Option 1

TITLE

Women in Wage Labour: A systematic review of effects.

BACKGROUND

METHODS

RESULTS

DISCUSSION

OTHER

0. Abstract

This paper aims to report on the systematic review of the evidence base for the effect of interventions that seek to improve the participation of women in wage labour. The systematic review is of all the available impact evaluation evidence on the effects of interventions aiming to support women's wage labour participation in higher-growth/male-dominated sectors in LMICs. In addition to finding a heterogeneous evidence base and one that is largely of a low quality, combined training and job placement interventions is the only intervention category in which a rigorous synthesis could be executed. Our meta-analysis provided cautious evidence that combined training and job placement interventions could yield positive results insofar as the participation of women in more prestigious sectors is concerned. Our findings suggest that such interventions increased women's participation by 7.8% while female income improved by 7.2%. For all the other labour market interventions identified in this systematic review, the size and nature of the underlying primary research evidence base does not allow for a rigorous synthesis of effects.

Two main features of this study constitute a key research contribution. First, the best evidence in this regard comes from two intervention models: the World Bank's AGI and the Jovenes approach. These two intervention models for combining training and placement services come closest to a design template for promising interventions. In terms of more granular intervention design implications, seven promising design attributes were identified: (1) exposure to labour market participation enhancing social norms; (2) labour demand-led intervention design; (3) gender-sensitive intervention design; (4) provision of soft/life skills and social empowerment training; (5) participant profiling and targeting; (6) clear governance structures for intervention providers; (7) flexibility and responsiveness in intervention implementation and design. Finally, these are important findings for both policymakers and researchers in the pursuit of narrowing the gendered disparities that exist in LMICs.

Option 2

1. Introduction

In recent times, there have been global efforts towards a more equitable labour market for both men and women. Despite this, female labour market participation has remained relatively low with 50% of women participating in the labour market while their male counterparts contribute a relatively higher 76%. In addition to this, wage and salary benefits are skewed favourably towards men as women earn 24% less than them (United Nations, 2013). Moreover, the problem is compounded in low and middle income countries (LMICs) where women are more likely to dominate sectors that are not particularly well-remunerated.

There have been a slew of interventions to curtail the gendered nature of labour markets. This study aims to review and synthesise the impact evaluation evidence base on the effects of interventions aiming to support women's wage labour participation in higher-growth/male-dominated sectors in LMICs. Furthermore, while there has been some labour market evidence synthesis work in LMICs (seeKluve, 2016; Tripney 2013, 2015), none have looked at deconstructing the participation of women in male-dominated industries specifically. In the next section, we discuss the methodological approach taken to conduct this systematic review.

2. Review Methods

Prior to presenting the findings of the review, we briefly discuss the review process. This review followed a two-stage process consisting of an evidence map in the first stage and an analytical review and synthesis in the second stage. The former has a broader scope than a full systematic review. It essentially maps evidence from impact evaluations and systematic reviews on interventions that are designed to overcome barriers to women's labour market participation by adopting an intervention-outcome matrix to elucidate the size and nature of the evidence for different configurations of interventions and outcomes. The framework adopted is available online and the complete map was published online as an open access resources. The bodies of evidence presented in the map aided the identification of the interventions and outcomes of most interest to review stakeholders. The evidence map also guided our initial discussions about the most effective way to synthesise the evidence in answer to the review questions. The systematic review is published as a full technical report (Langer et al 2018) and a summary report (Langer and Tannous 2018). This paper is dedicated to exploring the effectiveness systematic review conducted (Snilstveit 2012).

2.1. Scope

For a study to be considered for inclusion in the systematic review, it had to indicate primary evidence on (i) the effectiveness of interventions supporting women's participation in the wage labour market in higher-growth and/or male-dominated sectors; and (ii) the design features of such interventions. Eligible studies had to be rigorous quantitative impact evaluation of women's wage labour interventions with a study sample of women aged 15 years or older. This study sample had to either be majority female (51% or more) or the study results had to be disaggregated by gender. Evaluations were required to apply a counterfactual-based design using random or quasi-random method of group assignment or non-random methods of assignment subject to specified conditions. Control or comparison conditions in eligible studies referred to the population receiving no treatment, treatment as usual, an alternative treatment, or pipeline treatment. Only studies conducted in low- or middle-income countries (LMICs) according to the World Bank classification at the time of data collection for the study were eligible for inclusion. Studies further had to evaluate the effects of relevant interventions applied in economic sectors with high or growing productivity *and/or* which were male-dominated.

2.2. Interventions and outcomes

2.2.1. Interventions

We included any intervention that aimed to overcome the barriers to women's wage labour market participation in LMICs. The term intervention in this context referred to a policy, programme, strategy, technology, device or any other type of deliberate action. We included multi-component (also known as bundled or combined) interventions only if: (i) all intervention components aimed to overcome a relevant barrier to women's wage labour participation and/or (ii) studies could attribute the observed outcome to an intervention component that aimed to overcome a relevant barrier.

2.2.2. Outcomes

To be included in the systematic review, studies had to evaluate the impact of interventions on one of these two final outcomes: (i) Participation in formal or informal wage labour employment (in higher-growth and/or male-dominated sectors), (ii) Economic empowerment (following outcome 1: participation in higher-growth and/or male-dominated sectors). Participation in formal and informal employment encompassed various indicators such as employment status, under-employment, nature of employment, progression and career prospects, and transitions from traditionally female-centric industries to more male-dominated ones. Economic empowerment could be indicated by poverty-related measures, asset and income control, saving levels, empowerment and investment in economic assets.

In cases where the majority of intervention participants were not female, studies were eligible for inclusion only if the impacts on women were assessed separately from those on men (i.e. in sub-group analysis) or in comparison to men. Studies were only included in the systematic review if they assessed one or both of the two broad outcomes. Studies investigating intermediate outcomes only were excluded.

2.3. Search strategy

2.3.1. Searching

A comprehensive search strategy was used to search the international research literature for qualifying studies. In order to identify relevant research studies, we conducted an exhaustive search of 74 academic and grey literature sources covering literature from international development, economics, sociology, psychology, education, and health care. The search of academic sources was led by an information scientist, who developed a detailed search query (or string) based on the inclusion criteria, relying on the database's index terms where available and/or free-text terms. There were no language restrictions to the search. A publication year filter to identify studies published since 1990 was used. This cut-off date was chosen as structural inequalities in women's economic empowerment only started to receive increasing attention in the 1990s (Kabeer, 2012). The Grey literature search was informed by our content experts and advisory group members and covered a large variety of organisational repositories and websites. The snowballing search included hand-searching the content pages of key journals of particular relevance to the review question; backward citation searches (i.e. searching the reference lists of included studies and seminal papers); forward citation searches (i.e. using Google Scholar to search for papers that cited included studies); requests to key authors and organisations to share studies with the review team; and a Twitter call for the wider community of practice to contribute relevant studies.

2.3.2. Data management

Review management software (EPPI-Reviewer 4) was used to manage the entire review process (Thomas et al 2010). All potentially relevant items identified through the academic database search were exported to EPPI-Reviewer and then manually screened for eligibility, with EPPI-Reviewer used to keep track of decisions made about each citation. Search hits from organisational repositories and snowballing were stored in MS Word, and only the details of studies deemed relevant for the map, plus those over which there was any doubt, were transferred to EPPI-Reviewer.

2.4. Screening

All the studies identified and recorded from the search process were screened against a defined inclusion criteria, leading to the exclusion of any studies that did not meet this criteria. As our approach was two-pronged, comprising g of an evidence map in the first stage and an analytical review and synthesis in the second stage, we firstly screened studies for inclusion in the evidence map on title and abstract only. To this end, we started by manually examining the titles and abstracts of all records entered into the reviewing software after removing duplicates. Where we were in any doubt as to a study's eligibility (e.g. because no abstract was available, or it did not provide enough information), we classified it as 'unsure'.

Secondly, we screened studies for inclusion in the in-depth review using full texts. These were obtained for studies included in the map and for the items marked as 'unsure', and detailed manual examination of the full reports was undertaken independently by pairs of reviewers. Disagreements between the reviewers' decisions were resolved by identification of the source of the disagreement, re-reading of the text and discussion. If a final decision could not be reached, a third reviewer was asked to reconcile the differences. In the event that we could not determine the eligibility of a study (e.g. because the full text was unavailable, insufficient information was provided, or the only version we had was in a language other than English), the study authors were contacted to request additional information about the study or access to English-language versions. In cases where no English-language version or additional information could be obtained, the study was excluded from the review.

2.5 Data extraction and critical appraisal

2.5.1 Data extraction

Data was extracted for all included studies. We used a predefined data extraction tool in order to systematically and transparently extract data from the included primary studies. The tool was translated into a coding set on EPPI-Reviewer to extract the information required for both the mapping exercise and the in-depth review. The data were entered directly into the EPPI-Reviewer database.

For the in-depth analysis and synthesis of study results, full-text reports were examined and studies coded on variables related to the study context, the characteristics of the study samples, details of the intervention design and its implementation, the study methods and the outcome variables and data.

Two members of the review team piloted the data extraction tool, working independently on a purposive sample of eligible studies selected to test the tool on the full range of included impact evaluation designs and methods. This process was repeated until a very high level of consistency in reviewers' application of the codes was achieved, at which point the tool was finalised. Thereafter, the

remaining studies were coded by individual reviewers, with a sub-set of full-texts being coded by different combinations of two reviewers independently extracting information from each study report and then coming together to compare their decisions. Any uncertainties and discrepancies were resolved by discussion, further review of the respective study reports, and where necessary, consultation with a third reviewer.

2.5.2 Critical appraisal

Each of the study was critically appraised to assess its methodological quality. A critical appraisal tool was applied to assess the impact of bias on the trustworthiness of primary study results, where trustworthiness refers to the confidence of the review team that the findings reported in the included studies used for the synthesis were rigorous and credible. In order to assess the risk of bias of the primary studies, we adapted the Cochrane risk of bias tool for non-randomised studies (Sterne et al 2016), which we had previously used and adapted in international development reviews (Stewart et al 2015). Sterne and colleagues used a domain-based risk of bias assessment covering the following six indications of trustworthiness: (i) selection bias; (ii) confounding bias; (iii) bias due to departures from applied interventions; (iv) bias due to missing data; (v) bias due to measurement of outcomes; and (vi) bias due to selection of the reported result. Each domain of bias received a low, moderate, high or critical risk of bias rating, allowing for a transparent calculation of an overall risk of bias score for each study. Studies with a critical risk of bias were included in the review but excluded from the synthesis.

The critical appraisal tool used to assess studies for the full review is presented in Appendix 6. It was piloted using a similar approach to that used for the data extraction tool. Two reviewers independently assessed each study and then came together to compare their decisions. Where these reviewers could not come to an agreement about the risk of bias rating for a particular study, a third reviewer was consulted.

2.6 Synthesis

This systematic review included an aggregative synthesis in order to: (i) distinguish between effective and non-effective interventions; and (ii) identify those configurations of intervention design features, participants and contextual characteristics that might be associated with a given outcome. The analysis focused particularly on identifying design features of interventions that were critical to their success. We used meta-analysis to identify overall effects and narrative synthesis and qualitative comparative analysis (QCA) for the disaggregated analysis according to the intervention design features (Gough et al 2012; Thomas et al, 2014).

2.6.1 Method of analysis (1): Meta-analysis

Meta-analysis is the most rigorous method to synthesise quantitative evidence (Borenstein et al 2009; Lipsey and Wilson 2001). As a statistical approach it aggregates the numerical findings, such as effect sizes, of primary research to report a pooled overall numerical value. This numerical value – the pooled effect size – expresses the overall finding derived from the combined primary research results. The pooled effect size reflects the direction and magnitude of the observed primary effect sizes, which are allocated different weights in the analysis depending on sample sizes and variance.

We report calculated effect sizes in tabular format as well as using forest plots. Where sufficient contextual homogeneity prevailed, effect sizes were averaged across studies by using an inverse

variance weighting of the individual effect size. This weighting resulted in the individual effect sizes of studies with larger study samples being given more weight in the combined, pooled effect size. The meta-analyses were carried out using random effects statistical models.

The studies included in the review featured impact evaluation designs based on a random (or quasi-random) method of group assignment, and those that used non-randomised procedures. Some caution is needed when synthesising effect sizes from studies using different designs and estimation techniques. Where appropriate, we summarised across designs and made explicit our rationale for doing so. In addition, we calculated separate estimates of treatment effects for randomised and non-randomised studies.

To test the robustness of the results of the meta-analysis, a number of sensitivity analyses were conducted. Broadly, this involved collecting data on, and assessing sensitivity of findings to (i) the methods of the primary studies and (ii) the methods of the review. The included studies varied methodologically. We conducted sensitivity analyses to examine the influence of these variations on the summary measures, in order to offer possible explanations for the differences between studies when interpreting the results. We examined whether the results were sensitive to study design, the risk of bias associated with the study, the degree of missing/incomplete data, and the way outcomes were measured and the timing at which they were measured. We also conducted moderator analyses to try to explain variations in effect sizes. Moderator analyses were reported in tabular format below each meta-analysis. Analyses were calculated using the same one-way random effects ANOVA model as applied to the sensitivity analyses.

2.6.2 Method of analysis (2): Qualitative comparative analysis (QCA)

QCA investigates the configuration of different conditions and their association with intervention effectiveness (Thomas et al, 2014). A condition refers to different design features in the context of our review, for example the use of behavioural techniques, gender sensitive design or the intensity of the intervention. Using QCA enables us to analyse whether configurations of different design features are present (or not) when the intervention has been effective (or not) in increasing women's wage labour participation. In comparing different types of configurations, QCA applies Boolean logic to establish necessary or sufficient conditions for intervention effectiveness. We refer to these conditions as the active ingredients of women's wage labour participation interventions in our review.

The advantage of this synthesis approach is the presentation and investigation of overlapping pathways to causality, which is to be expected in our review, as it is unlikely that any single intervention or design feature is the single cause of positive labour market participation effects. While meta-analysis is restricted to evaluating the effects of different variables individually, QCA allows us to evaluate the effects of different configurations of conditions. For example, conducting a moderator analysis as part of our meta-analysis, we could examine whether intervention intensity had a significant impact on intervention effectiveness *or* whether using behavioural approaches had a significant impact on intervention effectiveness, and so forth. In contrast, QCA allows us to compare different configurations of these design features (e.g. using a behavioural approach *and* a gender-sensitive design) and how these configurations – not the conditions themselves – are associated with intervention effectiveness. We assumed that providing different configurations of intervention design features could provide more policy-relevant synthesis results that could inform future policy and programme designs.

2.7. Overall assessment of the quality of the evidence

We applied the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) tool to assess the overall quality of the evidence included in the review and the strengths of the review's findings. GRADE evaluates the quality of the primary evidence included in the review based on a range of factors, including: primary study limitations, inconsistency of the identified effect sizes, indirectness of the evidence (e.g. not evaluating final outcomes), and risk of bias ratings (Guyatt et al 2008). The quality of evidence rankings ranged from high to moderate, low and very low. In addition to assessing the quality of the primary evidence, GRADE further assigns a strength of recommendation ranking to review findings. In this, GRADE combines the quality of evidence scores with additional variables such as uncertainty about the balance between desirable and undesirable effects, or uncertainty or variability in values and preferences. Strength of recommendation rankings are divided into 'strong' and 'weak'.

2.8. Stakeholder engagement in producing the review

In this review, we applied a range of mechanisms to support stakeholder engagement in designing and producing the review. These mechanisms included the creation of a review advisory group, regular joint meetings between the review team, the EPPI-Centre as the co-ordinating body, and DFID, and the production of an interactive evidence map. Each of these mechanisms aimed to ensure that the review's scope was of relevance to decision makers and that it addressed a relevant policy question. Two main features of this approach stand out for its implication for the design and production of the review: (i) using the interactive evidence map to jointly decide on the review scope; and (ii) identifying intervention design features associated with women's increased labour market participation.

3. Findings

3.1. Descriptive findings

The systematic search for impact evaluations on the effects of interventions supporting women's labour market participation in LMICs identified 16,091 citations (Figure 1). These were derived from 74 different sources compromising scientific databases of research studies as well as a range of Grey literature sources. Our exhaustive search for studies indexed in scientific databases was conducted between December 2016 and January 2017 and updated in April 2017. The search of Grey literature sources was conducted in February 2017. All search hits were imported into EPPI-Reviewer (V.4.6.4.1) to facilitate data storage and management, except for the Grey literature searches; because they had to be entered manually, only those that were found to be relevant after initial screening (or of which the reviewers were doubtful) were captured to EPPI-Reviewer.
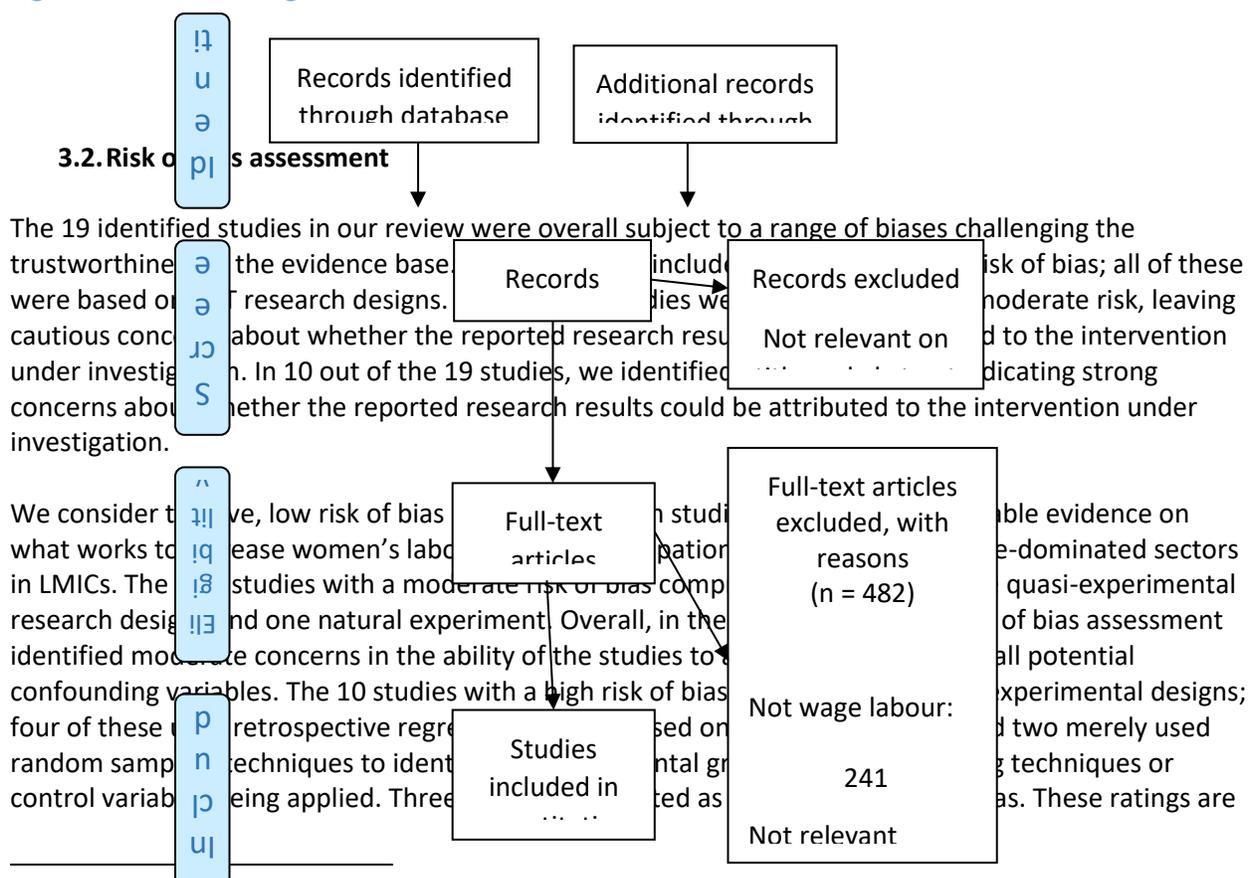
After screening on title and abstract, the majority of citations were excluded as not relevant to the review question (n=15,590). Full texts of the remaining 501 studies were then retrieved and screened against our inclusion criteria. This led to the exclusion of a further 482 studies. The main reason for exclusion on full text was that studies did not evaluate the effects of interventions on wage labour outcomes (n=241), a common example being examining small-business development outcomes. Further notable reasons for exclusion were irrelevant economic sectors (n=112) and studies that did not apply rigorous impact evaluation designs as outlined above (n=80). Studies included varied by study type. Of the 19 studies included in the review, the majority were published as working papers (n=13), followed

by journal articles (n=5), and a single evaluation report. Most studies were published between the years 2013 and 2015. Regionally, studies were dominantly from Asia (n=9), followed by Latin America (n=5) and Africa (n=5). The most researched country was India (n=4).

A variety of socio-economic indicators were investigated using different dimensions.[1] According to the World Bank classification of economies, a majority (n=13) of our included evaluations were conducted in countries classified as LMICs, with an average GNI/per capita of USD 2,400. There was no clear trend among the included studies regarding urban and rural socio-economic categorisation; majority of studies spanned both rural and urban areas (n=10), and only one study targeted employment in rural labour markets exclusively.

In terms of population, included studies featured a diverse group of women. When averaging out the characteristics of female participants, we developed an idea of the average woman targeted in interventions that are part of this review. The average woman is unmarried, 25.9 years old, out of work or has recently completed or dropped out of school, enjoys a higher than average level of schooling and lives in a disadvantaged setting. In summary, the reviewed labour market interventions therefore did target more economically mobile and skilled women and developed explicit participant profiling criteria in that regard. Information regarding sectors highlighted a significant inclination for interventions targeted at manufacturing (n=10)[2].

Figure 1 Flow chart diagram of search results and identification of studies

3.2. Risk of bias assessment

The 19 identified studies in our review were overall subject to a range of biases challenging the trustworthiness of the evidence base. included isk of bias; all of these were based on T research designs. dies we moderate risk, leaving cautious conc about whether the reported research resu d to the intervention under investig n. In 10 out of the 19 studies, we identified dicating strong concerns about hether the reported research results could be attributed to the intervention under investigation.

We consider t ve, low risk of bias studi ble evidence on what works to ease women's labo ation e-dominated sectors in LMICs. The studies with a moderate risk of bias comp quasi-experimental research desig nd one natural experiment. Overall, in the of bias assessment identified mo te concerns in the ability of the studies to all potential confounding variables. The 10 studies with a high risk of bias xperimental designs; four of these retrospective regre sed on two merely used random samp techniques to ident ntal gr techniques or control variab eing applied. Three ted as as. These ratings are

---

[1] See the technical report for a description of all the socio-economic and political dimensions investigated.
[2] See technical report for detailed breakdown of the number of studies covering different sectors.

mainly a consequence of non-compliance during the selection of participants, resulting in randomisation failure as well as a problem with the completeness and quality of the collected data. Lastly, a single natural experiment was also judged to have a high risk of bias. This rating resulted mainly from an absence of the methodological information required to transparently assess the study's risk of bias. Figure 2 provides a summary of the full risk of bias ratings of the 19 included studies.

3.3. Overall quality of the evidence

In addition to the risk of bias calculations, we also conducted a formal assessment of the overall quality of the evidence included in the review and how this influenced the review's findings. For this purpose, we applied the GRADE strengths of the evidence assessment tool (**Rather insert reference**). Upon applying the GRADE framework, we established that the overall quality of the evidence included in our five syntheses was low. Only the evidence included in the meta-analysis on combined training and placement interventions was of moderate quality. All the other syntheses were based on either low-quality evidence (n=1) or very low-quality evidence (n=3). In summary, the small size and heterogeneous, low quality nature of the evidence base limit the findings of our systematic review. Findings from syntheses based on intervention groupings with low- to very low-quality evidence have to be treated with particular caution.

Figure 2 Overview of the risk of bias assessment
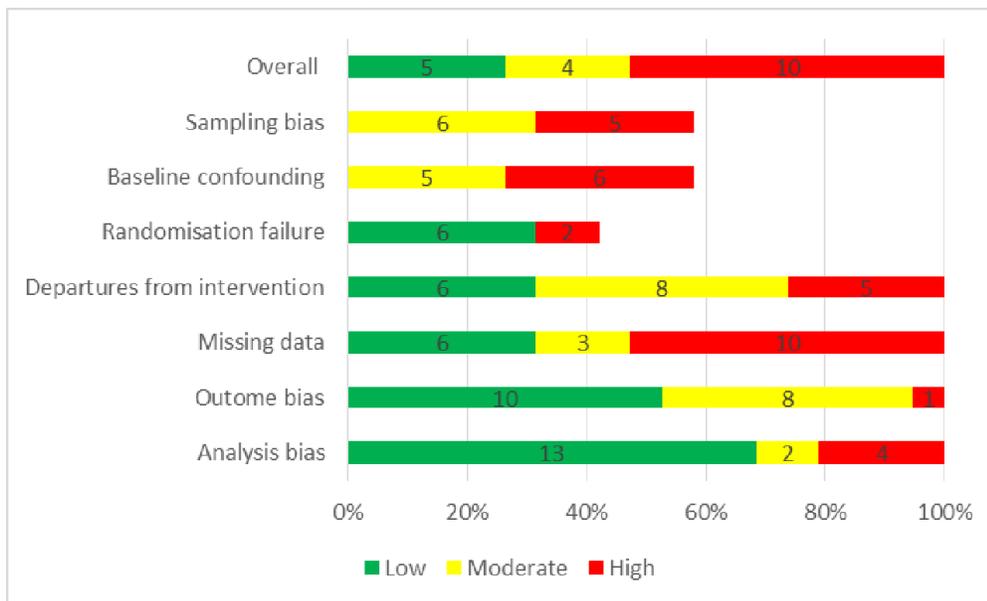
3.4. Synthesis Results

This section presents the results of our synthesis of what works to increase women's participation in the wage labour market in LMICs. The synthesis is based on the results of the 19 impact evaluations assessing the impact of labour market interventions included in the review. A total of 20 interventions reported in the 19 included studies were assessed for their effectiveness on women's wage labour,

income, and empowerment outcomes. The interventions were grouped into five homogeneous groups of programmes to facilitate a meaningful synthesis, namely; (i) Interventions combining training with job placement services (n=9) ; (ii) Interventions providing soft skills training only to address vertical occupational segregation (n=2); (iii) Interventions providing job placement services only (n=3); (iv) Interventions providing national labour subsidies (n=2); (v) Intervention altering changes to macro structures to support women's empowerment (n=2).

Statistical meta-analysis could only be conducted for the effects of interventions combining training with job placement services. Remaining intervention groupings were covered by too few studies paired with a lack of necessary information to calculate effect sizes. Effects of these interventions are thus reported using narrative synthesis based on structured summary of findings tables. In total, we therefore report one meta-analysis and four narrative syntheses for the abovementioned intervention groupings. From a methodological perspective, our meta-analysis on interventions combining training with job placement services presents the most rigorous synthesis of the evidence base as it draws on a significantly larger number of primary studies.

3.5. Effects of combining training and job placement service interventions

Nine studies were identified that evaluated the effects of combining training with job placement services as an intervention approach to enhance women's participation in wage labour markets in higher-growth/male-dominated sectors in LMICs. These interventions represented three types of labour market programmes and one outlier. First, three studies evaluated the effects of the Latin America Youth Training model Jovenes (Elías 2004; Galdo 2008; Ñopo 2007). The Jovenes programmes combine technical and vocational skills training with subsequent internships and job placements for youth in a range of Latin American countries. Second, three studies evaluated the effects of programmes conducted under the World Bank's Adolescent Girl's Initiative (AGI).[3] These were the Economic Empowerment of Adolescent Girls and Young Women (EPAG) programme in Liberia (Adoho 2014), the Adolescent Girls Employment Initiative (AGEI) programme in Nepal (Chakravarty 2016), and The Jordan New Opportunities for Women (Jordan NOW) programme (Groh 2012). Third, two studies evaluated the effects of Kenya's Youth Empowerment Program Ninaweza (de Azevedo 2013; Honorati 2015). The studies evaluated the impact of the programme in different sites and at different times, with the programme design varying between the evaluations. Lastly, Chen (2017) was an outlier, as the study evaluates the effects of the Barefoot Mechanics programme in India. Barefoot Mechanics is a programme designed by the Self-Employed Women's Association, an Indian NGO focused on the empowerment of women; it trains women to repair water pumps in rural areas and further guarantees them wage employment to conduct these repairs on behalf of the NGO. Notwithstanding this outlier, all in all, the included interventions present a homogeneous programme approach and design.

3.6. Combined effects of training and job placements on labour market outcomes

---

[3] http://www.worldbank.org/en/programs/adolescent-girls-initiative

Meta-analysis was conducted aimed at identifying the overall effects of the described training and job placement services on women's wage labour market outcomes in higher-growth/male-dominated sectors in LMICs. Considering that the true effect across studies was likely to differ due to various socio-economic backgrounds, intervention designs and other contextual factors, the approach followed was the random effect meta-analysis model.

Overall, we were able to calculate effect sizes for all studies except Ñopo (2007), which only reported frequencies and percentage change of employment rates without any measure of variance or significance being stated. The Galdo (2008) study is not included in this meta-analysis on employment outcomes as we were only able to calculate effect sizes for income outcomes, given a lack of statistical information on the disaggregated employment outcome for females. For the study by de Azevedo (2013), we calculated two effect sizes, as the study compared the effects of two different types of training + job placement services interventions: first, a training that combined ICT and life/soft skills followed by job placement services (T1); and second, a training that only provided ICT skills followed by job placement services, but no life/soft skills (T2).

The results of our meta-analysis of combined training and placement interventions are presented in Figure 3. Applying the GRADE framework, the evidence on the effects of these intervention on women's wage employment is of moderate quality (Appendix 1). The meta-analysis includes eight different interventions and a total of 10,207 participants. Overall, we find that training and placement interventions lead to a positive combined effect size of 0.159 (0.09, 0.23). This identified effect can be described as small to moderate and represents a 7.8% difference in changes in employment rates in favour of the women participants. The meta-analysis results are subject to a moderate degree of heterogeneity (Q=14.8; p= 0.04; i2=52.5%; tau2=0.005)[4], and the confidence intervals of all but one study overlap.

Figure 3 Meta-analysis of training and placement on wage labour outcomes

---

[4] We took an observational approach to uncover possible sensitivities that we then formally assessed statistically using a one-way random effects ANOVA model. The same process applies to all sensitivity analyses reported in this review, the full description of which can be found in the technical report.
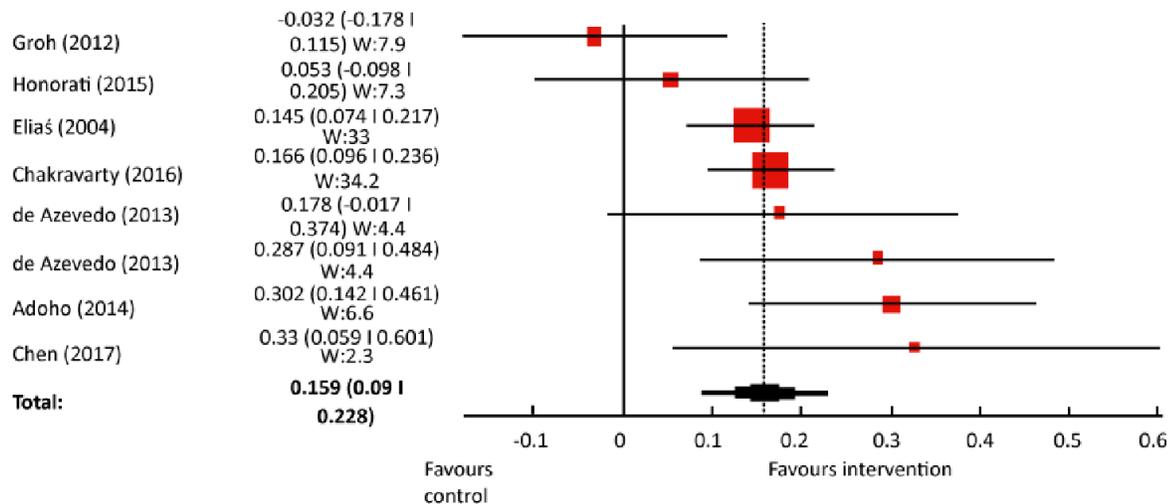
Heterogeneity: Q=14.8; p=0,0393; i²=52,5%; tau²= 0,0045; total participants: 10,207

To assess whether the observed overall effect might be driven by variables other than the applied labour market intervention, we ran sensitivity and moderator analyses.[5]
Our sensitivity analysis found that the effect sizes are not sensitive to the following variables:
• Applied study design
• Risk of bias
• Period of follow-up.

Since only one study did not use household survey data as an outcome measure of employment, we could not run a formal sensitivity analysis for this variable. However, on observation there does not seem to be a significant difference in results. In sum, we rule out the possibility that variances related to study design systematically influence the results of our meta-analysis.
Using our predefined list of potential moderator variables[6] (Langer et al 2017), we used the same structure as in the sensitivity analysis, based on an observational overview table followed by a one-way random effects ANOVA model, and found no moderator variable that systematically influenced the results of our meta-analysis. The moderators considered included:
• Intervention design
• Population characteristics
• Intervention setting.

Lastly, due to a lack of diversity in intervention designs, we could not assess whether the meta-analysis results were moderated by the type of implementing organisation (i.e. public vs private provider) or whether younger women experienced significantly different labour market effects from older women. In terms of the economic sector that the intervention targeted for women's wage labour employment, we observe that the effects of programmes in the ICT and electronics sector and the service sector showed the largest effects (g=0.195 and g=0.207). The finance and business administration sector

---

[5] The detailed results from these analyses are available in tabular format in the technical report.

[6] Two moderator categories were not reported in the included studies, and thus we were unable to assess the potential influence of these variables on the identified effects. These variables were the PROGRESS-plus categories (age, religion, social capital of female participants) and the UN categories of women (poor women vs extremely poor and non-poor).

showed the smallest effects (g=0.009 and g=0.135). We could not formally test these results as most interventions targeted employment in multiple higher-growth sectors simultaneously and the pooled effects per sector were therefore not based on independent effect sizes.

3.7. Combined effects of training and job placements on income

Following the meta-analysis on the effects of training combined with job placements on women's wage labour employment, we next investigated whether these employment outcomes translated into an increased income for women. The results of this meta-analysis are presented in the forest plot in Figure 4. Applying the GRADE framework, the evidence on the effects of combined training and placement interventions on women's income is of moderate quality (Appendix 1). This meta-analysis includes nine different interventions, due to the inclusion of the Galdo (2008) study and features a total of 12,084 participants. The analysis indicated a positive effect of training combined with job placements on women's income of 0.145 (0.07, 0.22). This pooled effect size is slightly smaller than the effect identified in the meta-analysis on employment outcomes (c.f. g=0.159) and can similarly be described as a small to moderate effect size. Expressed in terms of percentage change, the pooled effect indicated a 7.2% difference in changes in income in favour of the women taking part in combined training and placement interventions. The meta-analysis results on income are subject to a larger degree of heterogeneity (Q=27,9; p=0.0001; i2=71,4%; tau2=0.009), which is driven by the inclusion of the Galdo (2008) study. As per above, we assess the robustness of the identified combined effect through moderator and sensitivity analyses.

We assessed the sensitivity of our meta-analysis results against the same criteria relating to study design. Again, the meta-analysis results are not sensitive to the following variables:
• Applied study designs
• Period of follow-up
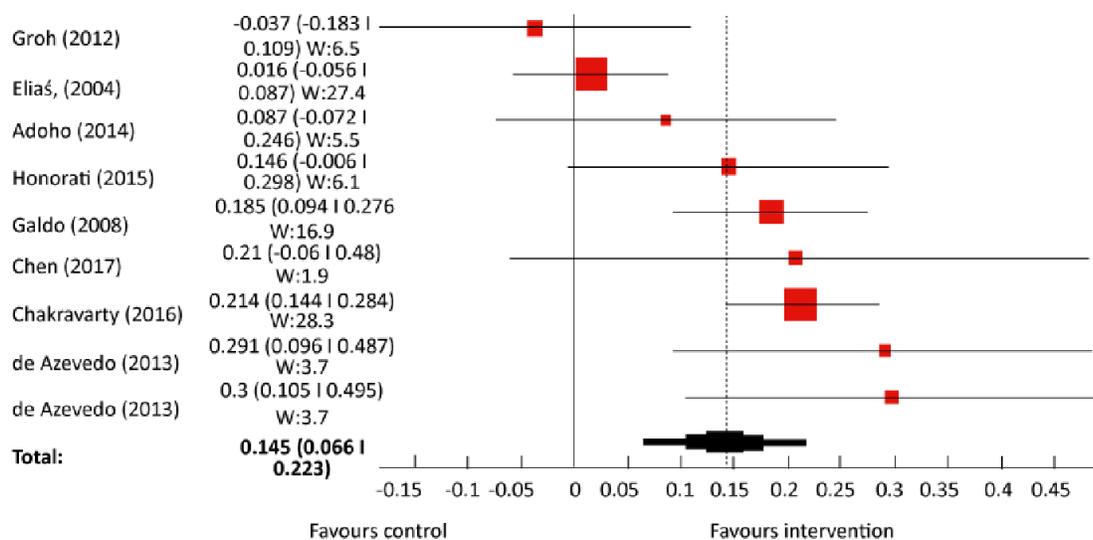• Applied outcome measure.

There was, however, a statistically significant difference in the effects of studies of different risks of bias ratings. Studies with a high risk of bias reported a significantly higher average effect (g=0.204) than studies with a low risk of bias (g=0.021). There was no statistically significant difference between studies with a low risk of bias and studies with a moderate risk of bias, and between studies of moderate and high risk of bias.

Our analysis of the effects of training and placement interventions on women's income finds no significant differences between the following moderator variables:
• Socio-economic setting
• Intervention design characteristics
• Programme design

Based on observation only, we find that the ICT and electronics, manufacturing and engineering sectors display larger than average income

Figure 4 Meta-analysis of training and placement on income

Groh (2012) -0.037 (-0.183 | 0.109) W:6.5
Eliaś, (2004) 0.016 (-0.056 | 0.087) W:27.4
Adoho (2014) 0.087 (-0.072 | 0.246) W:5.5
Honorati (2015) 0.146 (-0.006 | 0.298) W:6.1
Galdo (2008) 0.185 (0.094 | 0.276 W:16.9
Chen (2017) 0.21 (-0.06 | 0.48) W:1.9
Chakravarty (2016) 0.214 (0.144 | 0.284) W:28.3
de Azevedo (2013) 0.291 (0.096 | 0.487) W:3.7
de Azevedo (2013) 0.3 (0.105 | 0.495) W:3.7
Total: 0.145 (0.066 | 0.223)

Favours control   Favours intervention

Heterogeneity: Q=27,9; p=0,00049; i^2=71,4%; tau^2=0.009; total participants: 12,084

3.8. Combined effects of training and job placements on women's empowerment

Following the meta-analysis of the effects of training combined with job placement services on wage employment and income outcomes, we also investigated whether the observed positive changes in employment and income translated into women's empowerment. However, this was only measured as an outcome in five training + placement interventions: two AGI-modelled interventions, one Ninaweza programme, one Jovenes-modelled intervention, and the Barefoot mechanics programme.

Overall, the five studies included in this synthesis are subject to a range of biases: two studies each have a high and moderate risk of bias, and one study has a low risk of bias. Applying the GRADE framework, the overall quality of the evidence was rated to be of low quality. Table XX presents a structured breakdown of the empowerment related findings in the primary studies.

Empowerment was measured at an individual level using measures of economic empowerment[7] (n=5) (control over household spending) and psychological empowerment (n=2) (e.g. confidence levels, outlook on life). In all studies where this theory of change was observed, the intervention design included a specific focus on the individual empowerment of women in the intervention design. Three out of the four studies identified positive effects on measures of economic empowerment, while a single study found no effects.

The AGI intervention in Nepal, as well as the Barefoot Mechanics programme in India, led to women having greater control over spending of household financial resources (Chakravarty 2016; Chen 2017), while the Ninaweza programme in Kenya led to an increase in women's reported confidence and positive attitude towards looking for a job in the ICT sector (de Azevedo 2013). Only in Chakravarty (2016) were these observed improvements in economic empowerment accompanied by improvements in psychological empowerment.

All interventions found to be effective in increasing wage labour participation and women's income also increased the indicators of women's economic empowerment, while ineffective interventions had no effect on women's empowerment. This is the case in Groh's (2012) RCT on the effects of the AGI-

---

[7] 12 One study measured empowerment at a macro-economic level using the gender occupational segregation index.

modelled Jordan NOW programme: the programme failed to support female employment or increase income, which directly translated to non-significant effects on the economic and psychological empowerment of women.

Table 5.6 Narrative synthesis of training and placement on empowerment

Study	Programme	Context	Findings
Chakra varty (2016) Moder ate risk of bias	The Employment Fund and the Adolescent Girls Employment Initiative (Adolescent Girls Initiative) TVET training + Life skills training + Job placement	Asia: Nepal UN: Declining fertility, formalizing economies GAP: 0.661 (#110) GII: 0.497 (#115) Programme has positive effect on employment and income.	**Economic empowerment:** EF training participants report having more money of their own, more control over household spending, and more access to mentors who can advise them on work-related matters. **Psychological empowerment:** Strong gains on psychological empowerment, including significant increases in self-confidence both in life and with regard to entrepreneurial activities.
Chen (2017) High risk of bias	Barefoot mechanics program TVET training + job reservation	Asia: India UN: Declining fertility, formalising economies GAP: 0.683 (#87) GII: 0.530 (#125) Programme has positive effect on employment and income.	**Economic empowerment:** Program participants spend more on female-favoured consumption goods and have a greater say in household spending decisions.
de Azeved o (2013) High risk of bias	The Kenya Youth Empowerment Program: Ninaweza TVET training + Life skills training + Job placement	Africa: Kenya UN: High fertility agrarian societies GAP: 0.702 (#63) GII: 0.565 (135) Programme has positive effect on employment and income (not statistical significant).	**Subjective economic empowerment:** The RCT established that, among the young women who were not confident in their skill set, the Ninaweza program was successful in bolstering their confidence. Treatment group participants saw gains on the kinds of life skills items that pertained to workplace behavior and searching for a job.

Groh (2012) Low risk of bias	The Jordan New Opportunities for Women (Jordan NOW) (Adolescent Girls Initiative) Soft skill training + Wage subsidy voucher	Asia: Jordan UN category: declining fertility, formalising economies GAP index (#134): 0.603 GII index (#111): 0.478 Programme has no effect on employment and income.	**Economic empowerment:** We see that there is no significant impact on attitudes towards the role of women. Part of the reason for the lack of effect might be that there is already strong agreement on attitudes related to work – 97 percent think women should be allowed to work outside the home. **Psychological empowerment:** There is a significant and large negative interaction effect for receiving both treatments, so that graduates assigned to both treatments have no better current life evaluation than the control group.
Nopo (2007) Moderate risk of bias	Youth labor training (PROJOVEN) TVET training + Job placement	Latin America and Caribbean: Peru UN category: declining fertility, formalising economies GAP index (#80): 0.687 GII index (#86): 0.385 Programme has positive effect on employment and income (not statistical significant).	**Economic empowerment:** Eighteen months after participation in the program, gender occupational segregation reduces by 30 percent.

3.9. Cost analysis of training and job placements on labour market outcomes

Lastly, we investigated the extent to which the included studies assessed the cost and potential cost-effectiveness of the applied training and placement interventions. Unfortunately, information on cost and cost-effectiveness was not reported consistently. Out of the nine studies included in the meta-

analysis, only five reported data on intervention costs, and of those, only three conducted a formal cost-benefit analysis. We therefore cannot comment on the overall cost-effectiveness of interventions due to a lack of information reported in the included studies.

Table 2 Cost analysis of training and job placements on labour market outcomes

Study	Intervention cost	Cost-effectiveness
Adoho (2014)	USD 1,650 per participant for a 12-months programme.	**Not cost-effective** For the job skills training, it would take approximately 12 years to recoup the training costs, making it less cost-effective.
Elías (2004)	USD 2,000[8] per participant for an approximately 90-day program.	**Cost-effective** The programme's rates of return vary from 7% over five years to 17.5 % over a woman's life time.
Galdo (2008)	USD 514 per participant for a 3-month programme compromising 300 hours of classes and subsequent job placement.	**No cost-effectiveness calculations**
Groh (2012)	USD 400 per participants for a 9-day training course comprising of 45 hours of classes and a job voucher of USD 210 per month for a maximum of six months (USD 1260) resulting in a total programme cost of ≈ USD1660.	**No cost-effectiveness calculations**
Honorati (2015)	USD 1,150 per participant for a 6-month programme comprising a 2-week lifeskills training, 5-week business training, 5-week TVET training, and 12-week job placement.	**Cost-effective** These estimates show a KES 9,623 gain for females which implies that it would take about 10 months to offset the costs of the programme.

Table 1 presents an overview of the cost data available in the included studies. In the five studies that did report cost data, the average programme cost per participants was USD 1,143 (range: USD 400–2,000) but programme lengths and intensity varied greatly. In the three studies that conducted a formal cost-benefit analysis, two found the combined training and placement interventions to be cost-effective

---

[8] Study dates from 2004 but it is unclear what year the price data is based on.

(Elías 2004; Honorati 2015). These were the Peruvian Jovenes programme and the Ninaweza programme in Kenya. The evaluation of the AGI-modelled EPAG programme in Liberia, on the other hand, found that the programme was not cost-effective (Adoho 2014).

3.10. Effects of soft skills training to address vertical occupational segregation

We identified two interventions that provided soft skills training as the only intervention component to address vertical occupational segregation (Table 3) (Adhvaryu 2016; Macchiavello 2015).

Two interventions applying an RCT design with a moderate and high risk of bias respectively provided soft skills training as the only intervention component in order to address vertical occupational segregation (Adhvaryu 2016; Macchiavello 2015). The studies focused on female factory workers in the garment sector in India and Bangladesh, where the majority of supervisors were male, and consisted of soft skills programmes to enhance women's chances of being promoted to managerial posts.

Table 3 Narrative synthesis of soft skills training on addressing vertical segregation

Study	Programme	Context	Findings
Adhvaryu (2016) Moderate risk of bias	Personal Advancement and Career Enhancement (P.A.C.E.) programme Soft skills training (for career progression)	Asia: India UN: Declining fertility, formalizing economies GAP: 0.683 (#87) GII: 0.530 (#125 Majority of the work force in factories are females but a majority of line managers and supervisors are male.	**Career progression:** Treated workers are less likely to leave during the program, and exhibit substantially higher productivity up to nine months after program completion. This leads to being assigned to more complex tasks and a greater likelihood of promotion. **Empowerment:** First, treatment workers exhibit greater acquisition and use of information; second, treatment workers were more likely to be saving for children's education; third, increase in self-regard.

	Management skills training (for career progression)	Asia: Bangladesh UN category: declining fertility, formalising economies	**Career progression:**
Macchi avello (2015) High risk of bias	Soft skills training (for career progression)	GAP index (#72): 0.698 GII index (#119): 0.520 80% of the work force in factories are females but 95% of line managers and supervisors are male.	When the trainees are deployed in supervisory roles, production line workers initially judge females to be significantly less effective, and there is some evidence that the lines on which they work underperform. But after around four months of exposure, both perceptions and performance of female supervisors catch up to those of males. **Empowerment:** Prior to the training, we find that workers at all level of the factory believe males are more effective supervisors than females. We document evidence that the exposure to female supervisors changes the expectations of male production workers with regard to promotion and expected tenure in the factory.

Adhvaryu (2016) evaluated a training programme for female production line workers, focusing on a variety of life skills (e.g. communication, time management), while Macchiavello (2015) reported on a training programme for sewing machine operators, focusing on the skills necessary to become line supervisors (e.g. production planning and technical knowledge).

Three commonalities in the findings of each study exist. First, both programmes identified a positive effect on women's promotion to managerial posts, achieved through increased productivity and retention of female workers (which then led to their promotion). Second, both studies changed both men's and women's perceptions of females in managerial positions through exposure to women in management positions, which in the short term created conflict with men (Macchiavello 2015) but ultimately resulted in women reporting an increased sense of self-efficacy, accessing a greater range of professional development opportunities, and investing more in the education of their children (Adhvaryu 2016). Finally, both studies found that investment in females to assume managerial positions was beneficial to the factories themselves: it led to increased productivity and retention rates (Adhvaryu 2016) and in the long term to a more diverse managerial base (Macchiavello 2015).[9]

3.11.   Effects of job placement services only

---

[9] Questions about whether entering a low-value sector is a rational choice for some women due to a greater chance of promotion were not assessed in the review.

Three studies investigated the effects of placement services as the only intervention component (Groh 2012; Groh 2014; Jensen (2010). These covered heterogeneous types of placement services. As part of the Jordan New Opportunities for Women (Jordan NOW) labour market interventions, Groh (2012) provided women with job vouchers (valid for six months only, equal to the minimum wage of 150 JD / USD 210 per month) that they could take to a firm while searching for jobs. In a follow-up experiment under the Jordan NOW programme, Groh (2014) also tested the effectiveness of a screening and matching service to pair employers with prospective employees. Lastly, Jensen (2010) provided recruiting services for employment opportunities in the business outsourcing industry in India, where women in rural communities were visited by recruitment agents over the course of three years and informed about employment opportunities and how to access them.

In terms of the underlying programme mechanisms, Jensen (2010) and Groh (2014) both attempted to overcome a lack of information and matching between employers and prospective employees. Groh (2012), in contrast, attempted to overcome a barrier in the price and rigidity of women's labour.

The identified effects of the three job placement programmes varied greatly. First, Groh's (2014) screening and matching service was found to be ineffective and did not influence any of the three assumed outcomes. Women were held back by a strong reservation prestige and rejected the vast majority of matched employment opportunities. However, Jensen's (2010) evaluation of addressing informational and matching barriers through recruitment services found significant positive effects on women's employment and empowerment. Three years after the intervention commenced, women were 2.4% more likely to be employed in the business outsourcing industry, and this increase in wage employment further translated into increased investment in girl children in the rural communities. Third, Groh's (2012) wage vouchers were found to have large, short-term effects on women's employment in higher-growth sectors such as finance and business administration which had disappeared by four months after the vouchers had expired; the long-term employment gains were not significant. Within Groh's voucher experiment were two additional interventions arms: (i) a soft skills programme and (ii) vouchers combined with soft skills. The vouchers-only intervention design outperformed soft skills training and a combination of vouchers and soft skills; it is unclear why these results are observed.

Table 5 Narrative synthesis on job placement services only

Study	Programme	Context	Findings
Groh (2012) Low risk of bias	The Jordan New Opportunities for Women (Jordan NOW) (Adolescent Girls Initiative) Soft skill training + Wage subsidy voucher	Asia: Jordan UN category: declining fertility, formalising economies GAP index (#134): 0.603 GII index (#111): 0.478	**Employment:** The analysis finds that the job voucher led to a 40% increase in employment in the short- run, but that most of this employment is not formal, and that the average effect is much smaller and no longer statistically significant 4 months after the voucher period has ended. **Income:** The job voucher group earns 64 JD more per month than the control group at midline. However, by the endline the difference has

			fallen to only 6 JD per month and is not statistically significant. **Empowerment:** We see that there is no significant impact on attitudes towards the role of women. Further, there is a significant and large negative interaction effect for receiving both treatments, so that graduates assigned to both treatments have no better current life evaluation than the control group.
Groh (2014) Low risk of bias	The Jordan New Opportunities for Women (Jordan NOW) (Adolescent Girls Initiative) Search Frictions & Matching Services	Asia: Jordan UN category: declining fertility, formalising economies GAP index (#134): 0.603 GII index (#111): 0.478	**Employment:** No effects **Income:** No effects **Empowerment:** No effects
Jensen (2010) Low risk of bias	Business outsourcing recruiting services Job search assistance and placement support	Asia: India UN category: declining fertility, formalising economies GAP index (#87): 0.683 GII index (#125): 0.530	**Employment:** In villages that received the recruiting treatment, paid employment was 2.4 percentage points higher for women aged 18–24. **Income:** No effects **Empowerment:** Girls aged 5-15 in villages that received the recruiting services were 6 to 7 percentage points more likely to be in school and experienced an increase in Body Mass Index, reflecting greater nutrition and/or medical care.

3.12. Effects of national labour subsidies

Only two studies, using a retrospective regression design and rated as having a high risk of bias, evaluated the effects of national labour subsidies on women's wage labour participation in higher-growth/male-dominated sectors (Ayhan 2013; Broecke 2013), both of which were implemented in the

context of macro-economic crises[10]. In Turkey, this was a national labour stimulus referred to as the 'Employment Package Law', while in Tunisia, it was the labour law enacting the SIVP15 programme. The employment package can be described as a national-level demand-side labour subsidy, while Tunisia's SIVP programme[11] is best described as a mix between demand- and supply-side labour subsidies.

Table 6 Narrative synthesis on national labour subsidies

Study	Programme	Context	Findings
Ayhan (2013) High risk of bias	Labour law: Employment Package Law No. 5763 2008. Demand-side macro-economic subsidy (reduction in employers' social security contributions)	Europe: Turkey UN: Ageing societies GAP: 0.623 (#130) GII: 0.328 (#69) Labour market intervention necessitated by financial crisis of 2008/09	**Employment:** The results suggest a positive effect of the reduction in non-wage costs on employment creation for the targeted group (women). The probability of being hired for a woman aged 30 to 34 increased by 3%.
Broeke (2013) High risk of bias	Labour law: The Stage d'Initiation à la Vie Professionelle (SIVP) programme Supply & Demand-side macro-economic subsidy (employees' salary paid by government & reduction in employers' social security contributions)	Africa: Tunisia UN: Ageing societies GAP: 0.636 (#126) GII: 0.289 (# 58) Labour market intervention necessitated by persistent high graduate unemployment.	**Employment:** The analysis shows clearly that SIVP beneficiaries have lower joblessness and unemployment rates, and that they are much more likely to be hired in the private sector.

Both studies identified positive effects on women's wage labour participation in higher-growth/male-dominated sectors following the introduction of the labour subsidies. Ayhan (2013) reported that per

---

[10] The 2008/09 global financial crisis leading to a contraction of employment in Turkey (Ayhan 2013) and persistent high graduate unemployment in Tunisia (Broecke 2013).
[11] French: The Stage d'Initiation à la Vie Professionelle.

quarter over a two-year period, women were 1.4% – 3% more likely to be hired in the industry and construction sector than men. Broecke's (2013) finding overlaps with this: the study established that women's employment in professions related to the natural sciences and engineering increased by 11% following the introduction of the graduate subsidy.

The studies recommend that in future design of the subsidy programmes those graduates most at risk of unemployment should be targeted, the subsidy should be combined with other labour market interventions such as training (Broecke 2013), and clear governance structures to be in place (Ayhan 2013).

3.13.   Effects of macro-level policies targeting women's empowerment

Two interventions aimed to alter macro structures to support women's labour market participation and empowerment through a change in state-level implementations of political reservations for women in India (Ghani 2014) and a revision of the national family law in Ethiopia (Hallward-Driemeier 2013).

Table 7 Narrative synthesis of macro-level women empowerment policies

Study	Programme	Context	Findings
Ghani (2014) High risk of bias	National law: Political reservations for women in India	Asia: India UN category: declining fertility, formalising economies GAP index (#87): 0.683 GII index (#125): 0.530	**Employment:** No effects on wage labour employment: While overall employment of women in manufacturing does not increase after the reforms, there is significant evidence that more women-owned establishments were created in the unorganized/informal sector. **Empowerment:** Reserved leadership positions can influence the allocation of local resources in line with a greater general provision of infrastructure and public goods towards women. A second candidate explanation is that the political reservations inspired women to start their own businesses by providing positive role models and nurturing aspiration.

	National Law: Family Code	Africa: Ethiopia	**Employment:**
Hallward-Driemeier (2013) Moderate risk of bias	Economic rights (making employment decisions without the consent of the husband) Increased property rights for women	UN category: declining fertility, formalising economies GAP index (#109): 0.662 GII index (#116): 0.499	A strong effect (15-24%) in increasing women's share in occupations that are non-home based, paid work, year-round employment, and those with average higher educational requirement. **Empowerment:** The revised family law increased the bargaining power of women and increased the age of marriage from 15 to 18.

In India, the 73rd and 74th Constitutional Amendment Acts legislated a large-scale devolution and decentralisation of power to local government bodies, amongst which a third of all seats at each local governance level were reserved for women. In Ethiopia, the 2000 revision of the national family law legislated spouses' shared decision making regarding marital property as well as removing restrictions that prevented females from working outside the home.

Both interventions were evaluated using retrospective regression designs that relied on longitudinal survey and labour market data and were judged to have a moderate (Hallward-Driemeier 2013) and high risk of bias (Ghani 2014) respectively.

Both studies identified positive long-term effects on women's participation in the economy; only the change in family laws (Hallward-Driemeier 2013) led to an increase of women's employment in higher-skills sectors. The political quota (Ghani 2014) led to an increase in women's ownership of businesses in higher-growth sectors, but not in female wage employment.

3.14.   Qualitative comparative analysis

Our systematic review applied narrative synthesis and QCA in order to identify the design features of interventions aiming to support women's wage labour participation in higher-growth/male-dominated sectors in LMICs. Given the heterogeneity of the included interventions in our review, we limit our investigation of intervention design features to a homogeneous sub-set of studies: combined training and job placement interventions. Zooming in on this sub-set of women's wage labour interventions, the narrative synthesis led to the identification of seven design features reported in the included studies namely: Labour market; enhancing norms; Demand-led Gender-sensitive; Soft/Life skills; Participant profiling; Provider governance; Flexibility & iteration.

Following the identification of these seven intervention design features associated with programme effectiveness, we further attempted to unpack the specific configurations of these design features and their correlation with programme effects. The subsequent attempted QCA, however, was inconclusive and we therefore cannot comment on specific configurations of design features. All in all, this leaves us to conclude that the seven individual intervention design features constitute the active ingredients of combined training and placement programmes to support wage labour participation in higher-

growth/male-dominated sectors in LMICs. Further research is required to assess the specific configurations and combination of individual design features and their resultant effects.

4. Conclusion

This systematic review provided an evidence map and structured synthesis on the effects of interventions aiming to support women's participation in wage labour in highergrowth/male-dominated sectors in LMICs. We identified a small evidence base that was heterogeneous in terms of the applied labour market interventions and of low quality in terms of the methodological trustworthiness of the studies and consistency of effects.

Appendix

appendix 6. quality appraisal tool

Quality domain	Quality criterion	QUALITY APPRAISAL *For each criterion, rate the study 0 (does not meet), .5 (partially meets), or 1 (fully meets)*	BIAS ASSESSMENT *For each quality domain, describe any bias considerations*
Quality domain	**Quality criterion**	**Guidance and criteria for informing judgements**	**Within-study assessment**
1. RIGOR IN SAMPLING	Q1a. Sampling strategy described	What locations are sampled? What organizations are contacted to help recruit respondents? Or what other strategies are used to recruit respondents? What types of respondents were sampled and what kind of information/perspective are they expected to contribute?	**B1. Bias considerations include** (look first at what authors indicate as limitations and then add on based on purpose and scope of overall review): Individual Characteristics - Gender - Age - Marital status - Child age/status Community Characteristics - Country (ies) - Other key characteristics (prevailing norms) Work Characteristics - Sector - Type/Nature of Work - Workplaces
	Q1b. Sampling strategy justified	Why were specific locations chosen? Why were specific types of respondents chosen? Why was the specific sampling strategy used (i.e. what are the strengths of the sampling strategy)?	
	Q1c. Sample characteristics presented	Are relevant characteristics of the sample presented in a table or in the narrative?	
2. RIGOR IN DATA COLLECTION	Q2a. Steps to strengthen rigor in data collection described	Was data collection comprehensive, flexible, and/or sensitive enough to provide a complete and/or vivid and rich description of people's perspectives and experiences (e.g. did the	**B2. Bias considerations include** (look first at what authors indicate as limitations and then add on based on purpose and scope of overall review):

		researchers spend sufficient time at the site/with participants? did they keep 'following up'? Was more than one method of data collection used?	- ways of building trust - ways of addressing power differentials - ways of addressing language and culture
	Q2b. Steps taken to ensure consent and participation	Were steps taken to ensure that all participants were able and willing to contribute (e.g. processes for consent – language barriers, power relations between adults and children/young people?	
3. RIGOR IN DATA ANALYSIS	Q3a. Evidence of careful analysis methods presented (e.g., using multiple coders, validation methods, qualitative software, or discussions of data validity)	Is there evidence of careful analysis methods? At a minimum, there should be a discussion of what the analysis method is and the steps that were taken to do the analysis. Note that not all methods require multiple coders. For example, ethnographic studies are often conducted by one person and involve 'thick description'. The highest rating should include a specific discussion of the steps that were taken to ensure the dependability, confirmability, and validity of the data and results.	B3. Bias considerations include (look first at what authors indicate as limitations and then add on based on purpose and scope of overall review): - focus of analysis - exploration of diverse perspectives - guiding principles of interpretation
	Q3b. Balance of researcher preconceptions and emergent themes	Was the analysis balanced in the extent to which it was guided by preconceptions or by the data?	
	Q3c. Effort to rule out alternative explanations	Did the analysis seek to rule out alternative explanations for findings (in qualitative research this could be done by e.g. searching for negative cases/exceptions, feeding back preliminary results to participants, asking a colleague	

		to review the data, or reflexivity?	
4. FINDINGS SUPPORTED BY THE DATA	Q4a. Findings clearly connected with evidence (e.g., direct quotes or detailed descriptions of observations, not just opinion)	Are claims supported with specific examples, quotes, or other detailed evidence rather than just opinion? Is it clear what evidence was used to support significant findings?	**B4. Bias considerations include** (look first at what authors indicate as limitations and then add on based on purpose and scope of overall review): - Authors' indications of study limitations - Reviewers' assessments of additional limitations summarizing from above factors
	Q4b. Study limitations and generalizability described	Does the study point out its limitations and the populations to which it could be reasonably generalized? Is there discussion of any limitations, weaknesses, biases, or challenges with the data, analysis, conclusions, or populations to which conclusions can be generalized?	
5. BREADTH AND DEPTH OF FINDINGS	*Incorporated into criteria above*	Consider whether (it may be helpful to consider 'breadth' as the extent of description and 'depth' as the extent to which data has been transformed/analysed): - **a range of issues are covered** - **the perspectives of participants are fully explored in terms of breadth (contrast of two or more perspectives) and depth (insight into a single perspective)** - **richness and complexity have been portrayed (e.g.**	*Incorporated into criteria above*

		variation explained, meanings illuminated) - **there has been theoretical/conceptual development**	
6. PRIVILEGING PARTICIPANT PERSPECTIVES/ EXPERIENCES		Consider: - **whether there was a balance between open-ended and fixed response questions** - **whether participants were involved in designing the research** - **whether there was a balance between the use of an a priori coding framework and induction in the analysis** - **the position of the researchers (did they consider it important to listen to the perspectives of children?)** - **whether steps were taken to ensure confidentiality and put young people at ease.**	**B6. Bias considerations include** (in addition to assessments above): - balance between open-ended and closed-ended questions - participant involvement in research design

Source: Adapted from Bangpan, M., Dickson, K., Felix, L. and Chiumento, A. (2017). *The Impact of Mental Health and Psychosocial Support Interventions on People Affected by Humanitarian Emergencies: A Systematic Review*. Oxford: Oxfam GB.

Note: Grey cells indicate criteria that were rated in the first phase of quality appraisal, and other cells were added in the second phase full appraisal of final studies selected for inclusion.

Discussion

Conclusion