement									
Mandatory	47	0.044	0.030	0.058	.000	137.6 (46)	66.6	0.525 (2)	.769
Voluntary	14	0.056	0.027	0.084	.000	23.9 (13)	45.6		
Work experience									
Yes	49	0.051	0.037	0.064	.000	120.9 (48)	60.3	1.927 (1)	.165
No	14	0.029	0.000	0.057	.046	42.9 (13)	69.7		
Education									
Yes	28	0.043	0.025	0.060	.000	72.9 (27)	62.9	0.231 (1)	.631
No	35	0.049	0.032	0.066	.000	97.2 (34)	65.0		
Time limits									
Yes	11	0.026	-0.006	0.058	.116	45.7 (10)	78.1	1.899 (1)	.168
No	52	0.050	0.038	0.062	.000	104.9 (51)	51.4		
Financial incentives									
Yes	18	0.047	0.017	0.078	.002	64.6 (17)	73.7	0.013 (1)	.910

Moderator	k	Hedges' g	Lower limit	Upper limit	P	Q _w (df)	I ²	Q _B (df)	p _B
No	45	0.045	0.033	0.058	.000	105.2 (44)	58.2		
Job search									
Yes	39	0.051	0.038	0.065	.000	73.5 (38)	48.3	1.248 (1)	.264
No	24	0.037	0.015	0.058	.001	85.3 (23)	73.0		
Skills training									
Yes	19	0.045	0.022	0.069	.000	51.2 (18)	64.8	0.005 (1)	.944
No	40	0.044	0.029	0.060	.000	112.6	65.4		
Child care									
Yes	19	0.046	0.021	0.071	.000	40.2 (18)	55.2	0.000 (1)	.986
No	44	0.046	0.032	0.060	.000	129.1 (43)	66.7		
Era									
Before 1988	24	0.050	0.033	0.067	.000	42.3 (23)	45.6	3.219 (2)	.200
1988-1996	34	0.049	0.031	0.066	.000	97.1 (33)	66.0		
After 1996	5	0.008	-0.036	0.052	.735	14.7 (4)	72.9		
Administration									
Canada	2	0.075	0.026	0.124	.003	0.725 (1)	0.000	4.672 (5)	.457
Ford	1	0.118	0.011	0.225	.030	---	---		
Reagan first period	10	0.057	0.037	0.076	.000	9.0 (9)	0.416		
Reagan second period	15	0.041	0.020	0.063	.000	30.1 (14)	53.6		
Clinton	22	0.038	0.015	0.062	.002	73.5 (21)	71.4		
George Bush	13	0.044	0.017	0.070	.001	43.0 (12)	72.1		
Evaluator*									
MDRC	37	0.062	0.050	0.074	.000	50.9 (36)	29.3	20.257 (4)	.000
Mathematica	5	0.065	0.036	0.094	.000	2.5 (4)	0.0		
Abt Associates	8	0.004	-0.031	0.040	.808	23.4 (7)	70.0		
University	11	0.011	-0.017	0.039	.444	29.1 (10)	65.6		
Other	2	0.005	-0.060	0.069	.891	5.8 (1)	82.7		

Table 17: Effects of design quality moderators on welfare payment

Moderator	k	Hedges' g	Lower limit	Upper limit	p	Q _w (df)	I ²	Q _B (df)	P _B
Random allocation*									
Met	9	0.055	0.031	0.078	.000	10.5 (8)	23.6	9.096 (3)	.028
Unclear	31	0.052	0.034	0.070	.000	80.4 (30)	62.7		
Not met	2	-0.014	-0.104	0.075	.753	4.36 (1)	77.1		
Allocation concealment*									
Met	12	0.049	0.033	0.065	.000	12.2 (11)	9.6	6.377 (2)	.041
Unclear	30	0.047	0.027	0.067	.000	87.7 (29)	66.9		
Performance bias									
Met	20	0.055	0.038	0.073	.000	31.8 (19)	40.4	0.948 (2)	.623
Unclear	14	0.040	0.011	0.068	.006	49.6 (13)	73.8		
Not met	9	0.045	0.007	0.082	.002	21.8 (8)	63.2		
Detection bias									
Met	40	0.047	0.032	0.062	.000	101.5 (39)	61.6	0.995 (1)	.319
Unclear	3	0.076	0.020	0.132	.008	2.48 (2)	19.3		
Attrition bias									
Met	26	0.048	0.025	0.071	.000	77.9 (25)	67.9	1.261 (2)	.532
Unclear	13	0.053	0.034	0.072	.000	20.9 (12)	42.7		
Not met	4	0.030	-0.005	0.065	.091	5.7 (3)	47.6		
Intention-to-treat*									
Met	16	0.027	0.003	0.051	.003	44.7 (15)	66.4	9.467 (2)	.009
Unclear	26	0.065	0.048	0.083	.000	47.3 (25)	47.1		
Not met	1	0.024	-0.005	0.052	.0100	---	---		

Table 18. Effects of categorical moderators on welfare payments

Moderator	k	Hedges' g	Lower limit	Upper limit	p	Q _w (df)	I ²	Q _B (df)	p _B
Focus of intervention									
Employment	30	0.053	0.034	0.072	.000	82.2 (29)	64.7	0.954 (2)	.621
Education	6	0.033	-0.006	0.073	.095	12.5 (5)	60.1		
Other	7	0.043	0.021	0.066	.000	9.4 (6)	36.2		
First activity									
Job-search-first	14	0.050	0.023	0.077	.000	40.0 (13)	67.5	0.677 (3)	.879
Varied first activity	15	0.043	0.014	0.071	.003	41.3 (14)	66.1		
Other	12	0.057	0.036	0.078	.000	16.7 (11)	34.0		
Unclear	2	0.047	-0.078	0.0173	.460	2.46 (1)	59.4		
Enforcement									
Mandatory	30	0.041	0.025	0.057	.000	77.0 (29)	62.4	3.383 (2)	.184
Voluntary	11	0.081	0.041	0.121	.000	24.3 (10)	58.9		
Unclear	2	0.057	0.008	0.107	.024	1.56 (1)	36.0		
Work experience									
Yes	29	0.050	0.032	0.068	.000	71.8 (28)	61.0	0.022 (1)	.882
No	14	0.047	0.022	0.072	.000	33.8 (13)	61.6		
Education									
Yes	18	0.039	0.019	0.059	.000	40.3 (17)	57.8	1.363 (1)	.243
No	25	0.056	0.036	0.076	.000	62.4 (24)	61.5		
Time limits									
Yes	12	0.033	0.002	0.065	.036	45.0 (11)	75.6	1.250 (1)	.264
No	31	0.053	0.038	0.069	.000	58.4 (30)	48.7		
Financial incentives									
Yes	16	0.057	0.019	0.095	.003	65.5 (15)	76.4	0.370 (1)	.543
No	27	0.045	0.032	0.058	.000	42.2 (26)	38.3		

Moderator	k	Hedges' g	Lower limit	Upper limit	p	Q _w (df)	I ²	Q _B (df)	p _B
Job search									
Yes	28	0.050	0.031	0.068	.000	70.5 (27)	61.7	0.021 (1)	.886
No	15	0.047	0.023	0.072	.000	35.2 (14)	60.3		
Skills training									
Yes	14	0.051	0.020	0.082	.001	41.1 (13)	68.4	0.018 (1)	.893
No	29	0.048	0.032	0.065	.000	64.1 (28)	66.3		
Child care									
yes	15	0.059	0.034	0.083	.000	21.7 (14)	35.6	0.879 (1)	.348
No	28	0.044	0.027	0.062	.000	79.9 (27)	66.2		
Era									
Before 1988	10	0.046	0.025	0.066	.000	11.1 (9)	18.8	0.400 (2)	.819
1988-1996	28	0.051	0.030	0.071	.000	84.1 (27)	67.9		
After 1996	5	0.037	0.001	0.074	.045	10.1 (4)	60.5		
Administration									
Canada	2	0.174	-0.048	0.395	.124	6.82 (1)	85.3	2.349 (4)	.672
Reagan first period	6	0.055	0.030	0.080	.000	5.01 (5)	0.2		
Reagan second period	14	0.044	0.023	0.065	.000	25.5 (13)	49.0		
Clinton	15	0.035	0.004	0.066	.028	53.4 (14)	73.8		
George Bush	6	0.048	0.022	0.075	.000	8.49 (5)	41.1		
Evaluator									
MDRC	24	0.051	0.028	0.075	.000	71.9 (23)	68.0	7.127 (4)	.129
Mathematica	2	0.062	0.017	0.106	.007	0.02 (1)	0.0		
Abt Associates	7	0.047	0.016	0.077	.003	10.1 (6)	40.8		
University	9	0.050	0.024	0.076	.000	17.6 (8)	54.5		
Other	1	0.004	-0.028	0.036	.793	---	---		

Table 19: Effects of design quality moderators on welfare proportion

Moderator	k	Risk ratio	Lower limit	Upper limit	P	Q _w (df)	I ²	Q _B (df)	p _B
Random allocation									
Met	5	0.961	0.907	1.018	.178	7.79 (4)	48.7	0.199 (2)	.905
Unclear	32	0.962	0.942	0.983	.000	455.4 (31)	93.2		
Not met	2	0.927	0.787	1.092	.365	6.78 (1)	85.3		
Allocation concealment									
Met	9	0.955	0.917	0.996	.031	247.1 (8)	96.8	0.130 (1)	.718
Unclear	30	0.963	0.945	0.983	.000	269.0 (29)	89.2		
Performance bias									
Met	15	0.967	0.947	0.987	.001	158.3 (14)	91.2	3.653 (2)	.161
Unclear	15	0.987	0.946	1.030	.544	265.9 (14)	94.7		
Not met	9	0.936	0.902	0.971	.000	42.4 (8)	81.2		
Detection bias									
Met	34	0.959	0.944	0.975	.000	464.1 (33)	92.9	3.051 (1)	.081
Unclear	5	1.038	0.951	1.132	.404	61.8 (4)	93.5		
Attrition bias*									
Met	26	0.974	0.960	0.989	.001	113.5 (25)	78.0	114.440 (2)	.000
Unclear	10	0.968	0.937	1.000	.051	175.8 (9)	94.9		
Not met	3	0.852	0.834	0.869	.000	0.60 (2)	0.0		
Intention-to-treat									
Met	16	0.970	0.953	0.988	.001	73.3 (15)	79.5	2.062 (2)	.357
Unclear	22	0.962	0.936	0.989	.007	390.6 (21)	94.6		
Not met	1	1.084	0.920	1.278	.336	---	---		

Table 20. Effects on categorical moderators on welfare proportion

Moderator	k	Risk ratio	Lower limit	Upper limit	P	Q _w (df)	I ²	Q _B (df)	p _B
Focus of intervention*									
Employment	28	0.960	0.942	0.978	.000	440.7 (27)	93.9	14.172 (3)	.003
Education	5	0.948	0.925	0.971	.000	1.38 (4)	0.0		
Other	5	1.037	0.969	1.110	.298	35.4 (4)	88.7		
Unclear	1	0.910	0.881	0.941	.000	---	---		
First activity									
Job-search-first	13	0.954	0.929	0.979	.000	270.3 (12)	95.6	2.090 (3)	.554
Varied first activity	17	0.971	0.947	0.997	.029	127.0 (16)	87.4		
Other	8	0.955	0.898	1.016	.145	95.8 (7)	92.7		
Unclear	1	1.024	0.910	1.152	.694	---	---		
Enforcement									
Mandatory	28	0.965	0.950	0.982	.000	386.3 (27)	93.0	0.196 (1)	.658
Voluntary	11	0.954	0.908	1.003	.065	126.7 (10)	92.1		
Work experience									
Yes	28	0.958	0.941	0.976	.000	392.7 (27)	93.1	0.375 (1)	.541
No	11	0.971	0.934	1.009	.137	123.9 (10)	91.9		
Education*									
Yes	16	0.980	0.958	1.003	.086	103.3 (15)	85.5	4.239 (1)	.040
No	23	0.947	0.926	0.969	.000	414.7 (22)	94.7		
Time limits									
Yes	4	0.938	0.819	1.074	.354	86.4 (3)	96.5	0.195 (1)	.659
No	35	0.967	0.953	0.981	.000	409.6 (34)	91.7		
Financial incentives									
Yes	14	0.949	0.906	0.994	.025	146.8 (13)	91.1	0.583 (1)	.445

Moderator	k	Risk ratio	Lower limit	Upper limit	P	Q _w (df)	I ²	Q _B (df)	P _B
No	25	0.967	0.951	0.984	.000	364.8 (24)	93.4		
Job search									
Yes	24	0.958	0.940	0.976	.000	370.7 (23)	93.8	0.230 (1)	.632
No	15	0.967	0.933	1.002	.066	157.0 (14)	91.1		
Skills training*									
Yes	13	0.996	0.980	1.013	.655	39.9 (12)	69.9	12.020 (1)	.001
No	26	0.947	0.926	0.970	.000	489.4 (25)	94.9		
Child care									
yes	13	0.970	0.937	1.003	.078	63.5 (12)	81.1	0.230 (1)	.631
No	26	0.961	0.943	0.978	.000	453.7 (25)	94.5		
Era									
Before 1988	16	0.974	0.946	1.002	.073	317.8 (15)	95.3	1.313 (2)	.519
1988-1996	19	0.952	0.926	0.979	.001	157.9 (18)	88.6		
After 1996	4	0.969	0.932	1.009	.125	17.3 (3)	82.6		
Administration*									
Johnson second period	1	1.084	0.920	1.278	---	---	---	14.512 (6)	.024
Carter	1	0.908	0.767	1.075	---	---	---		
Reagan first period	9	0.954	0.908	1.003	.063	222.3 (8)	96.4		
Reagan second period	13	0.970	0.944	0.996	.025	102.7 (12)	88.3		
Clinton	9	0.979	0.951	1.007	.144	32.3 (8)	75.2		
George Bush	4	0.984	0.941	1.030	.491	12.2 (3)	75.4		
Canada	2	0.796	0.708	0.895	.000	3.96 (1)	74.7		
Evaluator*									
MDRC	23	0.968	0.955	0.982	.000	208.5 (22)	89.4	16.301 (4)	.003
Mathematica	2	0.981	0.947	1.016	.287	0.02 (1)	0.0		
Abt Associates	3	0.984	0.954	1.015	.298	0.90 (2)	0.0		
University	8	1.005	0.920	1.098	.910	189.3 (7)	96.3		

Moderator	k	Risk ratio	Lower limit	Upper limit	P	Q _w (df)	I ²	Q _B (df)	p _B
Other	3	0.910	0.882	0.940	.000	0.02 (2)	0.0		

Table 14. Grading of the outcomes

GRADE Evidence Profile

Author(s): Geir Smedslund

Date: 22.06.2006

Question: Should work programmes be used for welfare recipients?

Patient or population: Welfare recipients being potentially able to work

Settings: USA, Canada

Systematic review: Smedslund, et al. (2006). Work programmes for welfare recipients. Campbell review....

		Quality assessment					Summary of findings			
No of studies	Design	Limitations ¹	Consistency	Directness ²	Other considerations	No of patients		Effect		Quality
						work programmes		Relative (95% CI)	Absolute (95% CI)	
Employment 2 years after start of intervention (State and county administrative records Follow up: 2 years)										
10	Randomised trials	Serious limitations (-1)	Important inconsistency (-1) ³	Major uncertainty (-2)	None	14935/29011 (51,5%)	12401/25779 (48,1%)	RR 1.09 (1.03 to 1.16)	30/1 000 (to)	⊕○○○ Very low
Earnings 2 years after start of intervention (State and county administrative records Range: to . Better indicated by: higher scores)										
26	Randomised trials	Serious limitations (-1)	Important inconsistency (-1) ⁴	Major uncertainty (-2)	None	71416	59396	-	SMD: 0.04 (0.02 to 0.07)	⊕○○○ Very low
Welfare payments 2 years after start of intervention (State and county administrative records Range: to . Better indicated by: lower scores)										
9	Randomised trials	Serious limitations (-1)	Important inconsistency (-1) ⁵	Major uncertainty (-2)	None	19017	21558	-	SMD: 0.05 (-0.01 to 0.11)	⊕○○○ Very low
Proportion on welfare 2 years after the intervention started (State and county administrative records Follow up: 2 years)										

10	Randomised trials	Serious limitations (-1)	Important inconsistency (-1) ⁶	Major uncertainty (-2)	None	12172/27570 (44,1%)	8030/13061 (61,5%)	RR 0,946 (0.886 to 1.010)	37/1 000 (to)	⊕○○○○ Very low
----	-------------------	--------------------------	---	------------------------	------	---------------------	--------------------	---------------------------	-----------------	-------------------

Footnotes:

1. 19 % of studies described the method of random allocation, 28 % described allocation concealment, 36 % had same co-interventions, 59 % had lost more than 20 % of participants, 50 % had intention to treat analysis
2. The fact that all included studies have been conducted in the United States and Canada, makes it impossible to generalise to other countries with different social welfare systems.
3. There was significant heterogeneity (Q = 61.478, df=37, p=0.007). One study result was in the opposite direction of the others.
4. Serious unexplained heterogeneity, I-squared = 91%.
5. Serious unexplained heterogeneity, I-squared = 82%
6. Serious unexplained heterogeneity, I-squared = 93%

Figure 1: Selection process of eligible randomised controlled trials from all identified citations

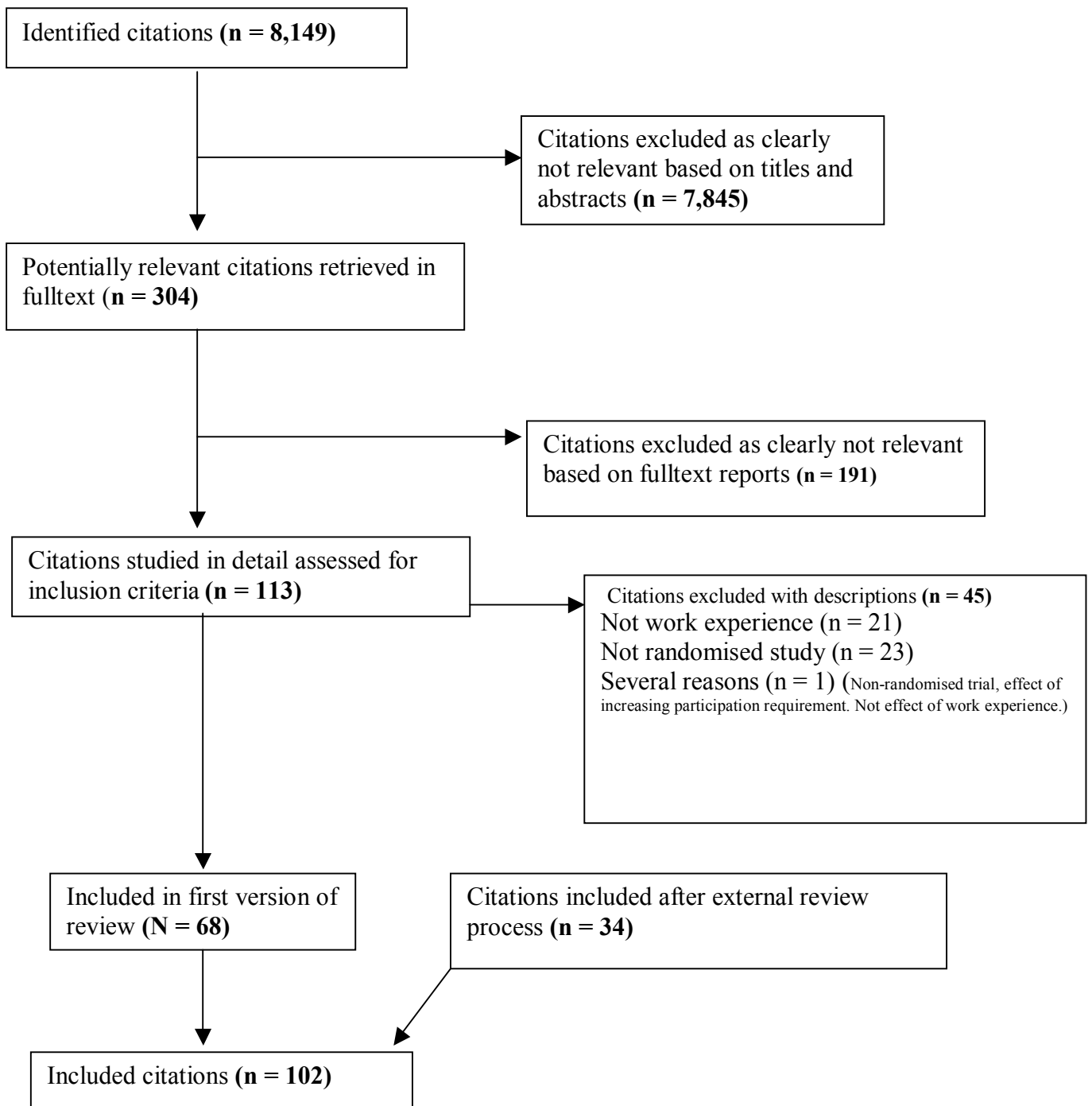


Figure 2. Year when data collection started (n = 58 intervention sites)

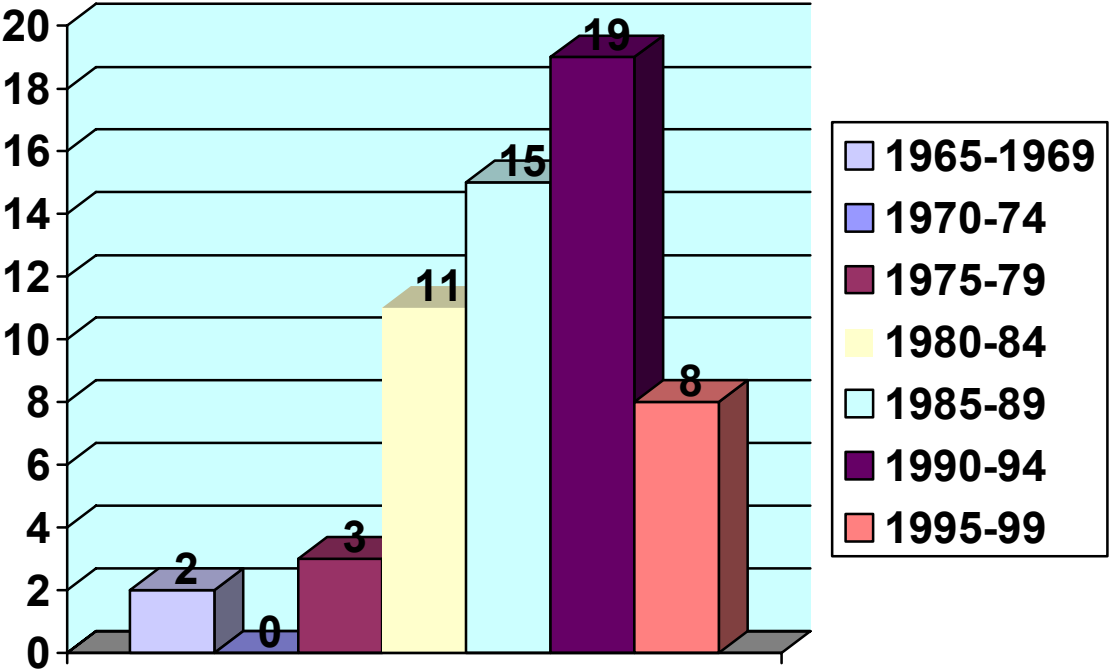


Figure 3. Effects on employment

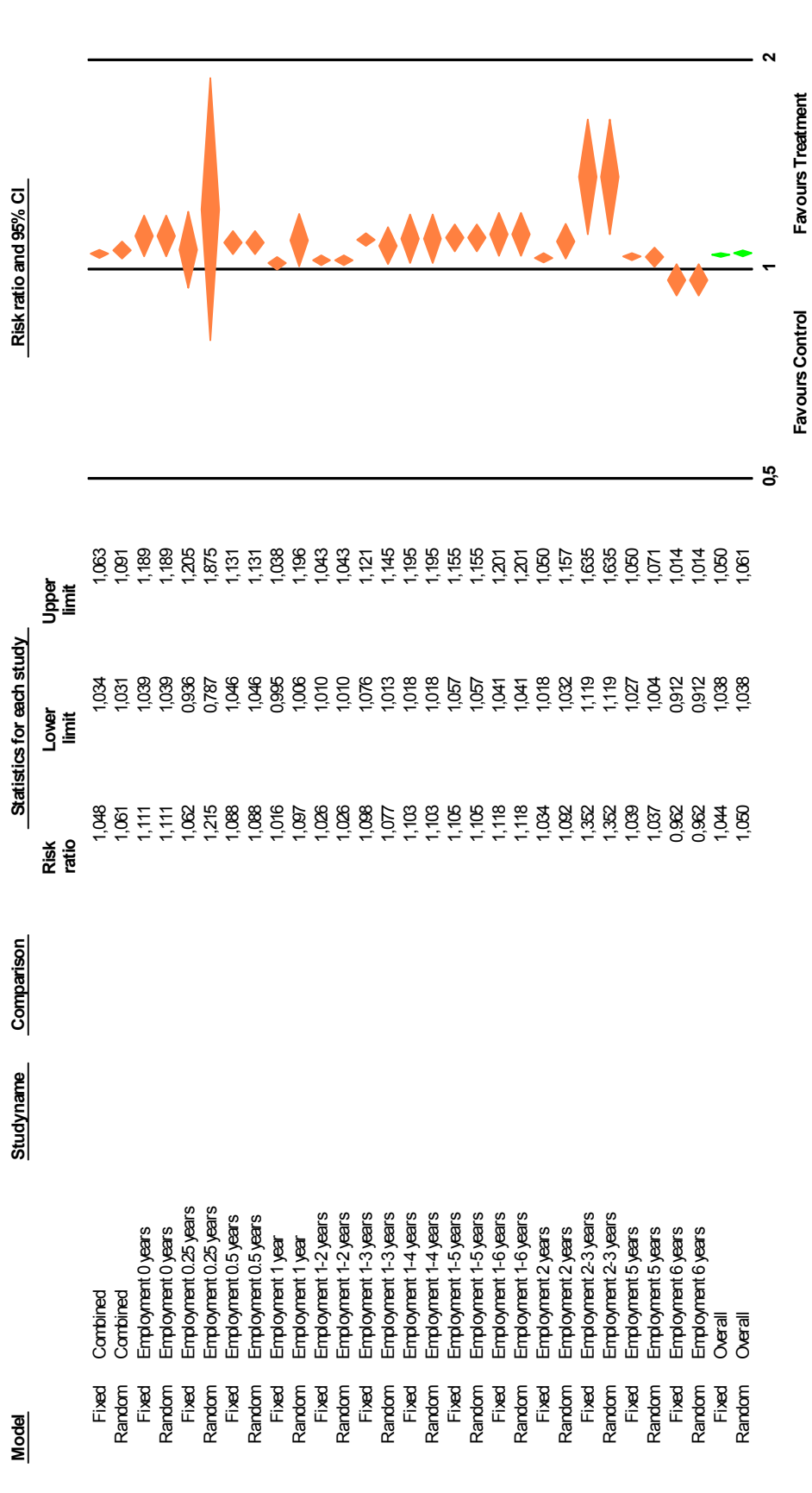


Figure 4. Effects on earnings

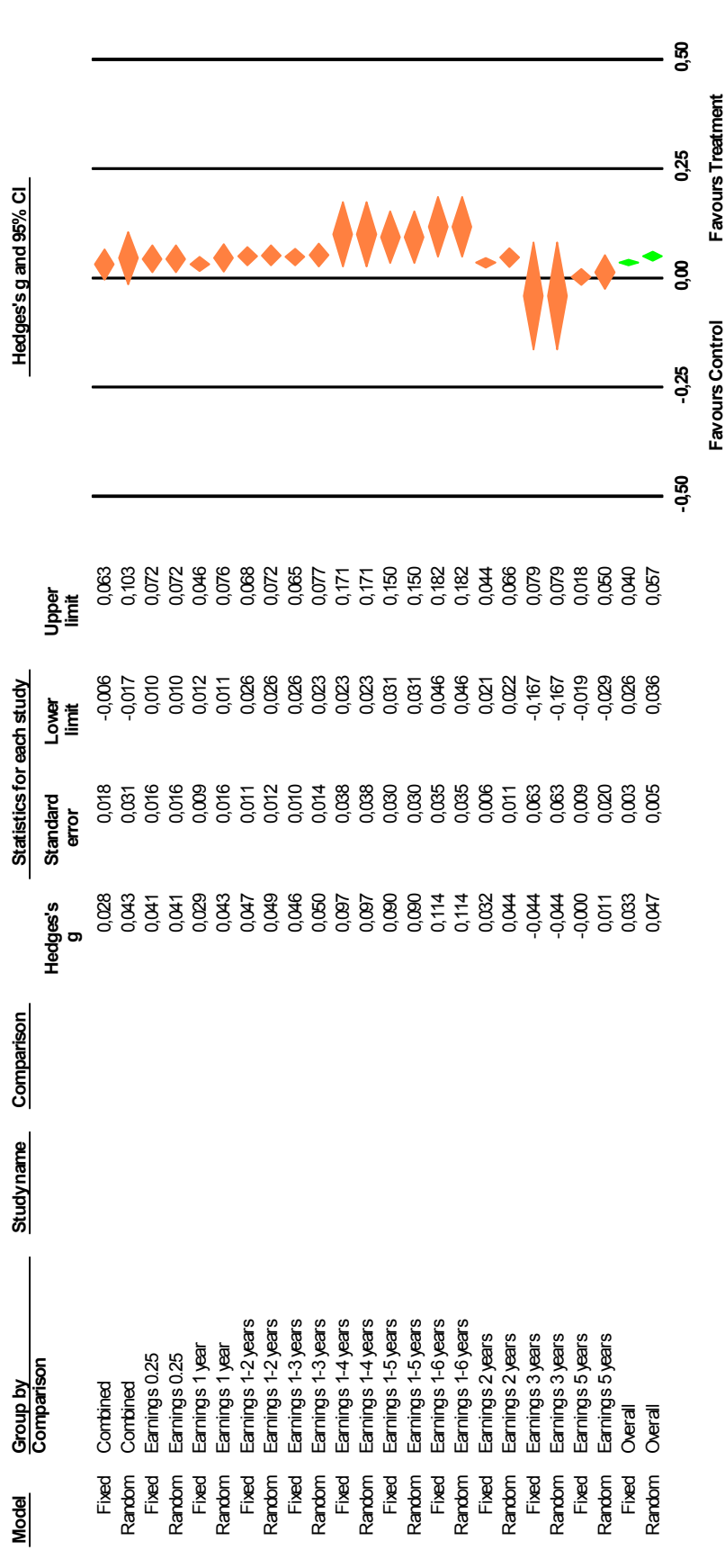


Figure 5. Effects on welfare payments

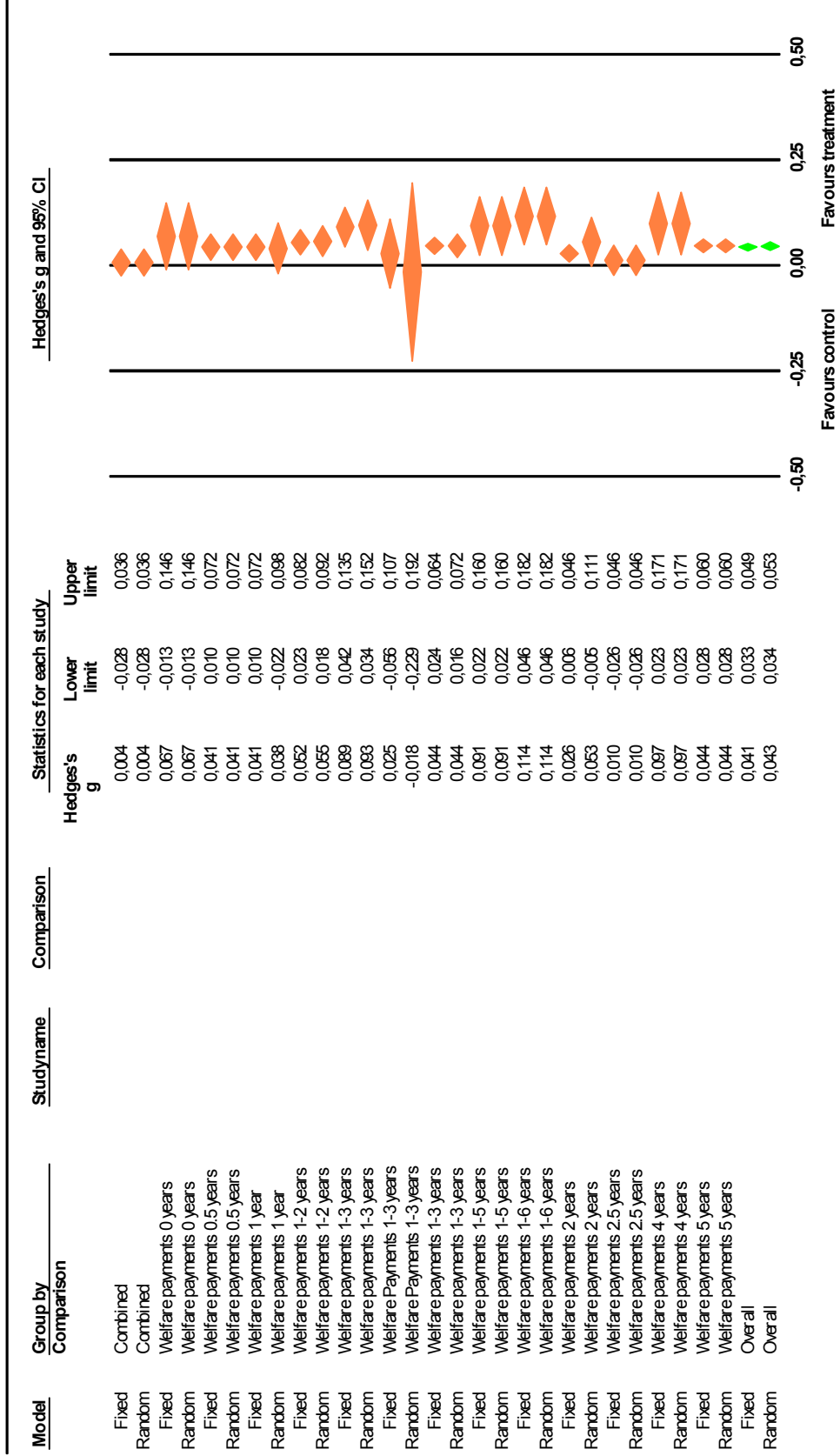
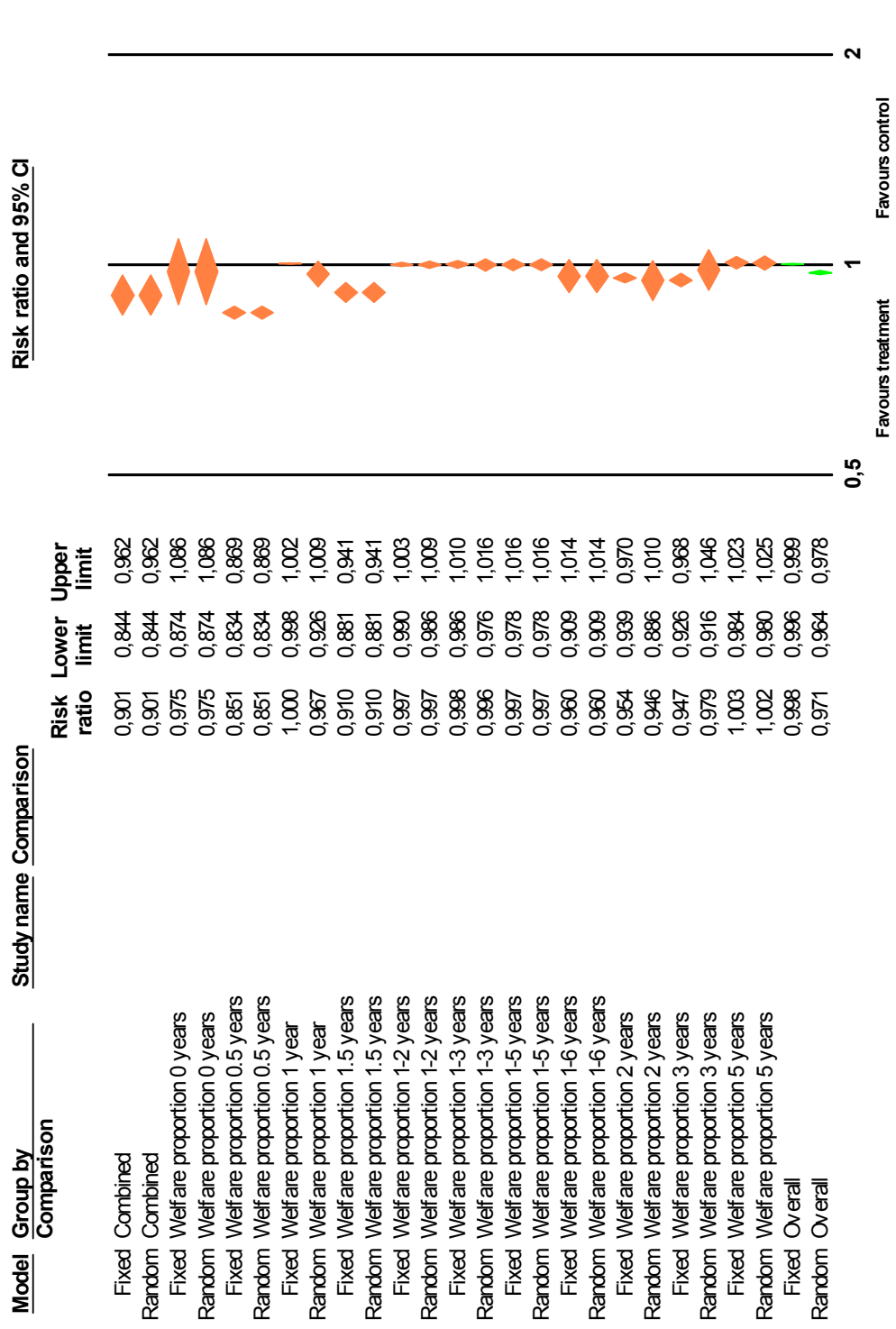


Figure 6. Effects on welfare proportion



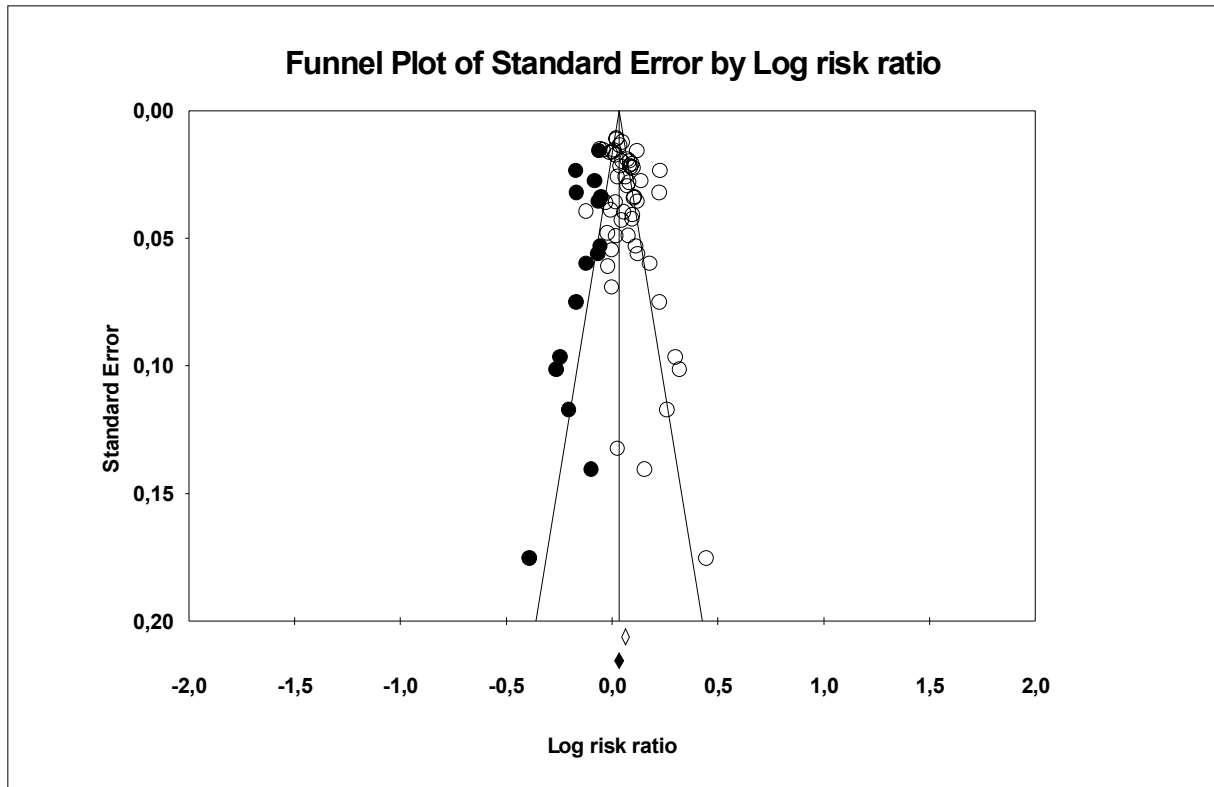


Figure 7. Employment publication bias.

Funnel plot of log risk ratio against its standard error. Open circles signify the actual studies. Closed circles are virtual imputed studies that would be expected if there were no publication bias present in the data.

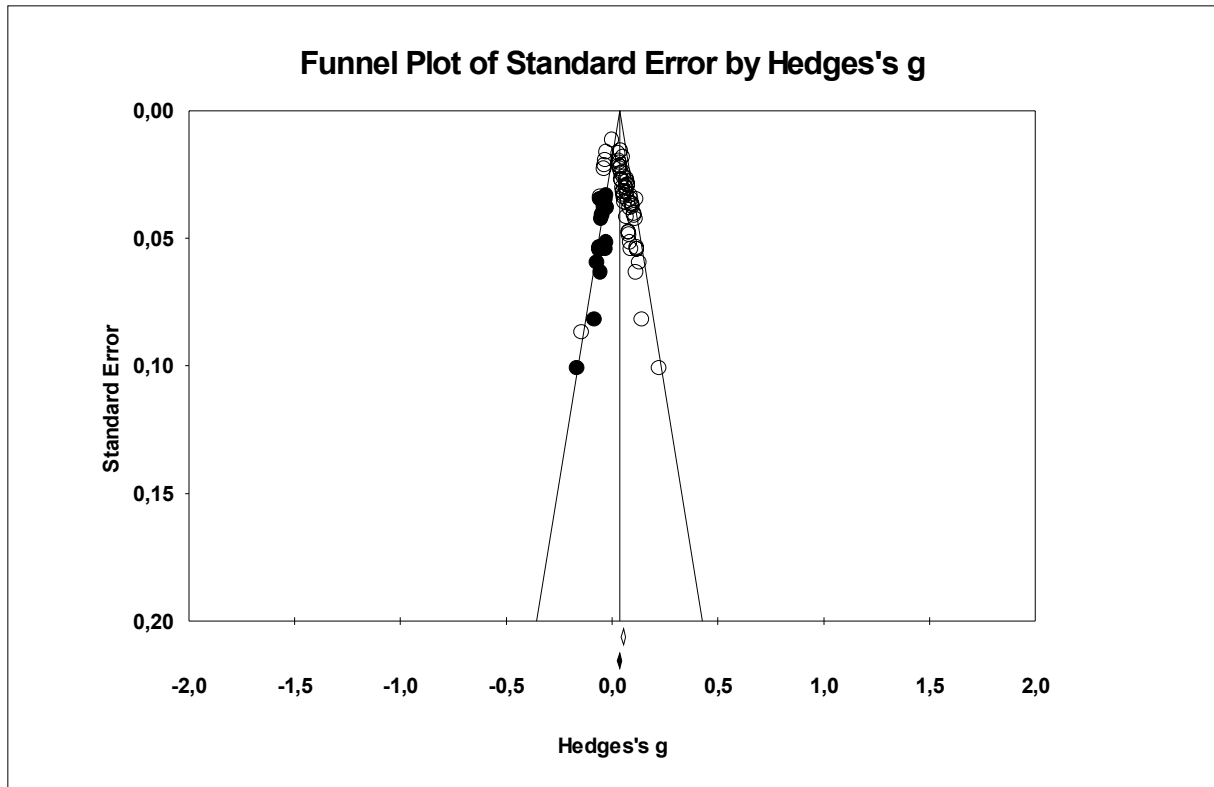


Figure 8. Earnings publication bias.

Funnel plot of Hedges' g against its standard error. Open circles signify the actual studies. Closed circles are virtual imputed studies that would be expected if there were no publication bias present in the data.

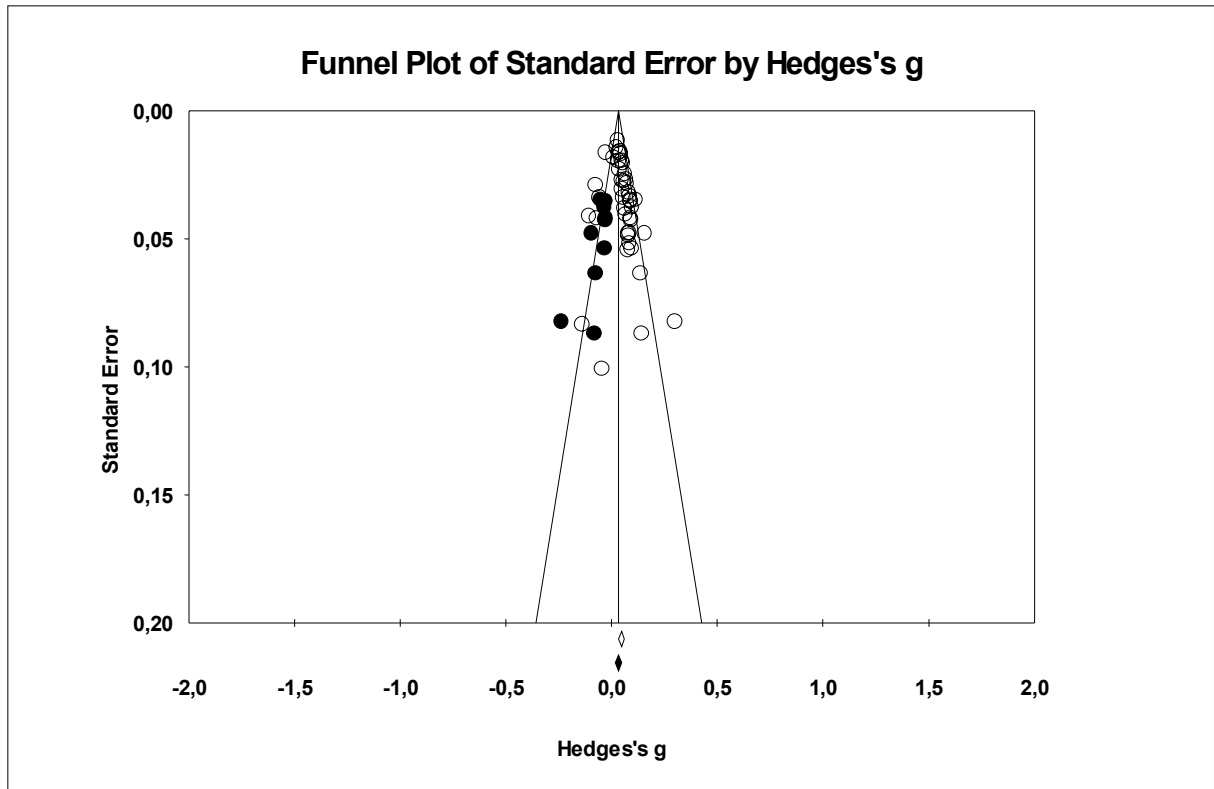


Figure 9. Welfare payments publication bias.

Funnel plot of Hedges' g against its standard error. Open circles signify the actual studies. Closed circles are virtual imputed studies that would be expected if there were no publication bias present in the data.

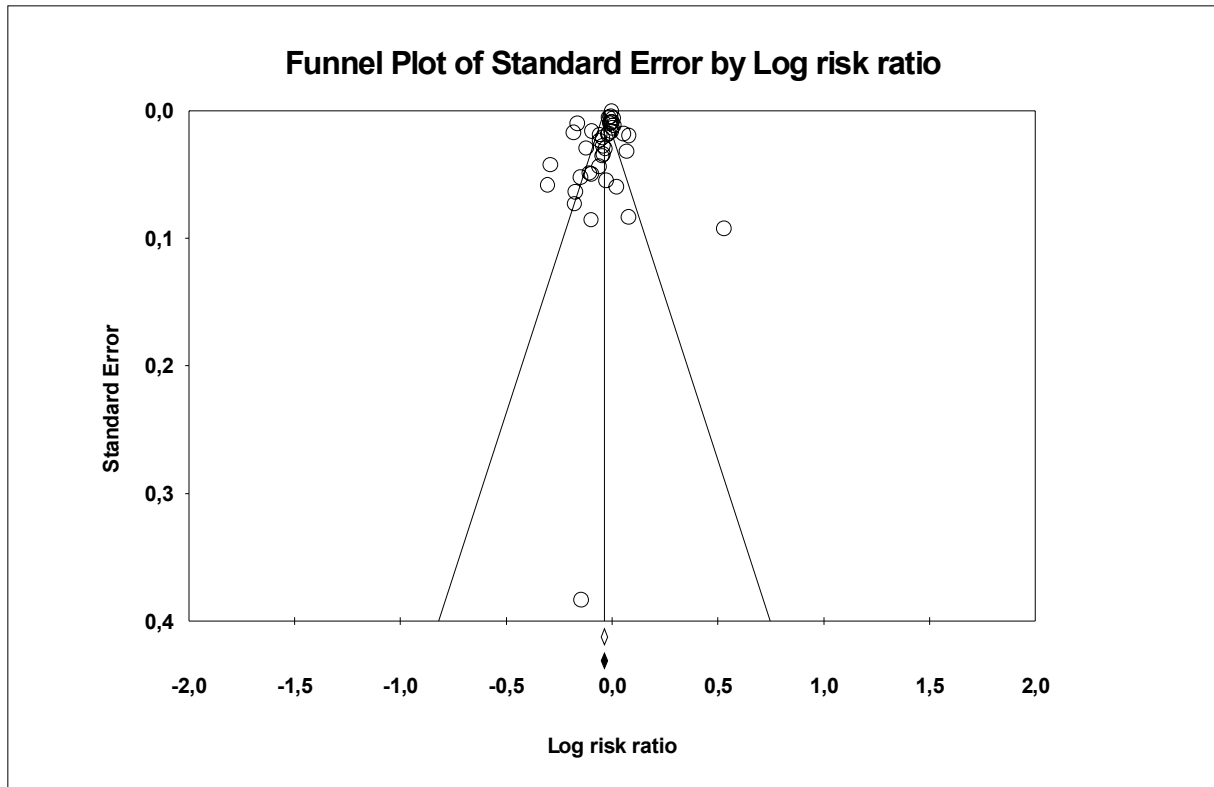


Figure 10. Welfare proportion publication bias.

Funnel plot of the log risk ratio against its standard error. Open circles signify the actual studies. Closed circles are virtual imputed studies that would be expected if there were no publication bias present in the data.

Appendix A: Details of the search strategy

The following is the search strategy that we used to search Sociological Abstracts:

- 1 Social Security/
- 2 Social Welfare/
- 3 Social Support/
- 4 Benefits/
- 5 Welfare Services/
- 6 Welfare Recipients/
- 7 Financial Support/
- 8 (social adj1 assistance\$.tw.
- 9 (social adj1 securit\$.tw.
- 10 (social adj1 welfare).tw.
- 11 (social adj1 allowance\$.tw.
- 12 (insurance\$ adj1 benefit\$.tw.
- 13 (social adj1 benefit\$.tw.
- 14 (welfare adj1 benefit\$.tw.
- 15 (welfare adj1 payment\$.tw.
- 16 (welfare adj1 recipient\$.tw.
- 17 (welfare adj1 support\$.tw.
- 18 (economic adj1 support\$.tw.
- 19 (public adj1 assistance\$.tw.
- 20 (public adj1 support\$.tw.
- 21 (financial adj1 support\$.tw.
- 22 (welfare adj1 service\$.tw.
- 23 (direct\$ adj1 payment\$.tw.
- 24 TANF.tw.
- 25 AFDC.tw.
- 26 temporary assistance to needy families.tw.
- 27 aid to families with dependent children.tw.
- 28 eetc.tw.
- 29 earned income tax credit.tw.
- 30 food stamps.tw.
- 31 (general adj1 assistance).tw.
- 32 (cash adj1 assistance).tw.
- 33 (income adj1 assistance).tw.
- 34 wic.tw.
- 35 (special supplemental food program for women infants and children).tw.
- 36 or/1-35
- 37 Vocational Rehabilitation/
- 38 Workfare/
- 39 Employee Assistance Programs/
- 40 Job Training/
- 41 Employability/
- 42 (vocation\$ adj1 rehab\$.tw.
- 43 (occupation\$ adj1 rehab\$.tw.
- 44 employability.tw.
- 45 (subsidized adj1 employment).tw.
- 46 (employ\$ adj1 incentive\$.tw.
- 47 (employ\$ adj1 program\$.tw.

48 (employ\$ adj1 scheme\$.tw.
 49 (employ\$ adj1 training).tw.
 50 (support\$ adj1 employ\$.tw.
 51 (employ\$ adj1 rehab\$.tw.
 52 (target\$ adj1 employ\$.tw.
 53 (subsidi?ed adj1 job\$.tw.
 54 (job\$ adj1 incentive\$.tw.
 55 (job\$ adj1 program\$.tw.
 56 (job\$ adj1 scheme\$.tw.
 57 (job\$ adj1 training).tw.
 58 (job\$ adj1 creation\$.tw.
 59 (support\$ adj1 job\$.tw.
 60 (job\$ adj1 rehab\$.tw.
 61 (job\$ adj1 search\$.tw.
 62 (job\$ adj1 applicat\$.tw.
 63 (subsidi?ed adj1 work).tw.
 64 (work\$ adj1 incentive\$.tw.
 65 (work adj1 program\$.tw.
 66 (work adj1 scheme\$.tw.
 67 (work adj1 training).tw.
 68 (support\$ adj1 work\$.tw.
 69 (work\$ adj1 rehab\$.tw.
 70 (work adj1 approach\$.tw.
 71 (work\$ adj1 relief).tw.
 72 (training adj1 program\$.tw.
 73 human capital development.tw.
 74 hcd.tw.
 75 WIN.tw.
 76 JOBS.tw.
 77 (job opportunity and basic skills program).tw.
 78 (employment adj1 initiative\$.tw.
 79 (employment adj1 experience\$.tw.
 80 (employment adj1 experiment\$.tw.
 81 (work\$ adj1 initiative\$.tw.
 82 (work\$ adj1 experience\$.tw.
 83 (work\$ adj1 experiment\$.tw.
 84 (GAIN adj1 program\$.tw.
 85 (independence\$ adj1 program\$.tw.
 86 (independence\$ adj1 demonstration\$.tw.
 87 FTP.tw.
 88 family transition program\$.tw.
 89 FIP.tw.
 90 family investment program\$.tw.
 91 (welfare adj1 restructuring).tw.
 92 (welfare adj1 reform).tw.
 93 (ABC adj1 program\$.tw.
 94 (better chance adj1 independence program\$.tw.
 95 or/37-94
 96 36 and 95
 97 welfare to work.tw.

98 workfare\$.tw.
99 or/96-98

Appendix B: Inclusion form and data extraction form

NB: IF NOT OBVIOUS, INDICATE WHERE YOU FOUND THE DATA (PAGE, TABLE NO, ETC.)

ASSESSMENT AND DATA EXTRACTION FOR RANDOMISED STUDIES (April 3, 2006)

RefMan ID of this report: _____

SRS ID of this report: _____

Reviewer (three letters): _____ Date of completing form: _____

PUBLICATION/STUDY

(1 study can have several publications and 1 publication can report results from several studies)

Study identifier: (study name, site) (e.g. GAIN, Miami) _____

Publication type:

report 1 journal article 2 book 3 book chapter 4 dissertation 5 conference 6
other 7

Publication year:

First year of data collection:

Last year of data collection:

Country or countries of where study was conducted: _____

TYPE OF DESIGN (use flow chart)

Randomised controlled trial 1

Cluster randomised trial 2

Quasi randomised trial 3

TRIAL QUALITY

Random generation of allocation

Met 1 (Resulting sequences are unpredictable (explicitly stated use of either computer-generated random numbers, table of random numbers, drawing lots or envelopes, coin tossing, shuffling cards, or throwing dice)).

Unclear 2 (Vague statement that the study was randomised but not describing the generation of the allocation sequence.)

Not met 3 (Explicit statement that the study was not randomised OR explicit description of inadequate generation of sequence, e. g. (e.g., using case record numbers, alternation, date of admission, date of birth).

Allocation concealment

Met 1 (Participants and investigators cannot foresee assignment, e.g. central randomisation performed at site remote from trial location, sequentially numbered, sealed, opaque envelopes).

Unclear 2 (Vague statement that the study was randomised but not describing the concealment of the allocation sequence.)

Not met 3 (Explicit statement that allocation was not concealed OR statement indicating that participants or investigators can foresee upcoming assignment (e. g., open allocation schedule, unsealed or non-opaque envelopes)).

Performance bias

Met 1 (Interventions other than the present one avoided or used similarly across comparison groups.)

Unclear 2 (Use of interventions other than welfare-to-work programmes not reported and cannot be verified by contacting the investigators.)

Not Met 3 (Dissimilar use of interventions other than welfare-to-work programmes across comparison groups, i. e. differences in the care provided to the participants in the comparison groups other than the intervention under investigation.)

Detection bias

Met 1 (Assessor unaware of the assigned treatment when collecting outcome measures. Score as met if outcome is questionnaire data or register data.)

Unclear 2 (Blinding of assessor not reported and cannot be verified by contacting investigators.)

Not met 3 (Assessor aware of the assigned treatment when collecting outcome measures.)

Attrition bias

Met 1 (Losses to follow up less than or equal to 20% and equally distributed between comparison groups (proportion of total loss to follow-up equal to or less than 60% in group with the highest loss to follow-up).

Unclear 2 (Losses to follow up not reported.)

Not met 3 (Losses to follow up greater than 20% or not equally distributed between comparison groups.)

Intention-to-treat

Met 1 (Intention to treat analysis performed or possible with data provided.)

Unclear 2 (Intention to treat not reported, and could not be undertaken by contacting the investigators.)

Not met 3 (Intention to treat analyses not done and not possible for reviewers to calculate independently.)

PARTICIPANTS (Use data for total group. If not reported, use data for intervention group).

Age Data

Age not reported

If data are only reported for subgroups, use the Excel files “combine two groups” or “combine three groups”

	Total group
Mean age	
St. dev.	
Median age	
Description of age data:	

Gender

Gender not reported

Percent males total group..... (round to nearest whole percent)

Ethnicity

Ethnicity not reported

Total group

Percent Whites (Caucasians)..... (round to nearest whole percent)

Percent African Americans..... (round to nearest whole percent)

Percent Hispanics (round to nearest whole percent)

Other, specify..... (round to nearest whole percent)

Education Level Data

Education not reported

	Total group
Percent with GED or High school diploma	
Mean number of years of education	
St. dev.	
Median number of	

	Total group
years of education	
Describe education data:	

Number of children ,

Unemployment rate in the area: ^{Percent} **Go to**
<http://data.bls.gov/PDO/outside.jsp?survey=la>
and search the county, city or state. Use average unemployment for the study years.

INTERVENTION

Table/figure/page where interventions are described _____

Short description of intervention: _____

—

Employment focused or education focused (mark one)
 Employment-focused 1 Education-focused 2 Other 3 Unclear 4

Job search first or varied first activity (mark one)
 Job search first 1 Varied first activity 2 Other 3 Unclear 4

Enforcement
 Mandatory 1 Voluntary 2 Unclear 3

Components of intervention (mark all that apply)

Work experience job search assistance job club

Time limits financial incentives child care assistance

Transportation assistance education

Other
(describe) _____

Duration of intervention (period when intervention and comparison groups face different conditions) _____

Participation

	Percent		
In any activity:	<input type="text"/>	<input type="text"/>	<input type="text"/>
In job clubs:	<input type="text"/>	<input type="text"/>	<input type="text"/>
In job search:	<input type="text"/>	<input type="text"/>	<input type="text"/>
In education and training:	<input type="text"/>	<input type="text"/>	<input type="text"/>
In work experience:	<input type="text"/>	<input type="text"/>	<input type="text"/>
In vocational training:	<input type="text"/>	<input type="text"/>	<input type="text"/>

Proportion sanctioned: ^{Percent} (round to nearest whole percent) **Not reported**

Short description of comparison group: _____

OUTCOME

Table/figure and page where results are reported _____

Sample Size at baseline

Number in intervention group 1	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Number in intervention group 2	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Number in intervention group 3	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Number in control group 1	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Number in control group 2	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Number in control group 3	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Type of outcome:

Register data 1

Report by social worker 2

Self-report 3

Reporting of adverse outcomes? Yes 1 No 0

Description of adverse outcomes: _____

Comparison _____ **vs.** _____

Outcome measure	Follow up time	Group 1 Events/total number	Group 2 Events/total Number	Group 1 Mean (SD)/ median (1)	Group 2 Mean (SD)/ median (1)	Summary statistic (2) Adjusted?	CI	P-value	Comment (3)	Favours treatment (TX) or Control (C)

Comparison _____ **vs.** _____

Outcome measure	Follow up time	Group 1 Events/total number	Group 2 Events/total Number	Group 1 Mean (SD)/ median (1)	Group 2 Mean (SD)/ median (1)	Summary statistic (2) Adjusted?	CI	P-value	Comment (3)	Favours treatment (TX) or Control (C)

Comparison _____ **vs.** _____

Outcome measure	Follow up time	Group 1 Events/total number	Group 2 Events/total Number	Group 1 Mean (SD)/ median (I)	Group 2 Mean (SD)/ median (I)	Summary statistic (2) Adjusted?	CI	P-value	Comment (3)	Favours treatment (TX) or Control (C)

Comparison _____ **vs.** _____

Outcome measure	Follow up time	Group 1 Events/total number	Group 2 Events/total Number	Group 1 Mean (SD)/ median (I)	Group 2 Mean (SD)/ median (I)	Summary statistic (2) Adjusted?	CI	P-value	Comment (3)	Favours treatment (TX) or Control (C)

Comparison _____ **vs.** _____

Outcome measure	Follow up time	Group 1 Events/total number	Group 2 Events/total Number	Group 1 Mean (SD)/ median (1)	Group 2 Mean (SD)/ median (1)	Summary statistic (2) Adjusted?	CI	P-value	Comment (3)	Favours treatment (TX) or Control (C)

- (1) Note if SE or other measure was used
- (2) Note what summary statistics were used. Note if adjusted analysis and if adjusted, which variables.
- (3) Note statistical analysis used