

This article was downloaded by: [115.85.25.194]

On: 31 March 2015, At: 20:27

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



[Click for updates](#)

Journal of Development Effectiveness

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/rjde20>

Meta-analysis of the impact of microcredit on women's control over household decisions: methodological issues and substantive findings

Maren Duvendack^a, Richard Palmer-Jones^a & Jos Vaessen^b

^a School of International Development, University of East Anglia, Norwich, UK

^b Department of Foundations and Methods of Law, Maastricht University, Maastricht, The Netherlands

Published online: 04 Apr 2014.

To cite this article: Maren Duvendack, Richard Palmer-Jones & Jos Vaessen (2014) Meta-analysis of the impact of microcredit on women's control over household decisions: methodological issues and substantive findings, *Journal of Development Effectiveness*, 6:2, 73-96, DOI: [10.1080/19439342.2014.903289](https://doi.org/10.1080/19439342.2014.903289)

To link to this article: <http://dx.doi.org/10.1080/19439342.2014.903289>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &

Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Meta-analysis of the impact of microcredit on women's control over household decisions: methodological issues and substantive findings

Maren Duvendack^{a*}, Richard Palmer-Jones^a and Jos Vaessen^b

^a*School of International Development, University of East Anglia, Norwich, UK;* ^b*Department of Foundations and Methods of Law, Maastricht University, Maastricht, The Netherlands*

Systematic reviews and meta-analysis have risen in popularity in international development to provide evidence on 'what works'. This paper reports the findings of a meta-analysis to assess the impact of microcredit on women's control over household spending to illustrate the challenges of conducting meta-analysis in the case of a diverse evidence base. We provide an assessment of methodological quality and present the findings of a meta-analysis. The results suggest that the effect sizes are small. Furthermore, the confidence that we can place in these findings is limited by the high level of heterogeneity within and between studies and the general reliance on non-experimental studies and statistical analyses which are not reported in sufficient detail to enable confident judgement as to their robustness.

Keywords: meta-analysis; microcredit; women's empowerment

Introduction

This paper reflects our experience with meta-analysis of the impacts of microcredit, an intervention that has been very fashionable for more than three decades, on women's empowerment. Microcredit has largely been targeted on women for intrinsic as well as instrumental reasons. The intrinsic purpose is purported to be achieved by, among other pathways, enabling women to participate in markets, gaining income and bargaining power in the family, exposure to modernising tendencies, and participation in supportive collective action. These experiences are supposed to be empowering, leading to better outcomes for the women themselves, and their families (Kabeer 2001a, 2001b). Many studies have been conducted on the impacts of microcredit, many of which purport to lend empirical support to these claims, but there have been strongly argued contrary indications as well and no consensus has emerged. Meta-analysis, including meta-regression, can synthesise multiple studies, and, to some extent, analyse characteristics of the literature, including publication bias (Higgins and Green 2011). However, meta-analysis and meta-regression need to be carefully conducted and interpreted (Higgins and Thompson 2004), and are not always suitable (Thompson and Higgins 2002).

Since the late 1990s, policy-makers have become increasingly interested in learning 'what works' in social and economic policy, often focusing on quantitative studies which attribute outcomes to particular interventions. The medical model of generating and synthesising evidence is seen as successful and now forms the basis for assessing the evidence base in international development. The research designs of many studies assessing the impact of major development interventions often employ quasi-experimental designs involving the use of sophisticated statistical techniques to establish quantitative

*Corresponding author. Email: m.duvendack@uea.ac.uk

attribution. Hence, research synthesis on ‘what works’ using systematic reviews and meta-analysis have risen in popularity in international development. This drive towards adapting tools from the medical sciences such as systematic reviews and meta-analysis has led to debates among researchers revolving around the constitution of methodological quality, the ability of statistical methods to control for biases and the assessment of risk of biases.

While many systematic reviews on major development interventions have now been conducted and much has been written about the practical challenges of conducting systematic reviews in international development (Mallett et al. 2012), less has been written about the challenges of conducting meta-analysis in contexts where studies are conceptually and methodologically highly diverse as is often the case in international development research.

Meta-analysis is based on the claim that conclusions of a range of studies on a given topic can be enhanced by ‘the statistical combination of results from two or more separate studies’ (Green and Higgins 2011, 97). Since Smith and Glass (1977) meta-analysis has been increasingly used to systematically synthesise research evidence mainly in the medical sciences where many studies are similar in terms of context, intervention and outcomes (for example, randomised control trials (RCTs) assessing the impact of the same intervention addressing the same condition in similar populations, using clearly defined outcomes measured by simple effect size statistics as a measure of impact). Many researchers argue that comparability on a conceptual level, that is, in terms of constructs, relationships and statistical methods, is a prerequisite for a successful meta-analysis. Furthermore, methodologically low quality studies should be excluded from meta-analysis as they can affect the overall findings, thus it is essential to first assess the methodological quality of studies before proceeding with meta-analysis. However, some systematic reviews and meta-analyses include diverse studies, suggesting an ‘apples and oranges’ problem (Lipsey and Wilson 2001; Sharpe 1997), and include studies which would be assessed as of relatively poor quality compared to well-conducted RCTs. In international development, studies are highly diverse with a range of research designs, interventions and/or outcomes, and would be assessed as of poor quality. Another problem is publication bias, which derives from bias in the availability of studies for synthesis due to research and publication practices (Stanley and Jarrell 1989).

In this paper, we report the findings of a meta-analysis that assesses the impact of microcredit on women’s control over household spending.¹ We illustrate the difficulties of conducting meta-analysis when faced with a diverse evidence base, and make an estimate of impact and publication bias. Apart from Yang and Stanley (2012) who conduct meta-analysis to investigate the impact of microcredit on income extracting effect sizes from Duvendack et al. (2011) and Stewart et al. (2012), no meta-analysis has been conducted investigating microcredit.

Duvendack et al. (2011) and Stewart et al. (2010, 2012) illustrate the diversity of interventions and outcome measures in microfinance studies. They also show that much of the quantitative evidence on microfinance impact is based on quasi-experimental² designs with few convincing RCTs. Quasi-experimental designs are highly susceptible to selection bias caused by non-random assignment to treatment and control groups and require the application of sophisticated statistical techniques to address selection bias issues. However, doubts exist about the ability of these techniques to account for selection due to observables as well as unobservables, and their success heavily depends on the quality of the model specifications and data used (Leamer 1983). Quasi-experimental designs are more reliant on theoretical assumptions which are harder to assess objectively compared to experimental designs. As a result, the inclusion of quasi-experimental designs in

meta-analysis is controversial and the assessment of their methodological quality prior to inclusion is extremely important (Deeks et al. 2003).

We first provide a background to microcredit and women's empowerment. We briefly describe the systematic review process with a focus on assessing methodological quality and conducting meta-analysis. We then present the meta-analysis and a discussion of its findings. The final section concludes.

Background

Microcredit has been heralded by many as the silver bullet to empower women. Supposedly, it enables the poor to access low cost credit, from which they were previously excluded, to undertake profitable entrepreneurialism. However, the impact of microcredit has become a highly contested arena, with critical reviews tending to controvert earlier claims that there existed good evidence of highly beneficent impacts (Bateman 2010; Roy 2010; Sinclair 2012).

There have been many reviews of microcredit and its effects (for example, Hulme and Mosley 1996; Goldberg 2005; Odell 2010), including three recent systematic reviews on microcredit in sub-Saharan Africa (Stewart et al. 2010) and microcredit in general (Duvendack et al. 2011; Stewart et al. 2012). Overall, these reviews suggest that the effects of microcredit on women's empowerment are at best mixed and that there are doubts about the methodological quality of the evidence base. The mixed results related to women's empowerment can partly be explained by the heterogeneity of microcredit interventions, contexts and target groups. Kabeer (2001a) argues that the divergence of results between different impact studies is in large part due to differences in interventions and in conceptualisation and measurement of empowerment. Regarding the latter, the literature on microcredit conceives and measures empowerment in such diverse ways as participation in household decision-making (for example, Mizan 1993; Kabeer 2001a), control over assets (for example, Goetz and Sen Gupta 1996), women's physical mobility, political and legal awareness and participation in public protests and political campaigns (for example, Hashemi, Schuler, and Riley 1996; Zaman 1998; Pitt, Khandker, and Cartwright 2006) and knowledge of accounting practices (for example, Ackerly 1995). The diversity in conceptualisation of the construct of empowerment makes it more difficult to generate generalisable findings through a systematic review.

The literature also conceives multiple potential causal pathways between microcredit and empowerment. A commonly studied causal relationship is the potential link between microcredit as a resource and the opportunity for women receiving loans to act more independently from men. Credit, and the income that it may generate, potentially strengthens the bargaining position of a woman vis-à-vis a man within the household and provides a basis for gradual changes in the power balance within the household, the social status of the woman within the household and the community, and ultimately gender relations. Thus, bargaining power or decision-making power (often studied in the context of decisions on household expenditure) is a key intermediate variable which lies at the basis of many conceptualisations of empowerment and its empirical operationalisation. Decision-making may have intrinsic as well as instrumental value, as it may both manifest and enhance self-esteem; hence, decision-making can be a direct indicator of empowerment, although a distinction can be made between decisions which are conventionally seen as the 'responsibility' of women, and those that are 'self-indulgent' and hence perhaps indicative of greater empowerment (Basu and Koolwal 2005). This is one of the main reasons why our study focuses on women's control over household expenditure.³

In terms of methodological quality, microcredit impact studies often struggle to convincingly overcome threats to internal and external validity (see Duvendack et al. 2011). The few RCTs conducted in the context of microcredit interventions that cover aspects of women's empowerment (for example, Kim et al. 2007; Banerjee et al. 2009) are no exception. The majority of studies examining the relationship between microcredit and women's empowerment (including women's control over household spending) use quasi-experimental regression-based approaches. Quasi-experiments and even more so regression-based analyses have been criticised regarding their limitations in addressing selection bias issues. Selection bias is likely to arise in microcredit programmes targeting the poor, as it may be expected that women who choose to participate in the programme are on average more empowered than those who do not (Goldberg 2005). When assessing the evidence base, it is important to distinguish studies with a high degree of internal validity (in terms of being able to attribute findings to microcredit controlling for other factors) from studies that are weaker in this regard. Keeping these qualifications in mind, in this review we extract quantitative effect size information on variables that relate to women's control over household spending without critiquing the construction of the metric of control, and attempt a meta-analysis.

Approach to quantitative synthesis

We adapted standard systematic reviews procedures and developed a protocol setting out inclusion and exclusion criteria, a search process and a synthesis methodology. We included studies with participants from poor, lower and upper-middle income countries as defined by the World Bank. Only studies which examined microcredit and women's control over household spending, or relevant proxies, as the main outcome variables, were included. In addition, we only included quantitative studies which had statistical controls and/or a comparison group.

Next, numerous academic databases, web-based search engines and microfinance-related websites were searched. We initially identified 310 papers that were selected for full text analysis. Many of these were non-experimental studies not reporting estimates of quantitative impacts and were screened out; 113 studies remained of which 56 were within the scope of this systematic review; that is, they focus on the relationship between microcredit and women's control over household spending. Finally, a methodological assessment was conducted which brought the number of studies down to 29 which were included in this review. While Vaessen et al. (forthcoming) report on qualitative synthesis (as well as further details of the systematic reviews process), we report only on the quantitative synthesis.

The quantitative synthesis employed meta-analysis. As argued above, meta-analysis should be conducted only when studies can be meaningfully compared on a conceptual level as well as in their statistical approach (Lipsey and Wilson 2001). Petticrew and Roberts (2006) stress that:

meta-analysis should only be applied when a series of studies has been identified for review that address an identical conceptual hypothesis. (205, box 6.13)

The question then arises as to what 'identical' means and how identity can be established. A particular case arises when, although addressing an 'identical' conceptual hypothesis that microfinance empowers women, some studies treat intermediate or

indicator variables as outcome variables. This is particularly likely for empowerment since it is a latent (and contested) concept which is not directly observable.⁴

The mainstream literature and application of quantitative research synthesis assume that both treatment and outcome in different studies are equivalent in the relevant way, and generally fairly standardised in the field; they also employ common and fairly simple analytical methods of analysis and reporting. However, it is not uncommon that measures of outcomes in different studies differ and there are many suggestions as to how to combine studies which use different outcome indicators. Alternatively, estimates of effect are reported separately for different metrics, and/or different interventions, by these categories assuming they correspond to the same construct of the ‘true’ outcome (Hedges and Olkin 1985; Cooper and Hedges 1994). This means that one is dealing with strictly comparable treatments and outcomes, or with comparable treatments with multiple indicators of outcomes. We also may have results which can be considered ‘intermediate’ or ‘instrumental’ in attaining the true outcome. Thus some studies compare treatment A with ‘intermediate’ ‘outcomes’ B, and they, or others compare B with ‘final’ outcomes C. It may be possible to indirectly link A with C through their common links with B.

There are also cases where treatments may be considered conceptually identical but superficially differ. Thus, Raudenbusch (2009, 296) refers to ‘definitions of treatment’ as a characteristic of studies which might account for why effects vary. In the present case, we are faced with not only diverse (definitions and metrics of) outcomes and diverse (definitions and or metrics of) treatments, but also with diverse research designs, methods of analysis, and reported statistics parameters and statistics. A sanguine view of this situation might argue the case for pooling and extensive sub-group analysis or meta-regression. However, this would suggest a need for many studies with at least statistical homogeneity among some groups of estimations – this is not the case here.

Given the studies we are dealing with here are highly heterogeneous in terms of treatments, outcomes and research designs the ‘apples and oranges’ problem is likely to affect us as distinctly different studies are pooled in the same meta-analysis (Lipsey and Wilson 2001). For example, we find a high degree of heterogeneity of treatment. Treatment indicators can be dichotomous (membership, participation⁵), continuous (length of membership, number or amount of loans taken), and so on. We pool studies with membership and participation treatment indicators, in part because these terms are often used interchangeably; however, it is important to bear in mind that microcredit borrowers may or may not receive microcredit, and may or may not receive other dimensions of treatment such as group discussions, technical assistance or social support from peers or the microcredit institution. We do not include studies with treatment indicators which cannot be represented by a dichotomous membership variable, because they are few, and cannot be put on a comparable basis. The majority of treatment indicators could be treated as dichotomous (92%).

We further argue that studies with methodological flaws and/or low quality have no place in meta-analysis as they will have adverse effects on the overall findings (Slavin 1986). However, there is a debate among researchers as to what constitutes high methodological quality, and a lot depends on the judgement of the researchers and the study context. We found that tools for assessing the quality of studies relied heavily on ‘expert opinions’ as much as they do on ‘expert systems’ (for example, World Cancer Research Fund 1997). Other tools such as the Scottish Intercollegiate Guidelines Network – SIGN (n.d.) – relegate research studies characterised by the methods of the studies included in this review to the lowest three (out of eight) levels, largely because

there are no RCTs or ‘high quality’⁶ quasi-experimental studies. Under these circumstances, detailed assessment of risk of bias (for example, using the approach promoted by Higgins and Green (2011), as reported in Duvendack et al. (2011)), will be excessively time-consuming, since the evidence is often not clearly reported, and is in any case irrelevant since the research quality assessed using the simpler tool-combining research design and analytical method reported above, rule these studies out of pooling for meta-analysis.

When many studies of a topic have been undertaken, it is generally supposed that confidence in the conclusions can be enhanced by considering all the relevant analyses and pooling their results and this is what research synthesis aims to achieve (Chalmers, Hedges, and Cooper 2002); effect size calculations and meta-analysis are important ways to achieve research synthesis of quantitative analyses. Since our studies are of slightly different interventions, use different research methods, (and to a limited extent) assess different indicators of outcomes, and occur in different contexts (time, place, social group), all of which may be expected to influence the specific conclusions reached, there may well be considerable heterogeneity in the precise relationships between intervention, context and outcome variables. As mentioned above and as further argued below, the studies included in this systematic reviews are highly heterogeneous, and in the case, it is often recommended to conduct subgroup analysis. Thus we considered pooling the various studies using particular characteristics such as treatment indicators, method (which is combination of research design and analytical method) and others. However, as the number of included studies here is very small, subgroup analysis was rendered meaningless. Further to this, pooling the effect size calculated from such studies needs to be carefully examined because of the pervasive threats to internal and external validity and the dangers of compounding confounded estimates (Egger, Schneider, and Smith 1998; Petticrew and Roberts 2006; Shrier et al. 2007; Manchikanti et al. 2009). In fact, Shapiro (1994) ‘propose[s] that the meta-analysis of published non-experimental data should be abandoned’ (777).

In order to combine different analyses, it is first necessary to put estimates of impact on a common scale. Duvendack et al. (2012) describe the different types of effect sizes:

There are three types of effect size scales; *d*-values and *r*-values for continuous variables, and odds or risk ratios for categorical variables. *d*-values are differences in outcomes standardised by their variability measured by their standard deviation or a close approximation, and are consequently similar to *z*-scores. *R*-values are correlation coefficients, which range from -1 to $+1$. Odds and risk ratios are relevant to categorical data such as the occurrence of a disease, achievement of an educational status, or survival. Other common metrics of outcome are scales or ordinal values, which can be transformed into incidence density ratios with similar interpretation to risk ratios. (460)

Given the diversity of estimation methods found in the selected studies, we use 12 different effect size formulae; these and the challenges of calculating the different size types are set out in Duvendack et al. (2012). A final question that arises is whether and how to extract a single effect size when there are multiple estimates in a given study of impact of a given indicator of treatment on a given outcome variables. There appear to be four alternative approaches – to include all estimates, to drop some for which there can be a clear justification, to drop ‘outliers’ based on an arbitrary rule, for example, (observation which are more than ± 2 standard deviations of the weighted mean), or to model the diversity. Lipsey and Wilson (2001) recommend averaging effects, but whether this is appropriate when effect sizes are heterogeneous remains unresolved.

Risk of bias assessment

Before conducting meta-analysis, we assess the methodological quality using the risk of bias tool developed by Duvendack et al. (2011).⁷ We used this scheme to categorise each of the studies by scoring their reported research design and analytical method; these scores were then combined into an index. A cut-off point of 2 was applied; for example, a study with a score of 2 and above was considered to have high threats to validity. Studies with scores of less than 2 had lower threats to validity.

Studies are ranked on two dimensions – research design and method of statistical analysis – from low threat to validity (1) to high threat to validity (5 in the case of research design, and 3 for statistical analysis). As indicated in Table 1, more than half of the included studies had a score of 2 and above, that is, a high score, in each dimension, and can thus be considered to have high threats to validity. Ideally we would have excluded those studies from further analysis and synthesis but this would have left us with a very small number of studies for meta-analysis. The parameters of our tool are subjective and pragmatic as explained in Note 7.

Another risk of bias assessment tool developed by 3ie could be useful here since it also includes risks due to spill-overs/contamination and reporting bias in addition to assessing selection bias. We apply a modified version of the 3ie tool to the studies screened through the tool used in Table 1 (Duvendack et al. 2011). Table A1 provides details of the modified 3ie tool and concludes that the majority of studies included in this systematic review have a moderate to high risk of bias.

Meta-analysis results

This section presents an account of the effect size calculations, subsequent meta-analysis and analysis of publication bias of the outcomes that represent ‘women’s control over household spending’. Due to substantial gaps in available data for effect size calculation, we were only able to generate effect size estimates from 17 studies. Several studies have more than one treatment variable; from the 17 studies, 22 effect size estimates could be recovered where the treatment indicator is a binary variable representing membership or participation in micro-finance organisation. Following Lipsey and Wilson (2001), we only include one effect size

Table 1. Distribution of studies according to research design and statistical methods of analysis.

| | | Statistical methods of analysis | | |
|---|-----------------|---------------------------------|--------------|------------|
| | | IV,PSM,2SLS/LIML, DID, RD | Multivariate | Tabulation |
| Research design | Scores | 1 | 2 | 3 |
| RCT | 1 | | 4 | |
| Pipeline | 2 | 1 | 0 | 1 |
| Panel or before/after and with/without | 3 | 0 | 1 | 0 |
| Either before/after and with/ without | 4 | 5 | 14 | 3 |
| Natural experiment | 5 | 0 | 0 | 0 |
| | Legend | | High score | |
| | Low score | 11 | | 17 |
| | Medium score | 1 | | |

Downloaded by [115.85.25.194] at 20:27 31 March 2015

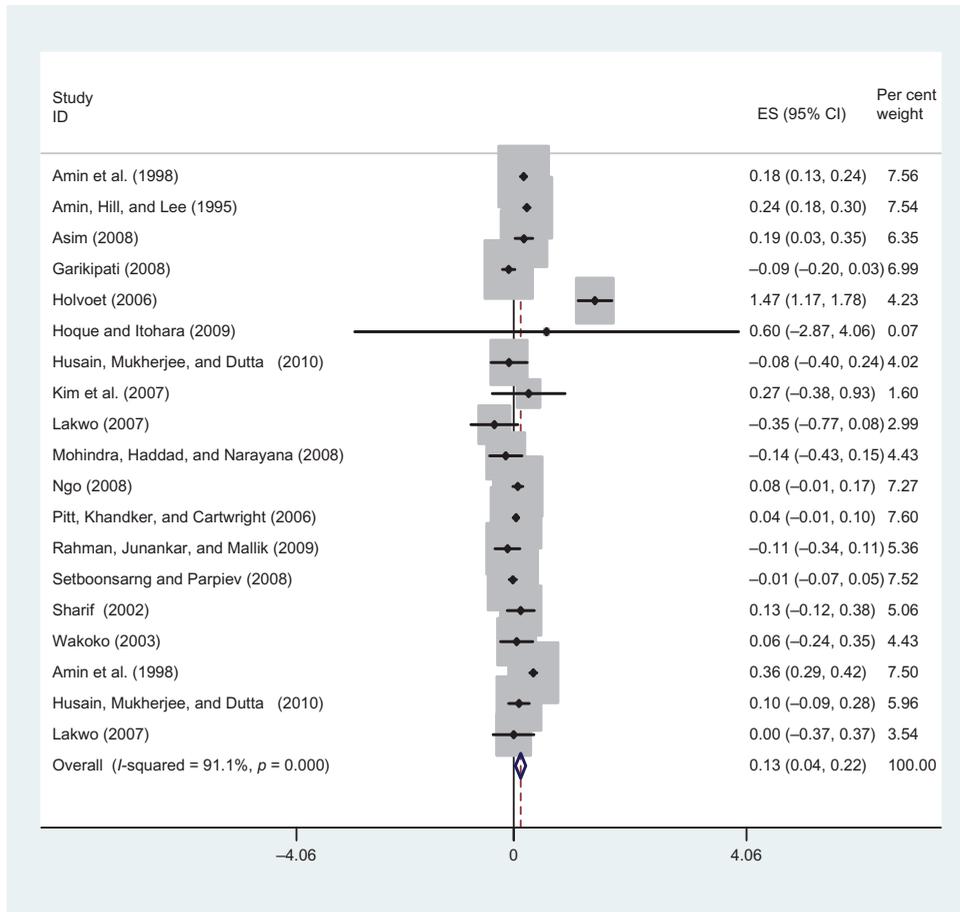


Figure 1. Forest plot of Effect of Women's control over household expenditures.
Note: Weights are from random effects analysis.

per combination of study and empowerment construct. Where multiple estimations are reported, we compute a weighted average using an inverse variance weight.

First we conducted a meta-analysis for effect size estimates using both means and regression based 'd' type effect sizes of continuous outcome variables with a dichotomous treatment variable denoting membership of a microcredit institutions.⁸ Of the 22 study/construct estimates extracted, only 19 were for a membership treatment indicator. Figure 1 shows the forest plot for all resource control outcomes, and Table 2 for the random effects model. This shows a high degree of heterogeneity which is confirmed by formally testing. The pooled effect size is positive and statistically significant (95% CI 0.035–0.222, $z = 2.69$, $p < 0.007$ from a random effects model), although small in size (< 0.2 , Cohen 1988) at 0.129, and with a high degree of heterogeneity ($\chi^2 = 123.11$ (d.f. 18) $p < 0.000$).⁹

Sensitivity analysis

All studies included in the meta-analysis are quasi-experimental with and without studies, where the treatment cases are due to independent interventions without any attempt at

Table 2. Meta-analysis of microcredit and women's control over household spending.

| Panel A | | | | |
|---------------------------------------|--------|----------------------|-------|-----------------|
| Study | ES | [95% Conf. Interval] | | Per cent weight |
| Amin, Becker, and Bayes (1998) | 0.181 | 0.126 | 0.236 | 7.56 |
| Amin, Hill, and Li (1995) | 0.239 | 0.180 | 0.298 | 7.54 |
| Asim (2008) | 0.187 | 0.028 | 0.346 | 6.35 |
| Garikipati (2008) | -0.086 | -0.198 | 0.026 | 6.99 |
| Holvoet (2006) | 1.472 | 1.165 | 1.779 | 4.23 |
| Hoque and Itohara (2009) | 0.596 | -2.870 | 4.062 | 0.07 |
| Husain, Mukherjee, and Dutta (2010) | -0.081 | -0.405 | 0.244 | 4.02 |
| Kim et al. (2007) | 0.273 | -0.385 | 0.931 | 1.60 |
| Lakwo (2007) | -0.348 | -0.773 | 0.077 | 2.99 |
| Mohindra, Haddad, and Narayana (2008) | -0.139 | -0.431 | 0.153 | 4.43 |
| Ngo (2008) | 0.077 | -0.011 | 0.165 | 7.27 |
| Pitt, Khandker, and Cartwright (2006) | 0.045 | -0.006 | 0.095 | 7.60 |
| Rahman, Junankar, and Mallik (2009) | -0.110 | -0.335 | 0.115 | 5.36 |
| Setboonsarng and Parpiev (2008) | -0.013 | -0.074 | 0.048 | 7.52 |
| Sharif (2002) | 0.130 | -0.115 | 0.376 | 5.06 |
| Wakoko (2003) | 0.055 | -0.236 | 0.346 | 4.43 |
| Amin, Becker, and Bayes (1998) | 0.356 | 0.293 | 0.419 | 7.50 |
| Husain, Mukherjee, and Dutta (2010) | 0.098 | -0.087 | 0.283 | 5.96 |
| Lakwo (2007) | 0.001 | -0.367 | 0.368 | 3.54 |
| D + L pooled ES | 0.129 | 0.035 | 0.222 | 100.00 |

| Panel B | |
|--|--------------------------------|
| Heterogeneity chi-squared | 123.11 (d.f. = 18) $p = 0.000$ |
| I^2 -squared (variation in ES attributable to heterogeneity) | 91.1% |
| Estimate of between-study variance Tau-squared | 0.0293 |
| Test of ES = 0 | $z = 2.69, p = 0.007$ |

Note: There are two entries for some studies because of different treatment indicators within the study.

matching or randomisation. Kim et al. (2007) is the only exception; however, this study does not warrant treatment as a randomised experiment since it merely compares four matched pairs of villages with the treatment randomised to one village of each pair. The three other RCT studies (Karlan and Zinman 2010; Banerjee et al. 2009; Crepon et al. 2011) did not present sufficient information to compute effect sizes. Hence we cannot compare experimental with quasi- or non-experimental results.

However, it is clear that both the positive effect size and to some extent the heterogeneity are both driven by two particular studies (Holvoet 2006; Hoque and Itohara 2009), which also show large positive effect size. Hoque and Itohara (2009) report logit odds ratio coefficients which we transformed into logs, and adjusted to standardised mean

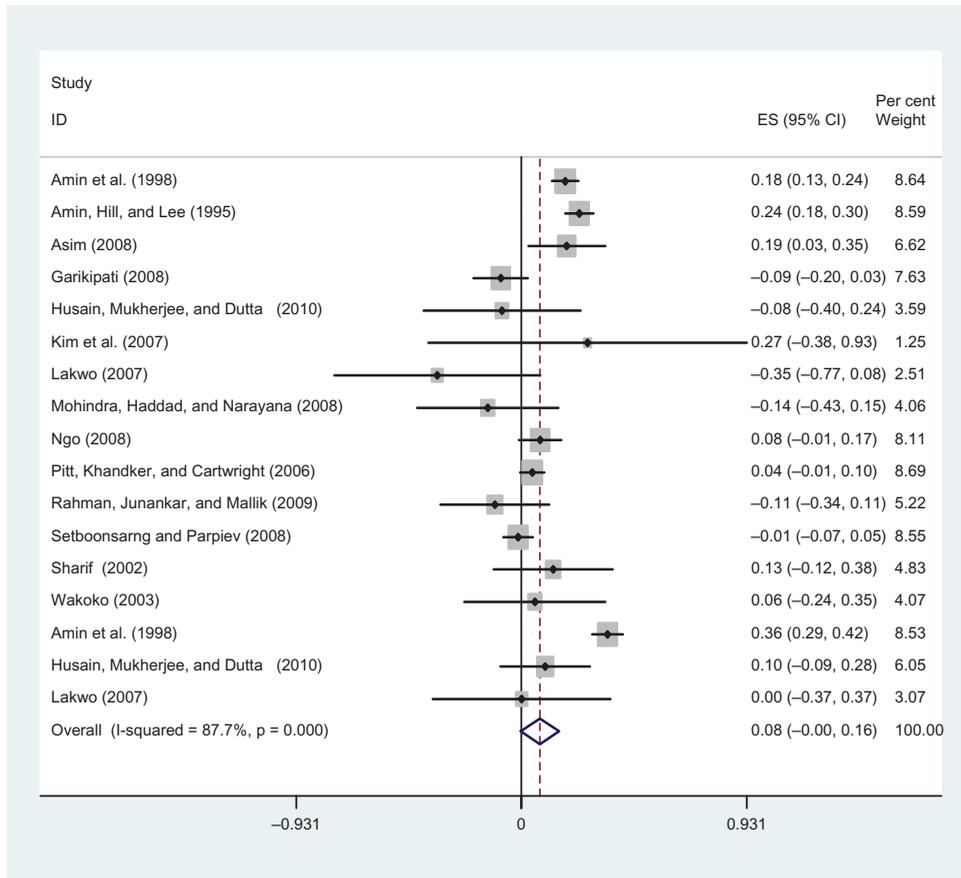


Figure 2. Forest plot of microcredit and women's control over household spending. Note: There are two entries for some studies because of different treatment indicators within the study. Weights are from random effects analysis.

difference following Chinn (2000); Holvoet (2006) seems to be an outlier. Hence, we next report the results dropping these two exceptional studies (Figure 2 and Table 3). These results indicate that the small positive effect size is not significant by conventional standards ($p \leq 0.06$), but heterogeneity remains large.

Bias in selection for extraction

We tested whether selection for extraction (inclusion in meta-analysis) was associated with a number of study characteristics including country, year of publication, year of fieldwork, type of microcredit product, individual or group lending, research design or method of analysis. In no case was there a statistically significant association of extraction with selection, except for publication in a peer-reviewed journal. Studies were more likely to be extracted if they were peer-reviewed. It is notable that for most of the characteristics describing the studies, there are many categories; this is testimony to the diversity of research approaches represented in these studies. We return to this point when we report on our meta-regressions of effect sizes.

Table 3. Meta-analysis of microcredit and women's control over household spending (excluding outliers).

| Panel A | | | | |
|---------------------------------------|--------|----------------------|-------|-----------------|
| Study | ES | [95% Conf. Interval] | | Per cent weight |
| Amin, Becker, and Bayes (1998) | 0.181 | 0.126 | 0.236 | 8.64 |
| Amin, Hill, and Li (1995) | 0.239 | 0.180 | 0.298 | 8.59 |
| Asim (2008) | 0.187 | 0.028 | 0.346 | 6.62 |
| Garikipati (2008) | -0.086 | -0.198 | 0.026 | 7.63 |
| Husain, Mukherjee, and Dutta (2010) | -0.081 | -0.405 | 0.244 | 3.59 |
| Kim et al. (2007) | 0.273 | -0.385 | 0.931 | 1.25 |
| Lakwo (2007) | -0.348 | -0.773 | 0.077 | 2.51 |
| Mohindra, Haddad, and Narayana (2008) | -0.139 | -0.431 | 0.153 | 4.06 |
| Ngo (2008) | 0.077 | -0.011 | 0.165 | 8.11 |
| Pitt, Khandker, and Cartwright (2006) | 0.045 | -0.006 | 0.095 | 8.69 |
| Rahman, Junankar, and Mallik (2009) | -0.110 | -0.335 | 0.115 | 5.22 |
| Setboonsarng and Parpiev (2008) | -0.013 | -0.074 | 0.048 | 8.55 |
| Sharif (2002) | 0.130 | -0.115 | 0.376 | 4.83 |
| Wakoko (2003) | 0.055 | -0.236 | 0.346 | 4.07 |
| Amin, Becker, and Bayes (1998) | 0.356 | 0.293 | 0.419 | 8.53 |
| Husain, Mukherjee, and Dutta (2010) | 0.098 | -0.087 | 0.283 | 6.05 |
| Lakwo (2007) | 0.001 | -0.367 | 0.368 | 3.07 |
| D + L pooled ES | 0.076 | -0.003 | 0.155 | 100.00 |

| Panel B | |
|--|--------------------------------|
| Heterogeneity chi-squared | 129.72 (d.f. = 16) $p = 0.003$ |
| I -squared (variation in ES attributable to heterogeneity) | 87.7% |
| Estimate of between-study variance Tau-squared | 0.0182 |
| Test of ES = 0 | $z = 1.88$ $p = 0.06$ |

Publication bias

Analysis using the Egger test (Egger et al. 1997) and a funnel plot for the full sample suggests the presence of significant publication bias but no small sample effects; however, the effect size estimated by Hoque and Itohara (2009) is small but of low precision. An Egger test of the effect size against its standard errors excluding outliers gives an insignificant negative intercept (-1.27 , $p \leq 0.287$), and significant positive slope ($p \leq 0.006$), again suggesting publication bias may be present. This is supported by the funnel plot (Figure 3).

Meta-analysis by risk of bias

In this section, we report tests of pooled effect size by risk of bias. Because we are now dealing with a smaller number of studies, we used a more elaborate procedure to classify

Table 4. Meta-analysis of effect of microcredit on women's control over household spending by risk of bias.

| Study | ES | [95% Conf. Interval] | | | Per cent Weight |
|---------------------------------------|-------------------------|----------------------|----------|---------------------|-----------------|
| Unclear risk of bias | | | | | |
| Amin, Becker, and Bayes (1998) | 0.181 | 0.126 | 0.236 | | 8.64 |
| Amin, Hill, and Li (1995) | 0.239 | 0.180 | 0.298 | | 8.59 |
| Garikipati (2008) | -0.086 | -0.198 | 0.026 | | 7.63 |
| Mohindra, Haddad, and Narayana (2008) | -0.139 | -0.431 | 0.153 | | 4.06 |
| Pitt, Khandker, and Cartwright (2006) | 0.045 | -0.006 | 0.095 | | 8.69 |
| Rahman, Junankar, and Mallik (2009) | -0.110 | -0.335 | 0.115 | | 5.22 |
| Amin, Becker, and Bayes (1998) | 0.356 | 0.293 | 0.419 | | 8.53 |
| Sub-total | | | | | |
| D + L pooled ES | 0.099 | -0.019 | 0.217 | | 51.35 |
| Moderate risk of bias | | | | | |
| Asim (2008) | 0.187 | 0.028 | 0.346 | | 6.62 |
| Husain, Mukherjee, and Dutta (2010) | -0.081 | -0.405 | 0.244 | | 3.59 |
| Kim et al. (2007) | 0.273 | -0.385 | 0.931 | | 1.25 |
| Lakwo (2007) | -0.348 | -0.773 | 0.077 | | 2.51 |
| Ngo (2008) | 0.077 | -0.011 | 0.165 | | 8.11 |
| Setboonsamg and Parpiev (2008) | -0.013 | -0.074 | 0.048 | | 8.55 |
| Sharif (2002) | 0.130 | -0.115 | 0.376 | | 4.83 |
| Wakoko (2003) | 0.055 | -0.236 | 0.346 | | 4.07 |
| Husain, Mukherjee, and Dutta (2010) | 0.098 | -0.087 | 0.283 | | 6.05 |
| Lakwo (2007) | 0.001 | -0.367 | 0.368 | | 3.07 |
| Subtotal | | | | | |
| D + L pooled ES | 0.049 | -0.015 | 0.114 | | 48.65 |
| Overall | | | | | |
| D + L pooled ES | 0.076 | -0.003 | 0.155 | | 100.00 |
| Test(s) of heterogeneity: | | | | | |
| | Heterogeneity statistic | Degrees of freedom | <i>p</i> | <i>I</i> -squared** | Tau-squared |
| Unclear | 93.00 | 6 | 0.000 | 93.5% | 0.0211 |
| Moderate | 12.02 | 9 | 0.212 | 25.1% | 0.0023 |
| Overall | 129.72 | 16 | 0.000 | 87.7% | 0.0182 |

Notes: Between group heterogeneity not calculated; only valid with inverse variance method: significance test(s) of ES = 0; unclear $z = 1.65$, $p = 0.100$; moderate $z = 1.50$, $p = 0.132$; overall $z = 1.88$, $p = 0.060$. There are two entries for some studies because of different treatment indicators within the study. ***I*-squared: the variation in ES attributable to heterogeneity).

publication practices which can be expected to be associated with differences in effect sizes and their variance. Some of the latter can be characterised as publication biases (Stanley and Jarrell 1989), which our results suggest are present. Nevertheless, meta-regression can be a way in which effect size estimates may be improved from studies with diverse context, outcome variables, intervention and quality, through the use of statistical controls reflecting these variables.

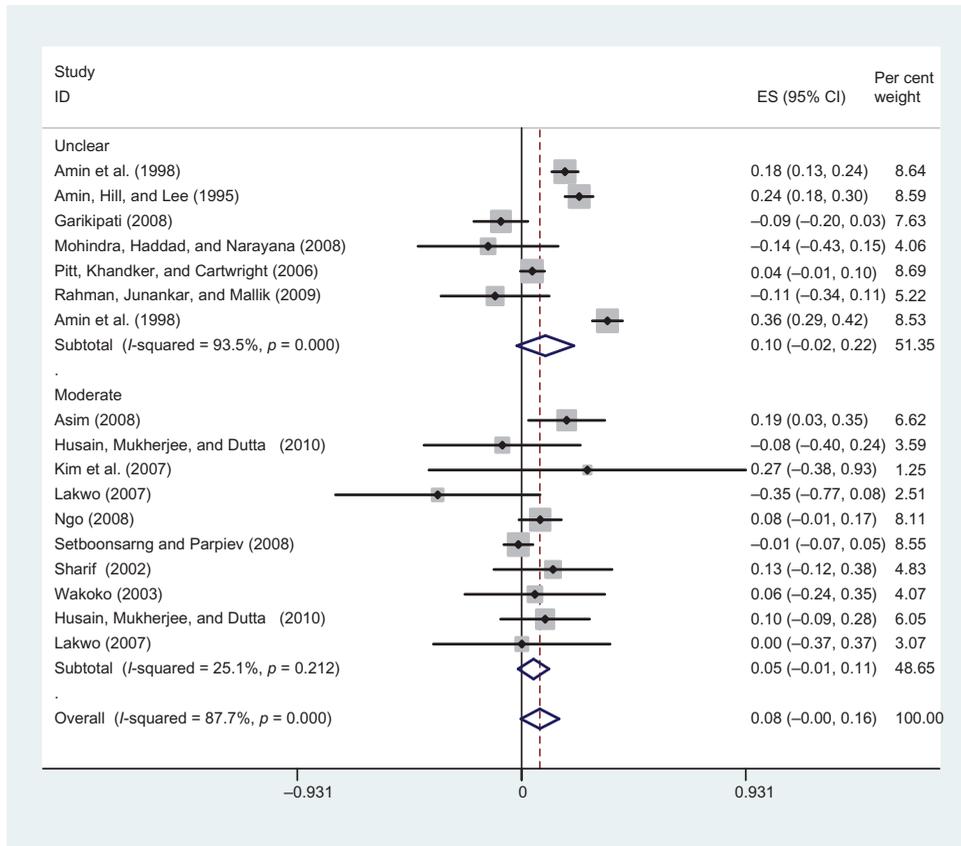


Figure 4. Forest plot of impacts of microcredit on women's empowerment by risk of bias. Note: Weights are from random effects analysis.

However, there is little general guidance as to what variables to include in meta-regression (cf. Stanley and Jarrell 1989; Stanley et al. 2013), and little consensus as to how useful it is to use meta-regression to control for diversity in the content, contexts and quality of studies. In our case, the small number of studies deemed worthy of extraction, the dangers of over-fitting, data-dredging and so on (Thompson and Higgins 2002), and what would be post-hoc specification of the meta-regression analysis leads to the distinct possibility of coming to spurious conclusions (Higgins and Thompson 2004), and it is unlikely that anything very meaningful would come from an attempted meta-regression. Furthermore, there are several sources of sources of diversity in our research topic and only a limited range of variables characterising the studies. The sources of diversity include the type of outcome variable (dimensions and indicators of empowerment), the type of intervention (broadly the extent of services in addition to microcredit and the specific microcredit contracts, of each intervention), the country, region and time, in which the study was conducted, and the quality of the research design and estimation strategies. Several of these variables have more than two categories (see Table 5). We do not, as a consequence, and in view of the limited space for explication, report meta-regression results even with the most obvious covariates.¹¹ The results of the meta-analysis suggest that the effect sizes are small, not attaining the lower end of the moderate

Table 5. Association of study extraction with study characteristics.

| Study characteristic | <i>N</i> | Rows | Columns | χ^2 | <i>p</i> |
|-----------------------------|----------|------|---------|----------|----------|
| Country | 29 | 8 | 2 | 5.31 | 0.62 |
| Year of publication | 29 | 16 | 2 | 13.72 | 0.54 |
| Year of data collection | 29 | 14 | 2 | 11.86 | 0.53 |
| Location type (rural/urban) | 29 | 4 | 2 | 1.88 | 0.59 |
| Microcredit product | 29 | 4 | 2 | .84 | 0.83 |
| Group lending | 29 | 3 | 2 | 1.39 | 0.49 |
| Research design | 29 | 8 | 2 | 8.38 | 0.30 |
| Statistical analysis | 29 | 17 | 2 | 13.63 | 0.62 |
| Peer-reviewed | 29 | 2 | 2 | 5.09 | 0.024 |

level suggested by Cohen (1988) (>0.2). Given the risks of bias in most studies, conducting a meta-analysis on all 29 studies is problematic. Analysis separating studies with moderate risk from the others showed that results are relatively consistent between the two sets of studies, which could be taken as an indication of a certain degree of robustness of the meta-analysis.

However, a high level of heterogeneity within and between studies (in terms of treatment indicators, outcome variables and analytical methods), and the general reliance on non-experimental studies and statistical analyses which are not reported in sufficient detail to enable confident judgement as to their robustness, limits the confidence that we can place in these findings. Quite a number of the studies come from works which have not been peer-reviewed, and our general assessment of vulnerability to bias is moderate or unclear.

Under these circumstances, meta-analysis risks inflating statistical significance by combining poor quality studies each of which at best yields only marginally significant results, and are vulnerable to unknown biases which are likely to inflate both effect size and confidence levels. While we show that there is some publication bias, we choose not to report attempts to control for this by meta-regression in view of the number of plausible controls it would be desirable to include to account for both this and substantive heterogeneity.

Conclusion

Over the past decades, microfinance activities have spread across the globe, reaching millions of poor households with tailored financial services. This growing importance has resulted in many research, evaluation and impact studies. Women's empowerment in relation to microcredit has been studied extensively; many of these studies have been carried out in the context of South Asia. The main objective of this study was to conduct a meta-analysis to assess the effects of microcredit on women's control over household spending in developing countries, which is an underlying key dimension of the more complex construct of empowerment. In the course of the meta-analysis, we came to conclusions that there were significant risks of over-interpretation when studies are highly heterogeneous, and there appeared to be publication bias.

We adjusted standard systematic review procedure as detailed in Vaessen et al. (forthcoming) and focused on the methodological assessment of the included studies. Our risk of bias assessment indicated that most of the included studies displayed serious methodological weaknesses, an aspect which inevitably compromised the synthesis phase. Thus,

the findings of the meta-analysis have to be treated with caution as it relied on quasi-experimental designs which have high levels of heterogeneity within and between studies and diverse conceptual framings. The meta-analysis suggests that the effect sizes are small, at best barely attaining the lower end of the moderate level suggested by Cohen (1988) (>0.2). Consequently, we conclude that the evidence on the effect of microcredit on women's control over household spending is weak, but that what evidence we have implies that any effect of microcredit on women's empowerment is also weak. As discussed earlier, women's control over household resources may constitute an important intermediary dimension in processes of women's empowerment, and/or a proxy for empowerment, but cannot be considered empowerment itself. Further, given the overall weak effect of microcredit on women's control over household resources, it is very unlikely that our evidence can be taken to mean microcredit has a meaningful and substantial impact on empowerment processes. While the latter type of impact may have occurred in particular contexts, for some women, overall the evidence suggests that this is not so for most. The model on which control of resources is associated with empowerment is based (broadly a bargaining model of the household) can itself be questioned (Lundberg and Pollak 2008), and none of the studies tested their model of empowerment against alternatives. These weaknesses militate against over-interpretation, and in favour of reporting our uncertainties rather than emphasising the substantive empirical results.

Our conclusions on the effects of microcredit are corroborated by Duwendack et al. (2011) and Stewart et al. (2010, 2012), who report (to a limited extent) on empowerment effects. The main difference between our review and the others (including Stewart et al. (2012)) lies in the scope of the review. The previous three systematic reviews were broader in scope than the present one, both in terms of interventions (microcredit, micro-savings and micro-leasing) as well as outcome measures. As a concluding remark, there appears to be a gap between the often optimistic (societal) belief in the capacity of microcredit to ameliorate the position of women in decision-making processes within the household on the one hand, and the empirical evidence base on the other hand, a gap which our meta-analysis should not be thought to have bridged.

Acknowledgements

The authors thank Hugh Waddington for extensive comments on the considerably longer working paper version of this paper. This paper is based on a forthcoming 3ie working paper.

Funding

The authors express many thanks to 3ie for funding this study.

Notes

1. This paper is based on the systematic review by Vaessen et al. (forthcoming).
2. For the purpose of this paper, quasi-experimental designs are those which have a control group that has not been assigned by randomisation by the investigator (for example, Shadish, Campbell, and Cook 2010). Randomisation by the investigator does not of course guarantee validity of the control group.
3. For a comprehensive discussion of the different dimensions underlying women's empowerment and their relation to microcredit, see Vaessen et al. (forthcoming).
4. Empowerment is a latent variable for which there can be presumed indicators such as responses to questions about control over resources; however, it is an assumption that this

construct is in fact a measure of empowerment. This construct may be validated by association with presumed causes of empowerment such as microcredit (for example, Schuler, Islam, and Rottach 2010, 842). For an insightful discussion of the different dimensions of empowerment and underlying causal linkages, see Kabeer (2001b). See also Vaessen et al. (forthcoming) for a more comprehensive discussion of the different conceptualisations of empowerment.

5. A member of a group may or may not participate in, for example, microcredit.
6. This does not mean that the studies do not contribute to knowledge in significant ways either substantively or methodologically, only that they do not have the qualities required for the type of cumulative contributions to understanding that can be legitimately achieved through meta-analysis.
7. Consensus on methods for assessing the quality of non-experimental studies has not been reached, and readily usable tools are not available. Deeks et al. (2003) identified nearly 200 tools, with much overlap, report that quality of studies was assessed in only 33 per cent of their sample of systematic reviews that included non-experimental studies, and that bias was found to be pervasive and significant. All the tools, and the GRADE system emanating from Cochrane Collaboration (Higgins and Green 2011, Sections 12.2 and 13.5.2) were burdensome, perhaps contributing to their limited use. We developed this tool for preliminary screening of a large – given our resources – number of studies of impacts of microfinance characterised by multitudinous methods the assessment of which using established risk of bias tools would have been impossible. The approach is in the spirit of Stanley and Jarrell (1989), who comment that ‘.. focusing ... on the relative merits of methods used across studies rather than the relative merits of the studies themselves ... largely avoids implicit valuation of researchers’ work and (by implication) their reputations’ (61). The abstracts, where they exist, of social science studies generally, and in international development in particular, are not informative as to the methods used or conducive to the use of these conventional tools. The whole paper may have to be read, often very carefully, and sometimes including other material on the original study, before a proper assessment of quality can be produced. This is very different to reports of modern studies in medicine, for example. Consequently, we built a database of the included studies containing information on study design, method of analysis, country, type of intervention, effect size calculation formula, measures of statistical significance as well as other relevant quantitative information. This database was used in the quality assessment and subsequent meta-analysis. The scale reported in Table 1 uses only the research design and estimation method. The risk of bias categories reported in Table A1 uses in addition the other variables.
8. The analysis is done in Stata 12 using the user-written ‘metan’, ‘metabias’ and ‘metafunnel’ commands.
9. The fixed effects model yielded a much larger and highly significant pooled effect size (0.136, $z = 11.63$, $p = 0.000$), but given the heterogeneity between studies, and their non-experimental method, it is not advised to use a fixed effects model which is a key cause of false positives when heterogeneity is present (Higgins and Thompson 2004, 1672).
10. The risks of bias categories, reported in Table A1, are not derived from the values reported in Table 1, which are objective but mechanical, and were produced to enable preliminary screening of a large number of studies. However, the scores reported in Table 1 is nearly significantly different between the moderate risk of bias category the unclear category ($p < 0.053$), which are the only categories of study present – that is, there are no low risk of bias studies according to the criteria reported in Table A1.
11. Concerned readers can approach the authors for the results we produced.

References

- Ackerly, B. A. 1995. “Testing the Tools of Development: Credit Programmes, Loan Involvement and Women’s Empowerment.” *IDS Bulletin* 26 (3): 56–68. doi:10.1111/j.1759-5436.1995.mp26003007.x.
- Amin, R., S. Becker, and A. Bayes. 1998. “NGO-Promoted Microcredit Programs and Women’s Empowerment in Rural Bangladesh: Quantitative and Qualitative Evidence.” *The Journal of Developing Areas* 32 (2): 221–237.

- Amin, R., R. B. Hill, and Y. Li. 1995. "Poor Women's Participation in Credit-Based Self-Employment: The Impact on Their Empowerment, Fertility, Contraceptive Use, and Fertility Desire in Rural Bangladesh." *The Pakistan Development Review* 34 (2): 93–119.
- Asim, S. 2008. "Evaluating the Impact of Microcredit on Women's Empowerment in Pakistan." CREB Working Paper 2–9, DECRG World Bank Group, Lahore School of Economics.
- Banerjee, A., E. Duflo, R. Glennerster, and C. Kinnan. 2009. *The Miracle of Microfinance? Evidence from a Randomized Evaluation*. Boston: Massachusetts Institute of Technology. Accessed August 27, 2013. <http://economics.mit.edu/files/4162>
- Basu, A. M., and G. B. Koolwal. 2005. "Two Concepts of Female Empowerment – Some Leads from DHS Data on Women's Status and Reproductive Health." Collected papers on gender using DHS data, ORC Macro, Calverton MD 15–33.
- Bateman, M. 2010. *Why Microfinance Doesn't Work? The Destructive Rise of Local Neoliberalism*. London: Zed Books.
- Chalmers, I., L. V. Hedges, and H. Cooper. 2002. "A Brief History of Research Synthesis." *Evaluation and the Health Professions* 25 (1): 12–37. doi:10.1177/0163278702025001003.
- Chinn, S. 2000. "A Simple Method for Converting an Odds Ratio to Effect Size for Use in Meta-Analysis." *Statistics in Medicine* 19 (22): 3127–3131. doi:10.1002/1097-0258(20001130)19:22<3127::AID-SIM784>3.0.CO;2-M.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioural Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum Associates.
- Cooper, H. M., and L. V. Hedges, eds. 1994. *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.
- Crépon, B., F. Devoto, E. Duflo, and W. Parienté. 2011. *Impact of Microcredit in Rural Areas of Morocco: Evidence from a Randomized Evaluation*. Boston: Massachusetts Institute of Technology. Accessed August 27, 2013. <http://economics.mit.edu/files/6659>
- Deeks, J. J., J. Dines, R. D'Armico, A. J. Sowden, C. Sakarovich, F. Song, M. Petticrew, and D. G. Altman. 2003. "Evaluating Non-Randomised Intervention Studies." *Health Technology Assessment* 7: 1–183.
- Duvendack, M., J. G. Hombrados, R. Palmer-Jones, and H. Waddington. 2012. "Assessing 'What Works' in International Development: Meta-Analysis for Sophisticated Dummies." *Journal of Development Effectiveness* 4 (3): 456–471. doi:10.1080/19439342.2012.710642.
- Duvendack, M., R. Palmer-Jones, J. G. Copestake, L. Hooper, Y. Loke, and N. Rao. 2011. *What is the Evidence of the Impact of Microfinance on the Well-being of Poor People?* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Egger, M., M. Schneider, and G. D. Smith. 1998. "Meta-Analysis Spurious Precision? Meta-Analysis of Observational Studies." *British Medical Journal* 316: 140–144. doi:10.1136/bmj.316.7125.140.
- Egger, M., G. D. Smith, M. Schneider, and C. Minder. 1997. "Bias in Meta-Analysis Detected by a Simple, Graphical Test." *British Medical Journal* 315 (7109): 629–634. doi:10.1136/bmj.315.7109.629.
- Garikipati, S. 2008. "The Impact of Lending to Women on Household Vulnerability and Women's Empowerment: Evidence from India." *World Development* 36 (12): 2620–2642. doi:10.1016/j.worlddev.2007.11.008.
- Goetz, A. M., and R. Sen Gupta. 1996. "Who Takes the Credit? Gender, Power and Control over Loan Use in Rural Credit Programs in Bangladesh." *World Development* 24 (1): 45–63. doi:10.1016/0305-750X(95)00124-U.
- Goldberg, N. 2005. *Measuring the Impact of Microfinance: Taking Stock of What We Know*. Washington, DC: Grameen Foundation.
- Green, S., and J. P. T. Higgins. 2011. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* (updated March 2011), edited by J. P. T. Higgins and S. Green. Chichester: The Cochrane Collaboration. Accessed August 27, 2013. <http://www.cochranehandbook.org>
- Hashemi, S. M., S. R. Schuler, and A. P. Riley. 1996. "Rural Credit Programs and Women's Empowerment in Bangladesh." *World Development* 24 (4): 635–653. doi:10.1016/0305-750X(95)00159-A.

- Hedges, L. V., and I. Olkin. 1985. *Statistical Methods for Meta-Analysis*. San Diego, CA: Academic Press. Accessed August 27, 2013. http://www.jameslindlibrary.org/trial_records/20th_Century/1980s/hedges/hedges-kp.pdf
- Higgins, J. P. T., and S. Green. 2011. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: The Cochrane Collaboration. Accessed August 27, 2013. <http://www.cochrane-handbook.org>
- Higgins, J. P. T., and S. G. Thompson. 2004. "Controlling the Risk of Spurious Findings from Meta-Regression." *Statistics in Medicine* 23: 1663–1682. doi:10.1002/sim.1752.
- Holvoet, N. 2005. "The Impact of Microfinance on Decision-Making Agency: Evidence from South India." *Development and Change* 36 (1): 75–102. doi:10.1111/j.0012-155X.2005.00403.x.
- Holvoet, N. 2006. "The Differential Impact on Gender Relations of 'Transformatory' and 'Instrumentalist' Women's Group Intermediation in Microfinance Schemes: A Case Study for Rural South India." *Journal of International Women's Studies* 7 (4): 36–50.
- Hoque, M., and Y. Itohara. 2009. "Women Empowerment through Participation in Micro-Credit Programme: A Case Study from Bangladesh." *Journal of Social Sciences* 5 (3): 244–250. doi:10.3844/jssp.2009.244.250.
- Hulme, D., and P. Mosley. 1996. *Finance Against Poverty*. London: Routledge.
- Husain, Z., D. Mukherjee, and M. Dutta. 2010. "SHG and Empowerment of Women: Self-Selection or Actual Benefits?" MPRA paper 20765.
- Jamal, H. 2008. *Exploring the Impact of Microfinance in Pakistan*. Research Report No. 77. Karachi: Social Policy and Development Centre. Accessed March 21, 2014. <http://www.spdc-pak.com/publications/Research%20Reports/RR77.pdf>
- Kabeer, N. 2001a. "Conflicts over Credit: Re-Evaluating the Empowerment Potential of Loans to Women in Rural Bangladesh." *World Development* 29 (1): 63–84. doi:10.1016/S0305-750X(00)00081-4.
- Kabeer, N. 2001b. "Resources, Agency, Achievements: Reflections on the Measurement of Women's Empowerment." In *Discussing Women's Empowerment – Theory and Practice*, edited by A. Sisask, Sida Studies No. 3. Stockholm: Novum Grafiska.
- Karlan, D., and J. Zinman. 2010. "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts." *Review of Financial Studies, Society for Financial Studies* 23 (1): 433–464.
- Kim, J. C., C. H. Watts, J. R. Hargreaves, L. X. Ndhlovu, G. Phetla, L. A. Morison, J. Busza, J. D. H. Porter, and P. Pronyk. 2007. "Understanding the Impact of a Microfinance-Based Intervention on Women's Empowerment and the Reduction of Intimate Partner Violence in South Africa." *American Journal of Public Health* 97 (10): 1794–1802. doi:10.2105/AJPH.2006.095521.
- Lakwo, A. 2007. *Microfinance, Rural Livelihoods, and Women's Empowerment in Uganda*. Research Report 85. African Studies Centre.
- Lastarria-Cornhiel, S., and Y. Shimamura. 2008. "Social Welfare Outcomes of Micro-credit in Malawi." Prepared for The World Bank Agriculture and Rural Development Department.
- Leamer, E. E. 1983. "Let's Take the Con out of Econometrics." *The American Economic Review* 73 (1): 1–43.
- Lipsey, M. W., and D. B. Wilson. 2001. *Practical Meta-Analysis*. Thousand Oaks, CA: Sage.
- Lundberg, S., and R. A. Pollak. 2008. "Family Decision Making." In *The New Palgrave Dictionary of Economics*, 2nd ed., edited by L. Blume, and S. Durlauf. New York: Palgrave Macmillan.
- Mallett, R., J. Hagen-Zanker, R. Slater, and M. Duvendack. 2012. "The Benefits and Challenges of Using Systematic Reviews in International Development Research." *Journal of Development Effectiveness* 4 (3): 445–455. doi:10.1080/19439342.2012.711342.
- Manchikanti, L., S. Datta, H. S. Smith, and J. A. Hirsch. 2009. "Evidence Based Medicine: Systematic Reviews and Guidelines in Interventional Pain Management: Part 6. Systematic Reviews and Meta-Analyses of Observational Studies." *Pain Physician* 12: 819–850.
- Mizan, A. N. 1993. "Women's Decision-making Power in Rural Bangladesh: A Case Study of Grameen." In *The Grameen Bank, Poverty Relief in Bangladesh*, edited by A. N. M. Wahid. Boulder, CO: WestviewPress.
- Mohindra, K., S. Haddad, and D. Narayana. 2008. "Can Microcredit Help Improve the Health of Poor Women? Some Findings from a Cross-Sectional Study in Kerala, India." *International Journal for Equity in Health* 7 (1): 2–14. doi:10.1186/1475-9276-7-2.

- Ngo, T. 2008. "Microfinance and Gender Empowerment in Kyrgyzstan." Prepared for the World Bank Agriculture and Rural Development Department (ARD).
- Odell, K. 2010. *Measuring the Impact of Microfinance: Taking Another Look*. Washington, DC: Grameen Foundation.
- Petticrew, M., and H. Roberts. 2006. *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford: Blackwell.
- Pitt, M., S. R. Khandker, and J. Cartwright. 2003. "Does Micro-Credit Empower Women? Evidence from Bangladesh." World Bank Policy Research Working Paper 2998.
- Pitt, M., S. R. Khandker, and J. Cartwright. 2006. "Empowering Women with Micro-Finance: Evidence from Bangladesh." *Economic development and Cultural Change* 54 (4): 791–831. doi:10.1086/503580.
- Rahman, S., P. N. Junankar, and G. Mallik. 2009. "Factors Influencing Women's Empowerment on Microcredit Borrowers: A Case Study in Bangladesh." *Journal of the Asia Pacific Economy* 14 (3): 287–303. doi:10.1080/13547860902975648.
- Raudenbusch, S. W. 2009. "Analyzing Effect Sizes: Random-Effect Models." In *The Handbook of Research Synthesis and Meta-Analysis*, edited by H. Cooper, L. V. Hedges, and J. C. Valentine, 295–316. New York: Russell Sage.
- Roy, A. 2010. *Poverty Capital: Microfinance and the Making of Development*. London: Routledge.
- Schuler, S. R., and S. M. Hashemi. 1994. "Credit Programs, Women's Empowerment, and Contraceptive Use in Rural Bangladesh." *Studies in Family Planning* 25 (2): 65–76. doi:10.2307/2138085.
- Schuler, S. R., F. Islam, and E. Rottach. 2010. "Women's Empowerment Revisited: A Case Study from Bangladesh." *Development in Practice* 20 (7): 840–854. doi:10.1080/09614524.2010.508108.
- SIGN (Scottish Intercollegiate Guidelines Network). n.d. "Guidelines Methodology, Appendix B." Accessed August 27, 2013. <http://www.sign.ac.uk/methodology/index.html>
- Setboonsarng, S., and Z. Parnpiet. 2008. "Microfinance and the Millennium Development Goals in Pakistan: Impact Assessment Using Propensity Score Matching." ADB Institute Discussion Paper 104.
- Shadish, W. R., D. T. Campbell, and T. D. Cook. 2010. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.
- Shapiro, S. 1994. "Meta-Analysis/Shmeta-Analysis." *American Journal of Epidemiology* 140 (9): 771–778.
- Sharif, N. R. 2002. "Poor Female Youth and Human Capital Development in Bangladesh: What Role for Micro-Credit Programmes?" In *Women and Credit, Researching the Past, Refiguring the Future*, edited by B. Lemire, R. Pearson, and G. Campbell. Oxford: Berg.
- Sharif, N. R. 2004. "Microcredit Programs and Women's Decision-Making Status: Further Evidence from Bangladesh." *Canadian Journal of Development Studies/Revue canadienne d'études du développement* 25 (3): 465–480. doi:10.1080/02255189.2004.9668989.
- Sharpe, D. 1997. "Of Apples and Oranges, File Drawers and Garbage: Why Validity Issues in Meta-Analysis Will Not Go Away." *Clinical Psychology Review* 17 (8): 881–901. doi:10.1016/S0272-7358(97)00056-1.
- Shrier, I., J.-F. Boivin, R. J. Steele, R. W. Platt, A. Furlan, R. Kakuma, J. Brophy, and M. Rossignol. 2007. "Should Meta-Analyses of Interventions Include Observational Studies in Addition to Randomized Controlled Trials? A Critical Examination of Underlying Principles." *American Journal of Epidemiology* 166: 1203–1209. doi:10.1093/aje/kwm189.
- Sinclair, H. 2012. *Confessions of a Microfinance Heretic: How Microlending Lost Its Way and Betrayed the Poor*. San Francisco, CA: Berrett-Koehler.
- Slavin, R. E. 1986. "Best-Evidence Synthesis: An Alternative to Meta-Analytic and Traditional Reviews." *Educational Researcher* 15 (9): 5–11. doi:10.3102/0013189X015009005.
- Smith, M. L., and G. V. Glass. 1977. "Meta-Analysis of Psychotherapy Outcome Studies." *American Psychologist* 32 (9): 752–760. doi:10.1037/0003-066X.32.9.752.
- Stanley, T. D., H. Doucouliagos, M. Giles, J. H. Heckemeyer, R. J. Johnston, P. Laroche, J. P. Nelson, et al. 2013. "Meta-Analysis of Economics Research Reporting Guidelines." *Journal of Economic Surveys* 27: 390–394. doi:10.1111/joes.12008.
- Stanley, T. D., and S. B. Jarrell. 1989. "Meta-Regression Analysis: A Quantitative Method of Literature Surveys." *Journal of Economic Surveys* 3: 161–170. doi:10.1111/j.1467-6419.1989.tb00064.x.

- Stewart, R., C. van Rooyen, K. Dickson, M. Majoro, and T. de Wet. 2010. *What is the Impact of Microfinance on Poor People? A Systematic Review of Evidence from Sub-Saharan Africa*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Stewart, R., C. van Rooyen, M. Korth, A. Chereni, N. Rebelo Da Silva, and T. de Wet. 2012. *Do Micro-Credit, Micro-Savings and Micro-Leasing Serve as Effective Financial Inclusion Interventions Enabling Poor People, and Especially Women, to Engage in Meaningful Economic Opportunities in Low- and Middle-Income Countries – A Systematic Review of the Evidence*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Thompson, S. G., and J. P. T. Higgins. 2002. "How Should Meta-Regression Analyses be Undertaken and Interpreted?" *Statistics in Medicine* 21: 1559–1573. doi:10.1002/sim.1187.
- Vaessen, J., A. Rivas, M. Duvendack, R. Palmer Jones, F. L. Leeuw, R. Lukach, N. Holvoet, and J. Bastiaensen. forthcoming. "The Effects of Microcredit on Women's Control over Household Spending in Developing Countries." Forthcoming as a 3ie Working Paper.
- Wakoko, F. 2003. "Microfinance and Women's Empowerment in Uganda: A Socioeconomic Approach." PhD Thesis, Ohio State University.
- WCRF (World Cancer Research Fund). 1997. *Food, Nutrition and the Prevention of Cancer: A Global Perspective*. Washington, DC: American Institute for Cancer Research.
- Yang, M., and T. Stanley. 2012. "Micro-Credit and Income: A Literature Review and Meta-Analysis." Conway, AR: Hendrix College. Accessed August 27, 2013. http://www.hendrix.edu/uploaded/Files/Departments_and_Programs/Business_and_Economics/AMAES/Micro-Credit%20Meta-Analysis12-18%282%29.pdf
- Zaman, H. 1998. "The Links Between BRAC Input and Empowerment Correlates." Working Paper 28, BRAC-ICDDR,B Joint Research Project Dhaka, Bangladesh.
- Zaman, H. 1999. *Assessing the Poverty and Vulnerability Impact of Micro-Credit in Bangladesh: A case study of BRAC*. Washington, DC: World Bank.

Appendix. Risk of bias assessment

Table A1. Risk of bias assessment.

| Study | Selection bias and confounding | | | | | | | | | | Overall risk of bias assessment | | | | | |
|--|--------------------------------|-----|-----|--------------|--------------|-------------------------------|-------|-------------------------------|-------------------|------|---------------------------------|-----|----|-----|-------|----------|
| | Experimental approaches | | | | | Quasi-experimental approaches | | | | | | | | | | |
| | PSM | IV | OLS | Logit/probit | Multivariate | Panel | Other | Spill-overs and contamination | Outcome reporting | RCTs | | PSM | IV | OLS | Other | |
| Amin, Hill, and Li (1995) | | | Uc | | | | | | | Yes | No | No | No | Uc | Uc | Unclear |
| Amin, Becker, and Bayes (1998) | | | | Uc | Uc | | | Yes | No | No | No | No | No | Uc | Uc | Unclear |
| Asim (2008) | Yes | | | | | | | Yes | No | No | No | No | No | Uc | Uc | Unclear |
| Banerjee et al. (2009) | | Yes | | | | | | Yes | No | No | No | No | No | Uc | Uc | Moderate |
| Crepon et al. (2011) | | | | | | | | Yes | No | No | No | No | No | Uc | Uc | Unclear |
| Garikipati, Hashemi, Schuler, and Riley (1996) | | | | | | | Uc | | No | No | No | No | No | Uc | Uc | Unclear |
| Holvoet, (2005) | | | | | | | Uc | | Yes | No | No | No | No | Uc | Uc | Moderate |
| Holvoet (2006) | | | | | Uc | | | Yes | Uc | Uc | Uc | Uc | Uc | Uc | Uc | Moderate |
| Hoque and Itohara (2009) | | | | | | | | Yes | Uc | Uc | Uc | Uc | No | No | Uc | Moderate |

| | | | | | | | | | | | |
|---|-----|--|-----|----|----|----|--|----|--|----|----------|
| Husain, Mukherjee, and Dutta (2010) | Yes | | Yes | Uc | | Uc | | Uc | | Uc | Moderate |
| Jamal (2008) | | | Uc | No | | | | | | Uc | Moderate |
| Karlan and Zinman (2010) | | | Yes | No | No | | | | | Uc | Unclear |
| Kim et al. (2007) | | | Yes | No | | | | | | Uc | Moderate |
| Lakwo (2007) | Yes | | Yes | Uc | | | | Uc | | Uc | Moderate |
| Lastarria-Cornhiel and Shimamura (2008) | Yes | | Yes | No | No | | | No | | Uc | Moderate |
| Mizan (1993) | Uc | | Yes | No | | | | No | | Uc | Unclear |
| Mohindra, Haddad, and Narayana (2008) | Uc | | Yes | No | | | | No | | No | Unclear |
| Ngo (2008) | | | Yes | No | | | | | | Uc | Unclear |
| Pitt, Khandker, and Cartwright (2003) | Uc | | Yes | Uc | | | | No | | Uc | Moderate |
| Pitt, Khandker, and Cartwright (2006) | Uc | | Yes | No | | | | | | Uc | Unclear |
| Rahman, Junankar, and Mallik (2009) | | | Yes | No | | | | | | No | Unclear |

(Continued)

Table A1. (Continued)

| Study | Selection bias and confounding | | | | | | | | | | Overall risk of bias assessment | | | | | |
|--------------------------------|--------------------------------|----|-----|--------------|--------------|-------------------------------|-------|-------------------------------|-------------------|------|---------------------------------|--------------------|-----|----|-----|----------|
| | Experimental approaches | | | | | Quasi-experimental approaches | | | | | | Other risk of bias | | | | |
| | PSM | IV | OLS | Logit/probit | Multivariate | Panel | Other | Spill-overs and contamination | Outcome reporting | RCTs | | | PSM | IV | OLS | Other |
| Schuler and Hashemi (1994) | | | | Uc | | | | | Yes | Uc | | | | No | Uc | Moderate |
| Setboonsang and Parpiev (2008) | Yes | | | | | | | Yes | No | | No | | | | Uc | Moderate |
| Sharif (2002) | | | | Uc | | | | Yes | Uc | | | | | | Uc | Moderate |
| Sharif (2004) | | | | Uc | | | | Yes | No | | | | | | Uc | Moderate |
| Wakoko (2003) | | | | Uc | | | | Yes | Uc | | | | | | Uc | Moderate |
| Zaman (1998) | | | | Uc | | | | Yes | Uc | | | | Uc | | Uc | Moderate |
| Zaman (1999) | | | | Uc | | | | Yes | Uc | | | | Uc | | Uc | Moderate |

Notes: 'Unclear' is abbreviated Uc. 3ie's bias assessment tool is the basis for this table but was modified. We use parts of their checklist to reach a conclusion about how well the study design and analysis was reported and executed; full implementation of this tool (1) would require detailed engagement with each study because the abstracts and texts of the papers are generally insufficiently informative about all the components of the tool, and (2) seems unnecessary because of the self-evident high vulnerability to bias of these studies. For the columns under 'selection bias and confounding', we score 'no' when study design and analysis were reported and executed adequately, 'yes' if otherwise and 'unclear' when the evidence is mixed and we cannot reach a firm conclusion. Studies scoring 'no' in the spill-over column when, based on our judgement, no spill-overs can be expected from the treatment to the control group and the groups are isolated from other interventions; 'yes' is scored when spill-overs are likely and 'unclear' when we cannot reach any firm conclusion and have lingering doubts. For outcome and analysis reporting, a score of 'no' is achieved when the outcomes/analyses discussed in the methods section are also reported in the results/analysis section, score 'yes' if otherwise and 'unclear' when the paper has no information on this. For other risks of bias, 'no' does not suggest any other sources of bias, 'yes' suggests other potential sources of bias and 'unclear' when no firm conclusions can be reached. The last column then reaches an overall conclusion as to the overall risk of bias: high risk of bias when three or more columns report 'yes', low risk of bias when all columns report 'no' and moderate risk of bias when three or more report 'unclear' or two report 'yes' and 1 'unclear'. When there is a tie, for example, at least two columns report 'unclear' and 1 reports 'yes', then the overall risk of bias assessment is unclear. This assessment is very subjective and judgement-based but provides a rough guidance to assess the validity of these studies and the tool used here needs to be refined as it is rather crude at the moment.