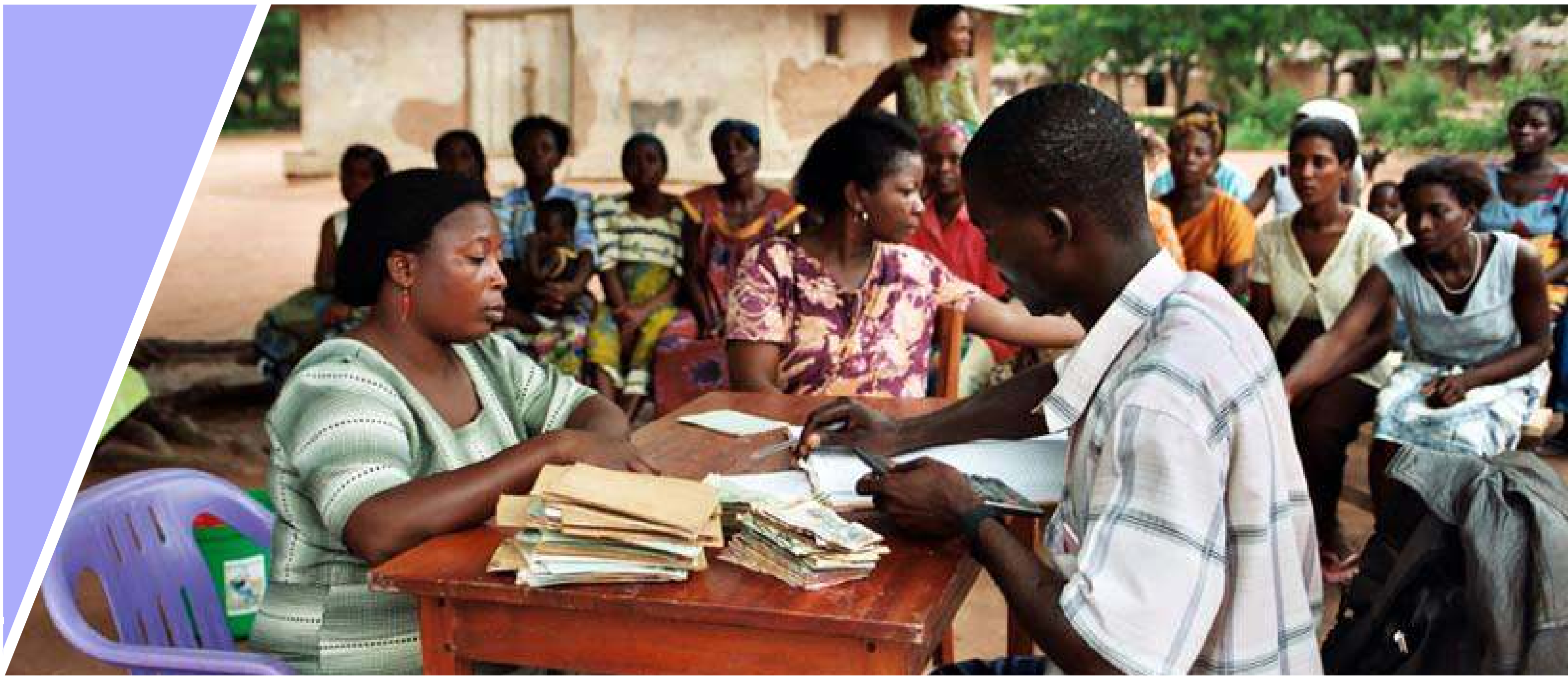
Systematic review

What is the evidence of the impact of microfinance on the well-being of poor people?



by Maren Duvendack
Richard Palmer-Jones
James G Copestake
Lee Hooper
Yoon Loke
Nitya Rao

August 2011

UKaid from the Department for International Development
DEV International Development University of East Anglia
3ie
EPPI Centre
IOE London
Leading education and social research
Institute of Education University of London

The EPPI-Centre reference number for this report is 1912.

This report should be cited as: Duvendack M, Palmer-Jones R, Copestake JG, Hooper L, Loke Y, Rao N (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.
ISBN: 978-1-907345-19-7

Table of contents

Abbreviations

2SLS	Two-stage least squares
A&M	Armendáriz de Aghion and Morduch 2005
ATE	Average treatment effect
ATT	Average treatment effect on the treated
CIA	Conditional independence assumption
DHS	Demographic and health survey
DID	Differences-in-differences
DIME	Development impact evaluation initiative
GB	Grameen bank
GDP	Gross domestic product
IE	Impact evaluation
IV	Instrumental variable
JGC	James G. Copestake
JLG	Joint liability group
LH	Lee Hooper
LIML	Limited information maximum likelihood
LSMS	Living standards measurement survey
MD	Maren Duvendack
MF	Microfinance
MFI	Microfinance institution
MICS	Multiple indicators cluster survey
MMR	Measles, mumps, rubella
NGO	Non-governmental organisation
NR	Nitya Rao
OLS	Ordinary least square
PnK	Pitt and Khandker 1998
PSM	Propensity score matching
RCT	Randomised controlled trial
RD	Regression discontinuity
RPJ	Richard Palmer-Jones
RnM	Roodman and Morduch 2009
ROSCA	Rotating savings and credit associations
SHG	Self-help group
SR	Systematic review
UPE	Universal primary education
YL	Yoon Loke

Executive summary

Background

The concept of microcredit was first introduced in Bangladesh by Nobel Peace Prize winner Muhammad Yunus. Professor Yunus started Grameen Bank (GB) more than 30 years ago with the aim of reducing poverty by providing small loans to the country's rural poor (Yunus 1999). Microcredit has evolved over the years and does not only provide credit to the poor, but also now spans a myriad of other services including savings, insurance, remittances and non-financial services such as financial literacy training and skills development programmes; microcredit is now referred to as microfinance (Armendáriz de Aghion and Morduch 2005, 2010). A key feature of microfinance has been the targeting of women on the grounds that, compared to men, they perform better as clients of microfinance institutions and that their participation has more desirable development outcomes (Pitt and Khandker 1998).

Despite the apparent success and popularity of microfinance, no clear evidence yet exists that microfinance programmes have positive impacts (Armendáriz de Aghion and Morduch 2005, 2010; and many others). There have been four major reviews examining impacts of microfinance (Sebstad and Chen, 1996; Gaile and Foster 1996, Goldberg 2005, Odell 2010, see also Orso 2011). These reviews concluded that, while anecdotes and other inspiring stories (such as Todd 1996) purported to show that microfinance can make a real difference in the lives of those served, rigorous quantitative evidence on the nature, magnitude and balance of microfinance impact is still scarce and inconclusive (Armendáriz de Aghion and Morduch 2005, 2010). Overall, it is widely acknowledged that no well-known study robustly shows any strong impacts of microfinance (Armendáriz de Aghion and Morduch 2005, p199-230).

Because of the growth of the microfinance industry and the attention the sector has received from policy makers, donors and private investors in recent years, existing microfinance impact evaluations need to be re-investigated; the robustness of claims that microfinance successfully alleviates poverty and empowers women must be scrutinised more carefully. Hence, this review re-visits the evidence of microfinance evaluations focusing on the technical challenges of conducting rigorous microfinance impact evaluations.

Methodology

Following the established medical and educational experience embodied in Cochrane and Campbell Collaborations, we assess the validity of available evaluations. Initially we focus on the intervention (e.g. provision of microcredit), the measurement of outcomes (e.g. income, expenditure, assets, health and education, empowerment, and so on) and contextual factors likely to affect differences in outcomes in different contexts, including other microfinance services. In addition, we consider different categories of persons (impact heterogeneity), and the potential existence, as well as the likely significance of factors which might confound observed relationships to undermine claims of a causal relationship with microfinance.

We search eleven academic databases, four microfinance aggregator and eight non-governmental (NGO) or aid organisation websites. We also consult bibliographies of reviewed books, journal articles, PhDs, and grey literature, using search terms given in section 2.1.2. We screen articles in two further stages, reducing 2,643 items to 58 which we examine in detail. In addition, we classify the research designs used in microfinance impact evaluations into five broad categories; in descending order of internal validity – randomised control trials (RCTs), pipeline designs, with/without comparisons (in panel or cross-section form), natural experiments and general purpose surveys. These five categories are cross-classified with three categories of statistical methods of analysis, which in descending order of internal validity are two-stage instrumental variables methods (IV) and propensity score matching (PSM), multivariate (control function) and tabulation methods.

As very few RCTs were available, we include in our review many studies with weak designs that have been analysed with sophisticated methods; however, we note that in general weak design cannot be fully compensated by sophisticated analysis (Meyer and Fienberg 1992, Rosenbaum 2002). Some articles used more than one method of analysis; actual designs, data production processes and analyses cannot be fully accommodated in such a basic two-way classification with limited numbers of categories. Nevertheless, we adopt a heuristic scoring of research designs and methods of analysis, combining these scores into a single value and defining a cut-off exclusion value. A few articles which we marginally exclude by this approach were included based on our judgement, resulting in a final count of 58 papers.

Our overall judgement draws mainly on RCTs and pipelines. However, we also devote considerable attention to the most prominent with/without studies which have been highly influential in validating orthodox favourable views of microfinance impacts. These earlier studies have turned out to have low validity with replicated analysis and critical assessment.

Results

There are only two RCTs of relevance to our objectives; neither has appeared in peer review form. In our judgement, one has low-moderate and the other high risk of bias; neither finds convincing impacts on well-being. We found nine pipeline studies reported in ten papers, all based on non-random selection of location and clients; most have only ex-post cross-sectional data, some with retrospective panel data, allowing only low validity impact estimates of change in outcome variables.

We find no robust evidence of positive impacts on women's status, or girl's enrolments - this may be partly due to these topics not being addressed in valid studies (RCTs and pipelines). Well-known studies which claim to have found positive impacts on females are based on weak research designs and problematic IV analyses which may not have survived replication or re-analysis using other methods, i.e. PSM.

Given their importance in validating perceptions of the beneficence of microfinance interventions, we devote considerable effort to the assessment of with/without studies which have low inherent internal validity notwithstanding analysis with sophisticated methods. In particular, we discuss the two historically most significant studies (Pitt and Khandker 1998 and USAID funded studies in India, Zimbabwe and Peru – see sections 3.4.1 and 3.4.2, which, partly as a result of their prominence, have been replicated. The replications fail to confirm the original beneficent findings, and conclude that there is no statistically convincing evidence in these studies to either support or contradict the main claims of beneficence of microfinance. This is partly because of their weak research design.

Conclusions

Thus, our report shows that almost all impact evaluations of microfinance suffer from weak methodologies and inadequate data (as already argued by Adams and von Pischke 1992), thus the reliability of impact estimates are adversely affected. This can lead to misconceptions about the actual effects of a microfinance programme, thereby diverting attention from the search for perhaps more pro-poor interventions. Therefore, it is of interest to the development community to engage with evaluation techniques and to understand their limitations, so that more reliable evidence of impact can be provided in order to lead to better outcomes for the poor.

1 Background

1.1 Aims and rationale for current review

While systems of credit provision for poor people have a long history (Shah et al. 2007), a new wave of microcredit provision has emerged in the past thirty years, inspired by pioneering innovations in Bangladesh, Bolivia, Indonesia and elsewhere. Microcredit has subsequently innovated in many ways, and is now more commonly viewed as one component of microfinance, along with savings, insurance and payment services for poor people. Microfinance institutions (MFI) have become important in the fight against poverty, growing worldwide in number of organisations and clients, and amount of donor funding [www.mixmarket.org/]. The sector continues to develop and innovate (Collins et al. 2009). A common feature of microcredit has been the targeting of women on the grounds that, compared to men, women both perform better as MFI clients and that their participation can have more desirable development outcomes (e.g. Pitt and Khandker 1998, Garikipati 2008).

At the time of writing only one systematic review (SR) on the impact of microcredit has been completed (Stewart et al. 2010), who focus on the impact of microcredit and/or microsavings on the poor in Africa. Another review focuses on the impact of microcredit on women's empowerment ((Vaessen et al. pending, personal correspondence with the authors) but apart from the protocol, no details are available yet.

The original objective of this SR was to assess the impact of microcredit. However, our study reviews not only 'credit', but also 'credit plus' and 'credit plus plus' interventions as set out in 1.3 and 2.1.1 on the social and economic well-being of people living in developing countries who are poor, excluded or marginalised within their own society. We exclude studies that solely look at microsavings and have no geographical focus although we use only works reported in English.

As set out in the protocol, we suggest adjusting the original review question from

'*What is the evidence of the impact of micro-credit on the incomes of poor people?*, to

'*What is the evidence of the impact of microfinance on the well-being of poor people?*

We include the following sub-questions:

1) What is the evidence of the impact of microfinance on other money metric indicators such as microenterprise profits and revenues, expenditure (food and non-food), assets (agricultural, non-agricultural, transport and other assets) and housing improvements?
2) What is the evidence of the impact of microfinance on other human development indicators such as education (enrolment and achievements for adults and children), health and health behaviour as well as nutrition?
3) What is the evidence of the impact of microfinance on women's empowerment?

As a secondary objective, investigating each of these questions requires us to examine, where the evidence allows, whether the impact of microfinance on any of these outcomes is modified by a) gender of borrower, b) poverty status of household, c) rural/urban setting, d) geographical location, e) presence of second income earner in the household, and f) type of product.

1.1.1 Research designs and analytical methods

The major problem identified in the review is that few, if any, studies provide reliable evidence of impact using the criteria normally adopted in systematic reviews[1]; we found only two randomised controlled trials (RCTs) and could not conduct a meta-analysis. In this section we look at randomised versus non-randomised research designs and briefly discuss the need for better designs and better quality data (research designs and analytical methods are discussed further in Section 3 and Appendix 7, section 6.7) to frame the substance of the review.

There are several studies comparing randomised controlled studies with other methodologies, and systematic reviews of such studies. The basic assumption underlying these comparisons is that the RCT will produce the 'correct' estimate of effect; the assessment is whether the non-randomised studies reproduce this RCT estimate satisfactorily. Cook et al. (2008) non-systematically summarised several recent social and educational studies that compared causal estimates from a high quality randomised experiment with those from high quality non-randomised intervention or observational data. In three cases that used regression-discontinuity analysis, they found comparable causal estimates to the RCTs, and in five cases with carefully matched comparison groups in quasi-experimental studies, and clear selection criteria and processes, they reproduced experimental estimates. In four remaining cases, the non-randomised data reproduced the RCT data in two studies only, but the non-randomised studies were declared of poor design. They concluded that the results of some non-RCTs can be trusted provided they meet specific quality criteria in terms of comparison groups and data quality. However, it is not clear how they searched for or chose the included 'cases', so this may be a biased sample.

A systematic review by Deeks et al. (2003) found eight health-based studies that compared randomised and non-randomised studies across multiple interventions. Additionally they conducted empirical work generating non-randomised studies from the datasets of two large multi-centre RCTs. They found that the results of non-randomised studies sometimes differ from the results of corresponding RCTs of the same intervention. As such, non-randomised studies may give seriously misleading results even when key prognostic factors in non-randomised groups appear similar to those in RCTs. They also found that in some cases results adjusted for case mix factors can be more misleading than non-adjusted results. They suggest only relying on non-RCT data when RCTs are not ethical or infeasible.

A more recent Cochrane systematic review (see Kunz et al. 2007) found that in 15 of 22 identified studies, important differences were found between estimates of effect from randomised and non-randomised studies. It also found that allocation concealment appears to be a crucial component of validity in RCTs. Generally, randomised studies with adequate allocation concealment tend to provide smaller estimates of effect than non-randomised studies, or randomised trials without adequate allocation concealment.

[1] For example, somewhat convincing evidence would have a score above 2+ (a C Grade of Evidence) in the Scottish Intercollegiate Guidelines Network (SIGN, n.d.) levels of evidence ranking (SIGN, n.d.), or a ranking of above 'Possible Evidence' in the World Cancer Research Fund scale (WCRF 1997); see footnote 52 for further discussion.

1.2 Definitional and conceptual issues

This discussion of randomised versus non-randomised approaches suggests that evaluation problems persist and cannot easily be resolved. We find, however, that the majority of the studies which applied econometric techniques to data from non-RCT designs fail to provide adequate evidence that they control appropriately for placement and selection biases. Nevertheless, a wealth of evaluation studies continue to claim that their impact estimates are robust and provide definite answers to the evaluation problem (e.g. Pitt and Khandker 1998, Pitt 1999, 2011). This can be misleading. Heckman et al. (1999) argued that the results of an impact evaluation heavily depend on the quality of the underlying data. In other words, advanced econometric techniques will not be able to control for poor quality data[2]. Meyer and Fienberg (1992) stated that:

> *Care in design and implementation will be rewarded with useful and clear study conclusions... Elaborate analytical methods will not salvage poor design or implementation of a study. (Meyer and Fineberg 1992, p106)*

This point is reiterated by Caliendo and Hujer (2005, p1) who stated that many evaluations in the past did not provide particularly meaningful results because of the non-availability of rich and high quality data sets due to poor designs. This makes it important, we suggest, that those who are to analyse the data, or who properly understand the analytical techniques and their data dependence, should be involved in the design of the impact evaluation early on to ensure the collection of rich data; this is one way to avoid pitfalls in the subsequent analytical process (Rosenbaum 2002). Rosenbaum and Silber (2001), for example, suggested using ethnographic or other qualitative tools with the objective of improving data collection procedures and the overall design of an evaluation. Heckman, LaLonde and Smith (1999), Rosenbaum (2002), Rosenbaum and Silber (2001) and Caliendo and Hujer (2005) suggested that it is not necessary to introduce ever more sophisticated econometric techniques, but instead a focus on collecting better quality data can be part of the solution to the evaluation problem. Therefore, not only do the econometric techniques employed require scrutiny when assessing the quality of an impact evaluation in which they have been applied, but so also do the underlying data.

In this study, we do indeed find that the evidence adduced in support of microfinance lacks robustness; a number of studies appear to have failed to replicate crucial findings of microfinance evaluations. The literature is not conducive to SRs in being characterised by high heterogeneity, with little consistency, or indeed precision, in the interventions implemented, high diversity of contexts, designs of evaluation, the covariate and outcome variables used, and so on.

We continue this review in the next section with a description of the key features of microfinance interventions and then outline the challenges of measuring microfinance impact.

In order to conduct a SR it is important to have clear and precise definitions of the interventions being evaluated and the outcomes assessed. In both interventions and outcomes explored here there is much diversity in practice. Thus, although there are many cases of 'Grameen replications'[3] among MFIs

[2] See also Rosenbaum 2002, p334.

[3] Grameen Bank MFIs employ joint liability group lending with small groups formed into a centre and regular equal weekly repayments, following the original classic model (Grameen 1) set by the Grameen Bank in Bangladesh (Yunus 1999). In the late 1990s, the Grameen Bank adopted a second model (Grameen 2) which

there is little uniformity in the interventions (see below), even for those which classify themselves as Grameen type. Nor is there much uniformity in the outcome variables (see below) even when nominally the same – for example income or consumption[4], business profits, assets, and so on. This is even more the case with social outcome variables especially empowerment. Even educational enrolment or achievements can be defined in various ways, as can health outcomes apart perhaps from anthropometric measures. Since most studies use several, sometimes numerous, outcome variables, with various definitions, in diverse contexts, we do not attempt, in this report, to produce standardised tables of estimated impacts and their variability, or to undertake statistical meta-analysis.

1.2.1 Theory of microfinance

Microfinance has spawned a large theoretical literature, which can be divided into two. The first addresses the specific problems that poor people have in gaining access to financial services at an affordable cost, particularly as a result of their lack of collateral. Would-be lenders are also deterred by high costs of collecting reliable information about the actual, or projected, incomes that borrowers might be able to lend against, particularly for potential clients with low overall 'debt capacity' (von Pischke 1991). Section 1.2.2 elaborates on this literature with particular reference to the potential for reducing loan monitoring, screening and enforcement costs through group lending. The second strand of literature explores impact pathways of microfinance on enterprises, households, and individuals. We take account of the ways communities assign access to livelihood opportunities, and how problems of access to credit, other income and consumption smoothing opportunities can at least partially be overcome by engagements with MFIs. Section 1.2.3 elaborates on this literature.

1.2.2 Microfinance and imperfect financial markets

The concept of group lending is commonly heralded as the main innovation of microfinance and claims to provide an answer to the shortcomings of imperfect credit markets, in particular to the challenge of overcoming information asymmetries (Armendáriz de Aghion and Morduch 2005, 2010). Information asymmetries may lead to the distinct phenomena of adverse selection and moral hazard. In the case of adverse selection, the lender lacks information on the riskiness of its borrowers. Riskier borrowers are more likely to default than safer borrowers, and thus should be charged higher interest rates to compensate for the increased risk of default. Accordingly, safer borrowers should be charged less provided each type can be accurately identified. Since the lender has incomplete information about the risk profile of its borrowers, higher average interest rates are passed on to all borrowers irrespective of their risk profile (Armendáriz de Aghion and Morduch 2005, 2010). In 'moral hazard' generally refers to the loan utilisation by the borrower, i.e. the lender cannot be certain a loan, once disbursed, is used for its intended purpose, or that the borrower applies the expected amounts of complementary inputs, especially effort and entrepreneurial skill, that are the basis for the agreement to provide the loan. If these inputs are less than expected then the borrower may be less able to repay it (Ghatak and Guinnane 1999). In addition to adverse selection and moral

allowed more flexibility in terms of the repayment schedule and higher loan amounts. Grameen 1 also involved regular weekly meetings at which there were physical exercises and reiteration of commitment to the 16 decisions about behaviour.

[4] For example, Deininger and Liu (2009) point out that income is constructed from 116 variables in their raw data; see Grosh and Glewwe (2000), for issues in measurement of most relevant variables produced in social surveys in low income countries.

hazard, high transactions costs, the provision of incentives to borrowers for timely repayment as well as the design and enforcement of adequate loan contracts are further challenges that play a role in explaining the failure of rural credit markets. In this context microfinance and its group lending approach steps in. Microfinance advocates claim that the formation of joint liability groups (JLGs) with its focus on peer pressure and monitoring responds to these challenges. As a result, the theoretical microfinance literature has focused on developing models that explain the workings of the JLG concept and its success, in particular, in overcoming information asymmetries.

The standard model of lending commonly contains two mechanisms which address the issue of information asymmetries: assortative matching[5] or screening to deal with adverse selection, and peer monitoring to overcome moral hazard (Ghatak and Guinnane 1999). In this widely cited paper, Ghatak and Guinnane (1999) reviewed how the principle of group lending facilitates assortative matching or screening and peer monitoring. Early models were developed by Stiglitz (1990) and Varian (1990) and Banerjee et al. (1994). These models examined how group liability schemes resolve moral hazard and monitoring problems. Other models developed by Ghatak (1999 and 2000), Gangopadhyay et al. (2005) and Armendáriz de Aghion and Gollier (2000) were inspired by Stiglitz and Weiss (1981) and focused on adverse selection and screening mechanisms. Moreover, social ties among group members, i.e. social connections in the language of Karlan (2007), also referred to as social capital, appear to play an important role in the context of group liability schemes in terms of enhancing repayment behaviour, as theorised by Besley and Coate (1995) and Wydick (2001).

The overall thrust of the literature is that the concept of JLGs does indeed overcome adverse selection by introducing better screening mechanisms. In addition, peer monitoring helps to overcome moral hazard and provides group members with incentives to repay loans resulting in high repayment rates (Ghatak and Guinnane 1999). In spite of that, Hermes and Lensink (2007) argued that MFIs are gradually abandoning the group liability scheme in favour of individual liability schemes; however, the literature on theorising individual liability schemes is surprisingly scant. Thus it seems that theory has lagged behind recent developments in the sector and requires some attention. Banerjee and Duflo (2010) and Fischer (2010) have made recent contributions to the theoretical literature on microfinance but it is beyond the scope of this SR to contribute to the theoretical discussion since the focus is on the evidence of impact of microfinance.

1.2.3 Pathways of impact of microfinance

The simplest theories of microfinance impact assume the borrower is the sole operator of a single income generating activity, the output of which is constrained either by lack of capital or by the high marginal cost of credit relative to its marginal returns. Easing the capital constraint permits the operator to increase output, net income, profits, and hence their own welfare (de Mel et al. 2008). Ability to borrow, or debt capacity, depends on the capacity of actual or potential income from the business to meet borrowing costs. More realistic theories take into account that debt capacity is also bound up with business vulnerability, risk and uncertainty. In the absence of insurance

[5] In the event of joint liability group lending where individuals are faced with endogenously forming their own groups (Chowdhury 2010), safer borrowers commonly form groups with safer borrowers rather than with riskier ones, while riskier borrowers have no choice but to form groups with riskier ones. This is referred to as 'positive assortative matching' (Ghatak 1999).

services, credit not only eases the capital constraints but can also serve as a mechanism for spreading risks. For example, access to credit (even if not actually taken up) can raise income by reducing the management of risk through livelihood diversification (Zeller et al. 2001). Borrowers' imperfect knowledge and limited computational capacity means that new forms of credit may have an important impact on the mental models that guide their business decisions (Nino-Zarazua and Copestake 2009). More generally, research into the psychology of credit among poor people has undermined the view that credit is generally unlikely to have an adverse impact on borrowers. This argument is based on the assumption that if credit did make them worse off, they would not have borrowed in the first place (Rosenberg 2010). However, possibilities for negative impacts of microfinance were early and clearly recognised in the framework developed in Sebstad et al. (1995).

A further complication arises because poor people's management of livelihood related resource allocation, risk and uncertainty cannot be separated from decisions about household reproduction (e.g. Gertler et al. 2009). As a factor in the management of diversified and seasonally volatile 'household economic portfolios' (Sebstad et al. 1995), the impact of credit on the cost of consumption smoothing may be as important as its impact on enterprise promotion (Morduch 1995, Rutherford 2001, Collins et al. 2009). Because portfolios are co-produced by household members both credit transactions costs and the potential benefits of credit can also profoundly affect intra-household relationships, including the gender division of labour, income and power. Induced changes in social relations inside and beyond the household are also associated with important changes in individuals' aspirations and understanding (e.g. Mayoux 2001, Johnson 2005, Hoelvet 2005).

Since changes in credit relations have direct effects on all aspects of poor people's households (and indeed wider kinship and neighbourhood networks) theoretical pathways can readily be traced, at least in theory, from credit to almost any indicator of individual socio-economic status or human well-being with positive or negative outcomes (e.g. Kabeer 2005a). For example, improved access to credit for cash crop production controlled by men may result in reallocation of resources away from food crop production controlled by women, with adverse effects on their children's nutrition. Likewise, improved access to credit for women's trading activities raises the opportunity cost of women's time with possible adverse impact on child care. Empirical testing of multiple pathways (e.g. using structural equation modelling) is relatively rare, perhaps because the lines of causation are so complex, with many relevant variables having both intrinsic and instrumental value (Sen 1999). It cannot be assumed, for example, that credit impact is only mediated via its effect on business income: direct relational, attitudinal and cognitive effects on individuals can be equally profound (Chen and Mahmud 1995). One potential response to this suggested by Scheffer (2009) is to regard the household economy as a complex dynamic system and credit as a variable capable of triggering critical system transitions.

Despite these complications, most research into the impact of credit on poverty continues to be framed by relatively simplistic causal models that link credit as an exogenous 'treatment' on individual borrowers to one, or more, indicators of well-being mediated via induced effects on household livelihoods and inter-personal relations. An alternative approach (not covered by this review) is to explore the effect of aggregate changes in financial systems on higher units of social organisation, from villages to national states. For example, credit supply may be treated as a resource constraint on a multi-sector input-output model, with distributional effects on poor people identified through use of a social accounting matrix (e.g. Subramanian and Sadoulet 1990) Alternatively, simulation models or cross-country multiple regression analysis can be used to explore the link between credit and indicators of national performance such as gross domestic product (GDP), which have testable relationships with poverty (e.g. Honohan 2004). An important example of this approach established positive links between rural credit expansion in India, district level growth performance and associated changes in poverty incidence (Binswanger and Khandker 1995, Burgess and Pande 2005).

In summary, the theoretical case for microfinance rests on the potential for joint liability and other innovations by MFIs, including individual liability with joint monitoring, to resolve issues such as adverse selection and moral hazard and to reduce MFI transaction costs. Mitigating financial intermediation constraints could lead to expansion of economic activities, higher net returns to household assets, and higher income. Furthermore, subsequent theory could be expanded on positive and negative potential relational, cognitive and attitudinal impact of access to credit.

1.2.4 Gender empowerment

Higher net returns to household assets may, of course, be goods in themselves, and may also lead to human developments which are income elastic. Insofar as credit is successfully targeted to women, it may benefit women specifically by enhancing their status and empowering them; it may also beneficially affect the pattern of household resource allocation, particularly benefitting children, especially females, at least in some patriarchal societies (Hashemi et al. 1996). These assumptions can be contested on the grounds that improved returns to assets, especially labour, power and entrepreneurship, are neither necessary nor sufficient grounds for improvements in health and education developments, may not exist, or may anyway be captured by males (Goetz and Sen Gupta 1996, Kabeer 2001, 2005b).

1.3 Interventions

Building on sections 1.2.2 and 1.2.3 it is important to bear in mind the diversity of actual interventions, and the extensive manifest and hidden subsidies that have typically been involved in microfinance (Armendáriz de Aghion and Morduch 2005, 2010). A simple classification of microfinance interventions as 'credit', 'credit plus' and 'credit plus plus' fails to capture the complex ways in which interventions are initiated with perhaps a given model in mind. Microfinance interventions do not, indeed cannot, exactly replicate given models, and subsequently evolve along their own context-specific path, resulting a unique intervention. Nevertheless, we classify the interventions studied into categories according to the ***credit product*** and ***credit type***. The ***credit product*** refers to a credit only product or whether it involves additional services such as savings, other financial products, training and or inputs, conscientisation, and so on. ***Credit type*** refers to whether the intervention provides credit to individuals, groups (self-help group, Grameen style and so forth), or both individuals and groups – see section 3 details. For those familiar with the realities of MFIs, it will be clear that this is a very crude classification and many cases will not fit well in these boxes. Nevertheless, in order to report on the impact of microfinance in the format recommended for this report – to compare like with like - this classification seems appropriate. However, we have severe doubts as to whether we are really comparing like with like.

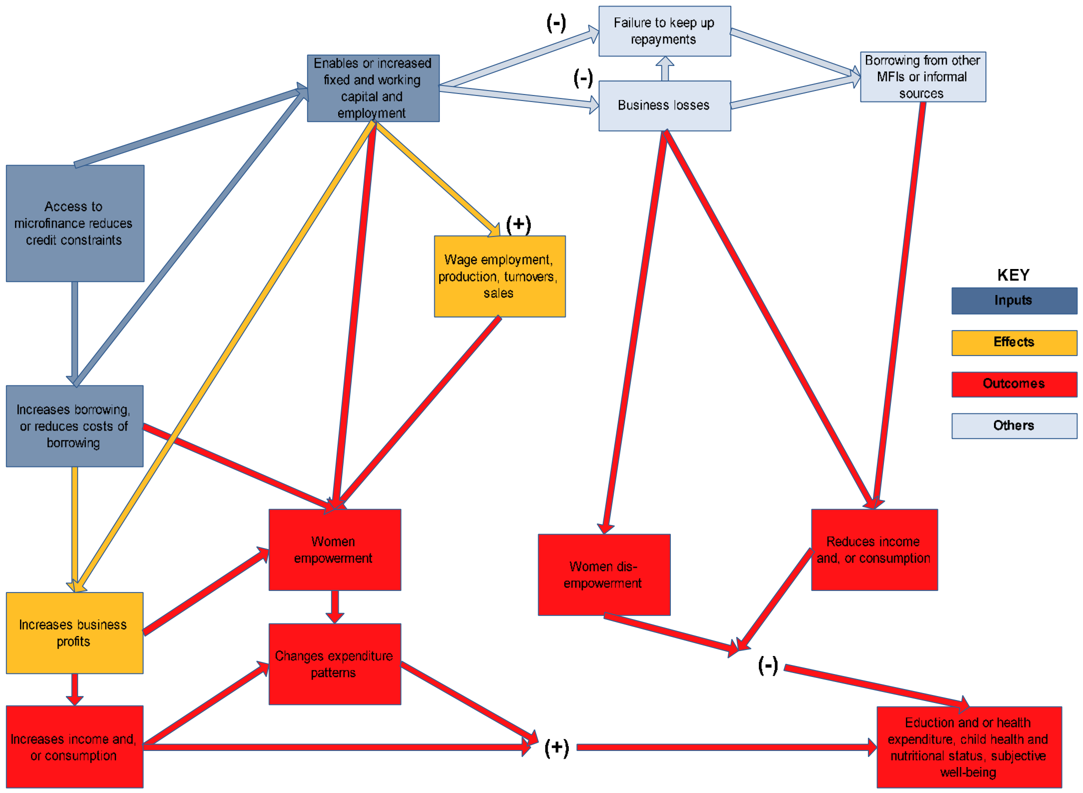
1.4 Outcomes

As noted above, there are many and varied pathways through which microfinance has been seen to have impacts, similarly diverse impacts that have been assessed. Figure 1 illustrates some of the more common positive and negative pathways and outcomes, but is far from exhaustively. Starting in the top left corner, we model the effect of access to microfinance; one route leads to increased borrowing, or reduced costs of borrowing if other more expensive loans are repayed. Through this route, or directly, access to loans provides insurance – enabling use of cash balances to increase business resources, and hence to increase output, sales and turnover. This, if the enterprise succeeds, in turn leads to increased profits, incomes and higher consumption. However, if the business fails, or cash has to be taken out of the business to meet emergencies, access to credit can lead to reduced enterprise activity, production, turnover, or sales, and ultimately in some cases, to business failure and increased indebtedness. The diagramme does not depict the fungibility of finance (Hulme 2000). Under this scenario increased borrowing substitutes other sources of cash. Borrowing may also be undertaken without adequate thought by the borrower, or under coercion from relatives, neighbours, or the microfinance organisers (Fernando 1997, Bateman 2010).

Because there are so many outcome variables tested in the papers we reviewed here, it is necessary to organise them into fewer groups. Some outcomes reflect direct effects of access to microfinance – particularly borrowing from microfinance. Therefore some outcomes have few specific implications for the value of microfinance, unless borrowing leads to outcomes which are more plausible indicators of welfare. Thus borrowing, increase in business assets, employment, sales or turnover means little unless they are positively associated with increased profits, household income, expenditure, and/or other indicators of welfare (housing or other consumer asset accumulation, education, nutritional or health status, and so on)[6]. Consequently, we can classify variables in these pathways according to a rough hierarchy of closeness to well-being. Borrowing and business assets are the indicators of effect and inputs of microfinance that are most distant from well-being. Business sales and turnover, profits, employment, agricultural, livestock or other production, are indicators of effect intermediate between inputs and well-being impacts. Income and expenditure, especially on food, education or health care (assuming the cause of ill-health is not related to the activity induced by microfinance), and indicators of education, nutritional status, and health, are subjective well-being indicators. We therefore classify outcomes both by category (economic, social and empowerment, and by position in the pathways between microfinance and well-being.

[6] Assets, sales and turnover may indicate increased business resilience, but this is only linked to well-being through other intermediate variables. Increased employment may indicate benefits for employees, but is not a direct indicator of well-being.

Figure 1: Pathways and outcomes of microfinance


(-)
Failure to keep up repayments
(-)
Enables or increased fixed and working capital and employment
Business losses
Borrowing from other MFIs or informal sources
Access to microfinance reduces credit constraints
(+)
Wage employment, production, turnovers, sales
KEY
Inputs
Effects
Outcomes
Others
Increases borrowing, or reduces costs of borrowing
Women empowerment
Women dis-empowerment
Reduces income and, or consumption
Increases business profits
Changes expenditure patterns
(-)
Increases income and, or consumption
(+)
Eduction and or health expenditure, child health and nutritional status, subjective well-being

Outcomes can be classified into three groups which we term economic[7], social[8] and empowerment[9]. The studies which report outcome variables that fall into these categories are given in section 3, Table 5 to 7, categorised by credit type and product. Table 8 and Table 9, section 3, report the number of tests reported by category of outcome variable, credit product and type.

It is important also to understand that different methods of analysis (and indeed research designs) have an influence on estimated impacts, as well as on the confidence one can have in impacts. Designs which do not accommodate selection and placement biases when analysed with naive methods are likely to overstate (positive or negative) impacts – this is discussed in detail further below.

1.5 Research background

1.5.1 Measuring impacts

The evaluation of social and economic programmes using experimental and observational methods has a long tradition. Interest in this area of work intensified in the early 1970s, and the evaluation of education and labour market programmes became popular (Imbens and Wooldridge 2008). The main concern of evaluations is to understand how programme participation affects the outcomes of individuals. Evaluators are trying to understand how outcomes

[7] Credit received from microfinance, business inputs and fixed and variable costs, production, sales, profits, expenditures by category of expenditure (excluding health and ducation), including food, non-food, total, housing, durables, and assets. These are sometimes nominal or deflated, and sometimes in logs.

[8] These are mainly health and education expenditures and outcomes; indicators of subjective well-being.

[9] Indicators of empowerment (exclusively of women).

differ when an individual participates in a programme compared with non-participation (Caliendo 2006, Caliendo and Hujer 2005). In other words, individuals can either participate or not in a given intervention, but they cannot do both at the same time. Constructing a counterfactual that would allow observing the potential outcomes of programme participants had they not participated is the main challenge of every evaluation study (Blundell and Costa Dias 2008, Heckman and Vytlacil 2007). Such comparison requires finding an adequate control group which would allow a comparison of programme participants with non-participants. However, this is a major challenge because programme participants commonly differ from non-participants in many ways, not just in terms of programme participation status. A simple comparison between participants and non-participants, i.e. analysing the mean differences of their outcomes after treatment, could highlight selection bias and therefore not provide any convincing impact estimates (Caliendo 2006, Caliendo and Hujer 2005). Selection bias occurs when individuals in a programme select themselves, or are selected by some criteria that make them differ from the general population with whom they are to be compared. Participants may self-select (or be selected) into a programme based on observable and/or unobservable characteristics; e.g. observable characteristics can be employment status, age, sex, educational attainment, and so on, while unobservable characteristics can be motivation, entrepreneurial ability, business skills, etc. (Armendáriz de Aghion and Morduch 2005).

The occurrence of selection bias can lead to errors in measurement of participation impact which, it is argued, can be dealt with by a wide range of experimental and observational methods[10]. However, many methods have drawbacks of one sort or another, and many fail to control for selection bias due to unobservable characteristics, thus potentially adversely affecting the accuracy of impact evaluation results. These shortcomings have been recognised by numerous government authorities, non-governmental organisations (NGOs), and academics. Therefore there has been a recent drive towards encouraging better impact evaluations, e.g. organisations such as 3ie (http://www.3ieimpact.org/) or the World Bank's Development Impact Evaluation (DIME) (http://go.worldbank.org/1F1W42VYV0) initiative encourage more rigorous approaches.

1.5.2 Challenges of measuring microfinance impact

The evaluation problem is pervasive, in particular in the context of microfinance. Despite the popularity of microfinance there is evidence that these programmes do not have uniformly positive impacts. Case studies and ethnographic evidence demonstrate that microfinance can have both positive and negative effects on the lives of the poor, but rigorous quantitative evidence on the nature, magnitude and balance of effects is scarce and inconclusive. There have been three major unsystematic reviews of microfinance impact, two of which are outdated (Sebstad and Chen 1996, Gaile and Foster 1996, Goldberg 2005). A very recent review has been published by Odell (2010), which is essentially a follow-up of Goldberg (2005). The first systematic review was

[10] Experimental data are produced when units of observation - usually individuals - are randomly allocated by the experimenter to treatment or to control groups (untreated, or placebo treatment). Observational data are produced when some attempt is made to find a comparable group, but without random allocation by the treater. Given the pervasive presence of placebo effects (Goldacre 2008, Imbens and Wooldridge 2008), a further level of complication arises with the issue of blinding, i.e. does the treated individual and/or treater know who is receiving the treatment and who is placebo treated. Observational data are not single or double-blinded, while experimental data may be, although this is generally very uncommon if not impossible in social experiments (compared say to pharmaceutical treatments) (Scriven 2008).

conducted by Stewart et al. (2010) to investigate microfinance impact evaluations in sub-Saharan Africa. There are several books reviewing microfinance (Hulme and Mosley 1996, Khandker 1998, Ledgerwood 1999, Robinson 2001, Johnson and Rogaly 1997, Armendáriz de Aghion and Morduch 2005 (A&M), Ledgerwood et al. 2006, Dichter and Harper 2007, Bateman 2010, Roy 2010). Numerous studies have assessed the impact of microfinance in different countries (e.g. Copestake et al. 2005, Copestake 2002). However, none of these constitute systematic reviews because they do not set out protocols for search, quality assessment, or analytical synthesis – the study by Stewart et al. (2010) is an exception in this regard. We know of one additional SR currently underway, but it does not attempt to assess impacts on a wide range of outcomes, which we consider likely to be interlinked in complex and context-specific ways in all major developing regions.

There have been a number of widely quoted studies that suggest positive social and economic impacts of microcredit (e.g. Pitt and Khandker 1998, Matin and Hulme 2003 on Bangladesh, Patten and Rosengard 1991, Robinson 2002 on Indonesia); others report that microfinance is not always beneficial (Adams and von Pischke 1992, Rogaly 1996). Hulme and Mosley (1996) imply that microfinance does on average have positive impacts but does not always reach the poorest; other studies claim that microfinance often can have positive impacts on the poorest (e.g. among others Rutherford 2001, Khandker 1998). There is no well known study that robustly shows any strong impacts (A&M p199-230); some recent RCTs (Banerjee et al. 2009, Karlan and Zinman 2009) may prove more convincing, although their restricted provenance limits external validity.

Most useful literature acknowledges two major problems with assessing microfinance impact using observational data – programme placement bias and self, peer, and lender selection of participants. Probably the most authoritative studies are by Pitt and Khandker (1998) (PnK), and Khandker (1998 and 2005) (see also related papers Pitt et al. 2006, 2003, 1999) These authors argue that microfinance has significant benefits for the poor, especially when targeted on women; 'PnK and Khandker (2005) thus remain the only high-profile economic papers asserting large, sustained impacts of microcredit', Roodman and Morduch (2009 p40-41) (RnM). The reliability of the PnK results has been contested (Morduch 1998), Pitt (1999) vigorously rebutted these criticisms, but neither paper was published, and although Goldberg (2005) clearly views PnK as unreliable (p17-20), he writes that Khandker (2005) is 'much less controversial' (p19). The matter rested until RnM replicated these four papers. RnM found that 'decisive statistical evidence in *favor* of [the idea that microcredit helps families smooth their expenditures, lessening the pinch of hunger and need in lean times … especially so when women do the borrowing] is absent' (RnM, p39; emphasis in original).

Another approach has been to exploit 'pipeline' quasi-experiments, in which control groups are constructed from randomly chosen people with apparently similar characteristics who have not yet participated in the MFI, but will join later (Coleman 1999, 2006). However, these designs may be vitiated if the persons joining later have different characteristics compared to the earlier participants (Karlan 2001), as is often the case (Goldberg 2005).

Partly in response to critical reviews of evaluations using observational (qualitative and quantitative) data there has been a trend towards conducting RCTs of many development interventions including MFIs (Karlan and Zinman 2009, Banerjee et al 2009, Banerjee et al. 2007). However, microfinance RCT interventions often lack some crucial characteristics of valid RCTs, particularly proper randomisation of microfinance allocation and/or double blinding. This has given rise to criticisms as there are possible effects of perceiving one is part of an experiment, or at least unusual set of circumstances, including Hawthorne and John Henry effects[11], so there is a continuing role for observational methods (Deaton 2009, 2010). The limited circumstances in which RCTs have been conducted affect their external validity.

The various research designs and analytical methods are discussed in more depth in Appendix 7, section 6.7 we provide a brief summary of the various approaches and methods of impact evaluation (IE) in the next section.

1.5.3 Research designs: RCTs, pipelines, with/without, panels, and natural experiments

This section discusses the characterisicts of the major research designs encountered in the microfinance IE studies encountered in this review. Readers with a good understanding of research designs and statistical/econometric analyses may proceed directly to section 2.

1.5.3.1 Randomised control trials

At the heart of every experimental design lies a natural, or artificially formulated, experiment which attempts to attribute the effects of an intervention to its causes (Hulme 2000). Evaluations applying a randomised design are generally believed to provide the most robust results. There is a long tradition of experimental methods in the natural sciences. Fisher (1935), Neyman (1923) and Cox (1958) were early proponents of randomised experiments.

Applying a randomised study design requires random assignment of potential clients to so-called treatment and control group; both groups must be drawn from potential clients whom the programme has yet to serve so that the impact of an entire programme can be evaluated (Karlan and Goldberg 2006). This random assignment to either treatment or control group ensures that potential outcomes are not contaminated by self-selection into treatment (Blundell and Costa Dias 2008). In other words, the potential outcomes or effects of the treatment are independent from treatment assignment. Proper randomisation ensures those individuals in treatment and control groups are equivalent in terms of observable and unobservable characteristics with the exception of the treatment status, assuming that no spill-over effects exist (Blundell and Costa Dias 2000, 2002, 2008). Hence, the mean differences in the outcomes of these individuals are understood to be effects of the treatment (Caliendo and Hujer 2005).

However, limitations exist in the case of randomised experiments, i.e. double-blinding, ethical issues, pseudo-random methods, attrition and the fact that behavioural changes caused by the experiment itself such as Hawthorne and John Henry effects cannot be ruled out. Also, spill-over effects cannot be eliminated (more details in Appendix 10, section 6.10) (Blundell and Costa Dias 2000, 2002).

RnM argue that the present drive towards encouraging RCTs also renews calls for taking a closer look at the value of observational studies which collect data through non-random processes. Like RCTs, observational studies are not uncontested as there are threats to both internal and external validity arising

[11] These are some of the more obvious biases that can arise in experiments with human subjects; see Miettinen and Cook (1981).

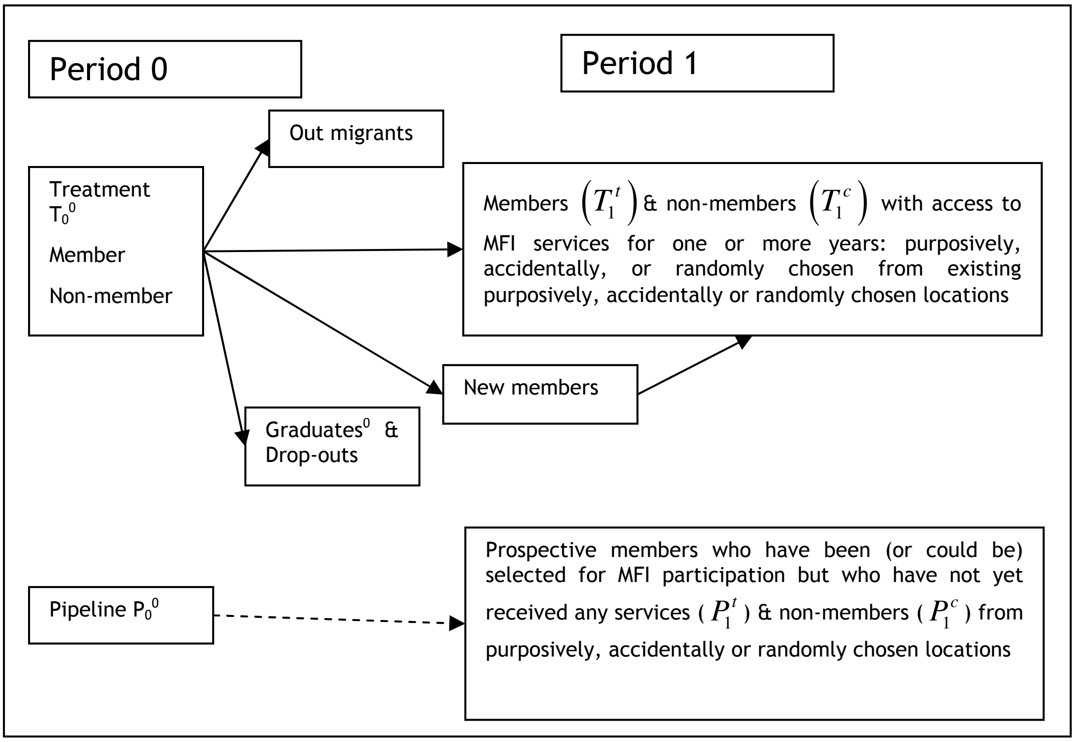
from observational data. There is a risk of confounding, i.e. confounding variables are both related to the outcome that is being measured and the exposure. Typically, observational data require the application of more complex econometric techniques, i.e. PSM, IV and DID estimations. However, many of these econometric techniques cannot deal adequately with selection bias due to unobservable characteristics as later sections argue.

1.5.3.2 Pipelines

Pipeline designs have become fashionable and are widely used in IEs (see Coleman 1999; Khandker, Koolwal, and Samad 2010), primarily for two reasons - they provide a convincing control group, and it is possible for them to be combined with randomised allocation; however, these are not universal characteristics of implementations.

The basic idea of the pipeline design is that it compares a representative sample drawn from the population which has had, or will have, access to the treatment together with a sample drawn from an equivalent population that is about to receive the treatment for the first time (the pipeline group). This is depicted in Figure 2; in this figure there are two periods, period 0 before there has been any intervention, and period 1 when interventions have been implemented in treatment locations only, while equivalent potential participants, yet to receive treatment, have also been identified in comparable locations. Ideally, the MFI enters both (sets of) locations, which are equivalent, preferably randomly chosen and sufficiently separated to minimise interactions, at the same time using the same recruitment procedures. The same prospective information is provided to potential recruits, although this will mean that one group – the pipeline – will be 'surprised' when their access is delayed. Nevertheless, this surprise is necessary if the groups recruited are to be equivalent, since if the pipeline group is aware when it is recruited that it will not have access for some

Figure 2: Set-up of pipeline design


Period 0
Period 1
Out migrants
Treatment T_0^0
Member
Non-member
Members $\left(T_1^t\right)$ & non-members $\left(T_1^c\right)$ with access to MFI services for one or more years: purposively, accidentally, or randomly chosen from existing purposively, accidentally or randomly chosen locations
New members
Graduates^0 & Drop-outs
Pipeline P_0^0
Prospective members who have been (or could be) selected for MFI participation but who have not yet received any services (P_1^t) & non-members (P_1^c) from purposively, accidentally or randomly chosen locations

time some who would otherwise have been recruited may decline[12]. If the control is recruited later, then the circumstances may have changed, meaning those who are recruited may not be all those who would have been recruited at the same time as the treatment group.

However, simultaneous recruitment did not happen in any of the pipelines. Treatment areas were mostly entered before the study was designed or implemented; control pipeline areas were selected and recruitment took place with some delay, and with the expectation of further delay in access to MFI resources. This means that in treatment areas not only was the selection process at a different time to that at which recruitment occurred in the control – pipeline – areas, allowing changed economic and social circumstances to potentially affect recruitment, but also opportunities would have existed for those originally selected in the treatment area to respond to their experiences, some either dropping out or graduating, meaning they were no longer equivalent, as a population, to those in the control areas. Of course, this may be addressed on an 'intention-to-treat' (ITT) basis, but it may no longer be possible to trace and interview these people. If this behaviour is based on unobservables, it will not possible in the control areas to identify those who would have also behaved in this way, leading to biased results[13].

In theory, pipeline designs allow a comparison of randomly assigned members and non-members versus randomly assigned future members and non-members, while controlling for selection and attrition (graduation and drop-out) bias in the process (see more on pipeline designs in Appendix 11, section 6.11). Members may be communities, if spill-overs are likely, in which case analysis will be made on the 'intention-to-treat' basis. In this case, randomly chosen communities are the units of study.

However, in many cases the ideal design is not implemented. Depending on the survey data available, impact in a pipeline design can be estimated from various formulae:

Where superscripts are time periods (0, 1)
Subscripts i, k = unit; j, l = location;
T is treatment
P is pipeline
X is covariates
V is village, ward, M is member, T is borrower (Treated)
Barred symbols are means

Simple ex-post difference:

1.1

$$\overline{Y^1} = \overline{T_{ij}^1} - \overline{P_{kl}^1}$$

In this case (1.1) we have data collected at the time point 1 when the treatment and pipeline groups are sampled; the treatment effect is the simple difference in means between these two groups. If the sample is of current borrowers, then the treatment group is net of graduates and drop-outs, and the control – pipeline – group is chosen at a different time to that at which the treatment group were

[12] The ethics of this 'deception' are of course debatable. One rationalisation that makes this design attractive is that if implementation capacity of the MFI is limited, then it becomes inevitable that some will only get access later. If those who get delayed access, and may be consequently somewhat disadvantaged, are randomly chosen in a greater public purpose, then this maybe considered 'ethical'.

[13] This problem is similar to that in the PnK studies where in the control villages it was not possible to identify those 'eligible' households who had more than 0.5 acres but whose 'value of land' was less than the normal value of 0.5 acres; see 0below and Morduch (1998), Duvendack and Palmer-Jones (2011).

recruited, with inevitable differences in the context of recruitment, and in all likelihood the recruitment process. The control – pipeline – group may also be selected for a different geographical location. A control function approach can be used if covariates are measured.

Ex-post Double Difference:

1.2 $$Y^{2a} = \overline{\left(T_{ij}^{t.1} - T_{ij}^{c,1}\right)} - \overline{\left(P_{kl}^{t,1} - P_{kl}^{c,1}\right)}$$

In this case (1.2) the difference between existing members and non-members in treatment communities is compared to the difference between pre-selected members and non-members in pipeline villages.

As in any DID estimation, it only modifies the simple difference estimate to the extent that there are differences in the non-members' metrics in the two locations.

Panel Double Difference:

1.3 $$Y^{2b} = \overline{\left(T_{ij}^{t.1} - T_{ij}^{t,0}\right)} - \overline{\left(P_{kl}^{t,1} - P_{kl}^{c,0}\right)}$$

In this case the change in outcome variable in the treatment locations is compared with the change in locations in the control locations. Ideally, as noted above, selection of members in both types of location, and sampling should be before loans are available in the treatment areas. If organisation, meetings and mobilisation among MFI clients are part of the intervention then it is moot whether this should be conducted also in the pipeline areas, since it is likely to affect the behaviour of the pipeline group and may affect dropout among them.

This estimation may be liable to attrition (dropping of cases from the baseline sample); also, of itself it does not address the issues of graduates or drop outs unless these are explicitly and successfully traced and the estimation is on an 'intention-to-treat' basis. In some cases the 'baseline' data are obtained by recall.

Control Function:

1.4 $$\mathrm{Y}_{ij}^{3} = \beta_1 X_{ij} + \beta_2 V_j + \beta_3 M_i + \beta_4 T_j + \varepsilon_{ij}$$

In this case the coefficient β_4 is the impact (since M=1 for all members, both existing and prospective, and 0 otherwise). Of course, since individuals select into membership this approach fails to address unobservable variables which are correlated with both membership and its impact. Also, since the data are ex-post there can be attrition bias unless there it is estimated on an ITT basis (with little attrition).

Panel:

1.5 $$\mathrm{Y}_{ijt} = \alpha_i + \delta_t + \beta M_{it} + \theta X_{it} + V_j + \varepsilon_{ijt}$$

In this case unobservable characteristics are supposed to be swept out of the estimation by fixed effects estimation. This may not occur if ex-ante equivalence of treatments and controls is conducted (and this can be demonstrated), or if the data are not a true panel with the baseline taking place after selection into treatment has already occurred - as is often the case with pipeline studies. In this case ex-ante equivalence cannot be demonstrated, graduation/dropout may already have occurred, and selection of the pipeline groups will have taken place under different circumstances to those of the treatment recruitment.

1.5.3.3 With/without (cross section)

With/without designs are the bases of most impact evaluations of microfinance. They involve the comparison of treated groups with comparable untreated groups and in the absence of randomisation are vulnerable to placement and selection biases. These may be mitigated by features of the data design, and by methods of analysis, which are discussed below. Key problems in these designs are that treatment groups may not include dropouts or graduates, and control groups may not come from the same population and sampling frame as the treatment group. Drawing the control group from the same community is risky since in most cases those who have chosen not to become members will clearly be different to those who have chosen to become members, generally as a result of an optimisation process (de Janvry et al., 2010). Also, since microfinance is likely to have spill-over effects to neighbours and to the local economy through general equilibrium effects, the comparison of MFI members and similar non-members from their own communities will be biased if it fails to account for these spill-overs.

Taking control groups from other communities risks placement bias unless the communities are demonstrably comparable (ex-ante). This can be achieved to some extent by matching the communities; it may not be achieved by random choice of control groups or communities from which to draw them unless the treatment communities were themselves randomly chosen from the same domain.

Often control groups are drawn from geographically separated areas because MFIs enter areas sequentially for administrative reasons. Thus several papers using data from with/without designs draw their control groups from different geographical domains; some provide descriptive statistics on observable characteristics, often with statistical tests of differences between treatment and control sub-samples, but this approach can not demonstrate equivalence on unobservables or variables for which there are no data.

Common analytical methods to mitigate biases due to non-comparable treatment and control groups include PSM, IV, fixed and random effect estimation, and control functions using community level variables.

1.5.3.4 Natural experiments

Natural experiments have been much sought after since the study by Duflo (1999) of a schooling programme introduced at different times in different geographical locations (see also Osili and Long 2008, for a very similar design based on the introduction of Universal Primary Education (UPE) in Nigeria).

Natural experiments exploit some difference between treatment and control groups to identify impact of a programme on the assumption that the difference is between statistically equivalent domains. Thus, for natural experiments to appropriately identify impacts the assumption is that the different domains are functionally equivalent – that is that there is no systematic difference between the treatment and control groups that interacts with the treatment which could account in part for the impacts.

1.5.3.5 Sample survey

In some cases, it is possible to make use of existing surveys which provide data that can be analysed to provide estimates of impact. Existing data sources have been used in the analysis of natural experiments (Demographic and Health

Surveys (DHS), Multiple Indicator Cluster Surveys (MICS), Living Standards Measurement Surveys (LSMS), etc.[14]). In principle, general surveys can be used wherever there are sufficient numbers of 'treated' and 'non-treated' units in the survey; since treatment may be fairly rare sufficient numbers will not occur in general surveys. However, in this case there may be sufficient comparable non-treated units to provide controls as in the classic LaLonde (1986) paper.

1.5.4 Analytical methods - PSM, IV, etc.

A number of econometric methods for overcoming, mitigating, or at least documenting the existence and consequences of selection bias have been developed. However, these econometric techniques have limitations and are often poorly executed or simply misunderstood as a review of the studies we included in this SR shows. A critique of econometric techniques is not new; in a landmark paper Leamer (1983) criticised the key assumptions many econometric methods are built on, however, despite his pessimistic view on the usefulness of econometric methods, there has been a trend towards ever more sophisticated techniques which has not necessarily provided the solution to the selection bias challenge. For recent expression of this debate see the symposium 'Con out of Economics' in the Journal of Economic Perspectives, 2010, 24 (2) (JEP 2010).

Apart from the technical challenges that impact evaluations have to grapple with, they are further hampered by the conflicting agendas of the various players involved. These agendas influence the design, execution and the results of an impact evaluation. Hence, Pritchett (2002) argues that it is not surprising that there are so few rigorous impact studies. Not only is that a phenomenon in the area of microfinance, but also health and education interventions are met with the same fate. Pritchett (2002) concluded that programmes usually have few incentives to be assessed seriously, and those that do are unusual.

1.5.4.1 Propensity score matching

We begin by explaining PSM, IV and DID to outline the best case scenario that studies should aspire to, and then discuss whether the included studies here have met these best case scenarios. We start with a brief introduction to PSM (more details in Appendix 6.7.2.1).

Matching has become a very popular technique in the area of development economics in recent years and has its roots in the experimental literature beginning with Neyman (1923). Rubin (1973a, 1973b, 1974, 1977, 1978) expanded on this literature and essentially laid the conceptual foundations of matching. The technique was further refined in particular by Rosenbaum and Rubin (1983, 1984). Econometricians got involved in advancing matching techniques in the mid-1990s; see studies by Heckman et al. (1997, 1998), Heckman et al. (1998) and Heckman et al. (1999).

The basic idea of matching is to compare a participant with one or more non-participants who are similar in terms of a set of observed covariates X (Caliendo and Kopeinig 2005, 2008, Rosenbaum and Silber 2001). In a next step, the differences in the outcome variables for participants and their matched non-participants are calculated, i.e. the average treatment effect on the treated (ATT) is the mean difference between participants and matched non-participants (Morgan and Harding 2006). The objective of this technique is to account for selection on observables. The drawback is that selection on unobservables remains unaccounted for.

[14] Osili and Long (2008) use DHS data sets to estimate a DID model of the effects of Universal Primary Education in Nigeria. See also Duflo (2001).

Despite this drawback, Dehejia and Wahba (1999, 2002) concluded that PSM results are a good approximation to those obtained under an experimental approach. They re-analysed the study of LaLonde (1986) and employed PSM to illustrate that PSM can in fact approximate the results obtained from an experimental setting.

However, Smith and Todd (2005) argued that the PSM estimates calculated by Dehejia and Wahba (1999, 2002) are sensitive to their choice of a particular sub-sample of LaLonde's (1986) data. They found evidence that a DID approach is in fact more appropriate as an evaluation strategy in this context than PSM as proposed by Dehejia and Wahba (1999, 2002). Overall, the outcome of this debate remains inconclusive with strong evidence provided by all parties involved.

The fact is that matching estimators are commonly not robust enough to deal with selection on unobservables. Hence, to test the likelihood that one or more unobservables could play a role in selection, which would explain unobserved differences, sensitivity analysis has become increasingly important. Sensitivity analysis attempts to gauge the vulnerability of the assignment process into treatment to unobservables (Becker and Caliendo 2007). In other words, the objective of sensitivity analysis is to explore whether the matching estimates are robust to selection on unobservables (Rosenbaum 2002).

Matching is a good choice when high quality data sets are available, but might not be an appropriate evaluation strategy if that is not the case (Smith and Todd 2005). Dehejia (2005) concluded that PSM is indeed not the panacea for solving the evaluation problem and pointed out that the correct specification of the propensity score is crucial, i.e. the balancing properties of the propensity score should be satisfied (see Appendix 0) – as emphasized by Caliendo and Kopeinig (2005, 2008) and Smith and Todd (2005) – and that the sensitivity of the results require testing – as advocated by Rosenbaum (2002), Becker and Caliendo (2007), Ichino et al. (2006) and Nannicini (2007).

1.5.4.2 Instrumental variables

The IV approach is widely used in the evaluation arena and claims to control for selection on observables as well as unobservables (Heckman and Vytlacil 2007b, Basu et al. 2007) which is in contrast to PSM which tries to construct an appropriate set of counterfactual cases to counteract selection on observables only. The main goal of the IV method is to identify a variable, or a set of variables, known as instruments, that influence the decision to participate in a programme, but at the same time do not have an effect on the outcome equation. Only when there are adequate instruments can the IV approach be an effective strategy for estimating causal effects (Morgan and Winship 2007).

In detail, a regressor qualifies as an instrument for Z*, which represents programme participation, when it is uncorrelated with the error terms and is not entirely influenced by *X*, the other right hand side variables which influence the outcome; i.e. the instrument has to be exogenous to be valid (Caliendo 2006, Caliendo and Hujer 2005). The main challenge of the IV method is to identify an adequate instrument which influences programme participation but at the same time does not directly influence the outcome equation.

Tests for instruments can be made, e.g. overidentification tests with the objective to assess the exogeneity of instruments. The Hansen-Sargan test for example is rather popular; it tests for overidentifying restrictions in a model. The main assumption is that the instruments are exogenous when the error terms are uncorrelated with a set of exogenous covariates *X*. If the null hypothesis is rejected then the instruments are considered to be weak, i.e. they are endogenous (Cameron and Trivedi 2005).

Deaton (2010) was cautious about the role of these tests and their ability to validate instruments; overidentification tests can be helpful but

> *acceptance is consistent with all of the instruments being invalid, while failure is consistent with a subset being correct (Deaton 2010, p431).*

He further argued that

> *passing an overidentification test does not validate instrumentation (ibid p431).*

Heckman and Vytlacil (2007b) argued that two-stage estimates are not necessarily better than simple ordinary least square (OLS) estimates. Studies using IV often fail to convincingly validate their instruments; weak instruments in turn can have adverse effects on the reliability of the two-stage estimates (Caliendo 2006, Caliendo and Hujer 2005). Instruments are often based on *a priori* arguments but these can be challenged.

1.5.4.3 Other methods (multivariate, control function and t-tests)

Because most studies included in this review are based on observational data they address identification problems associated with selection bias and require the use of advanced econometric methods, particularly the IV approach discussed above. However, many rejected and a few included papers use ordinary multivariate or bivariate statistics. These of course do attempt to deal with the identification problem, but as noted above this may not be done very well with more sophisticated methods, and, indeed may not be necessary if well-conducted surveys with carefully selected control groups have been employed. There can then be something to learn from studies using these designs and methods, although we have to bear in mind their vulnerability to selection bias. Well-conducted observational studies undermine a rigid hierarchy of methods approach to assessing the validity of studies (Pettigrew and Roberts 2006).

2 Methods used in the review

We based our methods on the Centre for Evidence Based Conservation and EPPI-centre guidelines as these are suited to the quantitative and mixed methods used in microfinance evaluations. In order to conduct an unbiased stakeholder relevant review, we set up an (unpaid) advisory group to support the SR and contacted a balanced team of reviewers by approaching representatives of major stakeholders and microfinance adepts. The main objective for this advisory group was to comment on the final outcome of our searches and to evaluate whether the relevant microfinance impact evaluation studies are included.

However, we received a response from only one member of the advisory group and proceeded with the SR without this potentially valuable feedback.

2.1 Identifying studies

2.1.1 Defining relevant studies: inclusion criteria

The included studies have the following characteristics:

Participants: Individuals living in poor, lower and upper-middle income countries (see appendix 2) with very few assets that could be used as collateral (Fernando 2006). Participants need to be classifiable as poor, excluded or marginalised within their society. The target group may include individuals, households or microenterprises.

Exposure or intervention: As mentioned in section 1.3 microfinance interventions are complex and diverse, we include microcredit and/or 'credit plus' programmes, including provision of credit of any sort to relevant participants plus savings; and/or 'credit plus plus' that includes savings, insurance and other financial services and that also combine financial services with complementary non-financial services such as business advice. Such services or programmes may be provided by basic, transformed or commercial MFIs, NGO-type MFIs (including those supporting informal or user-controlled financial services such as village banks), commercial banks, credit cooperatives and other public sector providers of financial services. Purely informal credit and savings associations such as Rotating Savings and Credit Associations (ROSCAs) are excluded since they are not classic microfinance providers. The duration of any microcredit program is at least three years.

Comparison groups: All included studies need to make use of some form of comparison or control group where microfinance has not been formally introduced. This may be a historic control (before/after comparison) or a concurrent control group where microfinance has not yet been introduced (by the assessed institution).

Outcomes: Primary outcomes include income, health and education. Secondary outcomes include microenterprise profits and/or revenues, expenditure (food and/or non-food), labour supply, employment, assets (agricultural, non-agricultural, transport and/or other assets), housing improvements, education (enrolment and/or achievements for adults and children), health and health behaviour, nutrition, women's empowerment.

Cut-off point: Studies published since 1970 are considered for review.

Methodologies: Controlled trials, before/after studies, action and observational studies and impact evaluations, and social survey datasets with pertinent indicators. Qualitative studies are assessed for inclusion but set aside and used to scope the literature in this area. Minimum sample sizes (subject to search outcomes) >100 (treatment and control combined) for quantitative and >10 for qualitative studies – these cut-off points have been set arbitrarily on the advice of our SR specialist.

Intervention studies including randomised controlled trials, controlled trials, before/after studies and action research assesses the impact of introduction of microfinance to some participants compared to the lack of such introduction in other participants (or at an earlier time).

Observational studies assess outcomes in populations served by microfinance and compare them to outcomes in areas not served by microfinance (or an earlier time before microfinance was introduced).

Qualitative research asks participants what they feel were the impacts of the introduction of microfinance to themselves and/or their family and/or community (compared to before such introduction or compared to nearby areas without such access to microfinance). It is beyond the scope of this SR to include qualitative studies and we merely collected them in a database for future research.

Publication status: Studies may be formally published or available in abstract, web-based, PhD thesis or organisational report form.

2.1.2 Identification of potential studies: search strategy

The electronic search strategy searched major on-line academic databases, systematic review databases, websites of relevant NGOs and funders as well as search for PhD thesis abstracts (see Table 1). The search included text and indexing terms, and Boolean operators in the format '[microfinance OR microcredit] AND [outcomes]'.

We experimented with these and other search terms until we obtained optimal results; then we saved these searches. Those saved searches at the same time left a documentation trail which allows others to reconstruct and validate our searches. Titles and abstracts were screened during these searches.

Table 1: Selected databases and websites that were searched:

Academic	External	NGO/Funder websites
EconLit (EBSCO)	British library of development studies (BLDS)	African development bank (AfDB), Asian development bank (ADB), Inter-American development bank (IDB)
informaworld	Eldis	Consultative group to assist the poor (CGAP)
ISI Web of Knowledge	Joint bank fund library network (JOLIS)	UK Department for international development (DFID)
Journal storage (JSTOR)	Google Scholar	Microfinance Gateway
AMED		MicroBanking Bulletin
SCOPUS (Elsevier)		Microfinance Network
Zetoc		United States Agency for international development (USAID)
The Cochrane Library		World Bank
Medline		
Embase		
PsychInfo		

We ran a draft search in ISI Web of Knowledge using the following search terms to assess the viability of our search strategy:

> #1 Topic=(evaluat* OR impact* OR benefit* OR poverty* OR empower* OR income* OR profit* OR revenue* OR employ* OR 'labour supply' OR job* OR expenditure* OR consume OR consumes OR consumed OR consumption OR asset* OR housing OR education* OR health* OR nutrition*) OR Title=(evaluat* OR impact* OR benefit* OR poverty* OR empower* OR income* OR profit* OR revenue* OR employ* OR 'labour supply' OR job* OR expenditure* OR consume OR consumes OR consumed OR consumption OR asset* OR housing OR education* OR health* OR nutrition*)
>
> #2 Topic=(microfinanc* OR microcredit* OR micro-credit* OR micro-financ* OR
> microenterprise* OR micro-enterprise* OR 'group lending' OR 'credit program*' OR 'credit plus*' OR credit-plus*) OR Title=(microfinanc* OR microcredit* OR micro-credit* OR micro-financ* OR microenterprise* OR

> micro-enterprise* OR 'group lending' OR 'credit program*' OR 'credit plus*' OR credit-plus*)
>
> #3 #1 AND #2

The draft search for Medline, EMBASE, AMED and Psychinfo (all on OvidSP) and the Cochrane Library (without the 'mp' at the end) was:

> (microfinanc* or microcredit* or micro-credit* or micro-financ* or microenterprise* or micro-enterprise* or 'group lending' or 'credit program*' or 'credit plus* or credit-plus*').mp.

This did not need limiting by outcome as few studies were located.
The reference lists of included quantitative studies and relevant reviews were checked for further relevant studies. Appendix 4 in section 6.4 has more details of our search strategy[15].

2.1.3 Screening studies: applying inclusion criteria

The searches were initially screened by MD who retrieved full text publications, reports or web-sites with potentially relevant text and data, which were then assessed independently in duplicate by the two lead reviewers (JGC & RPJ) using inclusion forms (see Appendix 2, section 6.2) developed for the review.

Quantitative studies: Data extraction and validity assessment of included quantitative studies (including all publications/reports etc. of a single dataset) were carried out by MD on forms developed for this review (see Appendix 2, section 6.2 and Appendix 5, section 6.5), then checked by either JGC or RPJ to create a final study dataset.

Data extracted and tabulated includes study characteristics: target group, exposure, comparison group and study relevance (distinguishing between those with different degrees of focus on our questions), validity criteria (developed to be relevant to each study's methodology and the review question), and outcome data (including sample sizes, data processing and analysis methods, values of categorical and ordinal impact variables, and parametric descriptive statistics of continuous data).

Qualitative studies: These are formally included in our database for a future systematic review, but are not included in the current review.

The remaining steps in the methodology and analysis refer solely to quantitative studies. It is beyond the scope of this SR to discuss the qualitative studies in depth.

2.1.4 Problems in searching and screening

We would like to draw attention to some differences between SRs in health and in development studies since our review team consists of researchers with backgrounds in natural as well as social sciences. Our colleagues from the School of Medicine have experience with reviews of RCTs, non-randomised studies and qualitative research and note that there were some major surprises in conducting a SR in the social sciences. The most immediate difference was the difficulty in searching incurred because abstracts are not structured, and do not always (or even often) address the question being addressed or the methodology employed. It was not possible to tell from most abstracts whether an article was

[15] The search records were managed in an Endnote library and the data extraction was later handled in Excel, Open database connectivity (ODBC), and Stata.

an informative narrative, a review article, or primary research. This lack of clarity in abstracts (along with a lack of methodological indexing terms) meant that it was difficult to run specific and sensitive searches, and also that assessing whether titles and abstracts may relate to relevant studies was extremely difficult. For these reasons, running the searches and assessing titles and abstracts for collection of full text articles was much more time-consuming than it otherwise would have been.

2.2 Assessing the validity and quality of studies

Criteria for judging validity used in this review are adapted from the Cochrane Handbook[16] and EPPI-Centre[17,18]. The Cochrane Collaboration suggests that the key components of bias (and therefore in assessment of validity) in any study are:

A. selection bias (systematic differences between baseline characteristics of the groups);

B. performance bias (systematic difference between care or support provided to the groups);

C. attrition bias (systematic differences between the arms in withdrawals from the study);

D. detection bias (systematic differences between groups in how outcomes are determined); and

E. reporting bias (systematic differences between reported and unreported findings).

EPPI-Centre formulates the risk of bias as being composed of the

F. trustworthiness of results (methodological quality, as discussed by Cochrane, including transparency, accuracy, accessibility and specificity of the methods);

G. appropriateness of the use of study design to address the review question (methodological relevance, including purposivity);

H. appropriateness of focus for answering the review question (topic relevance, including relevant answers and legal and ethical propriety); and

I. overall weight of evidence (a summary of the above).

See Appendix 6, section 6.6 for the full set of criteria which we attempted to use in this review. Following the established medical and educational experience embodied in Cochrane and Campbell Collaborations our assessment of validity initially focused on checking the delivery and adequacy of the intervention (e.g. provision of microfinance), reliability of the outcome measures (e.g. income, expenditure, assets, and so on), contextual factors affecting heterogeneity of outcomes (including other microfinance services), and potential existence and likely significance of confounding factors.

[16] Higgins JPT and Green S (2008) (eds) Cochrane handbook for systematic reviews of interventions version 5.0.0 [updated February 2008]. *Available at:* www.cochrane-handbook.org

[17] Gough D (2007) Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. In Furlong J and Oancea A (eds) *Applied and Practice-based Research. Special Edition of Research Papers in Education,* 22, (2): 213-228.

[18] EPPI-Centre website: 'Quality and relevance appraisal', http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=177 (accessed July 2010).

However, we found that in the context of microfinance evaluations there were few, if any, studies which met the rigorous standards of research design that this approach is based on. Hence, much of our work involved trying to assess validity of analyses of observational data which had further problems in their design. Many of these problems emerged during the review of papers, which, as noted above, generally did not have well-structured and methodologically informative abstracts; nor, on closer examination of the text was it easy to extract important details of the methodology (combination of design and analysis). Our initial selection criteria evolved during the course of the study, based on further consideration of validity of different designs and analyses, and their combinations as discussed next. However, in many cases the only way to be clear about the validity of results would have been to replicate[19] the studies in order to get a clearer understanding of the ways in which variables had been constructed and analyses undertaken; we illustrate this by reporting replications of two iconic MF evaluations (3.4.1 and 3.4.2). We elaborate these points in the next section.

2.2.1 Design and analysis - validity

As discussed above, where we discuss research designs and methods of analysis (sections 1.5.3 and 1.5.4), most microfinance IEs are based on with/without research designs which are of low inherent validity because it is difficult to control for selection and placement biases. Much of the earlier literature based on with/without designs, which purported support for beneficent development impacts of microfinance using sophisticated econometrics (PnK, USAID), have recently been shown to be questionable (Morduch 1998, RnM, Duvendack and Palmer-Jones 2011). We discuss these two examples in some detail to emphasise this point.

It has been argued that some recent microfinance IEs were based on appropriate designs (pipelines and RCTs), and well-considered with/without studies (see discussion below). While several of these studies have been published in peer reviewed journals, they have not been subject to rigorous criticism, (for example by thorough replication), or have evident deficiencies in implementation (for example, not reporting sensitivity analyses of PSM impact estimates[20], or tests of IV methods[21]). Others have only appeared quite recently as working papers. We discuss included RCTs and pipeline studies further below because their designs have generally accepted claims to validity, although their exclusive claims to validity are contested (e.g. Hausman and Wiese 1985, Heckman 1991, Scrivens 2008; Donaldson et al. 2009, Deaton 2010). For the studies based on with/without designs we discuss a sample, due mainly to limitations of time and resources, but also because of their lower inherent validity; we include the iconic PnK and USAID studies and a selection of others which seem to have higher validity among these designs. Brief summaries of individual papers using with/without designs, as well as RCT and pipeline designs are available in Appendix 15, section 6.15.

2.2.1.1 From a medical perspective

During the data extraction exercise, there were important and difficult issues in interpreting methodological descriptions to allow useful assessment of validity. There do not appear to be standard ways of describing methodology or laying out

[19] Replication is mentioned in 1.1.1 and discussed further in 2.2.3.

[20] See for example, Abera (2010), Imai et al. (2010), Takahashi et al. (2010).

[21] See for example, Diagne and Zeller (2001), Shimamura and Lastarria-Cornhiel (2010).

data in these studies; this differs from studies in the area of health, for example. Much work, in particular in the area of economics, takes previously generated data (that does not appear to be well described or considered very important) and carries out complicated and poorly described analytical techniques, making it difficult to reproduce findings. Within these analytical and statistical processes the information required to address validity issues - of pre-stating methodology, using sensitivity analyses to ensure robust results and reporting all analyses carried out (Higgins and Greene, 2008) - are not clearly present. Thus, reported data may well provide a biased set of analyses, and P-values may lose meaning if many analyses are 'tried out' before a subset is reported. Any one dataset may have several published analyses, so sets of reported results could represent de-facto sensitivity analyses. However, if many analyses are carried out using the same, possibly flawed methodology, or different analyses are made of a given dataset but only those supporting a particular position are reported (while others suggesting different interpretations are not reported), thenthis can result in serious bias. This is particularly likely when analyses, possibly of the same dataset, are carried out by the same group of researchers, and/or their students or researchers in personal, academic, consultancy or other relationship with institutions with interests in the work[22].

2.2.2 Summary of discussion of validity

Drawing on our discussion above, Table 2 summarises the threats to validity of varied research designs, and Table 3 summarises the threats to validity of varied analytical methods.

To summarise our arguments so far, it can be inferred from Table 2 and Table 3 that obtaining bias-free impact estimates for social experiments is a challenging task, mainly because of the limitations of the evaluation strategies available. Where possible, we checked the data to determine suitability for further evaluation by meta-analysis and/or meta-regression techniques to highlight outcome and contextual variability, and to appraise its usefulness for subsequent work, i.e. replications (see below). In some key cases, unit level data was accessed for data and data processing reliability but not to undertake replication (or re-analysis) of the study, because of resource constraints. Candidates for replication/reanalysis are merely identified, but not further discussed here. In fact, some replication exercises of studies which we include in this review are already underway. For example, Duvendack (2010a) replicated the IE conducted by USAID on SEWA Bank in India; further to this Duvendack (2010b) and Duvendack and Palmer-Jones (2011) replicated the IE conducted by PnK and related papers.

[22] These concerns parallel, but are more broad ranging than those leading to recent calls for a code of ethical conduct for economists (e.g. The Economist 2011).

Table 2: Threats to validity: research design

Research design	Score	Validity and threats	Comments
RCT	1	Blinding; failure to achieve random allocation, meaning (placebo) effects; adherence to treatment, randomisation/experiment effects; spill-overs	Much long standing discussion about validity of social intervention RCTs. In practice, surprisingly numerous threats to validity given cursory attention, or not reported in papers.
Pipeline	2	Random or non-random allocation; comparability of control groups; drop-outs & attrition	Relatively new design is simply a variant of with/without designs in which practice often falls short of precept.
Panel or before/after & with/without	3	Mostly non-random allocation, risk of confounding, selection & programme placement bias, panels no 'true' baseline	Many threats to validity, which cannot be fully compensated by 'elaborate analytic methods' (Meyer and Fienberg 1992, p106)
Either before/after or with/without	4	Mostly non-random allocation, risk of confounding, selection & programme placement bias	As above

Table 3: Threats to validity: methods of analysis

Methods of analysis	Score	Threat to validity	Comments
IV, PSM, 2SLS/LIML, DID	1	Weak instruments, poor/too few matches (limited common support); unbalanced covariates; small control groups, flawed data	Conduct simulation &/or sensitivity analysis to establish robustness; need good quality rich data sets; need clear account of data cleaning & variable construction
Multivariate	2	Control of endogenous variables	Use more advanced/sophisticated analytical methods; requires high quality dataset with many appropriate control variables.
Tabulation	3	Control of endogenous & exogenous variables	Use more advanced/sophisticated methods

See further Appendix 7 in section 6.7.

2.2.3 Evidence for replication or reinterpretation

Taken together, arguments so far (see statements by RnM, which are extended and supported by explorations of use of PSM by Duvendack and Palmer-Jones 2011), undoubtedly pose challenges to IE using observational and randomised controlled data, and draw attention to the need for replication (Hamermesh 2007). They also have implications for this SR. Firstly, results from papers published even in top rank peer-reviewed journals may not be reliable; secondly, replication is often highly advisable and requires access at least to original data. This suggests further criteria for quality assessment, preferably, the existence of supportive replications, or the availability of raw data to enable replication; repetition and replication in other locations are also desirable. This argument equally applies to qualitative data, where evidence (as opposed to assurances) of proper and ethical conduct of research (unbiased sampling, avoidance of leading questions, and so on) is not easy to provide. Qualitative study replication is substituted by reinterpretation, based on fully documented methods and texts giving assurance of ethical and professional conduct with regard to the production and interpretation of qualitative information (see http://www.timescapes.leeds.ac.uk/the-archive/ for UK best practice).

2.3 Process of synthesis and selection of studies

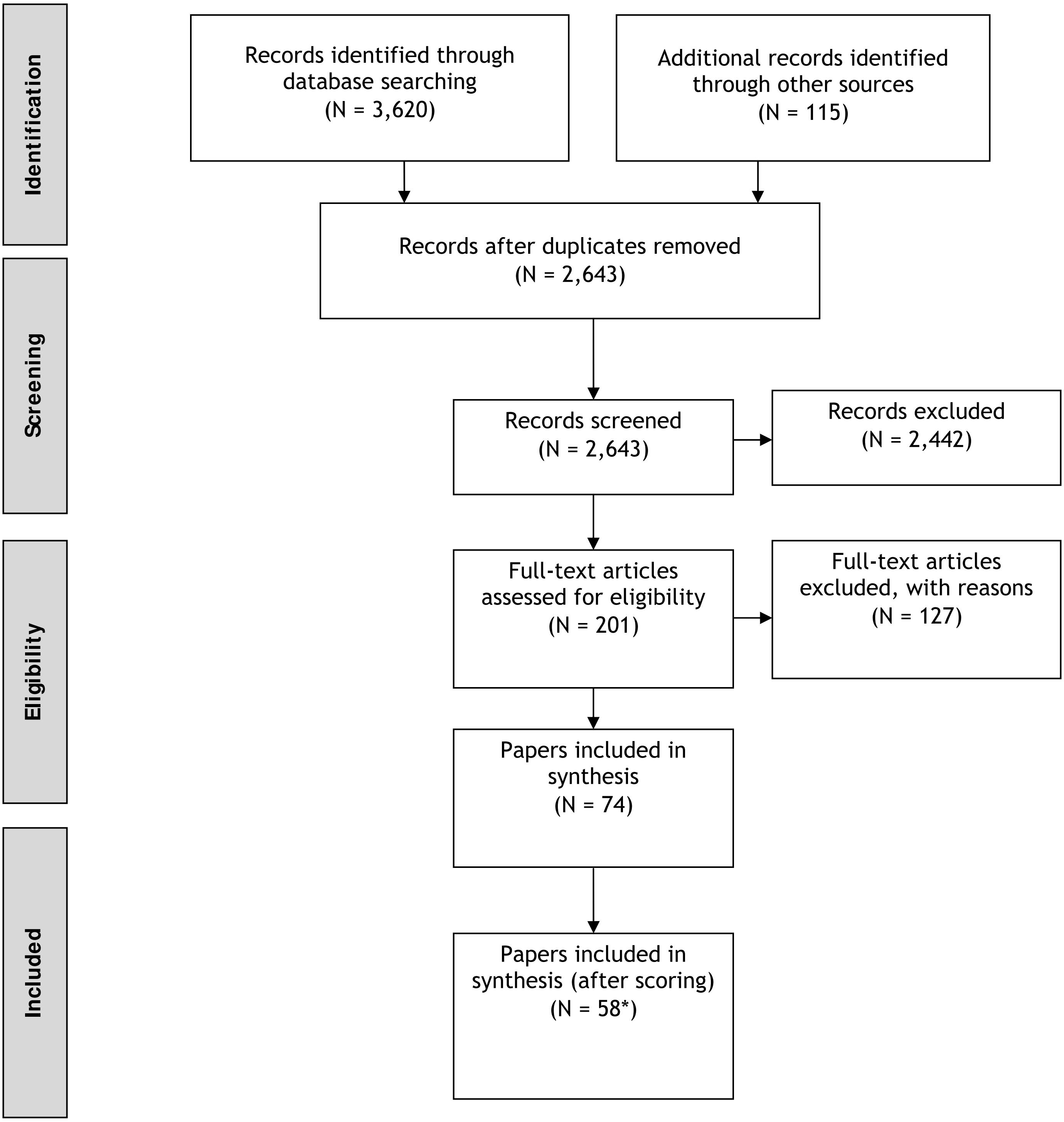
We have argued that evaluation of the impacts of microfinance is complex because of the difficulties of establishing causal relations in the presence of the challenges posed by the factors listed previously. These are particularly challenging for social science knowledge because of the impossibility of blinding in social experiments (compared to laboratory or field experiments conducted with inert subjects), and due to selection and placement biases. The criteria for selection and evaluation in SRs are demanding.

These challenges are addressed in social research by combinations of research design and analytical methods. To recap, by research design we mean the treatment allocation and sampling structure of data production; by analysis we mean the statistical (econometric) procedures used to interpret the data produced in the research design.

We refine our selection in three stages (see Figure 3); the first stage consists of searches of the databases listed in Table 1 that produced a list with 3,735 publications. 1,092 duplicates were found and removed leaving a final list of 2,643 publications. In the second stage we screen titles and abstracts of 2,643 publications, applying our inclusion criteria we shortlist 201 publications for which we obtain the full text to decide on inclusion or exclusion. In a third stage, we carefully screen shortlisted publications; this includes 74 papers which we analyse in depth.

The third stage requires more detailed explaination. This stage arises as the second stage resulted in very few papers that met rigorous selection criteria (this is discussed below). However, primarily there are only two RCT microfinance IEs, while all papers based on observational data have research design and analytical problems (more generally explored in Appendix 7, section 6.7). Consequently, we allowed inclusion of many papers based on observational data since they represent the bulk of the microfinance IE literature to date. In selecting among these papers, our logic is to score research design and data analysis approaches used in the papers, attaching weight to the quality of research design and statistical methods of analysis. These scores are then weighted and aggregated and a cut-off value specified to include papers judged to warrant further investigation. Scoring, weighting and aggregation were performed using Excel. The spreadsheet is available from the authors.

Figure 3: PRISMA flow diagram


Identification
Screening
Eligibility
Included
Records identified through database searching (N = 3,620)
Additional records identified through other sources (N = 115)
Records after duplicates removed (N = 2,643)
Records screened (N = 2,643)
Records excluded (N = 2,442)
Full-text articles assessed for eligibility (N = 201)
Full-text articles excluded, with reasons (N = 127)
Papers included in synthesis (N = 74)
Papers included in synthesis (after scoring) (N = 58*)

*Including only those papers with low and medium scores, see Table 4 for a breakdown. We excluded papers with high scores listed in the yellow fields in Table 4. For details of papers, see Appendix 8 and 9, section 6.

2.3.1 Process used to synthesise data

Our original plan was to scan papers by validity criteria as described in Appendix 5, section 6.5. This proved burdensome, if not impossible, in the absence of replication (as explained in section 2.2.3). Too many papers failed to provide sufficiently precise information to answer the questions in Appendix 5, section 6.5, such that some subjective judgements were required to complete these forms. We could not be confident about some assignments, as it was clear that few papers could achieve a 'low threat to validity' status.

Hence we adopt a slightly different approach to that originally intended, roughly scoring papers by their self-proclaimed research design, and by the analytical procedures. We combine scores into an index to which we apply a fuzzy cut-off to reduce the number of largely 'high threat to validity' papers to a manageable number.

The general principle we adopt is that weak research design requires more sophisticated methods of analysis in order to reach similar levels of validity. However, while this may in principle be true, it is also the case that more sophisticated analytical methods may not (fully) compensate for weak research design. Although individual studies can be roughly classified and ordered by research design and method of analysis (some papers use more than one of each), there is much further variation in the actual designs and analyses than can be accommodated in a simple two-way classification with limited numbers of categories. The final paper selection was based on a simplified ranking compiled from scores for design and analysis into a single index which varied from 0.0 (low threat to validity) to 2.78 (high threat to validity)[23]; we used a cut-off at 2 which excluded a significant number of studies with scores clustered just above 2 (see Figure 4). A few papers which were marginally excluded by this approach were included based on our judgement, resulting in a final count of 58 papers (see Figure 3 and Table 4).

[23] Score = ln(design) + ln(method), where design = 1 (RCT) to 5 (Observational), and method = 1 (IV etc.) to 3 (tabulation)

Figure 4: Distribution of 'validity scores'

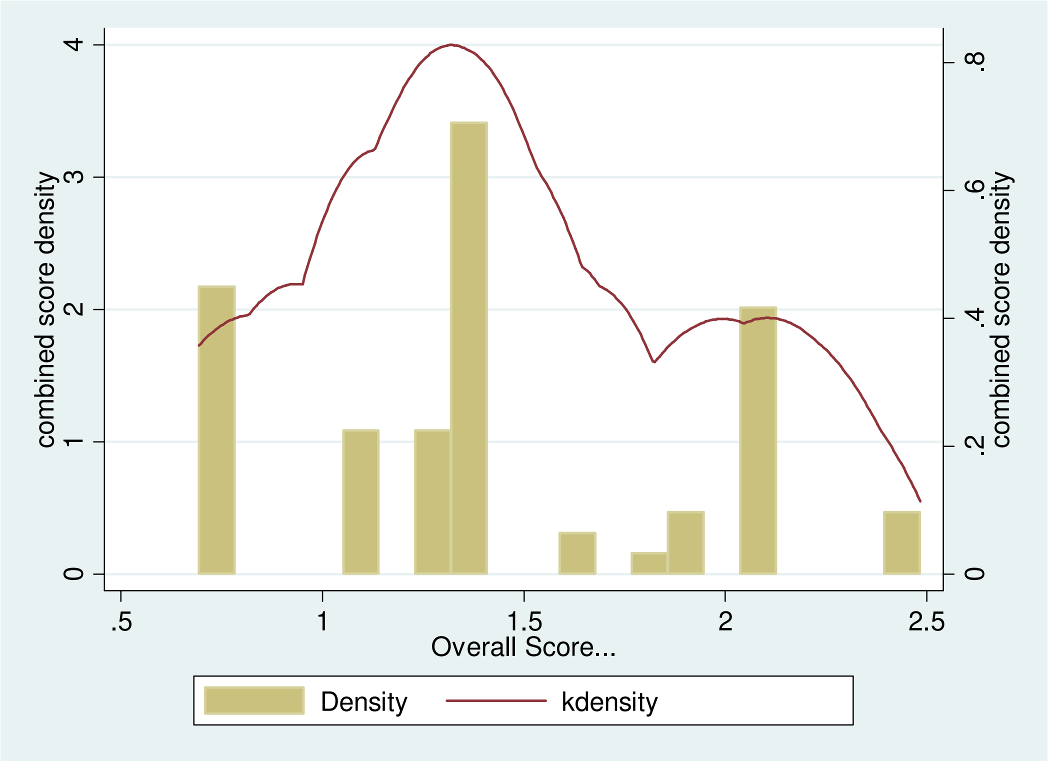


Table 4 provides a summary of included studies by scores. The red fields signify low scores and indicate that studies falling into this category are most certainly included in the SR. Studies in the yellow/orange field have a medium score which is still below 2, hence these studies are also included. Studies that fall within the bright yellow fields are excluded since their scores are high, i.e. above 2.

Table 4: Summary of included studies by scores; number of papers in each category*

		Statistical Methods of Analysis		
		IV,PSM,2SLS /LIML,DID	Multivariate	Tabulation
Research Design	Scores	1	2	3
RCT	1	🟥	🟥 2	🟥 1
Pipeline	2	🟥 9	🟥 0	🟥 0
Panel or before/after & with/without	3	🟥 14	🟧 6	🟨 0
Either before/after or with/without	4	🟥 22	🟨 13	🟨 3
Natural Experiment	5	🟥 2	0	0
Observation/ survey	6	🟥 0	0	0
		Legend		
	Low score	🟥 50	High score	🟨 16
	Medium score	🟧 6	Excluded	

* 2 papers (Chen and Snodgrass 1999, Dunn 1999) are included in our analysis but are missing from this table since they had a high score (2 and above). We include them in our synthesis because they were part of a group of papers that used the same dataset, i.e. the USAID data on India and Peru.

The majority of papers are from Bangladesh (31), including 21 from one dataset, and India (10), with a few from Thailand and Peru (4), Ethiopia and Pakistan (3) and several countries with two papers (e.g. Malawi) and many with just one. Appendix 15, section 6.15 and Appendix 16, section 6.16 provide further details of the studies included in this review.

3. Synthesis and discussion of results

In this section we frame and present summary evidence from the studies and papers selected, and draw together our findings in relation to the interventions with which they may be associated. In doing this, we frame the discussion in terms of a simplified representation of hypothesised pathways between microfinance and well-being outcomes, which puts the various 'outcome' variables in perspective.

We first list the included studies by intervention (credit type and product) and outcome, as explained in section 3.1 (Tables 5, 6 and 7). We then report, in Table 8 and Table 9, the numbers of results estimated, and the proportion that are statistically significant; a high proportion (more than half) of all the estimates of microfinance impact made - more than 2,800 in 58 papers (or 29 studies) - are not statistically significant. Even within the same study/paper estimates for a given impact can be significant or not depending on the sample, the estimation model, and the analytical technique. Replication, where done, does not always confirm the significance of the original results.

After this we discuss the included studies organised by **research design and method of analysis**. This does not mean that we adhere to a strict 'hierarchy of methods' (Guyatt et al. 1995, 2000), since low design validity can be partly, but not necessarily entirely, compensated by analytical sophistication, if well conducted. Not all included studies have received the same attention, notwithstanding meeting the inclusion criteria, since closer examination revealed some to be more valid than others. For reasons explained elsewhere we focus on the paradigmatic and highly influential PnK and USAID studies. These two studies generate a large proportion of the included papers, and they exemplify the theoretically powerful IV and panel methods of analysis. We also emphasise critical discussion of the emerging fashionable use of RCT, pipeline and PSM methods in microfinance IE.

We start the discussion with those research designs thought to provide most valid impact assessment, and those methods that are thought most convincing, proceeding to designs and methods of analysis that are less well regarded. Further details of these designs and methods are reported in Appendix 7, section 6.7.

3.1 The studies

Tables 5, 6 and 7 list the studies by outcome type, credit product and credit type, as set out in section 1.3. The majority of studies are on group lending with credit only products; however, few MFIs are really credit only since even the GB type institutions undertake other activities – recitation of the 16 principles, exercises, group meetings which provide motivation, advice, mutual solidarity, and so[24]. These nonfinancial aspects of microfinance can directly affect participants as well as, and beyond, income and consumption patterns, which are outcomes that most earlier microfinance IEs examined (Armendáriz de

[24] Also, note that we have classified the PnK studies as 'credit only' although they include BRAC and BRDB which encouraged savings, and GB (model 1) which has a compulsory savings element. It might have been better to classify PnK as 'credit plus' but this would create further presentation complications. GB type MFIs offer mainly credit only. Nevertheless, this highlights difficulties of classifying interventions into a small set of categories.

Aghion and Morduch 2010). The scores for all included studies are reported in Appendix 8[25].

3.1.1 Economic outcomes

Table 5 indicates that 26 out of 29 microfinance impact studies investigated economic outcome indicators. Within these 26 studies 18 are group lending programmes out of which nine provided credit only[26].

Table 5: Impacts of microfinance on economic outcome indicators by product and type of lending

	Type of lending		
Product	**Individual**	**Group & individual**	**Group**
Credit only	Abou-Ali et al. (2009) Cotler & Woodruff (2008)	Banerjee et al. (2009/10) USAID (Peru) (Year)	PnK (1998) Coleman (1999, 2002, 2006) Copestake (2001) Copestake (2002) Cuong (2008) Kondo (2008) Shimamura and Lastarria-Cornhiel (2010) Shirazi and Khan (2009) Tesfay (2009)
Credit plus (i.e. credit & savings)	Karlan and Zinman (2010) USAID (India)	Abera (2010) Diagne and Zeller (2001)	Takahashi et al. (2010) Zaman (1999) Zeller et al (2001)
Credit plus plus		Imai et al. (2010) USAID (Zimbabwe)	Copestake (2005) Deininger and Liu (2009) Imai and Azam (2010) Montgomery/Setboonsarng (2005/2008)

Economic outcome indicators include: business profits and revenues, sales, income p.c., consumption/expenditure, assets, employment, savings, debts, poverty indices, and other.

[25] The database deriving these scores is available upon request from the authors.

[26] As noted, many studies do not describe the interventions in detail. Other sources can provide in-depth and sometimes contradictory information on the intervention. Searching for this additional information significantly increases the time needed for the review. Note that we classify GB type interventions as 'credit only'. When there are multiple interventions some of which may be interventions different to the main intervention studied, we make arbitrary judgments as to the category to place this study in. Thus we classify the PnK study as 'credit only' although it includes BRAC and BRDB, which both have savings provision as well credit. In-depth information about interventions for the time and place of some studies is sometimes available to supplement what is provided in reviewed papers.

3.1.2 Social outcomes

As argued by Armendáriz de Aghion and Morduch (2010) and outlined in Figure 1, section 1.4, microfinance affects households beyond economic outcomes. We classify these as social and empowerment outcomes; these are more clearly outcomes of intrinsic as well as instrumental value, of which economic outcomes are better seen as instrumental (to achieving social and empowerment) outcomes (Sen, 1999). 19 out of the 29 studies assess the impact of microfinance on social outcomes. The majority of studies which include economic outcomes involve MFIs with a group lending approach and credit only products (see Table 6 for details).

This clustering of intervention types means that at least in principle, the information and understanding of the most frequently represented types should be better than for other types. However, this may not be the case because there are associations between outcome variables, intervention types studied and the period at which they were introduced, and the time and locations where they were studied. Thus early studies, which may have been methodologically weaker (smaller, less well-designed samples with reliance on complex analytical methods) were conducted in Bangladesh in the early 1990s. In contrast, at least some later studies were conducted in other countries (such as the Philippines, and Indonesia) with the benefit of understanding derived from earlier studies. We are also dealing with a small number of cases of many intervention types.

Table 6: Impacts of microfinance on social outcome indicators by product and type of lending

	Type of lending		
Product	**Individual**	**Group & individual**	**Group**
Credit only	Abou-Ali et al. (2009)	Banerjee et al. (2009/10) USAID (Peru)	PnK (1998) Coleman (1999, 2002, 2006) Copestake (2001) Copestake (2002) Shimamura and Lastarria-Cornhiel (2010)
Credit plus (i.e. credit & savings)	Karlan and Zinman (2010) USAID (India)	Diagne and Zeller (2001)	Steele et al (2001) Zaman (1999) Zeller et al (2001)
Credit plus plus		Imai et al. (2010) USAID (Zimbabwe)	Bhuiya and Chowdhury (2002) Deininger and Liu (2009) Montgomery/Setboonsarng (2005/2008)

Social outcome indicators include: children's school enrolment, school attendance, nutritional status, vulnerability to shocks, social capital, contraceptive use, and other.

3.1.3 Empowerment outcomes

It is often argued that microfinance empowers women (Armendáriz de Aghion and Morduch 2010), thus it is surprising that very few studies included in this review (merely 5 out of 29 studies) rigorously investigated the impact of microfinance on empowerment outcomes (see Table 7).

Table 7: Impacts of microfinance on empowerment outcome indicators by product and type of lending

	Type of lending		
Product	**Individual**	**Group & individual**	**Group**
Credit only		Banerjee et al. (2009/10) USAID (Peru)	PnK (1998)
Credit plus (i.e. credit & savings)	USAID (India)		Swain and Wallentin (2009) Zaman (1999)
Credit plus plus		USAID (Zimbabwe)	Deininger and Liu (2009) Montgomery/Setboonssarng (2005/2008)

Political outcome indicators include: empowerment.

Appendix 8 and 9, sections 6.8 and 6.9 have further details on the list of included and excluded studies. Appendix 16, section 6.16 details on the outcomes assessed listed by study.

3.1.4 Multiple outcome testing

As mentioned earlier, many studies have multiple impact estimates on the same dataset (see Table 8) using different sub-samples, estimation designs and methods. In many studies, estimates of impacts are made at several stages in the putative pathways between access and well-being impacts – as discussed further below. Many studies focus on the initial steps (or effects using the term used in Figure 1) – borrowing, business investments, activities and outputs, which are means to welfare improvement, before moving, if at all to impacts such as profits, household incomes, expenditure, which may be at least partly ends in themselves as well as means to further ends, or to health, nutrition and other indicators including subjective assessments of well-being or empowerment.

That many studies involve multiple tests is nowhere, as far as we can see, noted in relation to the dangers of multiple testing (Ioannidis 2005). Some studies test more than 100 outcome variables from various steps in the causal chain (see for example Coleman 1999, PnK). It is important to realise that a methodologically unsound study is liable to bias, and does not become any less liable as the tested number of outcome variables increases.

Table 8: Numbers of outcome variables tested

		Type of lending		
Outcome variable	**Product**	**Individual**	**Group & individual**	**Group**
Economic	Credit only	87	79	1114
	Credit plus (i.e. credit & savings)	396	143	63
	Credit plus plus	24	85	187
Social	Credit only	5	15	279
	Credit plus (i.e. credit & savings)	73	31	9
	Credit plus plus	0	0	26
Empowerment	Credit only	0	11	138
	Credit plus (i.e. credit & savings)	20	0	64
	Credit plus plus	0	0	20

Total number of impact estimates reported is 2,866.

In line with Tables 5, 6 and 7, we can see in Table 8 that most impact estimates involve group lending and credit only interventions; the focus is on economic outcomes rather than social and empowerment outcomes.

In section 1 we suggest examining whether the impact of microfinance on any of these economic, social and empowerment outcomes is modified by a) gender of borrower, b) poverty status of household, c) rural/urban setting, d) geographical location, e) presence of second income earner in the household, and f) type of product. This is not possible given the nature of the available evidence which is limited and rather diverse.

As in Table 8 and previous tables, the bulk of impact estimates in Table 9 listed are for group-lending and credit only studies with a focus on economic outcomes. Within this set of studies, most estimates are not significant indicating no impact of microcredit group lending on economic outcomes. It is puzzling how the view that microfinance is pro-poor and pro-women became so widespread given the evidence we have presented here so far. It appears that the microfinance hype has been shaped by very few influential studies that did not account for the diversity of the sector and the variety of the products. This is not the only problem as discussed further below. The majority of microfinance IEs discussed in this review suffer from shortcomings in research design and analytical method which casts further doubts on the credibility of the evidence they put forward.

Hence, having examined the included studies by intervention and outcomes, we now move on to discuss results by research design and analytical method; this is more meaningful given the diversity in types of interventions and outcomes.

Table 9: Significance of estimates of outcome variables by outcome, credit type and credit product

Outcome variable	Product	Individual		Group & individual		Group	
		sig	ns	sig	Ns	sig	ns
Economic	Credit only	61	26	28	48	471	641
	Credit plus (i.e. credit & savings)	129	267	91	52	14	46
	Credit plus plus	16	8	65	19	71	105
Total		206	301	184	119	556	792
Social	Credit only	3	2	3	11	57	229
	Credit plus (i.e. credit & savings)	37	36	3	22	4	4
	Credit plus plus	0	0	0	0	3	23
Total		40	38	6	33	64	256
Empowerment	Credit only	0	0	2	9	74	64
	Credit plus (i.e. credit & savings)	4	16	0	0	24	39
	Credit plus plus	0	0	0	0	0	13
Total		4	16	2	9	98	116

Total significant = 1160; not significant = 1680. Most studies do not report significance levels, only whether significant or not based usually on t-values of regression coefficients.

3.2 Randomised controlled trials (RCTs)

RCTs are considered the 'gold standard' for impact evaluations, they are widely and enthusiastically promoted in the development industry (Duflo et al. 2007, Banerjee and Duflo 2009), including for microfinance evaluations (Armendáriz & Morduch 2010)[27]. However, the advantages of RCTs for this purpose are moot, and to date very few have been conducted on microfinance interventions. We find eight studies which meet our selection criteria in stage 2, of which only two meet the higher hurdle in stage 3. One excluded study involved capital grants rather than microfinance and is not addressed in detail, although it aims to understand relaxation of credit constraints[28]. Two other studies are randomised comparisons of terms and conditions of microfinance (Giné and Karlan 2006, 2009, Field and Pande 2008), and another focused on South Africa (Karlan and Zinman 2005) and did not concern poor people as identified in this review. We do in fact discuss this paper as it accompanies another paper by authors included in this section. Duflo et al.'s (2008) study in Morocco concerned access to credit

[27] 'it [to answer the question how would borrowers have done without microfinance programmes] is a surprisingly difficult question to answer cleanly in studies that do not involve randomised research designs' (Armendáriz and Morduch 2010, p269).

[28] de Mel et al. (2008) assessed the impacts of capital grants rather than credit.

rather than the impact of credit; although the project research questions are given as: 'What is the impact of microcredit on individuals and their communities? What are the rates of returns of investments undertaken with microfinance loans? Does access to microcredit have a significant impact on household expenditures and activities?' Results for these questions do not seem to have been reported yet[29]. We note that there may have been many other RCTs in related areas, but since no register of such trials is kept, it is impossible to say whether such trials have been completed, or what results they have yielded. In the medical and related fields the failure to report on trials which yield insignificant or adverse outcomes is a major distortion limiting the usefulness of the RCT approach[30][31].

Of the eight studies we find that used RCT designs on microfinance interventions, only two were about impact of access to MFIs relative to no access and applied to relevant social domains[32]. The two RCT studies which meet the selection criteria are discussed in some detail in Appendix 15, section 6.15.1 because of the status ascribed to the RCT methodology in both the professional (Armendáriz and Morduch 2010, p293-308) and popular literature (The Economist 2009, Hartford 2009).

The validity and usefulness of RCTs has been extensively debated in many places as described above (see also example Donaldson 2009). While for policy purposes it has been argued that they 'did not yield credible evidence in a timely and useful manner' (ibid p3, paraphrasing Shadish et al. 1991), there continues to be widespread belief in their internal validity and the credibility of their findings. In general the main threats to internal validity pertain to (1) randomisation procedures; (2) adherence to treatment; (3) attrition (drop-outs and graduates); (3) behavioural responses of participants to randomisation, and in treatment and control contexts to masking/blinding or their absence; and (4) spill-overs and spill-ins.

We explain our understandings of these characteristics in the two included RCTs which lead to our overall judgement of their validity and what can be understood from them.

3.2.1 Randomisation

Failure to achieve randomisation are a common problem in RCTs, unless the randomisation is masked in the recruitment process leading to, preferably, double blinding – neither the treater (the MFI agents selecting and interacting with clients), nor the treatee (the clients) know they are treating, or being treated differently to some others in an experimental situation. This is not easy to achieve as the situation of being offered a loan is hard to disguise from that of not being offered a loan.

[29] At least not in a reviewable form; in addition to the report references, this study is reported in an unpublished Powerpoint presentation and a short www page presents some conclusions but without sufficient detail: www.philanthropyaction.com/nc/microfinance_impact_and_innovation_microfinance_impacts/

[30] See Song et al. (2010), 'Dissemination and publication of research findings: an updated review of related biases', available at: (www.hta.ac.uk/fullmono/mon1408.pdf)

[31] See also Petryna (2009), for a critical discussion of RCTs which indicates clearly how far below a gold standard they can fall. It is also insightful to refer to the talk presented by Kremer on 'Conducting Field Research in Developing Countries', in which he responds to the rhetorical question as to why one should conduct RCTs in developing countries with the comment 'because one can': www.streamingmeeting.com/webmeeting/matrixvideo/nber/20090724_1030_f/index.htm

[32] This excluded the Karlan and Zinman study in urban South Africa, although some details of this paper are discussed as they provide insight into the sister-study by these authors in the Philippines.

Banerjee et al. (2009) use 'slums' as units of study, and randomly chose one of a pair of slums that had been matched by minimum distance on a set of characteristics; the MFI entered the chosen slums but not the controls, but other MFIs could and did enter either during the course of the experiment. It appears that the randomised allocations were adhered to although the exact process of choosing one of each pair is not described. It is not clear how or whether selection of survey participants occurred prior to or subsequent to randomisation, (although it was almost certainly subsequent for the control survey sample).

Karlan and Zinman (2008) assigned randomly allocated applicants for consumer credit whose credit rating was 'marginally' below the normal cut-off. Loan officers were purportedly unaware of the credit ratings, although they conducted the survey on which the credit rating was determined. Furthermore, there were two groups among the marginal with different probabilities of being assigned to be offered a loan – those marginally below the cut-off and those somewhat further below. It appears that loan officers, whose remuneration depended on loan performance, did not always make the offers as instructed. This may well not have achieved randomisation since the loan officers would not have been blind to characteristics of the applicants and may well have selected on unobservables. While analysis was on an intention-to-treat basis using the original allocations, it seems likely that the sample of marginally creditworthy people actually being offered and taking up loans (which they did not all do even when offered), would have been biased by selection by loan officers and by self-selection. It is surprising that this design and its sibling in South Africa have been considered internally valid.

3.2.2 Adherence to treatment

This pertains to possible (unintended) dissimilarities of treatment between treated groups. There is no obvious issue in the Banerjee study in this regard, but doubts must exist about whether this could have been present in the Karlan and Zinman study - between those who surpassed the usual credit rating hurdle and those marginally and further below it. As argued above, loan officers would have been aware in at least some cases of the likely creditworthiness of those in different categories (as evidenced by their not offering loans to all those below the cut-off as instructed, and perhaps among those who declined to take up offers). As such, loan officers may well have acted differentially towards those to whom they were offered and who accepted loans. Loan officers may, for example, have visited them more frequently to raise the likelihood of repayments.

3.2.3 Attrition and response bias

The data analysed in Banerjee et al. (2009) was based on a random sample survey sampled from a population frame constructed by a census some time after the intervention started; an earlier base-line survey was not representative of the population. There is consequently some possibility that the population and sample miss some households who were present initially, and therefore would have had access but were unavailable during sample frame construction. There was no report of sample substitutions, but the impression is given that response rate were very high.

Karlan and Zinman (2008), on the other hand, achieved only a 70% response rate which is tantamount to a 30% attrition rate. Although attrition rate was not correlated with treatment there is no evidence about how characteristics affected attrition and whether this differed between treatments. Again this raises questions as to the validity of these studies.

3.2.4 Behavioural responses

These effects derive from participants knowing they are participating in an experiment. Details of the protocols used in the research to recruit participants are not explained, and there is no follow-up ethnographic or other evidence of perceptions of the participants (de-briefing). It the case of Banerjee et al. (2009) it is not clear why the presence or absence of the MFI in a slum would make much difference, since other MFIs were free to enter. However, it is shown that borrowing was greater in the treatment slums. Nevertheless, the cooperating MFI which did not enter control slums is one of the largest in India and is well known even among slum dwellers, there may well have been speculation in the control slums as to its absence. Indeed the authors of the study speculate that delayed entry may have affected business decisions, perhaps because of expectations about lower cost loans being available in the near future leading to some postponing taking loans for business expansion.

Karlan and Zinman (2008) explained that the 'stated purpose of the survey was to collect information on the financial condition and well-being of microentrepreneurs and their households. .... In order to avoid potential response bias in the treatment relative to control groups, neither the survey firm nor the respondents [were] informed about the experiment or any association with the Lender' (p9-10) It can be imagined, however, that speculation as to the meaning of being interviewed, and indeed thoughts that a less creditworthy person might have on being 'surprisingly' offered a loan in combination with being surveyed, can well have induced behavioural changes which would not be expected in a non-experimental context. Furthermore, there are ethical concerns about whether fully informed consent could have been obtained given the failure to inform participants of the nature of the study. Also, those with marginal credit scores who were not offered loans were discriminated against, raising further ethical concerns.

3.2.5 Spill-overs and spill-ins

In these studies there are no obvious problems in this regard Banerjee et al. (2009), explicitly acknowledge the effects of entrance of other MFIs in both treatment and control slums. Less clear is whether it is valid to assume that in control slums there was no 'surprise' that the MFI, which was well known, had not entered, not withstanding that surveys had been conducted in all slums but entrance had only occurred in one. This is to assume that information (about patterns of presence and absence of MFIs in particular locations) does not flow between slums, an assumption which is not in our view likely. Karlan and Zinman (2008) showed that there was not compensatory borrowing by rejected marginally creditworthy borrowers, but they did not discuss behaviour by other institutions which offer consumption credit in the locales.

3.2.6 Outcomes

Notwithstanding the randomised design of these two studies, very few significant impacts were found apart from borrowing amounts and sources – i.e. on the direct and intended effects of the intervention. But little can be learnt from increased MFI borrowing itself on well-being outcomes. A few coefficients implying impacts on business activities (inventories, profits, etc.), were also discernable, but these too have only indirect effects on well-being, even if some may consider increased business activity is (or portends) a good itself. Very few significant impacts on direct (health, education, subjective well-being) or indirect (income or consumption expenditures) indicators of well-being were found. Thus, while it may be the case, as Karlan and Zinman (2010, p19), argue, 'that it is important to measure impacts on a broad set of behaviours, opportunity sets, and outcomes', there is little evidence that RCTs have been able to deliver conclusive evidence of impacts on well-being.

More to the point is whether these studies deliver strong evidence that (short-term) impacts on well-being of MFIs are not present, as has been the popular interpretation of these studies – such as in the Economist Magazine[33] which considers whether based on these results, we can accept the null hypothesis of no impact with a high level of confidence. As noted elsewhere, some prominent academics involved in microfinance seem[34] to have preferred to not reject the alternate hypothesis. As such they imply that studies do not provide evidence leading to rejection of the hypothesis that MFIs have beneficent impacts.

Given the limitations on RCTs in this area by ethical considerations[35], avoidance of spill-overs, and intention to treat basis of evaluation, the potential contribution of RCTs at this stage may be very limited. One solution would be to find early indicators of later well-being, but this will often mean assuming what has yet to be proven (i.e. there are problems with finding convincing early indicators of later well-being).

[33] 'By being willing to take a risk on entrepreneurial sorts who lack any other way to start a business, microcredit may help reduce poverty in the long run, even if its short-run effects are negligible' (The Economist 16 July 2009).

[34] E.g. 'The study's relatively short time frame, however, limits the scope of the results and their implications for the short-term. Social outcomes, for example, may take longer to emerge. In the short-run, at least nothing big and positive leaps out from the evaluation' (Armendáriz de Aghion and Morduch 2010, p299). This response seems to echo that given by Roodman and Morduch (2009) to their replication of the PnK study, and in Roodman's interchange with Bateman in the former's blog (http://blogs.cgdev.org/open_book/2010/08/why-doesnt-milford-batemans-book-work.php) which echoes our private exchanges with Roodman about his interpretation of their replication.

[35] Undesirability of denying control groups access to a service that apparently has a high demand means that it is difficult to conduct long term trials in which control groups are denied access, and yet the impacts are likely to be long-term.

Table 10: Some Characteristics of RCT studies

	Treatment									Control								
Study	Design and analytical method	Random selection	Self-selection	MFI screening	Peer selection	Dropout included	Graduates included	Non-members?	Same population	Random selection	Same time	Same recruitment	Different area	Matching	Non-participants	Hawthorne	John Henry	Impact heterogeneity
Banerjee et al. 2009	Random choice of one of each pair of slums	Y[36](random allocation of areas to treatment)	Y	‘Not by Spandana’	Y	Intention-to-treat; ex-post survey will miss out-migrants who did not have access		Y	Y	Y[1]	Y	N	Y	Y	Y	N Selected hh knew of Spandana	N Control areas knew of Spandana but future access not clear?	Y
Karlan and Zinman 2008	Random allocation of marginally creditworthy	Y (but imperfect adherence to treatment)	Y	Y	N	Intention-to-treat analysis But many not found (figure 1? @600?			Y	Y	Y	Y	N	Y		Claims ‘double blind’ but low compliance suggests this is unlikely		
de Mel et al. 2009	Excluded because treatment is grants																	
Duflo et al. 2008	Excluded because does not have impact – only access outcome variables																	
Karlan and Zinman 2010	Similar to Philippines study, but excluded because of the context – urban SA is not a relevant domain.																	
Gine and Karlan ? 2008 or 2009	Treatments are different contractual forms – ‘Two randomized trials tested the overall effect, as well as specific mechanisms. The first removed group liability from pre-existing groups & the second randomly assigned villages to either group or individual liability loans.’																	
Field and Pande 2008	Ditto – ‘randomised client assignment to a weekly or monthly repayment schedule & find no significant effect of type of repayment schedule on client delinquency or default’																	
Dupas and Robinson 2009	Random allocation of negative interest bank account – no impacts on well-being – assess impact on savings																	

Notes: Columns are as follows:

[36] Endline survey has 500 households who already had loans retained from (non-random) baseline included to assess effect on

Random selection: are units of analysis (villages, enterprises, households, persons) randomly assigned to treatment/control

Self-selection: are units chosen by MFI or do they ‘volunteer’ or self-select?

MFI selection: does the MFI screen members or loans?

Peer selection: is the credit distributed with group liability or other group selection process?

Drop-outs: does the sample include drop-outs? (drop-outs are those who find microfinance is not for them or who fail in some way; they generally have not prospered)

Graduates: does the sample include members who have graduated? Graduates are usually successful clients who have repaid loans and prospered.

Same population: is the control group chosen by the same process from the same population as the treatment group? In other words, is the sampling frame the same?

Non-membership: includes those who borrow and those who do not in the treatment group (equivalent to ‘intention-to-treat’.

Same time: are the control group selected into microfinance/chosen at the same time as the treatment group? Selection at a different time can lead to different people/units being selected compared to those who would have been selected had the selection been at the same time as the treatment units were selected.

Same recruitment process: are the same procedures and conditions (and persons) used in the selection of control as treatment groups. Any deviation from the selection practices applied to the treatment units lead to differences among the control group.

Different area: are control units from the same locations as treatment units?

Matching of units: treatment and control slums are matched on minimum distance by a set of variables or characteristics reported.

Hawthorne: are Hawthorne effects discussed? N means no discussion in paper. What about likelihood of these effects? One can speculate that news of impending Spandana arrival spreads

John Henry: are John Henry effects discussed?

Both RCTs test many outcome variables using more or less complex statistical approaches. We do not report the results in detail, but Table 11 indicates that most 'impacts' (positive or negative) are early on in the causal chain, i.e. in inputs or effects (as outlined in Figure 1). The effects are mainly in the first stage of the causal chain, i.e. inputs, and the estimates are mostly insignificant and if significant, frequently negative[37] as well as positive. This is a different conclusion to that reached in, for example, The Economist Magazine (16 July 2009), where it was suggested that these results indicate 'overcoming the barriers posed by start up costs' of small businesses, and that 'there may well be some [beneficent effects on poverty] over a longer time-frame'. However, as also noted in Figure 1 businesses can make losses. For economic outcomes, there seem to be no impacts whatsoever. For social and empowerment outcomes, effects at the impact level are both positive and negative[38], and almost evenly split between significant (up to 10% level) and non-significant.

Table 11: Significance and sign of estimates by study, outcome variable and its location in causal chain (RCTs)

			Significance and sign					
			Ns			sig		
Study	Outcome category	Location in causal chain	+	-	total	+	-	total
Banerjee et al. (2009)	Economic	inputs	17	19	36	8	3	11
		effects	5	7	12	10	8	18
		impacts						
		total	22	26	48	18	11	29
	Social	inputs						
		effects						
		impacts	4	7	11	2	1	3
		total	4	7	11	2	1	3
	Empowerment	inputs						
		effects						
		impacts	7	2	9	2		2
		total	7	2	9	2		2
Karlan and Zinman (2010)	Economic	inputs	43	71	114	28	7	35
		effects	33	46	79	5	6	11
		impacts	-	-	-	-	-	-
		total	76	117	193	33	13	46
	Social	inputs						
		effects						
		impacts	6	1	7	8		8
		total	6	1	7	8		8
	Empowerment	inputs						
		effects						
		impacts	7	9	16	4	4	
		total	7	9	16	4	4	

[37] Some of the negative coefficients are on consumption of the presumptively bad 'temptation' goods – tobacco, for example – and could be presumed 'good'. However, if increased tobacco consumption reflects increased stress this might not be the case.

[38] For impacts on poverty the sign has been reversed so that reducing poverty is counted as a positive impact, while an increase in poverty is counted as a negative impact.

3.3 Pipelines

A number of papers (see Table 12 and Appendix 11, section 6.11 for further details and summaries of key papers Appendix 15, sections 6.15.2 and 6.15.3) included in the stage 3 selection use a pipeline design without random allocation of units of study (individuals, households, or communities), and there are no such studies, although Banerjee et al. (2009) could have achieved this had their baseline survey been satisfactory. Hence, treatments are confounded with locations, and they face problems in dealing with differences in characteristics or experiences of different locations during the study period (as argued in the discussion of individual papers in Appendix 15 sections 6.15.2 and 6.15.3. It should be noted that although most studies report making considerable efforts to match treatment and control locations, by definition, it is not possible to match on unobservables. The very fact that treatment areas were chosen first means that they are likely to have different characteristics to locations chosen later (closer to (further from) metropolitan areas, and, or good communications; more (or less) commercialised; with/without existing MF interventions; and so on). Other inevitable differences include delay in treatment and information in control areas.

Theoretically, more satisfactory studies include both treated and untreated units within both treatment and control areas, and have data in panel form from before any intervention and before the pipeline sample received any intervention; recruitment would take place at the same time and in identical ways. None of the pipeline studies achieved this design. Also, unfortunately, such a design is impossible because by its nature the pipeline sample will always have different interactions with researchers as they know that access to loans is delayed while the treated get earlier access. This may lead to behaviour by the pipeline sample which anticipates their future loan in some way, in ways that makes them different to the ideal counterfactual that would not have had reason to anticipate future loans from this source.

None of the included studies can claim to approximate even the ideal of recruitment at the same time, in locations appropriately similar initially, with similar shocks over the research period, and have true panel data. Whether using control functions (which without the use of IV must be considered not to have addressed endogeneity issues), panel methods (with the possible exception of Steele et al. (2001), who apply the recommended Hausman type tests), or PSM with sensitivity analysis, the analytical methods used fall short of best practices. As a consequence, whatever significant impacts are found in these pipeline studies must be held to be vulnerable to unobservables.

This is somewhat ironic as at least some pipeline studies have provided evidence that earlier microfinance evaluation impacts made by other methods have generally been overoptimistic about these impacts (Coleman 1999, 2005, Copestake 2001, 2002, 2005 and others).

Table 12: Pipeline studies without random allocation

		Treatment							Pipeline												
Study	Design & Method	Random selection	Self-selection	MFI screening	Peer selection	Drop-out included	Graduates included	Non-participants?	Same population	Random selection	Same time	Same recruitment	Different area	Matching informal	Non-participants	PSM	Sensitivity Analysis	Ex-post comparison	Outcome variables	Heterogeneity	Risk of bias
Coleman1999, 2002, 2006	Classic pipeline with non-participants in both	N	Y	Y	Y?	N	N	Y	Y?	N	N	N	N	Y	Y	N		Y	Loan size, wealth, assets, business , employment, expenditures	Y	Moderate
Copestake 2001	Pipeline & DID on growth of outcomes	N	Y	Y	Y	N	N		N	N	N	Y - putatively?	Not clear	N - multivariate analysis		N		N	Outreach & growth rates, & diversification, & household income growth	N	Moderate
Copestake 2002	Pipeline & DID	N	Y	Y	Y Village banking model	N high replacement rate among survey respondents	N	N	N since not same area-small businesses	Y - but had high number of replacements		Y	Y	N multivariate analysis	N	N			Household income & wellbeing	Y	Moderate
Copestake et al. 2005	Non-clients & DID control function	N	Y	Y	N? 'Village banking'	N	N	Y	Y?	N	N	Y?	Y?	N	Y	N		Y	Changes in sales, profits, family income & monthly per capita monthly income	Y	Moderate
Cotler and Woodruff 2008	Pipeline, panel with non-comparable control	N	Y	Y	N	?	?	N			No (838) -	Y/N (838)	Y	Y	N	N		N	Profits, revenues, inventories, assets	Y	Moderate
Deininger and Liu 2009	Pipeline & PSM	N	Y	Y	Y	N	N		Y?	N	N	Y?	Y	Y		Y	N		Empowerment variables, nutritional intake, income &	N	High

					Treatment								Pipeline							
																		expenditure, assets		
Kondo et al.– Philippines 2008	Pipeline with/without; matched baranguays; control function DID	N	Y	Y	Y	Y	Y	Y	N	N	N	Y	Y - see Coleman	Y but does not test between areas	Y	N		Income, expenditure, assets, health, education	Y	Moderate
Montgomery – Pakistan 2005	Pipeline DID	N	Y	Y		N	N	Y	N	N	N	?	Y	?				Consumption-expenditure, health and education, agriculture & enterprises	Y	Moderate-high
Setboonsarng and Parpiev – Pakistan 2008	Pipeline with PSM & DID	N	Y	Y		N	N	Y	N	N	N	Y	N	Y	Y	Y	N	Outreach and many agriculture, enterprise, employment, income, expenditure & human capital	Y	Moderate
Steele et al. 2001	Pipeline panel fixed/random	N	Y	Y	Y	N	N	N/Y[39]	Y	N	N	Y? eligibility	Y	Y	Y	N	N/Y	Modern contraception	Y (women's status)	moderate

Notes: Columns are as follows:
Random selection: Are units of analysis (villages, enterprises, households, persons) randomly assigned to treatment/control
Random selection: are treated randomly allocated – not in these pipelines – only in potential case of
Self-selection: are units chosen by MFI or do they 'volunteer' or self-select?
MFI selection: Does the MFI screen members or loans?
Peer selection: is the credit distributed with group liability or other group selection process?
Drop-outs: does the sample include drop-outs? (drop-outs are those who find microfinance is not for them or who fail in some way; they generally have not prospered)
Graduates: does the sample include members who have graduated? Graduates are usually successful clients who have repaid loans and prospered.
Non-participants: does the sample (treatment or controls) include non-participants?
Same population: is the control group chosen by the same process from the same population as the treatment group? In other words, is the sampling frame the same?
Same time: are the control group selected into microfinance/chosen at the same time as the treatment group?
Same recruitment process: are the same procedures and conditions (and persons) used in the selection of control as treatment groups.
Different area: are treatment and control groups are chosen from different sets of geographical units of the same overall population domains and if so are they random?

[39] See Appendix 12.15.3.1: there were three areas – old, new (expansion) and control. There were no non-participants in the 'old' area, but there are eligible non-members in the new and control areas.

Matching information:	is there information about the similarity of treatment and control samples?
PSM:	is propensity score matching conducted?
Sensitivity analysis:	is sensitivity analysis conducted?
Ex-post comparison:	evaluation when project is mature

Table 13: Significance and sign of estimates by outcome variable and its location in causal chain (pipelines)

Outcome category	Location in causal chain	Sign & significance					
		ns			sig		
		+	-	total	+	-	total
Economic	inputs	70	41	112	20	13	33
	effects	143	94	238	77	38	115
	impacts	8	2	10	27		27
	total	221	137	360	124	51	175
Social	inputs	12	7	19	5		5
	effects						
	impacts	57	65	122	10	6	16
	total	69	72	141	15	6	21
Empowerment	inputs						
	effects						
	impacts	8	11	20	11	1	12
	total	8	11	20	11	1	12

As in the case of RCTs, in Table 13 most effects occur in the early stages of the causal chain (as outlined in Figure 1) and not when significant estimates predominate. There are quite a number of impact estimates on outcomes more representative of welfare itself rather than putative means to welfare, however, the vast majority are not statistically significant. This supports arguments presented in this review so far, and further questions the apparent reasons for widespread perceptions of the success of microfinance.

Apart from Deininger and Liu's (2009) study, few of the pipeline studies have suggested strong positive impacts of microfinance, despite the large number of outcome variables tested, and different econometric specifications used (giving rise to a very large number of estimated impacts for assessment). Notably, the Deininger and Liu (2009) study has the highest vulnerability to bias, in large part because of the manifest difference in treatment and control locations; this repeats earlier findings from the medical literature that it is the studies with the weakest designs that tend to give the largest impacts. We would have preferred not to have included this study; it is included because there are unclear criteria in its research design or analytical procedures to distinguishing it from other pipeline studies[40]. Most of the negative and significant variables relate to inputs or effects of loans rather than well-being outcomes. Since no necessary connection can be claimed between business activities and well-being outcomes, and there are considerable difficulties in establishing business profits as opposed to business activities (de Mel et al. 2009), little could be inferred from these results.

[40] Within the health field many systematic reviewers and practitioners have accepted the hierarchy of evidence, and as a result accept that it makes sense to base clinical and policy decisions on the highest level of evidence available (Higgins and Greene 2008). In the area of development research it appears that this hierarchy of evidence is less discussed, less explored and less agreed. In the absence of this clear model of validity of different study methodologies, and with a paucity of evidence from the highest levels of evidence, it feels very difficult for new systematic reviewers to exclude studies from lower down the evidence hierarchy, especially when these are some of the very studies valued by the scientific community, regularly discussed and quoted as addressing important issues. This can lead to new reviewers wanting to include a wider selection of studies than they are capable of fully data-extracting, assessing for validity and reporting (given the time and funding provided for their review), making systematic reviewing a much bigger and more complicated job and making the conclusions harder to identify (Hooper, personal communication, 23 January 2011).

3.3.1 Conclusion of pipeline studies

Coleman's (1999) influential study has been followed by a number of studies employing pipeline designs. Colman's study has been noted in the literature as failing to provide strong evidence in support of the beneficence of microfinance for the poorer members of society; it found that benefits from microfinance were low and not much different to those from other borrowing sources, and to the extent that there were positive outcomes they were among the better off. Coleman warned that the area of Thailand in which the study was conducted might not be considered representative of the purported targets of microfinance in poorer countries; Thailand is a relatively high-income developing country and not particularly credit constrained. Coleman's study is notable for the very large number of variables assessed, and the relatively unsophisticated econometric analyses conducted. The latter characteristic is repeated in many of the subsequent pipeline designs, using basic DID estimations without many control variables or two-stage estimations, even thought a prior study using a very similar design by Steele et al (2001) (originally 1998), used more sophisticated methods. In the case of the latter study, with the exception of one area which was likely highly untypical, the results were also not significantly positive for the impact of microfinance on modern contraceptive adoption.

PSM has been applied by one recent study (Setboonsarng and Parpiev 2008) to pipeline data in the expectation that this would provide robustness to DID estimates; this somewhat illusory since in this case, all that occurs is a sub-setting of the control and treatment cases to the common support region. That is, with existing survey data produced from specific sampling of treatment and control groups, some cases, especially of controls, are dropped because they appear quite dissimilar to treatment cases. This may increase statistical precision, but at the cost of the variability that may be the loss of an important part of the data. No sensitivity analysis of the PSM result is conducted, further reducing the merits of this approach. PSM may be better used when there are other sources of data which may yield plausible control cases, as in Dehejia and Wahba (1999).

As our data extraction shows, pipeline studies have on the whole come to similar conclusions to Coleman, namely that microfinance has little or no statistically significant effect on well-being outcomes measured, even when there are positive and significant effects on variables such as borrowing and business activities. Several of these studies confirm that participants in microfinance are not among the poorest, and indeed some of the studies should perhaps have been excluded on these grounds. As an oeuvre these studies therefore fail to lend support to claims of beneficence of microfinance. The findings of the Deininger and Liu (2009) study are insufficiently reliable to warrant further examination.

The question arises, however, whether this finding reflects issues in the implementation of the studies – their geographical location and sampling units (for example micro-enterprises, or not the poorest households) – or their research design and its implantation, i.e. raising issues about the pipeline design itself. This design is attractive if only for ethical reasons that it does not entail denying poor people potential access to what might be a beneficial resource, that is discriminating against some. However, by its nature, and because of graduation, drop-outs, and spill-overs, the design has limitations to the gap in time between treatment in treatment areas then in pipeline areas means that there is only a small window of time within which the two groups can be considered differently treated. This will limit impacts to the relatively short term, and as noted above, it is likely that many social impacts may only be manifest in the longer term.

Identifying early indicators of later success[41] could be suggested, but is a problematic area; structural modelling approaches may prove useful - but the assessment of these is outside the purview of this report.

The majority of included studies in this SR apply a basic with/without design (see papers referred to in Table 14). This design itself is problematic as discussed elsewhere (see section 2.2.2, Table 2 and Appendix 7, section 6.7.1.3) mainly due to selection and programme placement bias. Various econometric techniques are available to deal with these biases but have shortcomings in one way or another.

Table 14 (with more details in Appendix 12 and 13, sections 6.12 and 6.13) provides a summary of the key characteristics of the with/without, before/after and non-pipeline panel studies included in this SR which made the cut-off point of the final (stage 3) scoring scheme, i.e. they have a score of < 2. It can be seen from the table that selection bias prevails in most studies, and hence advanced econometric techniques are often employed to control for selection bias, but with limited success as discussed further below.

3.4 With/without, before/after and panel

This section is organised as follows, we first discuss two iconic studies in depth: PnK and USAID, since these two studies have not only generated more than half of all microfinance IEs which survived to stage 3 (29 out of 58), but are also considered to be particularly influential evaluations in the context of microfinance and are widely quoted in support of the presumptive beneficence of microfinance. We closely examine these studies to illustrate the problems of with/without studies using advanced econometric methods (IV and panel techniques). We show that the more recently popular PSM method applied to these data does not overcome these limitations. To some extent limitations derive from the designs actually used in these studies. We then discuss the remaining with/without studies and focus on those that applied two-stage methods (including most significantly IV) and PSM.

[41] For example, the attempt to draw ‘Late lessons from early warnings’ as advocates of the precautionary principle suggest (Harremoës et al. 2001), as well as other searches for early signs of later values.

Table 14: Summary of studies using with/without, before/after and panel design

Study			Treatment					Control			
	Design	Method	Random selection sample	Self-selection	MFI selection	Peer selection	Drop-out and/or graduate	Random selection	Same time	Different area	Control larger than treatment
Abera 2010	with/without within community - panel	PSM, panel data analysis	Unclear	Unclear	Unclear	Unclear	N	N	Y	Y	Unclear
Abou-Ali et al. 2010	with/without	PSM	Y	Unclear	Unclear	Unclear	N	Unclear	Unclear	Unclear	Unclear
Bhuiya and Chowdhury 2002	with/without - before/after	Multivariate	N	Y	Y	Y	N	N	Unclear	Unclear	Unclear
Cuong 2008	panel	IV, panel data analysis	Y	Y	Y	Y	N	Y	Y	N	Unclear
Diagne and Zeller 2001	with/without	Two-stage LIML	Random within stratum	Y	Unclear	Y	Unclear	Random within stratum	Y	Unclear	Unclear
Imai et al. 2010	with/without	PSM	Random within stratum	Y	Unclear	Y	Y, sampled but unclear how that sample was used	Random within stratum	Y	N	N
Imai and Azam 2010	panel	PSM, panel data analysis	Random within stratum on village level	Y	Unclear	Unclear	Unclear	Random within stratum on village level	Y	Y	N
PnK[1]	with/without - panel	ML-IV, PSM, cmp, panel data analysis	Y	Y	Y	Y	N	Y	Y	Y	N
Shimamura and Lastarria-Cornhiel 2010	with/without	IV	Paired-site sampling	Y	Unclear	Y	N	Paired-site sampling	Y	Y	N
Shirazi and Khan 2009	with/without - before/after	DID	Unclear	Unclear	Unclear	Unclear	N	Unclear	Y	Unclear	Unclear
Swain and Wallentin 2009	with/without - panel by recall	RML (robust maximum likelihood) method	Y	Y	Unclear	Y	N	Unclear	Y	Unclear	N
Takahashi et al. 2010	b/a	PSM, DID	N	Y	Y	Y	Y	N on village level but Y within villages	Y	Y	Y
Tesfay 2009	panel	Panel data analysis	Y	Y	Unclear	Y	N	Y	Y	Y	N
USAID[2]	w/wo - panel	ANOVA, ANCOVA,	Y	Y	Y	N	Y - by some papers	Y	Y	N	N

		PSM, panel data analysis									
Zaman 1999	w/wo	Two-stage Heckman	Unclear	Y	Y	Y	N	Unclear	Y	N	N
Zeller et al. 2001	w/wo	IV	Y at village level and within stratum at household level	Y	Unclear	Y	Y, briefly discussed	Y at village level & within stratum at household level	Y	N	N

Notes:

1. Papers dealing with PnK data include: Chemin (2008), Duvendack (2010b), Duvendack and Palmer-Jones (2011), Khandker (1996, 2000, 2005), Khandker and Latif (1996), Khandker et al. (1998), Latif (1994), McKernan (2002), Menon (2006), Morduch (1998), Nanda (1999), Pitt (1999, 2000), Pitt et al (1999, 2003), Pitt and Khandker (1998), Pitt et al. (2006), Roodman and Morduch (2009).

2. Papers dealing with USAID data include: Augsburg (2006), Barnes (2001), Chen and Snodgrass (1999 and 2001), Dunn (1999), Dunn and Arbuckle (2001), Duvendack (2010a and 2010b), Tedeschi (2008), Tedeschi and Karlan (2010).

Explanation of column headings:

Design:	what research design
Method:	what statistical/econometric method
Random selection sample:	are units randomly sampled
Self-selection:	have individuals/households self-selected into microfinance
MFI selection:	does the MFI exercise control over selection of treated units (i.e. into treatment – not into sample)
Peer selection:	do peers play role in selection of treated units
Drop-out and/or graduate:	are dropouts and graduates included in treatment sample
Random selection:	are controls randomly allocated to control – i.e. self selected not to be treated
Same time:	are controls selected at same time as treatment
Different area:	do the controls come from the same geographical domain as treatment.
Control larger than treatment:	adequate sample size for matching

Table 15: Significance and sign of estimates of with/without papers by outcome variable and its location in causal chain using IV methods

Outcome category	Location in causal chain	Sign & significance					
		ns			sig		
		+	-	total	+	-	total
Economic	inputs						
	effect	17	16	33	31	10	41
	impacts	13	13	26	46	35	81
	total	30	29	59	77	45	122
Social	inputs						
	effects						
	impacts	9	5	14		6	6
	total	9	5	14		6	6
Empowerment	inputs						
	effects						
	impacts	13	1	14		8	8
	total	13	1	14		8	8

As opposed to the results presented in Tables 11 and Table 13, Table 15, which summarises the IV studies (excluding PnK and the remaining with/without studies), assesses a higher proportion of effects that occur later in the causal chain, i.e. at the impact stage, and few at the imput stage. Economic outcomes are frequently estimated to be statistically significant by the IV method, but rather more likely to be negative than positive. The 'impacts' on social and empowerment outcomes are less likely to be significant and as if not more likely to be negative as the economic impacts (recall that we have reversed the sign of impacts on poverty indicators so that an estimated fall in poverty appears as a positive sign in this table).

3.4.1 PnK studies (Bangladesh)

The study conducted by PnK uses cross-sectional data from a World Bank funded study which conducted a survey in 1991-2 on three leading microfinance group-lending programmes in Bangladesh, namely GB, BRAC and BRDB (PnK p959). According to Morduch, at the time these three programmes catered to more than four million microfinance clients in Bangladesh (p2). A quasi-experimental design was used which sampled target (having a choice to participate/eligible) and non-target households (having no choice to participate/not eligible) from villages with microfinance programme (treatment villages) and non-programme villages (control villages).

The survey was conducted in 87 villages in rural Bangladesh; 1,798 households were selected out of which 1,538 were target households (eligible[42]) and 260 were non-target households (not eligible). According to PnK, out of those 1,538 households, 905 effectively participated in microfinance (59%). Data were collected three times in the 1991-2 period in order to account for seasonal variations, i.e. various rice harvest seasons exist, namely Aman (November - February) which is the peak season, Boro (March - June) and Aus (July - October) which is the lean season (Khandker 2005 – henceforth Khandker p271). The study focuses on measuring the impact of microfinance participation by gender on

[42] Eligibility criteria are subject to debate. PnK deem any household with landholdings of less than 0.5 acres eligible.

indicators such as labour supply, school enrolment, expenditure per capita and non-land assets. PnK find that microcredit has significant positive impacts on many of those indicators and find larger positive impacts when women are involved in borrowing.

Further to this, Khandker investigated the long-term impact of microcredit and re-surveyed the same households as in the original PnK study in 1998-9. In addition, the follow-up survey

> *also added new households from the original villages, new villages in the original thanas, and three new thanas, raising the number of sample households to 2,599 (Khandker p271).*

Khandker argued that cross-sectional data only allows the measurement of short-term impacts of microcredit and that this is short-lived. Hence, he further argued that a panel data set is needed to gauge long-term impacts of microcredit programmes, because it allows control of unobservables. Based on the panel data analysis Khandker found that microfinance benefits the poorest and has sustainable impacts on poverty reduction among programme participants. In addition, positive spill-over effects were observed such as a reduction in poverty at the village level.

A number of studies, e.g. Morduch ((1998), henceforth Morduch) and RnM, have made an attempt to replicate the findings of the original PnK study, and Chemin ((2008), henceforth Chemin) has applied PSM, but with rather contradictory results. Morduch found hardly any impact, Pitt ((1999), henceforth Pitt) defended the original claims, but Chemin and RnM found rather negligible impacts of microcredit. Duvendack (2010b) and Duvendack and Palmer-Jones (2011) replicated the key studies related to PnK and applied PSM as well as sensitivity analysis concluding that PnK's original findings cannot be confirmed. It is beyond the scope of this report to discuss the drawbacks of the main studies involved in the re-examination of PnK in depth and the interested reader is referred to Duvendack (2010b) for a detailed discussion of this topic.

To conclude, the replication of PnK and associated studies posed a challenge due to the complex research design and poor documentation. All studies that dealt with the PnK data, i.e. Morduch, Chemin, RnM, Duvendack (2010b) and Duvendack and Palmer-Jones (2011), agree that PnK overstate the impacts of microcredit. PnK estimated positive and significant impacts for literally all of the six outcome variables with stronger impacts when women were involved in microcredit (PnK p987-988). Morduch argues that PnK overestimated the impact of microcredit because the eligibility criteria were not strictly enforced, i.e. he cannot support PnK's claims that microcredit increases per capita expenditure, school enrolment for children (Morduch p30) or labour supply. Chemin finds lower impact estimates than PnK, though for half of the outcome variables, such as male labour supply and children's school enrolment, he finds a significantly positive impact which contradicts Morduch's findings. Doubts about both Morduch and Chemin arise because of problems in replicating their data constructions. RnM's and Duvendack's (2010b) findings are mixed and mostly insignificant. The reasons for these discrepancies across studies can be explained by shortcomings in the empirical strategy that PnK put forward, e.g. the application of eligibility criteria was not strictly enforced, hence, a problem with mistargeting occurred.

Moreover, the studies by PnK, Morduch, Chemin and RnM neglect the role of multiple sources of borrowing which has implications for the nature of the control group (i.e. whether it is appropriate), the accuracy of the impact estimates as well as the appropriate definition of the counterfactual. As a result, Duvendack (2010b) and Duvendack and Palmer-Jones (2011) proposed novel treatment group comparisons to examine the impacts found using these more appropriate, and homogeneous, control groups. This strategy found mixed results when comparing microcredit participation with participation in other non-microcredit schemes, and so there is no clear evidence for or against microcredit as such. However, it appears that the utilisation of finance in general has significantly positive impacts across all outcome variables and indicates that other sources of finance can be as effective as microcredit. Many practitioners agree that individuals essentially need to borrow from multiple sources to obtain sufficient funds that would allow them to engage in more productive activities. Many microcredit loans are often too small to meet the needs of microentrepreneurs (Venkata and Yamini 2010). In addition, multiple sources of borrowing are often required to smooth income and consumption patterns as well as to cope with emergencies (Venkata and Yamini 2010). Moreover, Coleman (1999), Fernando (1997) and Venkata and Yamini (2010) find that it is common for individuals to use borrowing from one source to pay off the loans of another on time. Overall, criticisms of the more strident and unqualified claims about microfinance are becoming more common and further investigations as to the impact of microcredit versus other financial tools should be encouraged, i.e. using RCTs or carefully designed observational studies that allow the collection of rich and high quality datasets.

According to Armendáriz de Aghion and Morduch (2010), Morduch and RnM, PnK's 'econometric set-up [*here*] is not up to the task. We need to look elsewhere for reliable evidence' (Armendáriz de Aghion and Morduch 2010, p290). Furthermore, the re-analysis of PnK using PSM has raised doubts about the appropriateness of PSM in the context of PnK (see Chemin; Duvendack 2010b; Duvendack and Palmer-Jones 2011). There are often too few matches of low quality due to a small control group sample size; this adversely affects the reliability of the matching estimates. Rich, high quality and large data sets are needed which ideally contain more control than treatment observations (Smith and Todd 2005). Moreover, Duvendack (2010b) and Duvendack and Palmer-Jones (2011) apply sensitivity analysis which indicates that it is not unlikely that unobservables could result in over- or underestimating the impact of microcredit.

As to the panel data analysis, Khandker and RnM argued that longitudinal studies remedy the shortcomings of cross-sectional studies but this appears not always to be the case. The panel data results of the random effects model did not provide any new insights and generally confirmed the findings of the cross-section data analysis (Duvendack 2010b). Khandker, Koolwal and Samad (2010) among others, claim that the combination of PSM and DID is the way forward since it allows controlling for observable as well as unobservable characteristics assuming that the latter remain constant over time (as mentioned above). However, the results of the PSM/DID model did not offer anything different to what was found by the random effects model. As in the case of the USAID studies (discussed below), doubts are raised about the ability of techniques such as PSM and DID to account for selection on unobservables with the PnK data perhaps because it is not a 'true' panel which would allow a before/after comparison with a more demonstrably appropriate control group. What is compared is the change in outcomes between a group that was already participating in microfinance during the baseline and a control group surveyed at the same time, with both groups at a later date. This comparison is not adequate for reliably assessing the impact of microcredit and controlling for unobservables because any differences between the treatment and control groups before microfinance cannot be empirically observed in these data.

Overall, what can be learnt? The results provided by the numerous studies dealing with the PnK data are rather mixed ranging from significantly positive impacts to significantly negative ones depending on the econometric techniques applied. However, Morduch, Chemin, RnM, Duvendack (2010b) and Duvendack and Palmer-Jones (2011) agree that PnK and Khandker most likely overstated their impact estimates and the replication of their original findings is challenging[43]. Thus, methodological problems still remain particularly selection bias due to unobservable characteristics, inappropriate counterfactuals, and poor data quality as well as control groups that are contaminated and limited in size. In particular the control group is far too small to provide convincing matches, which hampers the usefulness of PSM in the context of PnK.

To sum up, poor quality data, poor research design and possibly inappropriately implemented econometric techniques fail to illuminate the role of the unobservables. Sensitivity analysis of the matching results indicated that the unobservables could readily confound impact estimates which are demonstrated to be not robust to unobservables (Duvendack 2010b, Duvendack and Palmer-Jones 2011).

Table 16 presents the results of all PnK studies and supports evidence presented in Tables 11 and Table 13; not only are the outcomes for which estimates are provided ones which occur early on in the causal chain, predominantely at the ‘effects’ stage in this case, but also while the effects are mainly positive, the majority are not statistically significant. This follows from the discussion above, namely that Chemin, RnM, Duvendack (2010b) and Duvendack and Palmer-Jones (2011) have cast doubts about the reliability of these estimates because of methodological flaws of the PnK study, and as replication has shown that the original authors seem to have overestimated the results that can be produced from the data of this study.

Table 16: Significance and sign of estimates by outcome variable and its location in causal chain (PnK)

Outcome category	Location in causal chain	Sign & significance					
		Ns			sig		
		+	-	total	+	-	total
Economic	inputs	63	57	120	68	47	115
	effects	111	107	218	193	39	232
	impacts		8	8			
	total	174	172	346	261	86	347
Social	inputs						
	effects		1	1			
	impacts	66	26	92	55	5	60
	total	66	27	93	55	5	60
Empowerment	inputs						
	effects						
	impacts	25	32	57	63	6	69
	total	25	32	57	63	6	69

[43] Thus Pitt (2011) challenges RnM (2009) on the grounds that they make a logical and a coding error in their replication, but these errors do not apply to Duvendack (2010b) and Duvendack and Palmer-Jones (2011) who use a different analytical method.

Note that, largely because of the failure to replicate the results of the earlier study, and the generally much less positive results that have emerged from replications of the PnK study that have been undertaken, we decided that it was beyond the scope of this section to review the results of other included papers based on the PnK dataset until they have been replicated. The interested reader is referred to Appendix 17, section 6.17.

3.4.2 USAID studies (Peru, India, Zimbabwe)

These three longitudinal studies aim to evaluate the impact of microfinance on poor people; USAID produced three panel datasets, each with two waves in the late 1990s 2-3 years apart. The studies seem to have employed a common dataset and the panels seem to contain the same or very similar variables, hence we only discuss USAID's India study as an example. The Indian study was on the microfinance operations of a well known NGO - the Self Employed Women's Association (SEWA) – based in Ahmedabad, Gujarat, in western India. There are data at the individual, household and enterprise levels which can be used for panel difference in difference and PSM estimation of impact (Duvendack, 2010a). The problems encountered in these three IEs are similar to those encountered in PnK. There are doubts about the robustness of the USAID sampling procedure of the control groups. Taking the example of the USAID study on SEWA Bank, Chen and Snodgrass (2001) argue that the neighbourhoods where most of SEWA Bank's clients reside are reasonably homogeneous in terms of caste, occupation and class (p53) and hence the control group is relatively similar to the treatment group. However, if the households in the control group are so similar, then why are they not clients of SEWA Bank? This question points towards a selection process that is driven by unobservable characteristics which account for why otherwise apparently eligible households did not belong to SEWA Bank. As a consequence, the control group sampling of USAID does not convince. Chen and Snodgrass (2001) admit that SEWA Bank members

> *are not chosen at random but are in fact purposefully selected from a larger population, both by themselves and by SEWA Bank. A woman must first self-select by deciding to open a savings account and later to apply for a loan. Once she does so, SEWA Bank decides whether to provide her with the financial service in question (p60).*

The headline findings of the USAID study provide evidence that microfinance leads to changes at the household level, i.e. higher household income in terms of total income and per capita income was observed. In addition, minor positive impacts could be observed on income diversification, food expenditure and the ability to cope with shocks. However, the evidence was rather mixed. Moreover, impact at the enterprise and individual levels were negligible. Chen and Snodgrass (2001) admit that measuring impacts at the enterprise and individual levels were rather challenging due to the fact that SEWA Bank clients are not classical micro-entrepreneurs per se. Most clients do not have microenterprises but are dependent sub-contractors or labourers, thus do not require microenterprise capital. SEWA Bank provides loans for a range of purpose, e.g. business, housing improvements/repairs, repayment of other debts and consumption but without a particular focus on microenterprise development.

Augsburg (2006) and Duvendack (2010a) re-visit the evidence of the USAID SEWA Bank panel dataset and subject it to PSM and DID to account for selection bias. The results presented by these two studies broadly confirm the findings of the USAID study if one ignores selection on unobservables. However, doubts remain as there are strong qualitative (see for example Ito 2003) and theoretical reasons to think that unobservables have not been fully controlled for, as argued by Duvendack (2010a). This is not too surprising at least in the case of PSM since its drawbacks are well-known although still debated (see debates between Dehejia and Wahba 1999, Smith and Todd, 2005). This notion is confirmed by the sensitivity analysis which shows that SEWA Bank's matching estimates are quite sensitive to selection on unobservables (Duvendack 2010a). Also, the quality of the matches is doubtful, considering PSM requires rich and large datasets in order to function properly (Heckman et al. 1997, Heckman et al. 1998, Smith and Todd 2005). Moreover, the panel does not resolve the issue because it is not a 'true' panel (it does not allow a before/after comparison), and, even if it were, might not control for the effects of unobservables. Microfinance clients might have been better off than non-clients even before participating in microfinance, i.e. in terms of access to social networks, wealth, skills or motivations (Armendáriz de Aghion and Morduch 2010). This may in turn have led them to self-select or to be selected into microfinance either by their peers or the staff of the microfinance organisation, and to be able to benefit more from membership than otherwise observationally similar households. The re-investigation of USAID's SEWA Bank study and PnK provides evidence that reduces the credibility of the quantitative support for microfinance and for lending to women in general. Furthermore, qualitative evidence (Fernando 1997) strongly suggests other less beneficent interpretations leading to an unraveling of the microfinance narrative.

We do not provide a separate table for the USAID studies on the significance and sign of the estimates by outcome variable and its location in the causal chain due to the sheer number of estimates provided by the USAID studies; we could not extract all to our database due to time and budget constraints.

3.4.3 Other with/without studies

First we discuss in general terms the characteristics of the papers which have applied PSM, IV, panel and other methods of analysis to with/without and panel studies. Summaries of selected individual papers can be found in Appendix 15, sections 6.15.5 and 6.15.6 with summaries of key characteristics in Appendix 12 and 13, sections 6.12 and 6.13, to allow an assessment of their reliability and the impacts that can be drawn from them.

The studies in Appendix 12 all use a two-stage approach in common; using IV, Heckman or LIML models which all require the selection of appropriate identification variables(s) commonly termed instruments. The two-stage methods applied to cross-sectional data (as also when applied to panel data) may be able to address some endogeneity and selection bias problems, but are crucially dependent on the availability of appropriate instruments. As discussed earlier, tests for assessing the validity of instruments can be made (e.g. Sargan-Hansen and Hausman tests), but not all studies report these tests (see Appendix 12 for a two-stage checklist).

Almost all of the studies in Appendix 12 are classified as having a high risk of bias since they have not sufficiently demonstrated the validity of their instruments using the various tests available and even if they have, doubts remain about the trustworthiness of these tests as argued by Deaton (2010).

The study by Cuong (2008) is the only one that is classified as moderate but only because it is a panel - the drawbacks of panels are discussed above and below. Ideally, to fully understand the studies listed in Appendix 12 and to verify their findings, they should be replicated. However, given the time and budget constraints of this SR and availability of data, it is beyond the scope of this study to attempt any replications.

Table 17: Significance and sign of estimates by outcome variable and its location in causal chain (other with/without studies)

Outcome category	Location in causal chain	Sign & significance					
		Ns			sig		
		+	-	total	+	-	total
Economic	inputs	1	1	2	3		3
	effects	53	43	96	125	16	141
	impacts	6		6	11	7	18
	total	60	44	104	139	23	162
Social	inputs						
	effects						
	impacts	12	14	26	4	4	8
	total	12	14	26	4	4	8
Empowerment	inputs						
	effects						
	impacts						
	total						

Table 17 summarises the directions and significances of the various outcomes of the other remaining (other than PnK and USAID) with/without studies by their location in the causal chain. As before, the majority of outcomes tested are early on in this chain, at the 'effects' stage, with a majority of significant and positive results. However, as we have indicated earlier, with/without designs can be problematic due to selection bias and limitations of the analytical methods commonly used to control for it. These methods have shortcomings; hence caution is required as to the trustworthiness of these results – a brief discussion of the pros and cons of the various techniques commonly used in the context of with/without studies follows below. We note that it is not unusual for the methodologically weaker papers to find more positive and significant results (Schulz et al. 1995), and it seems not unreasonable to apply this insight to these studies.

Since IV approaches are not free from controversy, many researchers advocate the use of panel data to deal with selection bias, since unobserved unit level confounders can be swept out of the analysis by differencing. Analysis of fixed or random effects models can test whether unobservables are likely to bias estimated coefficients. However, panels have shortcomings too as seen in the case of PnK and USAID; panel datasets require a 'true' baseline, i.e. the respondents should not have been microfinance participants at the time of the collection of the baseline dataset, which in the instance of most of the panel studies included in this section (e.g. PnK (Khandker. 2005), USAID, Abera 2010, Cuong 2008, Imai and Azam 2010, Tesfay 2009, Swain and Wallentin 2009 which established a panel by recall method) is not the case. In most of panel datasets the baseline included participants who had already been members of a microfinance programme for some years, and consequently cannot be shown to be indistinguishable from the control group. Moreover, it is assumed that the unobservables are time-invariant (as argued by Cuong 2008 for example), and that observables do not have diverse time varying effects. However, these assumptions usually do not hold and the changing nature of the unobservables cannot be accommodated by fixed or random effects models, and some studies do not include these analyses.

More recently, PSM has increasingly been used in the context of microfinance IEs (examples include Abera 2010, Abou-Ali et al. 2010, Imai et al. 2010, PnK (the Chemin and Duvendack papers) and USAID (the Augsburg and Duvendack papers) and Takahashi et al. 2010). However, PSM is not the wondrous tool advocated by many and researchers must follow certain procedures to ensure their matching estimates are robust, e.g. they must assess the quality of their matches amongst other things (see Appendix 13, section 6.13 for a PSM checklist). Caliendo and Kopeinig (2005, 2008) briefly outline the various procedures that are available to assess matching quality, e.g. t-tests and stratification tests to investigate whether the mean outcome values for both treatment and control groups are significantly different from each other. Only in the PnK, USAID and Takahashi et al. (2010) studies is matching quality assessed. One way to improve matching quality is to have more control than treatment households as well as reasonably homogenous groups, a requirement that only the study by Takahashi et al. (2010) fulfils (see Table 14 (last column) and Appendix 13).

Furthermore, Dehejia (2005) points out that the correct specification of the propensity score is crucial, i.e. the balancing properties of the propensity score should be satisfied. Appendix 13 shows that 5 of 7 studies have given evidence on whether there were good quality, balanced[44] covariates characterising participation. As discussed earlier, by itself PSM provides little information on the robustness of the estimates such as provided by confidence intervals in standard statistical estimation; hence sensitivity analysis should be applied since it can provide some (contested) evidence on robustness – as advocated by Rosenbaum (2002), Becker and Caliendo (2007), Ichino et al. (2006) and Nannicini (2007).

Appendix 13 shows that none of the microfinance studies that applied PSM (e.g. Abera 2010, Abou-Ali et al. 2010, Augsburg 2006, Chemin 2008, Deininger and Liu 2009, Imai et al. 2010, Imai and Azam, 2010; Setboonsarng and Parpiev, 2008; and Takahashi et al. 2010) used sensitivity analysis to assess the robustness of their matching estimates and hence the risk of bias for most of them is high. Exceptions in this regard are the studies by Duvendack (2010a, 2010b), Duvendack and Palmer-Jones (2011) and Abou-Ali et al. (2010), but Abou-Ali et al's interpretation of sensitivity analysis is flawed and they do not assess the sensitivity of their microfinance results (discussed in more detail in Appendix 15, section 6.15.5.1.

The studies by Bhuiya and Chowdhury (2002) and Shirazi and Khan (2009) are the only two remaining with/without studies that do not use two-stage or PSM techniques but rather basic methods of analysis, such as OLS multivariate analysis and DID with non-comparable control groups without control functions, which do not overcome the problems that commonly plague with/without studies. Hence, these two studies do not really warrant further discussion here.

3.4.4 Conclusion of with/without studies

The majority of microfinance IEs were applying a with/without design but this is gradually changing and pipeline studies as well as RCTs are slowly taking centre stage. These developments are hardly surprising due to the challenges of conducting with/without studies, i.e. the presence of placement and selection bias. Placement and selection biases can partially be mitigated by having a sound research design (e.g. appropriate control groups, inclusion of drop-outs – as outlined in Table 14) and applying sophisticated econometric methods such as PSM, IV, fixed and random effect estimations, etc.

[44] I.e. the characteristics of the covariates of treatment and control samples should have similar distributions.

The with/without studies we include in this SR, listed in Table 14, investigated a wide range of outcome variables providing rather mixed results from significantly positive to significantly negative - thus it is difficult to find a common thread. As outlined in Appendices 12 and 13, section 6.12 and 6.13, most studies are classified as having a high risk of bias. In the case of two-stage studies this is because they either did not test the validity of their instruments, or doubts about the rigor with which these tests were conducted remain. The majority of the PSM studies classified as having a high risk of bias did not sufficiently demonstrate high matching quality and/or did not apply sensitivity analysis, hence we assume that selection bias has not been accounted for. Panel studies in combination with either PSM, IV or DID have all been categorised as having a moderate risk of bias mainly because of the assumption that unobservables are time invariant and can therefore be accounted for by panel data techniques and that observables were controlled for through IV or PSM. However, doubts remain about the validity of panel studies presented here due to a lack of 'true' baselines, and doubts about the assumption that the unobservables are time invariant. Hence, a healthy scepticism about the reliability of results of with/without studies included here is in order, precisely because of the problems that commonly plague them.

The discussion earlier has stressed that various analytical methods that attempt to correct placement and selection bias have drawbacks in one way or another – as discussed above and further in Appendix 7, section 6.7.2. In addition, many studies in this section did not rigorously apply these techniques; hence risk of bias for almost all studies is relatively high. For example, in the case of the two-stage studies, testing the validity of instruments is often not done and even if conducted, doubts remain about the ability of these tests validate instruments; in the context of PSM, sensitivity analysis is rarely conducted. This has adverse implications for the robustness of impact estimates, therefore doubts about the reliability of outcomes presented by most with/without studies remain.

3.5 Gender empowerment

Some included studies address issues of female empowerment; one RCT assessed the impact of basic MFI focused on lending to women micro-entrepreneurs. This study concludes it 'appears to have no discernible effect on education, health, or womens' empowerment' (Banerjee et al. 2009, p30). Of course we have drawn attention to potential limitations of this study, which mean that it may not have adequate statistical power; the authors themselves also highlight the short study period during which effects could occur may be insufficient for such impacts to appear.

Among the pipeline studies Deininger and Liu (2009) are concerned with gender empowerment of a self-help group microfinance project in the state of Andhra Pradesh in India, and Steele et al. (year) are concerned with contraceptive use rather than indicators of intrinsic well-being. While Deininger and Liu (2009) find positive impacts, this study has high vulnerability to bias, and so cannot be taken as useful evidence of impact on gender empowerment[45].

[45] The Steele et al. (2001) study has higher credibility, and finds a positive impact of microfinance on contraceptive use in similar context and time period to Pitt et al (1999), who find no impact. Steele et al. (2001), rationalise this by arguing the broader indicator of microfinance empowerment that they use – membership – is more appropriate than the indicator used by Pitt et al. (1999) – borrowing – which identifies only those MFI members who borrow as 'empowered'. This is a narrower definition as it includes MFI group members who do not borrow but who may be 'empowered' through discussions at group meetings and so on (Steele et al. 2001, p280). It may be that the types of social interactions of the groups which were subjects of the Steele et al. (2001) study were somewhat different with different implications for empowerment; PnK assessed impacts of GB, BRDB and BRAC, while Steele et al. evaluated ASA and Save the Children (USA). There

There have been very few with/without studies taking a closer look at women's empowerment in the context of microfinance applying a mixture of methods, i.e. sample survey and case study methods, examples include Hashemi et al. (1996), Schuler and Hashemi (1994) and Goetz and Sen Gupta (1996). These three studies examine microfinance programmes in Bangladesh using a range of empowerment indicators such as mobility, economic security, decision-making, freedom from domination by family, political and legal awareness and participation in public and political life. These and other qualitative studies find negative as well as positive outcomes for women, especially if the enterprises in which they invest are not successful, or if notwithstanding being members of a MFI, they do not receive loans because their peers or the MFI judge them not creditworthy.

However, all three studies have been excluded from this SR since they either did not meet the inclusion criteria in the first instance or failed the scoring assessment. Of the included with/without studies, the PnK study and many related PnK papers (notably Pitt et al. 2006) investigate empowerment issues, as does the study by Zaman (1999). The USAID studies indirectly investigate women's empowerment by using decision-making as proxies for empowerment but with rather mixed results. Women's empowerment is notoriously difficult to measure and the few quantitative studies included in this report that do examine empowerment issues might lack credibility due to unresolved issues of measuring womens' empowerment. However, a wealth of qualitative studies exists (Todd 1996) suggesting that the perception of women within their communities changes due to their activities in microfinance. We are not examining the qualitative evidence further since that is beyond the scope of this SR and most are anecdotal. However qualitative evidence has often suggested that women become more involved in household and community decision-making or gain more control over resources. Much of this evidence has been assessed as 'inspiring stories', which do not amount to convincing positive evaluation in the face of more ambiguous quantitative evidence (Armendáriz de Aghion and Morduch 2010). In an otherwise sympathetic review, Kabeer (2005b), pointed out that the mainly qualitative evidence she reviewed suggested effects that are highly contingent on 'context, commitment and capacity' (p4709), the assessment of which would take us well beyond the scope of this study.

3.6 Discussion

Microfinance has recently become a highly contested arena in recent years, although for many years there have been some dissenters from the mainstream enthusiasm (Bateman 2010, Roy 2010). There have also been those in the health care arena who doubt the ability of SRs to resolve long standing disputes, even when the evidence of SRs are extremely robust (e.g. the MMR dispute, or the effectiveness of homeopathy). An SR may not convincingly resolve issues. As we

are differences of the location of the studies as well. However, while coming to a different conclusion to that derived by Pitt et al. (1999), from the PnK study using the two-stage approach used in most PnK papers, which we have argued is not robust based on the replication of PnK done by RnM and Duvendack and Palmer-Jones (2011), the argument of Steele et al. (2001) is of course at best indirect evidence on empowerment, since it is inferred as a latent causal variable of contraception adoption. Given the concerns we have about both studies, it would be best to replicate the Pitt et al. (1999) paper, which is possible since it would be relatively modest extension of our (i.e. Duvendack and Palmer-Jones, 2011) existing replication. It might also be desirable to replicate Steele et al. (2001), although the underlying problems with the design would still not raise it to have low bias vulnerability (and we do know what other problems might be encountered in an attempt at replication). Nevertheless, if both replications came to similar conclusions with regard to the effect of microfinance membership on contraceptive use, it would be suggestive evidence of whatever conclusion was reached.

argue throughout this SR, there seem to be two particular issues that readers are likely troubled by in evaluations of microfinance; firstly, the role of complex and sophisticated statistical, or econometric methods in prominent evaluations, and secondly why evidence and arguments that are convincing to the authors of this SR, such as that contesting the mainstream microfinance discourse, have not hitherto commanded more attention in the field[46].

Econometric methods are generally given high prestige in the economics literature and in policy analysis, at least by economists; we are more sceptical. We find them less than convincing in many cases, and refer the doubting reader to the literature that is critical of much of this work (e.g. Leamer 1983, and the symposium in the Journal of Economic Perspectives 2010, 24(2)). We present critical assessments of analyses above, but cannot elaborate due to constraints of time and space. We draw attention to the fragility of econometric results as shown by the relatively infrequent attempts to replicate results of econometric work (Hamermesh 2007). Few replications of microfinance evaluations have been undertaken (see above, section 2.2.3). Sceptical readers could consider the recent moves, for example by the American Economic Association[47], towards developing codes of ethical practice in economics requiring deposition of data and code allowing third party replication, as well as calls for economists to reveal potential conflicts of interest (Economist,2011[48]). Calls for better conduct with regards publication are also found in the medical literature[49] and in other academic fields[50].

It is also not unusual, in the health care arena, for SRs to conclude that there is limited robust evidence to support the continued application of existing, widely-used interventions, and that the current uncertainties ought to be subjected to further rigorous evaluation (Chalmers 2008). Indeed, there are examples of SRs that have led to considerable controversy because the widely adopted interventions reviewed were shown to be potentially harmful (Cochrane Albumin 1998, Singh et al. 2007). Ioannidis et al. (2008) point out that the aim of SR and meta-analysis should not be limited solely to generating a single pooled estimate, but should extend to examining ‘consistency of effects’ as well as promoting ‘understanding of moderator variables, boundary conditions, and generalisability’ (Ioannidis et al. 2008). Hence, evaluation of results according to different study designs and analytic methods as we have done in this microfinance review can prove informative. A similar example can be seen in a healthcare meta-analysis that found substantial heterogeneity in results of randomised trials, propensity score matched studies and multivariate adjusted observational studies (Kwok and Loke 2010).

In order to critically assess the quality of papers[51] our classification apparently differs from two other schemes used to classify studies – the Scottish Intercollegiate Guidelines Network (SIGN) and the World Cancer Research Fund (WCRF). Both classificatory schemes require judgements about the quality of execution of studies as well as design features[52]. Taking the SIGN classification

[46] For example, Fernando 1997, and some arguments marshalled in Bateman 2010.

[47] www.aeaweb.org/aer/data.php

[48] www.economist.com/node/17849319

[49] CONSORT (www.bmj.com/content/340/bmj.c332.full)

[50] See the Committee on Publication Ethics (COPE): (http://publicationethics.org/about).

[51] We discuss below the problem that an excessively critical approach to papers leads to a higher likelihood of making type 2 errors (failing to accept evidence of impact when there is in fact impact).

[52] The SIGN classification assesses ‘Levels of evidence’ as follows (www.sign.ac.uk/guidelines/fulltext/50/annexb.html):

listed below, it is clear that most of the studies included in our review would be rates 2- or below. These classificatory schemes apply to fields with a much higher proportion of rigorous research designs that are available in most social science arenas. Such classificatory approaches are not very helpful although they lend support to our argument that evidence to contradict the null hypothesis of no impact is of low quality.

In order to have more than a few studies we broadened the criteria of acceptability, but doing so opened the floodgates; we decided to use an intuitively plausible but arbitrary classificatory scheme with a cut-off that produced a good number of studies. The cut-off included most studies we had already graded and subjectively agreed in the screening phase as having some merit. We classified the research designs used in microfinance IEs into five broad categories; we put studies into descending (according to conventional assessments) order of internal validity these are – RCTs, pipelines, with/without comparisons (in panel or cross-sectional form), natural experiments and general purpose surveys. These five categories were cross-classified with three categories of statistical methods of analysis, which in descending (again according to conventional assessment) order of internal validity, are two-stage IV methods and PSM, multivariate (control function), and tabulation without controls methods.

Because there were very few RCTs, which in principle do not, and many pipeline and with/without studies which do, require sophisticated analytical methods, and because we hold to the view that validity of studies based on poor research designs (of lesser internal validity) should be compensated by sophisticated statistical analysis, we adopted the general principle that weak research design requires more sophisticated methods of analysis in order to reach levels of validity to warrant review, although in general weak design cannot be fully so compensated (Meyer and Fienberg 1992, Rosenbaum 2002). Some papers use more than one method of analysis, and actual designs, data production processes, and analyses are complex and diverse; hence, papers assessed cannot be fully accommodated in such a basic two-way classification with limited numbers of categories. Nevertheless, we adopted a heuristic scoring of research designs and methods of analysis, combining these scores into a single value, which allowed us to use a single cut-off score for exclusion. A few papers which were marginally excluded by this approach were included based on our judgement, resulting in a final count of 58 included papers.

Our overall judgement draws mainly on RCTs and pipelines, although we also devoted considerable attention to the most prominent with/without studies which have been highly influential in validating orthodox favourable views of microfinance impacts. These earlier studies occurred quite early in the

1++ High quality meta-analyses, SRs of RCTs, or RCTs with very low risk of bias

1+ Well-conducted meta-analyses, SRs, or RCTs with low risk of bias

1- Meta-analyses, SRs, or RCTs with high risk of bias

2++ High quality SRs of case control or cohort or studies
High quality case control or cohort studies with very low risk of confounding or bias and high probability that the relationship is causal

2+ Well-conducted case control or cohort studies with low risk of confounding or bias and moderate probability that the relationship is causal

2- Case control or cohort studies with high risk of confounding or bias and significant risk that the relationship is not causal

3 Non-analytic studies, e.g. case reports, case series

4 Expert opinion

microfinance phenomenon, but have turned out to have low validity when subject to critical appraisal and/or when their analysis has been replicated.

There are only two RCTs of relevance to our objectives; neither has appeared in peer-reviewed form, and our judgement is that one has low-moderate and the other high risk of bias. Neither finds convincing impacts on well-being. We found nine pipeline studies, which have been reported in ten papers. All pipeline studies were based on non-random selection of location and clients[53], and most have only ex-post cross-sectional data, some with retrospective panel data information allowing (low validity) impact estimates of change in outcome variables. Thus, we deal with a set of relatively low validity papers, from which it would be unwise to draw strong conclusions. In contrast to some recent reviews, this is the conclusion we wish to emphasise, in large part perhaps because of a preference for avoiding type 1 errors. That is, we come down on the side of 'there is no good evidence for', rather than 'there is no good evidence against the beneficent impact of microfinance'.

It might be argued that we have been too critical of studies; certainly our conclusions contrast with some (Armendáriz de Aghion and Morduch 2005, 2010, RnM) but not all (Bateman 2010, Roy 2009) other recent reviews of the impact of microfinance. It is harder to classify some of reviews into either camp (Odell 2010, Stewart el. 2010, Orso 2011).

RnM state

> *In our view, nothing in the present paper contradicts those [the view that microcredit is effective in reducing poverty generally, that extremely poor people benefit most especially so when women are borrowers] ideas (p39-40).*

The views of Odell (2010), Stewart et al. (2010) and Orso (2011), are more equivocal, and we cannot be sure what impression they will leave with the reader, but just such ambivalence leaves it open for readers to draw conclusions according to their preferences. Thus reviews, systematic or other, may not resolve policy issues to the satisfaction of policy makers (Pawson et al. 2005).

Our approach seeks to shine light on this situation by being rather more critical of the methodologies employed in studies reviewed. Petticrew and Roberts (2006), in their study of 'Systematic Reviews in the Social Sciences', point to the importance of critical assessment of the quality of methodologies used in papers, and also draw aattention to the way that an excessive emphasis on methodological rigour raises the risks of type 2 errors – rejecting on methodological grounds evidence that the intervention works. But, as is well known, reducing the likelihood of type 2 errors raises that of type 1 (accepting evidence that the intervention works, when in fact it does not). We try to steer a balanced course between the Scylla of type 2 and the Carybdis of type 1 errors, but are inclined to believe that recent reviews of microfinance IE steer too close to Carybdis (although now perhaps the tide has reversed, *pace* Pitt 2011) because of an overoptimistic view of what can be achieved by sophisticated econometric methods applied to data of questionable quality from research designs that are vulnerable to bias.

Failing to contradict the alternate hypothesis encourages one to believe there is a positive effect and therefore to tend to (continue to) reject the null (no effect) hypothesis even though it (no effect) may be true. This of course

[53] Banerjee et al. (2009) would have been a pipeline study with randomised allocation of communities to treatment and control, had the baseline data turned out to be usable.

depends on the decision procedure (see Neyman and Pearson 1933, for a detailed discussion on decision rules) and weighing the costs and benefits of an intervention. Even for critics of these evaluations the absence of robust evidence rejecting the null hypothesis of no impact has not led to a rejection of belief in the beneficent impacts of microfinance (Armendáriz de Aghion and Morduch 2010, p310; Roodman and Morduch 2009, p39-40), since it allows the possibility that more robust evidence (from better designed, executed and analysed studies) could allow rejection of this nul. However, given the possibility that much of the enthusiasm for microfinance could be constructed around other powerful but not necessarily benign, from the point of view of poor people, policy agendas (Bateman 2010, Roy 2010), this failure to seriously consider the limitations of microfinance as a poverty reduction approach, amounts in our view to a failure to take seriously the results of appropriate critical evaluation of evaluations.

4 Conclusion

In this concluding section we rehearse briefly the work done in this study before drawing some more general but tentative lessons for microfinance, for research on microfinance, and for SRs in the social science arena.

Following the established medical and educational experience embodied in Cochrane and Campbell Collaborations, we searched eleven academic databases, four microfinance aggregator and eight NGO or aid organisation websites; we also consulted reviewed book, journal article, PhD, and grey bibliographies, using search terms given in section 2.1.2. Articles were screened in two further stages, reducing 2,643 items to 58 which we examine in detail. Our assessment of validity initially focuses on assessing the intervention (e.g. provision of microfinance), the measurement the outcome measures (e.g. income, expenditure, assets, health and education, empowerment, and so on), contextual factors (including other microfinance services) affecting heterogeneity of outcomes, and potential existence and likely significance of confounding factors.

We investigate the included studies categorised by intervention and outcome; this was challenging due to the diverse nature of the microfinance interventions and the wealth of outcomes investigated. Table 18 summarises our findings reported in section 3 above. We find that most of the effects assessed occur in the early stages of the causal chain (Figure 1), with both positive and negative outcomes; the bulk of estimates reported were statistically insignificant even at the beginning of the causal chain, and a significant number of estimates suggest negative outcomes throughout the causal chain. These findings are not inconsistent with at least some of the qualitative literature (e.g. Fernando 1997).

The majority of microfinance IEs included investigate group lending and credit only interventions which do not reflect the diversity of the sector, hence this does not allow us to reach a conclusion as to the impact of the microfinance sector as a whole; individual lending is a more recent phenomenon that has not yet been evaluated widely. Paired with doubts about research designs and analytical methods used by various microfinance IEs, we can neither support nor deny the notion that microfinance is pro-poor and pro-women; what this might mean for policy makers is discussed further below.

Table 18: Combined research designs and analytical methods

Outcome category	Location in causal chain	Significance & sign					
		Ns			sig		
		+	-	total	+	-	total
Economic	inputs	194	189	383	127	70	197
	effects	362	313	675	441	117	558
	impacts	27	15	42	92	42	134
	Total	583	517	1,100	660	229	889
Social	inputs	12	7	19	5		5
	effects		1	1			
	impacts	154	118	272	79	22	101
	Total	166	126	292	84	22	106
Empowerment	inputs						
	effects						
	impacts	47	67	114	76	19	95
	Total	47	67	114	76	19	95

4.1 Policy recommendations

If indeed there is no good evidence to support the claim that microfinance has a beneficial effect on the well-being of poor people or empowers women, then, over the last decade or so, it might have been more beneficial to explore alternative interventions that could have better benefitted poor people and/or empowered women. Microfinance activities and finance have absorbed a significant proportion of development resources, both in terms of finances and people. Microfinance activities are highly attractive, not only to the development industry but also to mainsteam financial and business interests with little interest in poverty reduction or empowerment of women, as pointed out above. There are many other candidate sectors for development activity which may have been relatively disadvantaged by ill-founded enthusiasm for microfinance. Even within the microfinance sector, the putative success of basic models of lending such as the Grameen Bank and related models, may well have diverted attention from opportunities for alternatives; for example, recent studies (Collins et al. 2009) have pointed out that poor people do not just need credit but access to other financial products such as savings, and insurance. Also, the financial products offered by MFIs must become more flexible and adjust to rapidly changing circumstances faced by poor people. Many MFIs have already moved in that direction, providing more diverse and flexible products.

However, it remains unclear under what circumstances, and for whom, microfinance has been and could be of real, rather than imagined, benefit to poor people. Unsurprisingly we focus our policy recommendations on the need for more and better research. Thus, to have obtained a clearer picture on the impacts of microfinance, on whom, where, and when (e.g. under what circumstances), and the mechanisms which account for these effects, more and better quality quantitative evidence was required at an earlier stage in the diffusion of this intervention. While there is currently enthusiasm for RCTs as the gold standard for assessing interventions, there are many who doubt the universal appropriateness of these designs. Indeed there may be something to be said for the idea that this current enthusiasm is built on similar foundations of sand to those on which we suggest the microfinance phenomenon has been based.

Such research could have used a range of research designs (not just RCTs), and analytical methods, to assess both the short and longer-term impacts of microfinance. Thus, we suggest, the somewhat sorry tale of (mis-)evaluation of microfinance impacts leads us to advocate, specifically:

- Conduct of well designed experimental and observational studies, including longitudinal studies;
- Replication of highly regarded studies of whatever research design;
- Capacity building in the multi-disciplinary, mixed methods research, especially surveys drawing on ethnographically rich understanding of local context.

Such well designed and conducted studies[54] should be complemented with qualitative tools prior to, as well as during and after, embarking on quantitative studies. Also, and fashionably, orthodox social survey methods can be enhanced with coordinated behavioural and experimental economics research with microfinance participants (lenders and actual and potential borrowers) to gain a better understanding of the mechanisms underlying microfinance participation and conduct, and the role of the unobservables in this context. Exploring why what appears to have been inappropriate optimism towards microfinance became so widespread would also be a suitable subject for further research which would involve political scientists.

[54] Cook et al. 2008, report several conclusions which are likely to improve the quality of results of observational studies judged by their concordance with experimental results. These include using appropriate comparison groups, using the same measurement instruments (survey procedures) for both treatment and control and comparison subjects, and use of high quality field research designs and procedures building on the best, local, qualitative information (see also Rosenbaum 2002, 2010, Rosenbaum and Silber 2001).

5 Bibliography

Abera H (2010) Can microfinance help to reduce poverty? With reference to Tigrai, Northern Ethiopia. *Economics.* Mekele.

Abou-Ali H, El-Azony H, El-Laithy H, Haughton J, Khandker S (2010) Evaluating the impact of Egyptian social fund for development programmes. *Journal of Development Effectiveness,* 2 (4): 521 - 555.

Abou-Ali H, El-Azony H, El-Laithy H, Haughton J, Khandker SR, (2009) Evaluating the impact of Egyptian social fund for development programmes. World Bank Policy Research Working Paper No. 4993, July.

Adams DW, von Pischke JD (1992) Microenterprise credit programmes: déja vu. *World Development,* 20 (10): 1463-1470.

Ahlin C, Townsend RM (2007) Using repayment data to test across models of joint liability lending. *The Economic Journal,* 117 (517): F11-F51.

Ahmed SM, Adams AM, Chowdhury M, Bhuiya A (2000) Gender, socioeconomic development and health-seeking behaviour in Bangladesh. *Social Science & Medicine,* 51 (3): 361-371.

Aideyan O (2009) Microfinance and poverty reduction in rural Nigeria. *Savings and Development,* 33 (3): 293-317.

Alexander G (2001) An empirical analysis of microfinance: who are the clients? *Northeastern Universities Development Consortium Conference.* Boston, 28-30 September 2001.

Alkin MC (2004) *Evaluation roots: tracing theorists' views and influences.* Thousand Oaks, CA: Sage Publications.

Angrist JD, Imbens GW, Rubin DB (1996) Identification of causal effects using instrumental variables. *Journal of the American Statistical Association,* 91 (434): 444-455.

Angrist JD, Krueger AB (2001) Instrumental variables and the search for identification: from supply and demand to natural experiments. *The Journal of Economic Perspectives,* 15 (4): 69-85.

Arai L, Britten N, Popay J, Roberts H, Petticrew M, Rodgers M, Sowden A (2007) Testing methodological developments in the conduct of narrative synthesis: a demonstration review of research on the implementation of smoke alarm interventions. *Evidence & Policy: A Journal of Research, Debate and Practice,* 3 (3): 361-383.

Armendáriz de Aghion B, Gollier C (2000). Peer group formation in an adverse selection model. *The Economic Journal,* 110 (465): 632-643.

Armendáriz de Aghion B Morduch J (2005) *The economics of microfinance.* Cambridge: MIT Press.

Armendáriz de Aghion B, Morduch J (2010) *The economics of microfinance, 2nd edn* Cambridge: MIT Press.

Augsburg B (2006) Econometric evaluation of the SEWA bank in India: applying matching techniques based on the propensity score. Working Paper MGSoG/2006/WP003, Maastricht University, October.

Banerjee A, Besley T, Guinnane TW (1994) The neighbour's keeper: the design of a credit cooperative with theory and a test. *The Quarterly Journal of Economics,* 109 (2): 491-515.

Banerjee A, Duflo E (2010) Giving credit where it is due. Available at: http://econ-www.mit.edu/files/5415.

Banerjee A, Duflo E, Glennerster R, Kinnan C (2009) The miracle of microfinance? Evidence from a randomised evaluation. Available at: http://econ-www.mit.edu/files/4162.

Banerjee AV, Cole S, Duflo E, Linden L (2007) Remedying education: evidence from two randomised experiments in India. *Quarterly Journal of Economics* 122 (3): 1235-1264.

Barnes C (2001) Microfinance program clients and impact: an assessment of Zambuko trust, Zimbabwe. Report submitted to USAID assessing the impact of microenterprise services (AIMS), October.

Basu A, Heckman JJ, Navarro-Lozano S, Urzua S (2007) Use of instrumental variables in the presence of heterogeneity and self-selection: an application to treatments of breast cancer patients. *Health Economics,* 16 (11): 1133-1157.

Bateman M (2010) *Why microfinance doesn't work? The destructive rise of local neoliberalism.* London: Zed Books.

Bateman M, Chang HJ (2009) The microfinance illusion. *Available at:* http://www.econ.cam.ac.uk/faculty/chang/pubs/Microfinance.pdf.

Becker SO, Caliendo M (2007) Sensitivity analysis for average treatment effects. *The STATA Journal,* 7 (1): 71-83.

Besley T, Coate S (1995) Group lending, repayment incentives and social collateral. *Journal of Development Economics,* 46 (1): 1-18.

Bhuiya A, Chowdhury M (2002) Beneficial effects of a woman-focused development programme on child survival: evidence from rural Bangladesh. *Social Science & Medicine,* 55 (9): 1553-1560.

Binswanger HP, Khandker SR (1995) The impact of formal finance on the rural economy of India. *Journal of Development Studies*, 32(2): 234-64.

Blundell R, Costa Dias M (2000) Evaluation methods for non-experimental data. *Fiscal Studies,* 21 (4): 427-468.

Blundell R, Costa Dias M (2002) Alternative approaches to evaluation in empirical microeconomics. The Institute for Fiscal Studies, Department of Economics, University College London, Cemmap Working Paper No. CWP 10/02.

Blundell R, Costa Dias M (2008) Alternative approaches to evaluation in empirical microeconomics. The Institute for Fiscal Studies, Department of Economics, University College London, Cemmap Working Paper No. CWP 26/08.

Burgess R, Pande R (2005) Do rural banks matter? Evidence from the Indian social banking experiment. *American Economic Review*, 95(3): 780-94.

Caliendo M (2006) *Microeconometric evaluation of labour market policies.* Berlin: Springer.

Caliendo M, Hujer R (2005) The microeconometric estimation of treatment effects - an overview. Forschungsinstitut zur Zukunft der Arbeit (IZA) Discussion Paper No. 1653, July.

Caliendo M, Kopeinig S (2005) Some practical guidance for the implementation of propensity score matching. Forschungsinstitut zur Zukunft der Arbeit (IZA) Discussion Paper No. 1588, May.

Caliendo M, Kopeinig S (2008) Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys,* 22 (1): 31-72.

Cameron AC, Trivedi PK (2005) *Microeconometrics: methods and applications.* Cambridge: Cambridge University Press.

Cassar A, Crowley L, Wydick B (2007) The effect of social capital on group loan repayment: evidence from field experiments. *The Economic Journal,* 117: F85-F106.

Chalmers I (2008) Confronting therapeutic ignorance: tackling uncertainties about the effects of treatments will help to protect patients. *British Medical Journal,* 337: 841.

Chemin M (2008) The benefits and costs of microfinance: evidence from Bangladesh. *Journal of Development Studies,* 44 (4): 463-484.

Chen MA, Mahmud S (1995) Assessing change in women's lives: a conceptual framework. BRAC-ICDDRB Joint research project Working Paper 2. Dhaka.

Chen MA, Snodgrass D (1999) An assessment of the impact of SEWA bank in India: baseline findings. Report submitted to USAID Assessing the Impact of Microenterprise Services (AIMS), August.

Chen MA, Snodgrass D (2001) Managing resources, activities, and risk in urban India: the impact of SEWA bank. Report submitted to USAID assessing the impact of microenterprise services (AIMS), September.

Chowdhury IR (2010) Understanding the Grameen miracle: information and organisational innovation. *Economic and Political Weekly,* 45 (6): 66-73.

Cochrane Injuries Group Albumin Reviewers (1998) Human albumin administration in critically ill patients: systematic review of randomised control trials. *British Medical Journal,* 317: 235-240.

Coleman BE (1999) The impact of group lending in northeast Thailand. *Journal of Development Economics,* 60 (1): 105-141.

Coleman BE (2002) Microfinance in northeast Thailand: who benefits and how much? Economics and Research Department, Working Paper No. 9, Asian Development Bank, April.

Coleman BE (2006) Microfinance in northeast Thailand: who benefits and how much? *World Development,* 34 (9): 1612-1638.

Collins D, Morduch J, Rutherford S, Ruthven O (2009) *Portfolios of the poor: how the world's poor live on $2 a day.* Princeton: Princeton University Press.

Cook T, Campbell DT (1979) *Quasi-experimentation: design and analysis issues for field setting.* Chicago: Rand McNally.

Cook TD, Shadish WR, Wong VC (2008) Three conditions under which experiments and observational studies produce comparable causal estimates: new findings from within-study comparisons. *Journal of Policy Analysis and Management,* 27 (4): 724-750.

Copestake J (2002) Inequality and the polarising impact of microcredit: evidence from Zambia's copperbelt. *Journal of International Development,* 14: 743-755.

Copestake J, Bhalotra S, Johnson S (2001) Assessing the impact of microcredit: a Zambian case study. *Journal of Development Studies,* 37 (4): 81-100.

Copestake J, Dawson P, Fanning JP, McKay A, Wright-Revolledo K (2005) Monitoring the diversity of the poverty outreach and impact of microfinance: a comparison of methods using data from Peru. *Development Policy Review,* 23 (6): 703-723.

Cornfield J, Haenszel W, Hammond E, Lilienfeld A (1959) Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer Institute,* 22: 173-203.

Cotler P, Woodruff C (2008) The impact of short-term credit on microenterprises: evidence from the Fincomun-Bimbo programme in Mexico. *Economic Development and Cultural Change,* 56 (4): 829-849.

Cox DR (1958) *Planning of experiments.* New York: Wiley.

Cull R, Demirguc-Kunt A, Morduch J (2009) Microfinance meets the market. *Journal of Economic Perspectives,* 23 (1): 167-192.

Cuong NV (2008) Is a governmental microcredit programme for the poor really pro-poor? Evidence from Vietnam. *Developing Economies,* 46 (2): 151-187.

de Janvry A, McIntosh C, Sadoulet E (2010) The supply and demand side impacts of credit market information. *Journal of Development Economics,* 93 (2): 173-188.

de Mel S, McKenzie D, Woodruff C (2008) Returns to capital in microenterprises: evidence from a field experiment. *Quarterly Journal of Economics,* 123 (4): 1329-1372.

de Mel, S., D. J. McKenzie, Woodruff C (2009). Measuring microenterprise profits: must we ask how the sausage is made? *Journal of Development Economics* 88(1): 19-31.

Dawid AP (1979) Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B (Methodological),* 41 (1): 1-31.

Deaton A (2009) Instruments of development: randomisation in the Tropics, and the search for the elusive keys to economic development. Available at: http://www.princeton.edu/~deaton/downloads/Instruments_of_Development.pdf.

Deaton A (2010) Instruments, randomisation, and learning about development. *Journal of Economic Literature,* 48: 424-455.

Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F (2003) Evaluating non-randomised intervention studies. *Health Technology Assessment,* 7 (27).

Dehejia R (2005) Practical propensity score matching: a reply to Smith and Todd. *Journal of Econometrics,* 125: 355-364.

Dehejia RH, Wahba S (1999) Causal effects in nonexperimental studies: reevaluating the evaluation of training programmes. *Journal of the American Statistical Association,* 94 (448): 1053-1062.

Dehejia, R, Wahba S (2002) Propensity score-matching methods for nonexperimental causal studies. *The Review of Economic Studies,* 84 (1): 151-161.

Deininger K, Liu Y (2009) Economic and social impacts of self-help groups in India. World Bank Policy Research Working Paper No. 4884, March.

Desai J, Tarozzi A (2009) Microcredit, family planning programmes and contraceptive behavior: evidence from a field experiment in Ethiopia. *Available at:* http://ipl.econ.duke.edu/bread/papers/working/247.pdf.

Diagne A, Zeller M (2001) Access to credit and its impact on welfare in Malawi. Research Report 116. Washington, D.C., International Food Policy Research Institute.

Dichter T, Harper M (eds) (2007) *What's wrong with microfinance?* Warwickshire: Practical Action Publishing.

DiPrete TA, Gangl M (2004) Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociological Methodology,* 34 (1): 271-310.

Donaldson SI (2009) In search of the blueprint for an evidence-based global society. In Donaldson SI, Christie CA eds (2009) *What counts as credible evidence in applied research and evaluation practice?* Thousand Oaks: Sage.

Donaldson SI, Christie CA (eds) (2009) *What counts as credible evidence in applied research and evaluation practice?* Thousand Oaks: Sage.

Doocy S, Teferra S, Norell D, Burnham G (2005) Credit programme outcomes: coping capacity and nutritional status in the food insecure context of Ethiopia. *Social Science and Medicine,* 60 (10): 2371-2382.

Duflo E (2001) Schooling and labor market consequences of school construction in Indonesia: evidence from an unusual policy experiment. *The American Economic Review*, 91(4): 795-813.

Duflo E, Crépon B, Parienté W, Devoto F (2008) Poverty, access to credit and the determinants of participation in a new microcredit programme in rural areas of Morocco. J-PAL Impact Analyses Series, No 2.

Duflo E, Glenerster R, Kremer M (2007) Using randomisation in development economics research: A toolkit.

Duflo E, Glennerster R, Kremer M (2008) Using randomisation in development economics research: a toolkit. In Schultz TP, Strauss J eds *Handbook of Development Economics, Volume 4.* Amsterdam: Elsevier.

Duflo E, Kremer M (2005) Use of randomisation in the evaluation of development effectiveness. In Pitman GK, Feinstein ON, Ingram GK (ed) *Evaluating Development Effectiveness.* New Brunswick: Transaction Publishers.

Dunn E (1999) Microfinance clients in Lima, Peru: baseline report for AIMS core impact assessment. Report submitted to USAID assessing the impact of microenterprise services (AIMS), June.

Dunn E, Arbuckle JG (2001) The impacts of microcredit: a case study from Peru. Report submitted to USAID assessing the impact of microenterprise services (AIMS), September.

Dupas, P. & Robinson, J. (2009) Savings Constraints and Microenterprise Development: Evidence from a Field Experiment in Kenya. NBER Working Paper No. w14693.

Duvendack M (2010a) Smoke and mirrors: evidence of microfinance impact from an evaluation of SEWA bank in India. Working Paper 24, DEV Working Paper Series, The School of International Development, University of East Anglia, UK.

Duvendack M (2010b) Smoke and mirrors: evidence from microfinance impact evaluations in India and Bangladesh. *Unpublished PhD Thesis. School of International Development.* Norwich: University of East Anglia.

Duvendack M, Palmer-Jones R (2011) High noon for microfinance impact evaluations: re-investigating the evidence from Bangladesh. Working Paper 27, DEV Working Paper Series, The School of International Development, University of East Anglia, UK.

Fernando JL (1997) Non-governmental organisations, micro-credit, and empowerment of women. *The ANNALS of the American Academy of Political and Social Science,* 554 (1): 150-177.

Fernando JL (ed.) (2006) *Microfinance: perils and prospects.* London: Routledge.

Field E, Pande R (2008) Repayment frequency and default in microfinance: evidence from India. *Journal of the European Economic Association,* 6 (2-3): 501-509.

Fischer G (2010) Contract structure, risk sharing, and investment choice. Available at: http://personal.lse.ac.uk/fischerg/Research.htm.

Fisher RA (1935) *The design of experiments.* London: Oliver and Boyd.

Gaile GL, Foster J (1996) Review of methodological approaches to the study of the impact of microenterprise credit programmes. Report submitted to USAID assessing the impact of microenterprise Services (AIMS), June.

Galasso E, Ravallion M (2004) Social protection in a crisis: Argentina's plan Jefes y Jefas. *World Bank Econ Rev,* 18 (3): 367-399.

Gangopadhyay S, Ghatak M, Lensink R (2005) Joint liability lending and the peer selection effect. *The Economic Journal,* 115 (506): 1005-1015.

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.

Gertler P, Levine DI, Moretti E (2009) Do microfinance programmes help families insure consumption against illness? *Health Economics*, 18(3): 257-273.

Ghatak M (1999) Group lending, local information and peer selection. *Journal of Development Economics*, 60 (1): 27-50.

Ghatak M (2000) Screening by the company you keep: joint liability lending and the peer selection effect. *The Economic Journal,* 110 (465): 601-631.

Ghatak M, Guinnane TW, (1999) The economics of lending with joint liability: theory and practice. *Journal of Development Economics,* 60 (1): 195-228.

Giné X, Karlan DS (2007) Group versus individual liability: a field experiment in the Philippines. Available at: http://134.245.95.50:8080/dspace/bitstream/10419/26981/1/593239520.PDF.

Giné X, Karlan D (2008) Peer monitoring and enforcement: long term evidence from microcredit lending groups with and without lroup liability. Yale University Economic Growth Center Working Paper 940.

Giné X, Karlan DS (2009) Group versus individual liability: long term evidence from Philippine microcredit lending groups. Available at: http://www.econ.yale.edu/growth_pdf/cdp970.pdf.

Goetz AM, Sen Gupta R (1996) Who takes the credit? Gender, power, and control over loan use in rural credit programmes in Bangladesh. *World Development,* 24 (1): 45-63.

Goldacre B (2008) *Bad science.* London: Fourth Estate.

Goldberg N (2005) Measuring the impact of microfinance: taking stock of what we know. Grameen Foundation USA Publication Series, December.

Gough D (2007) Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. In Furlong J, Oancea A (eds) *Applied and Practice-based Research. Special Edition of Research Papers in Education,* 22, (2): 213-228.

Grosh ME, Glewwe P (2000) *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study, Volumes. 1-3.* The World Bank, Washington D.C.

Guha-Khasnobis B, Hazarika G (2007) Household access to microcredit and children's food security in rural Malawi: a gender perspective. World Institute for Development Economic Research (UNU-WIDER), Working Papers: UNU-WIDER Research Paper RP2007/87.

Guyatt G, Haynes RB. Jaeschke RZ. Cook DJ. Green L. Naylor CD. Wilson MC. Richardson WS (2000) A users' guide to the medical literature XXV: evidence-based medicine: principles for applying the users' guides to patient care: evidence-based medicine working group. *Journal of the American Medical Association*, 284: 1290-1296.

Guyatt G, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ (1995) Users' guide to the medical literature XI: a method for grading health care recommendations. *Journal of the American Medical Association*, 274: 1800-1804.

Hadi A, (2001) Promoting health knowledge through micro-credit programmes: experience of BRAC in Bangladesh. *Health Promotion International,* 16 (3): 219-227.

Hamermesh DS (2007) Viewpoint: replication in economics. *Canadian Journal of Economics,* 40 (3): 715-733.

Harremoës P, Gee D (2001) Late lessons from early warnings: the precautionary principle 1896-2000, Copenhagen, European Environment Agency.

Hashemi SM, Schuler SR, Riley AP (1996) Rural credit programmes and women's empowerment in Bangladesh. *World Development,* 24 (4): 635-653.

Hausman JA, Wise DA (1985) Social experimentation. Chicago, University of Chicago Press.

Heckman JJ (1974) Shadow prices, market wages, and labor supply. *Econometrica,* 42 (4): 679-694.

Heckman JJ (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Social and Economic Measurement,* 5 (4): 475-492.

Heckman JJ (1978) Dummy endogenous variables in a simultaneous equation system. *Econometrica,* 46 (4): 931-959.

Heckman JJ (1979) Sample selection bias as a specification error. *Econometrica,* 47 (1): 153-161.

Heckman JJ (1991) Randomisation and social policy evaluation. NBER Technical Working Paper No. 107.

Heckman JJ (1997) Instrumental variables: a study of implicit behavioral assumptions used in making programme evaluations. *Journal of Human Resources,* 32 (3): 441-462.

Heckman JJ, Ichimura H, Smith J, Todd P (1998) Characterising selection bias using experimental data. *Econometrica,* 66 (5): 1017-1098.

Heckman JJ, Ichimura H, Todd P (1997) Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *Review of Economic Studies,* 64: 605-654.

Heckman JJ, Ichimura H, Todd P (1998) Matching as an econometric evaluation estimator. *The Review of Economic Studies,* 65 (2): 261-294.

Heckman JJ, LaLonde R, Smith J (1999) The economics and econometrics of active labour market programmes. In Ashenfelter O, Card D (eds) *Handbook of Labor Economics, Volume 3A.* Amsterdam: Elsevier.

Heckman JJ, Urzua S (2009) Comparing IV with structural models: what simple IV can and cannot identify. NBER Working Paper No. 14706.

Heckman JJ, Vytlacil E (2007a) Econometric evaluation of social programmes, part I: causalmodels, structural models and econometric policy evaluation. In Heckman JJ, Leamer EE eds *Handbook of Econometrics, Volume 6B.* Amsterdam: North-Holland.

Heckman JJ Vytlacil E (2007b) Econometric evaluation of social programmes, part II: using the marginal treatment effect to organise alternative econometric estimators to evaluate social programmes, and to forecast their effects in new environments. In Heckman JJ, Leamer EE (eds) *Handbook of Econometrics, Volume 6B.* Amsterdam: North-Holland.

Hermes N, Lensink R (2007) The empirics of microfinance: what do we know? *The Economic Journal,* 117 (517): F1-F10.

Hidalgo-Celarie N, Altamirano-Cardenas R, Zapata-Martelo E, Martinez-Corona B (2005) Economic impact of microfinance targeted to women in the state of Veracruz, Mexico. *Agrociencia,* 39 (3): 351-359.

Higgins JPT, Green S (2008) Cochrane handbook for systematic reviews of interventions Version 5.0.0. Available at: www.cochrane-handbook.org.

Holvoet N (2005) The impact of microfinance on decision-making agency: evidence from south India. *Development and Change*. *36*(1): 75-102.

Honohan P (2004) Financial development, growth and poverty: how close are the links in financial development and economic growth: explaining the link. Goodhart C (ed.) London: Palgrave.

Hoque S (2004) Microcredit and the reduction of poverty in Bangladesh. *Journal of Contemporary Asia,* 34 (1): 21 - 32.

Hulme D (2000) Impact assessment methodologies for microfinance: theory, experience and better practice. *World Development,* 28 (1): 79-98.

Hulme D, Mosley P (1996) *Finance against poverty.* London: Routledge.

Ichino A, Mealli F, Nannicini T (2006) From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? Forschungsinstitut zur Zukunft der Arbeit (IZA) Discussion Paper No. 2149, May.

Imai KS, Arun T, Annim SK (2010) Microfinance and household poverty reduction: new evidence from India. *World Development,* 38 (12): 1760-1774.

Imai KS, Azam MS (2010) Does microfinance reduce poverty in Bangladesh? New evidence from household panel data. Discussion Paper, DP2010-24, Kobe University, September.

Imbens G (2009) Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). NBER Working Paper No. 14896.

Imbens GW, Angrist JD (1994) Identification and estimation of local average treatment effects. *Econometrica,* 62 (2): 467-475.

Imbens G, Wooldridge J (2008) Recent developments in the econometrics of programme evaluation. The Institute for Fiscal Studies, Department of Economics, University College London, Cemmap Working Paper No. CWP 24/08.

Ioannidis JP (2005) Why most published research findings are false. *PLoS Med*, 2(8): 124.

Ioannidis JP, Patsopoulos NA, Rothstein HR (2008) Reasons or excuses for avoiding meta-analysis in forest plots. *British Medical Journal,* 336 (7658): 1413-1415.

Ito S (2003) Microfinance and social capital: does social capital help create good practice? *Development in Practice,* 13 (4): 322-332.

JEP (Journal of Economic Perspectives), (2010) Symposium: con out of economics, *Journal of Economic Perspectives*, 24(2): 3-94.

Johnson S (2005) Gender relations, empowerment and microcredit: moving forward from a lost decade. *European Journal of Development Research.* 17(2).

Johnson S, Rogaly B (1997) *Microfinance and poverty reduction.* Oxford: Oxfam.

Kabeer N (2001) Conflicts over credit: re-evaluating the empowerment potential of loans of women in rural Bangladesh. *World Development,* 29(1): 63-84.

Kabeer N (2005a) Direct social impacts for the millennium development goals. Chapter 5 in Copestake J, Greeley M, Johnson S, Kabeer N, Simanowitz A *Money with a mission. Microfinance and poverty reduction.* London: Intermediate Technology Publications.

Kabeer N (2005b) Is microfinance the ‘magic bullet’ for women’s empowerment: analysis of findings from South Asia. *Economic and Political Weekly*, 29/20/2005: 4709-4718.

Kaboski JP, Townsend RM (2005) Policies and impact: an analysis of village-level microfinance institutions. *Journal of the European Economic Association,* 3 (1): 1-50.

Kaboski JP, Townsend RM (2009) The impacts of credit on village economies. SSRN eLibrary.

Karlan DS, Morduch J (2009) Access to finance. Available at: http://karlan.yale.edu/p/HDE_June_11_2009_Access_to_Finance.pdf.

Karlan D, Zinman J (2010) Expanding credit access: using randomised supply decisions to estimate the impacts. *Review of Financial Studies,* 23 (1): 433-464.

Karlan DS (2001) Microfinance impact assessments: the perils of using new members as a control group. *Journal of Microfinance,* 3 (2): 75-85.

Karlan DS (2007) Social connections and group banking. *The Economic Journal,* 117 (517): F52-F84.

Khandker SR (1996) Role of targeted credit in rural non-farm growth. *Bangladesh Development Studies,* 24 (3 & 4).

Khandker SR (1998) *Fighting poverty with microcredit: experience in Bangladesh.* New York: Oxford University Press.

Khandker SR (2000) Savings, informal borrowing and microfinance. *Bangladesh Development Studies,* 26 (2 & 3).

Khandker SR (2003) Microfinance and poverty: evidence using panel data from Bangladesh. World Bank Policy Research Working Paper No. 2945, January.

Khandker SR (2005) Microfinance and poverty: evidence using panel data from Bangladesh. *The World Bank Economic Review,* 19 (2): 263-286.

Khandker SR,Latif MA (1996) The role of family planning and targeted credit programmes in demographic change in Bangladesh. *World Bank Discussion Papers,* 337.

Khandker SR, Samad HA, Khan ZH (1998) Income and employment effects of microcredit programmes: village-level evidence from Bangladesh. *Journal of Development Studies,* 35 (2): 96-124.

Khandker SR, Koolwal GB, Samad HA (2010) *Handbook on impact evaluation: quantitative methods and practices.* Washington, DC: The World Bank.

Kondo T, Orbeta A, Dingcong C, Infantado C (2008) Impact of microfinance on rural households in the Philippines. *Ids Bulletin-Institute of Development Studies,* 39 (1): 51-70.

Kunz R, Vist GE, Oxman AD (2007) Randomisation to protect against selection bias in healthcare trials. In *Cochrane Database of Systematic Reviews*, 2007: Issue 2.

Kwok CS, Loke YK (2010) Meta-analysis: the effects of proton pump inhibitors on cardiovascular events and mortality in patients receiving clopidogrel. *Alimentary Pharmacology & Therapeutics,* 31 (8): 810-823.

LaLonde RJ (1986) Evaluating the econometric evaluations of training programmes with experimental data. *The American Economic Review,* 76 (4): 604-620.

Latif MA (1994) Programme impact on current contraception in Bangladesh. *Bangladesh Development Studies,* 22 (1): 27-61.

Leamer EE (1983) Let's take the con out of econometrics. *The American Economic Review,* 73 (1): 31-43.

Leatherman S, Metcalfe M, Geissler K, Dunford C (2011) Integrating microfinance and health strategies: examining the evidence to inform policy and practice. *Health Policy and Planning*: 1-17.

Ledgerwood J (1999) *Microfinance handbook: an institutional and financial perspective.* Washington DC: The World Bank.

Ledgerwood J, White V, Brand M (2006) *Transforming microfinance institutions: providing full financial services to the poor.* Washington, DC: The World Bank.

Levitt SD, List JA (2009) Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. NBER Working Paper No. 15016.

Little, R.J. A. and D.B, Rubin (1987) Statistical Analysis With Missing Data. New York: Wiley.

Marconi R, Mosley P (2004) The FINRURAL impact evaluation service - a cost-effectiveness analysis. Small Enterprise Development, 15 (3): 18-27.

Matin I, Hulme D (2003) Programmes for the poorest: learning from the IGVGD programme in Bangladesh. *World Development,* 31 (3): 647-665.

Mayoux L (2001) Tackling the downside: social capital, women's empowerment and microfinance in Cameroon', *Development and Change*. 32: 421-450.

McKernan SM (2002) The impact of microcredit programmes on self-employment profits: do non-credit programme aspects matter? *Review of Economics and Statistics,* 84 (1): 93-115.

Menon N (2006) Non-linearities in returns to participation in Grameen bank programmes. *Journal of Development Studies,* 42 (8): 1379 - 1400.

Meyer MM, Fienberg SE (eds) (1992) *Assessing evaluation studies: the case of bilingual education strategies.* Washington DC: National Academy Press.

Miettinen OS, Cook EF (1981) 'Confounding: essence and detection.' *American Journal of Epidemiology*, 114(4): 593-603.

Moerman D (2002) *Meaning, medicine and the 'placebo effect'.* Cambridge: Cambridge University Press.

Miguel E, Kremer M (2004) Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica,* 72 (1): 159-217.

Mohindra K, Haddad S, Narayana D (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: 2.

Montgomery H (2005) Serving the poorest of the poor: the poverty impact of the Khushhali bank's microfinance lending in Pakistan. *Poverty Reduction Strategies in Asia: Asian Development Bank Institute (ADBI) Annual Conference.* Tokyo, 9 December 2005.

Morduch J (1995) Income soothing and consumption smoothing. *Journal of Economic Perspectives*. 9(3): 103-14.

Morduch J (1998) Does microfinance really help the poor? New evidence from flagship programmes in Bangladesh. Unpublished mimeo.

Morduch J, Haley B (2002) Analysis of the effects of microfinance on poverty reduction. NYU Wagner Working Paper No. 1014, June.

Morgan SL, Harding DJ (2006) Matching estimators of causal effects. prospects and pitfalls in theory and practice. *Sociological Methods & Research,* 35 (1): 3-60.

Morgan SL, Winship C (2007) *Counterfactuals and causal inference. Methods and principles for social research.* Cambridge: Cambridge University Press.

Mosley P (1996) Metamorphosis from NGO to commercial bank: the case of BancoSol in Bolivia. In Hulme D, Mosley P (eds) *Finance against Poverty.* London: Routledge.

Nanda P (1999) Women's participation in rural credit programmes in Bangladesh and their demand for formal health care: is there a positive impact? *Health Economics,* 8 (5): 415-428.

Nannicini T (2007) Simulation-based sensitivity analysis for matching estimators. *The STATA Journal,* 7 (3): 334-350.

Neyman JS (1923) On the application of probability theory to agricultural experiments. essay on principles. Section 9. *Translated in Statistical Science,* 5 (4): 465-480, 1990.

Neyman J, Pearson ES (1933) On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London,* 231: 289-337.

Nino-Zarazua M, Copestake J (2009) Financial inclusion, vulnerability and mental models: from physical Access to effective use of financial services in a low-income area of Mexico City. *Savings and Development.* 32(4): 353-80.

Norwood C (2005) Macro promises of microcredit - a case of a local eSusu in rural Ghana. *Journal of International Women's Studies,* 7 (1): 1-7.

Odell K (2010) Measuring the impact of microfinance: taking another look. Grameen Foundation USA Publication Series, May.

Orso CE (2011) Microcredit and poverty. An overview of the principal statistical methods used to measure the programme net impacts. *POLIS Working Paper No. 180, February.*

Osili UO, Long BT (2008) Does female schooling reduce fertility? Evidence from Nigeria. *Journal of Development Economics,* 87: 57-75.

Patten R, Rosengard J (1991) *Progress with profits: the development of rural banking in Indonesia.* San Francisco: International Center for Economic Growth.

Pawson R, Greenhalgh T, Harvey G, Walshe K (2005) Realist review - a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy,* 10 (1): 21-34.

Petryna A (2009) *When experiments travel: clinical trials and the global search for human subjects.* Princeton, NJ: Princeton University Press.

Petticrew M, Roberts H (2006) *Systematic reviews in the social sciences: a practical guide.* Oxford: Blackwell Publishing.

Pisani MJ, Yoskowitz DW (2010) The efficacy of microfinance at the sectoral level: urban pulperias in Matagalpa, Nicaragua. *Perspectives on Global Development and Technology,* 9 (3-4): 418-448.

Pitt M, Khandker SR, Cartwright J (2006) Empowering women with microfinance: evidence from Bangladesh. *Economic Development and Cultural Change*: 791-831.

Pitt M, Khandker, SR, Chowdhury OH, Millimet DL (2003) Credit programmes for the poor and the health status of children in rural Bangladesh. *International Economic Review,* 44 (1): 87-118.

Pitt MM (1999) Reply to Morduch's 'Does microfinance really help the poor? New evidence from flagship programmes in Bangladesh'. Unpublished mimeo.

Pitt MM (2000) The effect of non-agricultural self-employment credit on contractual relations and employment in agriculture: the case of microcredit programmes in Bangladesh. *Bangladesh Development Studies,* 26 (2 & 3): 15-48.

Pitt MM (2011) Response to Roodman and Morduch's 'The Impact of microcredit on the poor in Bangladesh: revisiting the evidence'.pdf document processed 26 March 2011.

Pitt MM Khandker SR (1998) The impact of group-based credit programmes on poor households in Bangladesh: does the gender of participants matter? *Journal of Political Economy,* 106 (5): 958-996.

Pitt MM, Khandker SR (2002) Credit programmes for the poor and seasonality in rural Bangladesh. *Journal of Development Studies,* 39 (2): 1-24.

Pitt MM, Khandker SR, McKernan S-M, Latif MA (1999) Credit programmes for the poor and reproductive behavior of low-income countries: are the reported causal relationships the result of heterogeneity bias? *Demography,* 36 (1): 1-21.

Pritchett L (2002) It pays to be ignorant: a simple political economy of rigorous programme evaluation. *Journal of Economic Policy Reform,* 5 (4): 251-269.

Pritchett L (2009) The policy irrelevance of the economics of education: is 'normative as positive' just useless, or worse? In Cohen J, Easterly W (eds) *What Works in Development? Thinking Big and Thinking Small.* Washington DC: Brookings Institution Press.

Puffer S, Torgerson D, Watson J (2003) Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *British Medical Journal,* 327 (7418): 785-789.

Puhani PA (2000) The Heckman correction for sample selection and its critique. *Journal of Economic Surveys,* 14 (1): 53-68.

Rafiq RB, Chowdhury JA, Cheshier PA (2009) Microcredit, financial improvement and poverty alleviation of the poor in developing countries: evidence from Bangladesh. *Journal of Emerging Markets,* 14 (1): 24-37.

Rahman M, Davanzo J, Sutradhar SC (1996) Impact of the Grameen bank on childhood mortality in Bangladesh. *Glimpse,* 18 (1): 8.

Rahman S (2010) Consumption difference between microcredit borrowers and non-borrowers: a Bangladesh experience. *Journal of Developing Areas,* 43 (2): 313-326.

Ravallion M (2001) The mystery of the vanishing benefits: an introduction to impact evaluation. *The World Bank Economic Review,* 15 (1): 115-140.

Ravallion M (2008) Evaluating anti-poverty programmes. In Schultz TP, Strauss J (eds) *Handbook of Development Economics, Volume 4.* Amsterdam: Elsevier.

Robinson M (2001) *The microfinance revolution: sustainable finance for the poor.* Washington, DC: The World Bank.

Robinson M (2002) *The microfinance revolution volume 2: lessons from Indonesia.* Washington, DC: The World Bank.

Rodgers M, Sowden, A, Petticrew M, Arai L, Roberts H, Britten N, Popay J (2009) Testing methodological guidance on the conduct of narrative synthesis in systematic reviews: effectiveness of interventions to promote smoke alarm ownership and function. *Evaluation,* 15 (1): 49-73.

Rogaly B (1996) Microfinance evangelism, 'destitute women', and the hard selling of a new anti-poverty formula. *Development in Practice,* 6 (2): 100-112.

Roodman D, Morduch J (2009) The impact of microcredit on the poor in Bangladesh: revisiting the evidence. Center for Global Development, Working Paper No. 174, June.

Rosenbaum PR (1987) Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika,* 74 (1): 13-26.

Rosenbaum PR (2002) *Observational studies.* New York: Springer.

Rosenbaum PR (2005) Sensitivity analysis in observational studies. In Everitt BS, Howell DC (eds) *Encyclopedia of Statistics in Behavioural Science.* Chichester: John Wiley and Sons.

Rosenbaum PR (2010) *Design of observational studies.* New York: Springer.

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika,* 70 (1): 41-55.

Rosenbaum PR, Rubin DB (1984) Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association,* 79 (387): 516-524.

Rosenbaum PR, Silber JH (2001) Matching and thick description in an observational study of mortality after surgery. *Biostatistics,* 2 (2): 217-232.

Rosenberg R (2010) *Does microcredit really help poor people?* CGAP Focus Note, No. 59.

Roy A (2010) Poverty capital: microfinance and the making of development. Routledge, London.

Rubin DB (1973a) Matching to remove bias in observational studies. *Biometrics,* 29 (1): 159-183.

Rubin DB (1973b) The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics,* 29 (1): 185-203.

Rubin DB (1974) Estimating causal effects of treatments in randomised and non-randomised studies. *Journal of Educational Psychology,* 66 (5): 688-701.

Rubin DB (1977) Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics,* 2 (1): 1-26.

Rubin DB (1978) Bayesian inference for causal effects: the role of randomisation. *The Annals of Statistics,* 6 (1): 34-58.

Rutherford S (2001) *The poor and their money.* New Delhi: Oxford University Press.

Saretsky G (1975) The John Henry effect: potential confounder of experimental vs control group approaches to the evaluation of educational innovations. *The American Educational Research Association's Annual Meeting.* Washington, DC, 2 April 1975.

Scheffer M (2009) *Critical transitions in nature and society*. Princeton University Press.

Schuler SR, Hashemi SM (1994) Credit programmes, women's empowerment, and contraceptive use in rural Bangladesh. *Studies in Family Planning,* 25 (2): 65-76.

Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA,* 273(5): 408-412.

Scriven M (2008) A summative evaluation of RCT methodology: an alternative approach to causal research. *Journal of MultiDisciplinary Evaluation,* 5 (9): 11-24.

Scottish Intercollegiate Guidelines Network (SIGN), n.d. Guidelines methodology, Appendix B. Available at: http://www.sign.ac.uk/methodology/index.html, accessed June 2011.

Sebstad J, Chen G (1996) Overview of studies on the impact of microenterprise credit. Report submitted to USAID assessing the impact of microenterprise services (AIMS), June.

Sebstad J, Neill C, Barnes C, Chen G (1995) *Assessing the impact of microenterprise interventions: a framework for analysis*. Washington, DC, USAID.

Seiber EE, Robinson AL (2007) Microfinance investments in quality at private clinics in Uganda: a case-control study. *BMC Health Services Research,* 7: 168.

Sen AK (1999) *Development as freedom.* Oxford: Oxford University Press.

Setboonsarng S, Parpiev Z (2008) Microfinance and the millennium development goals in Pakistan: impact assessment using propensity score matching. Asian Development Bank Institute (ADBI) Discussion Paper No. 104, March.

Shadish WR, Cook TD, Leviton LD (1991) *Foundations of programme evaluation: theories of practice.* Newbury Park, CA: Sage Publications.

Shadish WR, Cook TD, Campbell DT (2002) *Experimental and quasi-experimental designs for generalised causal inference.* Boston: Houghton Mifflin Company.

Shah M, Rao R, Shankar PSV (2007) Rural credit in 20th century India: overview of history and perspectives. *Economic and Political Weekly,* 42 (15): 1351-1364.

Shimamura Y, Lastarria-Cornhiel S (2010) Credit programme participation and child schooling in rural Malawi. *World Development,* 38 (4): 567-580.

Shirazi NS, Khan AU (2009) Role of Pakistan poverty alleviation fund's microcredit in poverty alleviation: a case of Pakistan. *Pakistan Economic and Social Review,* 47 (2): 215-228.

Singh S, Loke YK, Furberg CD (2007) Long-term risk of cardiovascular events with rosiglitazone - a systematic review and meta-analysis. *JAMA,* 298: 1189-1195.

Smith SC (2002) Village banking and maternal and child health: evidence from Ecuador and Honduras. *World Development,* 30 (4): 707-723.

Smith JA, Todd P (2005) Does matching overcome LaLonde's critique of non-experimental estimators? *Journal of Econometrics,* 125: 305-353.

Snodgrass D, Sebstad J (2002) Clients in context: the impacts of microfinance in three countries: synthesis report. Report submitted to USAID assessing the impact of microenterprise services (AIMS), January.

Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, Hing C, Kwok CS, Pang C, Harvey I (2010) Dissemination and publication of research findings: an updated review of related biases. *Health Technology Assessment,* 14 (8): 1-215.

Ssendi L, Anderson AR (2009) Tanzanian microenterprises and microfinance: the role and impact for poor rural women. *Journal of Entrepreneurship,* 18 (1): 1-19.

Steele F, Amin A (1998) *The impact of an integrated microcredit programme on women's empowerment and fertility behavior in rural Bangladesh.* New York: Population Council.

Steele F, Amin S, Naved RT (2001) Savings/credit group formation and change in contraception. *Demography,* 38 (2): 267-282.

Stewart R, van Rooyen C, Dickson K, Majoro M, de Wet T (2010) What is the impact of microfinance on poor people? A systematic review of evidence from sub-Saharan Africa. Technical Report, EPPI-Centre, Social Science Research Unit, University of London.

Stiglitz, J. E., (1990) Peer monitoring and credit markets. *World Bank Economic Review,* 4 (3): 351-366.

Stiglitz JE, Weiss A (1981) Credit rationing in markets with imperfect information. *The American Economic Review,* 71 (3): 393-410.

Subramanian S, Sadoulet E (1990) The transmission of production fluctuations and technical change in a village economy: a social accounting matrix approach. *Economic Development and Cultural Change.* 39(1): 131-73.

Swain RB, Van Sanh N, Van Tuan V (2008) Microfinance and poverty reduction in the Mekong delta in Vietnam. *African and Asian Studies,* 7 (2-3): 191-215.

Swain RB, Wallentin FY (2009) Does microfinance empower women? Evidence from self-help groups in India. *International Review of Applied Economics,* 23 (5): 541-556.

Takahashi K, Higashikata T, Tsukada K (2010) The short-term poverty impact of small-scale, collateral-free microcredit in Indonesia: a matching estimator approach. *The Developing Economies,* 48 (1): 128-155.

Tedeschi GA (2008) Overcoming selection bias in microcredit impact assessments: a case study in Peru. *Journal of Development Studies,* 44 (4): 504-518.

Tedeschi GA, Karlan D (2010) Cross-sectional impact analysis: bias from drop-outs. *Perspectives on Global Development and Technology,* 9 (3-4): 270-291.

Tesfay GB (2009) Econometric analyses of microfinance credit group formation, contractual risks and welfare impacts in northern Ethiopia. *Agricultural Economics and Rural Policy.* Wageningen: Wageningen University.

Tesfay GB, Gardebroek C (2011) Does microfinance reduce rural poverty? Evidence based on household panel data from northern Ethiopia. *American Journal of Agricultural Economics,* 93: 43-55.

The Economist (2009) A partial marvel: microcredit may not work wonders but it does help the entrepreneurial poor. *Economist,* 16 July 2009.

The Economist (2011) Dismal ethics: an intensifying debate about the case for a professional code of ethics for economists. http://www.economist.com/node/17849319.

Todd H (1996) *Women at the centre.* Dhaka: University Press Limited.

Van der Weele KD, Van der Weele TJ (2007) Microfinance impact assessment: evidence from a development programme in Honduras. *Savings and Development,* 31 (2): 161-192.

Varian HR (1990) Monitoring agents with other agents. *Journal of Institutional and Theoretical Economics,* 146 (2): 153-174.

Venkata NA, Yamini V (2010) Why do microfinance clients take multiple loans? MicroSave India Focus Note 33, February.

von Pischke JD (1991) *Finance at the frontier: debt capacity and the role of credit in the private economy,* Washington DC, The World Bank.

World Cancer Research Fund (WCRF) (1997) *Food, nutrition and the prevention of cancer: a global perspective*, American Institute for Cancer Research.

Wydick B (2001) Group lending under dynamic incentives as a borrower discipline device. *Review of Development Economics,* 5 (3): 406-420.

Yunus M (1999) *Banker to the poor: microlending and the battle against world poverty.* New York: Public Affairs.

Zaman H (1999) Assessing the impact of microcredit on poverty and vulnerability in Bangladesh. The World Bank, Policy Research Working Paper Series: 2145.

Zeller M, Sharma M, Ahmed AU, Rashid S (2001) Group-based iinancial institutions for the rural poor in Bangladesh: an institutional- and household-level analysis. *Research Report of the International Food Policy Research Institute,* (120): 97-100.

Appendices

6.1 Appendix 1: Authors, funders and statement of conflict of interest

Professor James G Copestake (JGC), Reader in International Development, University of Bath. Before joining the University of Bath in 1991, James worked for development agencies in Bolivia, India and Zambia. He has carried out research in the fields of agrarian change, aid management, microfinance, measurement of poverty and social protection. Recent research projects include: Imp-Act ('Improving the impact of microfinance on poverty: an action research project') sponsored by the Ford Foundation. He has a degree in economics from Cambridge University, and an MSc and a PhD in agricultural economics from Reading University. In this project he is responsible for content and review.

Dr Maren Duvendack (MD), Postdoctoral Fellow at the International Food Policy Research Institute (IFPRI) and an Honorary Research Fellow at the University of East Anglia (UEA). She has an interest in rigorous impact evaluations, microfinance, micro-development economics and applied micro-econometrics. She completed her PhD at UEA entitled: 'Smoke and Mirrors: Evidence from Microfinance Impact Evaluations in India and Bangladesh'. MD has completed an impact evaluation of SEWA Bank in India and replicated the results of an impact evaluation of three major microfinance interventions in Bangladesh. MD is responsible for content, information retrieval and statistical analysis, i.e. survey, meta-analysis and replication.

Dr Lee Hooper (LH), Senior Lecturer in Evidence Synthesis and Nutrition, UEA. Lee is an editor of the Cochrane Heart Group and has published over 30 peer-reviewed publications, most of which are systematic reviews. She has a degree in biochemistry from UEA, a Diploma in Dietetics from Leeds and a PhD on systematic reviews in diet and cardiovascular disease from Manchester University. In this project, LH is responsible for providing advice and training in data search, development of a study inclusion and exclusion form; development of a data extraction and study validity assessment forms; data synthesis including narrative synthesis and data tables.

Dr Yoon Loke (YL), Senior Lecturer in Clinical Pharmacology, UEA. He is based in the School of Medicine at UEA which also houses the Campbell & Cochrane Economics Methods Group (www.c-cemg.org/) and Adverse Effects Methods Group of which YL is co-convenor. YL has published widely on systematic reviews and meta-analyses and is an expert in the field. He is a qualified MD and has a MB BS from London University. YL has an advisory role in this project on systematic review and meta-analysis methods.

Dr Richard Palmer-Jones (RPJ), Reader in the School of International Development, UEA. Richard is an economist with interests in poverty and inequality, microfinance, nutrition, health, education, agriculture, irrigation and natural resources, and impact evaluation. He has degrees in both economics and agriculture and has worked extensively in Malawi, Nigeria, Bangladesh and India. His recent research has focused on South Asia, in particular on the analysis of secondary data on poverty and ill-being, pro-poor growth, the attainment of millennium development goals (MDGs), and governance of natural resources, as reflected in recent research grants and publications. He has been working on impact evaluation of development projects and programmes; he has led teams to review non-governmental organisation (NGO) activities in Bangladesh (RDRS

and World Vision) and worked with Proshika, Grameen Bank and BRAC. He is an editor of the Journal of Development Studies. RPJ is responsible for content and statistical analysis, i.e. survey, meta-analysis and replication.

Dr Nitya Rao (NR), Senior Lecturer in Gender Analysis and Development at the School of International Development, UEA. Nitya's research interests include gendered changes in land and agrarian relations, migration, livelihood and well-being, equity issues in education policies and provisioning, gendered access and mobility, and social relations within environmental and other people's movements. She has degrees from Delhi University, an MA on Gender Analysis in Development and a PhD on land and livelihoods in India from UEA. In this project NR is responsible for content and particularly focuses on the role of microfinance on women's empowerment.

JGC, RPJ, NR and MD have extensive experience in literature reviews, expert commentaries, and, in the case of RPJ and MD, in critical replication studies. YL and LH are systematic reviews experts.

JGC has extensive professional involvement in microfinance policy analysis and evaluation.

LH and YL have no prior involvement with microfinance institutes.

RPJ has recently begun to work more intensively in the field of microfinance.

NR has gender expertise and has worked extensively on microfinance as part of broader resource access issues in relation to women's empowerment and family well-being, especially in India.

MD's PhD was on the impact evaluation of microfinance interventions, focusing on SEWA Bank in India and three major microfinance interventions in Bangladesh in the early 1990s.

We would like to thank DFID for funding this systematic review.

Thanks to Jake Lomax and Julie Humberstone for excellent research assistance. Thanks also to Birte Snilstveit and Hugh Waddington at 3ie for providing comments and organising the peer review.

Contact details

Maren Duvendack
University of East Anglia
School of International Development
Norwich
NR4 7TJ
UK
Email: m.duvendack@uea.ac.uk

Conflict of interest

There were no conflicts of interest in the writing of this review.

6.2 Appendix 2: Inclusion/exclusion form - microfinance systematic review

Study: author year journal ref/website
Reviewer: MD JGC RPJ

	Issue	Reviewer decision
1	**Participants**: Individuals living in poor, lower and upper-middle income countries with very few assets that could be used as collateral, and are poor, excluded or marginalised within their society. Participants can be individuals, households and microenterprises.	Yes / No/ ?
2	**Exposure or intervention**: Microcredit, 'credit plus' or 'credit plus plus' programmes of any sort that include one or more of loans, savings, insurance or other financial services. Provision is by basic, transformed or commercial NGO-type MFIs, commercial banks, credit cooperatives and other public sector providers of financial services.	Yes / No/ ?
3	**Duration of the microfinance programme at least 3 years.**	Yes / No/ ?
4	**Methodologies**: Controlled trials, before/after studies, action research, observational and qualitative research, impact evaluations, or social survey datasets (circle which)	Yes / No/ ?
5	**Sample size**: Quantitative studies >100 (treatment and control combined), qualitative studies >10.	Yes / No/ ?
6	**Comparison group**: Is the effect of microfinance compared to the effect of a lack of microfinance (a comparison group, such as a time before microfinance or another location without microfinance)?	Yes / No/ ?
7	**Outcomes**: At least one of the following is reported: income, microenterprise profits and/or revenues, labour supply, employment, expenditure (food and/or non-food), assets (agricultural, non-agricultural, transport and/or other assets), housing improvements, education (enrolment and/or achievements for adults and children), health and health behaviour, nutrition, women's empowerment.	Yes / No/ ?

If all 'yes's are circled the study is 'in'. If any 'no' is circled the study is 'out'. If any '?'s are circled the study is 'pending'. Decision (circle):

in out pending

6.3 Appendix 3: Low, lower middle and upper middle income countries - Microfinance systematic review

Income groups correspond to 2009 gross national income (GNI) per capita (World Bank Atlas method). Source: **http://data.worldbank.org/about/country-classifications/country-and-lending-groups**

1.1 Low income	1.2 Lower middle income		1.3 Upper middle income
Afghanistan	Angola	Sri Lanka	Albania
Bangladesh	Armenia	Sudan	Algeria
Benin	Belize	Swaziland	American Samoa
Burkina Faso	Bhutan	Syrian Arab Rep.	Antigua and Barbuda
Burundi	Bolivia	Thailand	Argentina
Cambodia	Cameroon	Timor-Leste	Azerbaijan
Central African Republic	Cape Verde	Tonga	Belarus
Chad	China	Tunisia	Bosnia and Herzegovina
Comoros	Congo, Rep.	Turkmenistan	Botswana
Congo, Dem. Rep.	Côte d'Ivoire	Tuvalu	Brazil
Eritrea	Djibouti	Ukraine	Bulgaria
Ethiopia	Ecuador	Uzbekistan	Chile
Gambia, The	Egypt, Arab Rep.	Vanuatu	Colombia
Ghana	El Salvador	Vietnam	Costa Rica
Guinea	Georgia	West Bank and Gaza	Cuba
Guinea-Bissau	Guatemala	Yemen, Rep.	Dominica
Haiti	Guyana		Dominican Republic
Kenya	Honduras		Fiji
Korea, Dem. Rep.	India		Gabon
Kyrgyz Republic	Indonesia		Grenada
Lao PDR	Iraq		Iran, Islamic Rep.
Liberia	Jordan		Jamaica
Madagascar	Kiribati		Kazakhstan
Malawi	Kosovo		Lebanon
Mali	Lesotho		Libya
Mauritania	Maldives		Lithuania
Mozambique	Marshall Islands		Macedonia, FYR
Myanmar	Micronesia, Fed. Sts.		Malaysia
Nepal	Moldova		Mauritius
Niger	Mongolia		Mayotte
Rwanda	Morocco		Mexico
Sierra Leone	Nicaragua		Montenegro
Solomon Islands	Nigeria		Namibia
Somalia	Pakistan		Palau
Tajikistan	Papua New Guinea		Panama
Tanzania	Paraguay		Peru
Togo	Philippines		Romania
Uganda	Samoa		Russian Federation
Zambia	São Tomé and Principe		Serbia
Zimbabwe	Senegal		Seychelles
			South Africa
			St. Kitts and Nevis
			St. Lucia
			St. Vincent and Grenadines
			Suriname
			Turkey
			Uruguay
			Venezuela, RB

6.4 Appendix 4: Log of search process

External databases

JOLIS

Date: 25 August 2010

Searched at:
http://external.worldbankimflib.org/uhtbin/cgisirsi/UQ7ccqhw8C/JL/29410072/60/495/X

Searches Performed:
Keywords anywhere '((microfinanc* OR microcredit OR micro-credit OR micro-financ* OR microenterprise OR micro-enterprise OR 'group lending') AND (evaluat* OR impact OR income OR expenditure OR consumption))'

ELDIS

Date: 27 August 2010

Searches Performed:
Note: Limited search functionality (single field only, OR unavailable, only default AND), no export facility

microfinanc* impact
111 Document results
http://www.eldis.org/index.cfm?Search_string=microfinanc*+impact&objectID=42B0EF43-E4B7-FB32-9CE720C904CB143A&submit_button=Go&RestrictToCategory=False&offerRestrictToCategory=true&Search_type=cDocument&DocCats=getDocCats&Search_author=&Search_publisher=&date_pub=6

microcredit* impact
67 results
http://www.eldis.org/index.cfm?Search_string=microcredit*+impact&objectID=42B0EF43-E4B7-FB32-9CE720C904CB143A&submit_button=Go&RestrictToCategory=False&offerRestrictToCategory=true&Search_type=cDocument&DocCats=getDocCats&Search_author=&Search_publisher=&date_pub=6

Google Scholar

Date: 27 August 2010

Searches Performed:
NOTE: Limited search functionality (no* allowed); no export facility

allintitle: microfinance AND evaluation
16 results
http://scholar.google.co.uk/scholar?hl=en&num=100&q=allintitle%3A+microfinance+AND+evaluation&as_sdt=2001&as_ylo=&as_vis=0

allintitle: microcredit AND evaluation
9 results
http://scholar.google.co.uk/scholar?hl=en&num=100&q=allintitle%3A+microcredit+AND+evaluation&as_sdt=2001&as_ylo=&as_vis=0

allintitle: microcredit evaluation OR impact OR income OR expenditure
91 results
http://scholar.google.co.uk/scholar?as_q=microcredit+&num=100&btnG=Search+Scholar&as_epq=&as_oq=evaluation+impact+income+expenditure&as_eq=&as_occt=title&as_sauthors=&as_publication=&as_ylo=&as_yhi=&as_sdt=1.&as_sdts=5&hl=en&num=100

allintitle: microfinance evaluation OR impact OR income OR expenditure
386 results
http://scholar.google.co.uk/scholar?hl=en&num=100&q=allintitle%3A+microfinance+evaluation+OR+impact+OR+income+OR+expenditure&as_sdt=2001&as_ylo=&as_vis=0

NGO/Funder websites – not all searched websites listed, sample only

DFID

Date: 25 August 2010

Searched at http://www.dfid.gov.uk/Media-Room/Publications/

Searches Performed:
http://www.dfid.gov.uk/Media-Room/Publications/?q=micro-credit
http://www.dfid.gov.uk/Media-Room/Publications/?q=microcredit
http://www.dfid.gov.uk/Media-Room/Publications/?q=micro+credit
http://www.dfid.gov.uk/Media-Room/Publications/?q=micro+finance
http://www.dfid.gov.uk/Media-Room/Publications/?q=microfinance
http://www.dfid.gov.uk/Media-Room/Publications/?q=micro-finance

CGAP

Incomplete

MicroFinance Gateway

Date: 27 August 2010

Search at
http://www.microfinancegateway.org/p/site/m/library/template.rc/advancedsearch

Searches Performed:
Keywords (all fields): impact, evaluat*, consumption, income, expenditure

MicroBanking Bulletin

Date: 27 August 2010

Search at: http://www.themix.org/publications/search

Searches Performed:
impact *OR* evaluation *OR* evaluate *OR* income *OR* consumption *OR* income *OR* expenditure

MicroFinance Network

Date: 25 August 20 10

Searches Performed:
All publications specified (see Annex1)
Criteria used: documents related to the evaluation of microfinance or microcredit
Link to page searched: MFNetwork publications @ http://www.mfnetwork.org/publications.html

USAID

Incomplete

World Bank

Incomplete

6.5 Appendix 5: Data extraction and validity assessment form - microfinance systematic review

Study details - Author(s): Year:
Journal ref:

Reviewer:

(Note: questions in ***bold italics*** are validity questions – see the validity section for details of how to answer these)

1	Study information			
1a	Research question as expressed in study			
1b	***Clarity of research question***	Done	Not done	
1c	Study design – describe			
1d	***Methodology – allocation***	Done	Not done	Unclear
1e	***Methodology – control for external circumstances***	Done	Not done	
1f	Describe the funding sources for the study, and financial or other issues declared			
1g	***Researcher bias***	Done	Not done	

	Participants	Microcredit group	No-Microcredit group
1h	Number of study participants		
1i	Ethnicity, religion and caste		
1j	Gender mix		
1k	Marital status		
1l	Age		
1m	Level of education		
1n	Household size and composition		
1o	Baseline income		
1p	Baseline assets		
1q	***Description of participants***	Done Partial Not done	
1r	***Similarity of participants***	Done Partial Not done	
1s	***Confounding re participants***	Done Partial Not done	
1t	Country of study		
1u	Country income	Low lower middle Upper middle	Low lower middle Upper middle
1v	Setting characteristics (eg urban/rural)		
1w	Numbers dropping out from baseline to outcome assessment and reasons		
1x	***Attrition bias***	Done Unclear Not done	

2	Microcredit and non-microcredit conditions	Microcredit group		No-Microcredit group
2a	Types of microcredit provided by study (e.g. credit plus, insurance, advice etc)			
2b	Types of microcredit available in area (outside of intervention if trial, generally if observational study)			
2c	Accessibility of microcredit to disadvantaged groups			
2d	Accessibility of microcredit to women			
2e	***Description of conditions***	Done	Partial	Not done
2f	Confounding interventions - describe (e.g. land reform, aid, employment intitiatives, new job opportunities, public-private partnership etc.)			
2g	***Confounding re interventions***	Done	Partial	Not done
2h	Duration of participants accessing microcredit			
2i	***Duration of microcredit***	Done	Partial	Not done
2g	Microcredit provider(s)			
2h	Other data on the microcredit provided			

Provide quantitative data as feasible, including mean and variance or median and inter-quartile range, units and descriptions of tools for assessment. Collect data at the latest time point available. For complex data use highlighter pen in original document (and page numbers below).

3	Outcomes	Microcredit group		Non-Microcredit group
3a	Time point for outcome assessment (study and participant points of view)			
3b	Income data			
3c	Profits or revenues			
3d	Labour supply/ employment			
3e	Expenditure (food/ non-food)			
3f	Assets (agricultural, non-agricultural, transport, other)			
3g	Housing changes			
3h	Education (enrolment/achievement, adults/children)			
3i	Health or health behaviours			
3j	Nutrition (intake or status)			
3k	Women's empowerment			
3l	Tools used to assess the outcomes above			
3m	***Outcome ascertainment***	Done	Partial	Not done

4	Additional information and summary			
4a	Additional validity problems:			
4b	**Any other validity problems?**	Done	Not done	
4c	Does the study offer additional information to help address any of the following (if so please describe here)? Is the impact of microcredit on any outcome modified by a) gender of borrower, b) poverty status of household, c) rural/urban setting, d) geographical location, e) presence of second income earner in the household, or f) type of product?			
4d	**Summary of validity**	Risk of bias: Low	Moderate	High

6.6 Appendix 6: Marking criteria for assessing validity

Criterion	Score as:
Clarity of the research question (F)	• 'done' when the question addressed by the research is clear, specific and addressed by the methods and results • 'not done' when there are any major problems with the above
Description of Participants (A, F)	• 'done' when the participants are in both groups are well described (e.g. gender, marital status, age, level of education, religion, caste, household size and composition, baseline (pre-microfinance) income and assets) • 'partial' when one or two of these ten factors are not well described or only in one group • 'not done' when three or more factors are not well described
Similarity of participants between microfinance and control sites (A, B, F)	• 'done' when before/after study or when populations in microfinance and control sites appear very similar (e.g. geographically close, similar participant characteristics (above), and no consistent trend that puts either group at greater risk of poor outcome) • 'partial' when there are both similarities and differences, and no consistent trend of disadvantage (or some factors are similar and some unclear) • 'not done' when the two sets of participants exhibit substantial differences (or several factors are unclear)
Methodology – allocation (A, F)	• 'done' when the intervention and control participants are allocated to microfinance or not randomly • 'unclear' when method of allocation is unclear • 'not done' when allocation to microfinance or not was by a non-random method (e.g. marketing decision, choice of an appropriate population for microfinance etc)
Methodology – control for external circumstances (F)	• 'done' where there is assessment of change between baseline and a time point at least 3 years later in the microfinance group, and this change is compared to change in the control group over the same time period • 'not done' where this design is not used (e.g. simple before after design with no separate control group or separate control group but no before/after assessment)
Duration of microfinance (G)	• 'done' when all the individual participants assessed have had access to microfinance for at least 5 years • 'partial' when the individual participants assessed have had access to microfinance for 3-5 years or at least 50% have had access for at least 5 years. • 'not done' when not either of the above
Confounding re participants (B)	• 'done' when the study attempts to account for and minimise the effects of any differences in gender, marital status, age, level of education, religion, caste, household size and composition, baseline (pre-microfinance) income and assets (or these are equivalent in both settings) • 'partial' when one or two of these factors are not

	equivalent, accounted for or minimised (or are unclear) • 'not done' when three or more factors are not equivalent, accounted for or minimised (or are unclear)
Confounding re interventions (B)	• 'done' when where there is a similar presence/absence of other poverty-alleviating interventions (such as land reform, public-private partnership, employment initiatives etc.) • 'partial' where there are some differences but they are not major • 'not done' where there are any major differences • 'unclear' where the presence or absence of these is not described
Description of conditions (F)	• 'done' when the microfinance <u>and</u> no microfinance conditions are well described (e.g. types available, from which providers, accessibility for disadvantaged, women etc.) • 'partial' when one or two of these factors are not well described • 'not done' when three or more factors are not well described
Researcher bias (A-E)	• 'done' when study funding and financial interests of authors are declared, and no bias is apparent • 'not done' when either funding or financial interests are not declared or there is potential bias apparent
Outcome ascertainment (D)	• 'done' when outcome measures are appropriate for both conditions, carried out the same way for both conditions, and appear valid and well executed • 'partial' when any one criteria above is not met • 'not done' in other cases
Attrition bias (C)	• 'done' when the participants who drop out are accounted for by study arm, and there do not appear to be big differences in the numbers dropping out, or their reasons, between arms (in before after studies the reasons for dropping out do not appear related to the outcomes assessed, 'done' for surveys without follow up • 'not done' when there are important differences in attrition • 'unclear' when not clearly described
Any other validity problems for this study?	• 'Done' if no further issues around validity • 'not done' if additional validity issues are raised
Summary of validity (I)	• Low risk of bias when all criteria above are 'done' • Moderate risk of bias when confounding of both participants and interventions is 'done' but one or two other criteria are partial, unclear or not done • High risk of bias for all remaining studies

6.7 Appendix 7: Research designs and methods

This appendix provides a more theoretical and in-depth discussion of the various research designs and analytical methods to provide the reader with more background information.

6.7.1 Research designs

The last two decades have seen advances in the improvement of putatively rigorous econometric techniques designed to account for selection bias. Given those developments experimental designs, e.g. RCTs slowly took centre stage in the area of microfinance. In experimental designs, data are derived from units of observation with individuals assigned randomly to treatment and control groups, hopefully without any bias in allocations. Since, in principle, other factors apart from treatment are equal between treatment and control groups, it is reasonable to attribute differences between groups after treatment has been applied to treatment itself. Many scientists believe that randomisation is the only method that can convincingly establish causality (Imbens and Wooldridge 2008). It is claimed that social experiments provide an accurate counterfactual and control for self-selection bias, provided that the experiment is properly implemented and individuals are randomly allocated to either treatment or control groups (Blundell and Costa Dias 2008). Furthermore, the analysis of experimental data is usually rather simple. Researchers commonly analyse the differences in mean values by treatment status. Alternatively, a regression-based approach can generate an unbiased estimator for the average treatment effect of a programme (Imbens and Wooldridge 2008).

However, limitations exist in the case of randomised experiments, i.e. double-blinding, ethical issues, pseudo-random methods, attrition and the fact that behavioural changes caused by the experiment itself, such as Hawthorne and John Henry effects, cannot be ruled out. Also, spill-over effects cannot be eliminated (Blundell and Costa Dias 2000, 2002).

On the other hand, non-experimental or observational designs have played a dominant role in the past; they are based on data that occur naturally, generally from surveys or censuses, direct observation or administrative data. In these observational situations subjects (individuals, households) generally choose what they do (or are chosen to do what happens to them), in particular, their 'treatment' status. Observational data are what happens in everyday life, and are generally characterised by non-random assignment; in everyday life people who participate are generally different to those who do not. There are numerous threats to both internal and external validity[55] that arise as a result (Shadish et al. 2002).

The discussion of experimental versus non-experimental approaches reveals that evaluation results heavily depend on the quality of the underlying data (Heckman et al. 1999). Data quality, for example, refers to the availability of a rich set of appropriate variables that are related to participation as well as outcomes (Smith and Todd 2005). Also, data on control groups located in the same environment as treatment groups greatly improve quality (Heckman et al, 1998). Many evaluations in the past provided results that were not particularly meaningful precisely because of the non-availability of rich datasets (Caliendo and Hujer 2005). Caliendo and Hujer (2005) further argued that researchers have

[55] Internal validity refers to the rigour with which one can assert that outcomes between treatment and control groups are different; external validity refers to whether the findings of this comparison are relevant to the broader population from which they are drawn (Shadish et al. 2002).

no control over the origination of data in the case of observational studies and can thus only observe outcomes for participants and non-participants after the intervention was implemented. In other words, the task of non-experimental techniques is often to restore comparability between treatment and control groups to allow solving the evaluation problem. Rosenbaum (2002) expanded on this and argued that researchers should ideally be involved in the design stages of an observational study and take part in the data collection process, in order to be able to avoid many pitfalls that later transpire in the analytical process. Rosenbaum (2002) further argued that ethnographic or other qualitative tools can be of great help in designing an observational study. In a paper published with Silber in 2001, the authors emphasised the importance of making use of qualitative tools, in particular during the design or pilot stages of a study, to improve data collection procedures and hence the overall quality of the quantitative study. Well-known statisticians and econometricians such as Heckman, LaLonde and Rosenbaum have advocated the collection of better quality data since this could possibly be the solution to the evaluation problem; they have not necessarily advocated the introduction of further even more sophisticated evaluation techniques (Heckman et al. 1999; Rosenbaum 2002, Rosenbaum and Silber 2001, Caliendo and Hujer 2005).

The data used in impact evaluations of microfinance are mainly from non-experimental quasi-experimental research designs; only two randomised designs have been used in impact evaluations, although there are other papers which draw on data from randomised designs to analyse other features pertinent to microfinance. Other designs include, in rough order of internal validity, pipeline (without randomisation), panel (before/after), with/without, and natural experiments. Each design has strengths and weaknesses in the evaluation of social programmes. We discuss briefly the background and characteristics of each in the following sections before proceeding to discuss the analytical methods used to attenuate the characteristic lacunae of the research designs.

6.7.1.1 RCTs

At the heart of every experimental design lies a natural or an artificially formulated experiment which attempts to attribute the effects of an intervention to its causes (Hulme 2000). Evaluations applying a randomised design are generally believed to provide the most robust results. There is a long tradition of experimental methods in the natural sciences. Fisher (1935), Neyman (1923) and Cox (1958) were early proponents of randomised experiments. However, few randomised experiments have been conducted in the social sciences in the past. Many of these early experiments were regarded with suspicion as to their credibility of establishing causality and their importance for researchers and policy-makers alike (Imbens and Wooldridge 2008). Moreover, in many cases the availability of experimental data are limited and therefore, observational studies have continued to capture the attention of many researchers.

Notwithstanding the critique of randomised studies, there has been a move towards RCTs in development economics driven by the so-called ‘randomistas’ (Banerjee et al. 2009, Duflo and Kremer 2005, Miguel and Kremer 2004).

Applying a randomised study design requires random assignment of potential clients to so-called treatment and control groups, whereby both groups must be drawn from potential clients whom the programme has yet to serve, so that the impact of an entire programme can be evaluated (Karlan and Goldberg 2006). This random assignment to either treatment or control group ensures that potential outcomes are not contaminated by self-selection into treatment

(Blundell and Costa Dias 2008). In other words, the potential outcomes or effects of the treatment are independent of treatment assignment. Proper randomisation ensures that individuals in treatment and control groups are equivalent in terms of observable and unobservable characteristics with the exception of the treatment status, assuming that no spill-over effects exist (Blundell and Costa Dias 2000, 2002, 2008). Hence, the mean differences in the outcomes of these individuals are understood to be the effects of the treatment itself (Caliendo and Hujer 2005).

However, limitations exist in the case of randomised experiments, i.e. double-blinding, ethical issues, pseudo-random methods, attrition and the fact that behavioural changes caused by the experiment itself, such as Hawthorne and John Henry effects, cannot be ruled out. Also, spill-over effects cannot be eliminated (Blundell and Costa Dias 2000, 2002).

Furthermore, Imbens and Wooldridge (2008) claimed that the identification problem that generally occurs when trying to establish causality cannot be solved by randomisation alone in particular when interactions between individuals or units are prevalent which is often the case. In many medical studies, such interactions are limited or non-existing, therefore randomised studies are an appropriate choice for providing robust results. The so-called double-blinding can commonly be enforced in medical studies, i.e. individuals participating in the experiment are generally not aware of their treatment status. This further enhances the robustness of the studies' results as well as improves external validity. However, double-blinding cannot necessarily be guaranteed in social science experiments and this raises serious concerns about the external validity of the results (Imbens and Wooldridge 2008).

Scriven (2008) emphasised that double-blinding is a prerequisite for a robust RCT; this is further reiterated by Goldacre (2008). As argued by Imbens and Wooldridge (2008), most medical RCTs can ensure double-blinding but the case is different for studies in the area of the social sciences. For example, RCTs evaluating the impact of education, social services or microfinance programmes are usually not even single-blinded and essentially 'zero-blinded' (Scriven 2008, p12). In other words, individuals usually discover whether they belong to treatment or control groups, which undermines the notion of double-blindedness. Hence, individuals in the treatment group may benefit from the programme

> *due either to the experimental treatment, or to the sum of that effect plus the effect of any interaction of that treatment with the psychological impact of knowing that one is part of an experiment… (Scriven 2008, p14).*

If non-interaction can be assumed, then the benefits reaped are due to the experimental treatment alone. However, RCTs in the social sciences generally do not assume non-interaction, hence the challenge to separate out the causal effects of a programme from all other factors that occur at the same time remains (Scriven 2008).

In addition, ethical questions are raised (Imbens 2009). The implementation of randomised studies is not always feasible, e.g. on which grounds can it be justified that certain individuals are assigned to treatment while others are excluded from a potentially beneficial treatment. However, it could be argued that these ethical concerns are not valid considering treatment will eventually become available to individuals in the control groups after a certain time delay.

Furthermore, Goldacre (2008) argued that pseudo-random methods are often used during the process of assigning individuals to the various treatment and control groups. It pays to investigate how exactly individuals were assigned to their respective groups; was the underlying process truly random? Many studies fail to accurately describe their assignment process. This can have consequences for the reliability of estimates obtained from an RCT.

Blundell and Costa Dias (2008) added to the limitations as follows:

> *[F]irst, by excluding the selection behaviour, experiments overlook intention-to-treat. However, the selection mechanism is expected to be strongly determined by the returns to treatment. In such case, the experimental results cannot be generalised to an economy-wide implementation of the treatment. Second, a number of contaminating factors may interfere with the quality of information, affecting the experimental results. One possible problem concerns drop-out behavior (p 19).*

Drop-out behaviour - or attrition - refers to individuals assigned to either treatment or control groups who then decide not to proceed with the experiment. It is not clear why those individuals drop out and this behaviour can have adverse effects on the results of the experiment (Blundell and Costa Dias 2008). Goldacre (2008) and Duflo et al. (2008) argued that the individuals dropping out would have been worse off than those remaining, so a risk of overstating impact estimates exists. To sum up, drop-outs change the composition of treatment and control groups thereby influencing results of the experiment since their outcomes cannot be observed (Blundell and Costa Dias 2008). Duflo, Glennerster and Kremer (2008) argued that attrition can be managed by tracking drop-out individuals to allow gathering information on them. However, this is a costly undertaking and might not always be feasible. More importantly, all randomised studies should report the level of attrition and compare drop-outs with the individuals that remain in the study to gauge whether there are systematic differences between these two groups – at least in terms of observable characteristics (Duflo et al. 2008).

Duflo, Glennerster and Kremer (2008) further argued that the generalisation and replication of randomised studies is further hampered by behavioural changes in treatment and control groups. To give an example, Hawthorne effects refer to behavioural changes in the treatment group while John Henry effects relate to behavioural changes in the control group. For example, individuals in the treatment group might positively change their behaviour for the duration of the study as they feel thankful for receiving treatment and as a response to being observed. The same behavioural changes might apply to members in the control group who might positively or in fact negatively alter their behaviour (Duflo, Glennerster and Kremer, 2008). However, a recent study by Levitt and List (2009) raised doubts about the existence of these Hawthorne effects. The authors' claimed that the evidence is not as convincing as previously thought. In fact, it cannot be said with certainty that changes in lightning led to an increase in workers' productivity. According to Duflo, Glennerster and Kremer (2008), Hawthorne and John Henry effects can be circumvented by continuing to collect data, even after the termination of the experiment, to confirm whether any behavioural changes were due to Hawthorne or John Henry effects, or due to the intervention itself.

Blundell and Costa Dias (2000, 2002), Duflo et al. (2008) reiterated that spill-over effects can have adverse effects on the impact estimates obtained from a randomised study. Spill-over effects refer to individuals in the control groups affected by treatment in physical ways or in the form of prices changes, learning or imitation effects. If spill-over effects are expected to be significant, then the experimental design can account for them. For example, the level of treatment exposure within groups can be adjusted to assess the magnitude of these spill-over effects (Duflo, Glennerster and Kremer 2008).

These issues do not exhaust the limitations of experimental studies yet. In addition, extensive cooperation from the programmes being evaluated is required. This is time and cost intensive. Researchers need to obtain the institution's consent for randomising the implementation of their microfinance services (Montgomery 2005). Moreover, for an experiment to work, the environment needs to be rigorously controlled, so that any difference in outcomes between the two groups can be adequately attributed to the impact of the intervention (Ledgerwood 1999).

The discussion so far has shown that there are threats to the internal and external validity of randomised studies, i.e. can the estimated impact be attributed to a particular intervention? Technical deficiencies such as a lack of double-blinding, pseudo-random methods as well as issues such as attrition and spill-over effects question the internal validity of experiments, while Hawthorne and John Henry effects commonly have consequences for the external validity.

RnM argued that the present drive towards encouraging RCTs also renews calls for taking a closer look at the value of observational studies which collect data through non-random processes. Observational studies are not uncontested, as there are threats to both internal and external validity that also arise in observational data. There is a risk of confounding, i.e. confounding variables are both related to the outcome that is being measured and the exposure. Typically, observational data require the application of more complex econometric techniques, i.e. PSM, IV and DID estimations. However, many of these econometric techniques cannot deal adequately with selection bias due to unobservable characteristics as mentioned earlier in this SR.

6.7.1.2 Pipelines

Following on from section 3.3, impact is estimated as the treatment effect in the existing locations less the treatment effect in control locations:

$$(1) \qquad \left(T_1^t - T_1^c\right) - \left(P_1^t - P_1^c\right)$$

Or the difference between current members who receive loans and prospective members yet to receive loans, less any difference between non-members in locations which have already received loans and non-members in locations selected to, but yet to, receive, loans.

$$(2) \qquad \left(T_1^t - P_1^t\right) - \left(T_1^c - P_1^c\right)$$

In ideal circumstances $T_1^c - P_1^c = 0$ and $T_0^t - P_0^t = 0$ because selected members and selected non-members in the two locations would be identical.

A major problem is to ensure that treatment and control groups would have had identical outcomes at the time the empirical data are produced had there been no MFI intervention, i.e.: $E\left(Y_i^T = \beta^T X_i + \mu_i^T\right) = E\left(Y_i^P = \beta^P X_i + \mu_i^P\right)$

Where Y is the outcome of interest, X_i is a vector of characteristics of the population, and □ is the error term. If one can assume that $E(\pi_i^T) = E(\pi_i^P) = 0$ then the impact is $\Delta = Y_i^T - Y_i^P = Average\ Treatment\ Effect\ or\ ATT$

Because we can never observe the true counterfactual for those with access to microfinance we are forced to seek surrogates. The surrogates (or counterfactual) will always be different individuals to the treatment sample but they should be as similar to it as possible.

To understand the threats to internal validity that can occur in this design we can consider an apparently ideal way to implement this idea. One way would be for the MFI to initiate recruitment in an area by inviting participation within a given time period; from those who present themselves for a randomly chosen set some would be allocated to treatment (get access to and make use of microfinance) (the treatment group T) and another randomly chosen group would be offered access in the future (the pipeline group P) (see Figure 2 for the set-up of a pipeline sign). Because both groups were recruited at the same time from the same population and were chosen at random, one can assume that they are equivalent, and that in the absence of any discriminatory intervention they would have equivalent outcomes under equivalent circumstances in future. Both groups would be selected under the same conditions so there should be no selection bias.

Even in this idea there are threats to validity. Clearly individuals would prefer access to potentially beneficial activities such as microfinance earlier rather than later, and at least some members of the pipeline group may be able to manipulate access. Secondly, lacking access they may seek other resources as compensation, and therefore, through a John Henry effect, become different to what the treatment group would have been had they not been treated. This can be controlled by rigid adherence to the randomised allocation.

Furthermore, there may be spill-overs from the treated to the controls with the same effect. This requires spatial, social and economic separation of the two groups which may be hard to attain while maintaining identicality of the populations from which the two samples are drawn. Random choice of villages from the same population of villages – subject to minimum 'distance' constraints - could achieve this, and would not be inconsistent with a cluster sampling design. However, there may be practical problems including the need for a quite large number of spatially separated villages (clusters) in each treatment which is likely to raise the heterogeneity among villages.

Choosing two different regions for treatment and control, as in the Copestake studies and also in Kondo et al. (2008)[56], runs into insuperable problems even if they can be claimed as similar in the appropriate way, it will be hard to defend this claim.

Thus the problems largely lie in the issue of how comparable the comparison group is to the treatment group. While most pipeline studies claim that the pipeline group is equivalent to the treatment group this claim has varying degrees of plausibility. Ideally equivalence would be demonstrated by comparisons of location and distribution of the important variables in the study (as done by Coleman 1999). But his comparison can only be done on observed variables, and is theoretically possible only for observables.

6.7.1.3 With/without (Cross section)

[56] 'The comparison *barangays*, on the other hand, are expansion areas where programme clients have been identified and organised into groups but no loans have yet been released to them' (p51).

With/without designs are the bases of most microfinance impact evaluations. They involve the comparison of treated groups with comparable untreated groups and in the absence of randomisation, are vulnerable to placement and selection biases. These may be mitigated by features of data design, and by methods of analysis. Key problems in designs are that treatment groups may not include drop-outs or graduates, and control groups may not come from the same population and sampling frame as the treatment group. Drawing the control group from the same community is risky since in most cases those who have chosen not to become members will clearly be different to those who have chosen to become members, generally as a result of an optimisation process (de Janvry et al. 2010). Also, since microfinance is likely to have spill-over effects to neighbours and to the local economy through general equilibrium effects, the comparison of MFI members and similar non-members from their own communities will be biased if it fails to account for these spill-overs.

Taking control groups from other communities risks placement bias unless the communities are demonstrably comparable (ex-ante). This can be achieved to some extent by matching the communities; it may not be achieved by random choice of control groups or communities from which to draw them unless the treatment communities were themselves randomly chosen from the same domain.

Many papers using data from with/without designs draw their control groups from different geographical domains; some provide descriptive statistics on observable characteristics, often with statistical tests of differences between treatment and control sub-samples, but this approach can not demonstrate equivalence on unobservables or variables for which there are no data.

Common analytical methods to mitigate biases due to non-comparable treatment and control groups include PSM, IV, and fixed and random effect estimation, and control functions using community level variables.

6.7.1.4 Natural experiments

Only one natural experiment was included in studies meeting our selection criteria (Kaboski and Townsend, 2005 and 2009). Natural experiments have been much sought after since the study by Duflo (1999) of a schooling programme introduced at different times in different geographical locations (see also Osili and Long 2008 for a very similar design based on the introduction of Universal Primary Education (UPE) in Nigeria).

Natural experiments exploit some difference to identify impact of a programme on the assumption that the difference is between statistically equivalent domains. Thus, in the case of Kaboski and Townsend (2005), an ongoing longitudinal survey in Thailand allowed variation in the presence of village level MFIs with different lending policies as part of an identification strategy to estimate their impacts. The spatial variation in incidence of institutions with different policies constitutes a ‘natural experiment’ in the impact of these institutions and their policies.

A later paper (Kaboski and Townsend 2009) used the data from the same survey to estimate the impact of a government programme (the Million Bhat Village Fund) which was introduced more or less simultaneously in all survey locations during the survey. Identification was achieved because the programme allocated the same fund (Thai Bhat 1million) to each ‘village’ regardless of size, thereby resulting in very different availability of loans - corresponding to the inverse of the village size.

Key assumptions required for natural experiments to appropriately identify impacts are that different domains are functionally equivalent – that is that there is no systematic difference between treatment and control groups that interacts with the treatment which could account in part for the impacts.

6.7.2 Methods of analysis

A number of econometric methods for overcoming, mitigating, or at least documenting the existence and consequences of selection bias have been developed. However, these econometric techniques have limitations and are often poorly executed or simply misunderstood, as a review of the studies we included in this SR shows. A critique of econometric techniques is not new; in a landmark paper Leamer (1983) criticised the key assumptions many econometric methods are built on; despite this pessimistic view on the usefulness of econometric methods, there has been a trend towards ever more sophisticated techniques which, however, did not necessarily provide the solution to the selection bias challenge.

Apart from technical challenges that impact evaluations must grapple with, they are further hampered by conflicting agendas of the various players involved. Such agendas influence the design, execution and the results of an impact evaluation. Hence, Pritchett (2002) argued that it is not surprising that there are so few rigorous impact studies. Not only is that a phenomenon in the area of microfinance, but health and education interventions are met with the same fate. Pritchett (2002) concluded that programmes usually have few incentives to be assessed seriously.

6.7.2.1 PSM

Matching has become a very popular technique in the area of development economics in recent years and has its roots in the experimental literature beginning with Neyman (1923). Rubin (1973a, 1973b, 1974, 1977, 1978) expanded on this literature and essentially laid the conceptual foundations of matching. The technique was further refined in particular by Rosenbaum and Rubin (1983, 1984). Econometricians got involved in advancing matching techniques in the mid-1990s; see studies by Heckman, Ichimura and Todd (1997, 1998), Heckman et al (1998) and Heckman et al. (1999).

The basic idea of matching is to compare a participant with one or more non-participants who are similar in terms of a set of observed covariates *X* (Caliendo and Kopeinig 2005, 2008, Rosenbaum and Silber 2001). In a next step, the differences in outcome variables for participants and their matched non-participants are calculated, i.e. the average treatment effect on the treated (ATT) is the mean difference between participants and matched non-participants (Morgan and Harding 2006). The objective of this technique is to account for selection on observables. The drawback is that selection on unobservables remains unaccounted for.

Despite this drawback, Dehejia and Wahba (1999, 2002) concluded that PSM results are a good approximation to those obtained under an experimental approach. They re-analysed the study of LaLonde (1986) and employed PSM to illustrate that PSM can in fact approximate the results obtained from an experimental setting. The author's argued that their results

> *'are close to the benchmark experimental estimate' (Dehejia and Wahba 1999, p1062), i.e. '[a] researcher using this method [Author's note:*

> *propensity score matching] would arrive at estimates of the treatment impact ranging from $1,473 to $1,774, close to the benchmark unbiased estimate from the experiment of $1,794' (ibid, p1062).*

However, Smith and Todd (2005) argued that the PSM estimates calculated by Dehejia and Wahba (1999, 2002) were sensitive to their choice of a particular sub-sample of LaLonde's (1986) data and found evidence that a DID approach was in fact more appropriate as an evaluation strategy in this context than PSM, as proposed by Dehejia and Wahba (1999, 2002). Overall, the outcome of this debate remains inconclusive with strong evidence provided by all parties involved.

The central assumption of PSM that needs to be observed is referred to as the Conditional Independence Assumption (CIA) or unconfoundedness. This assumption is denoted as follows, where the notation is taken from Heckman et al (1998) and Caliendo (2006) who in turn have taken it from Dawid (1979)[57]:

(3) $$Y^{0}, Y^{1} \amalg D \mid X$$ (Unconfoundedness)

Where $\amalg$ represents independence. If this is correct, it follows that

(4) $$E(Y^0 \mid X, D = 1) = E(Y^0 \mid X, D = 0)$$

and

(5) $$E(Y^1 \mid X, D = 1) = E(Y^1 \mid X, D = 0)$$

which implies that the outcomes of non-participants would have the same distribution as the outcomes of participants had they not participated given conditionality on *X* (Caliendo 2006, Caliendo and Hujer 2005). Caliendo (2006) explains that

> … matching balances the distributions of all relevant, pre-treatment characteristics *X* in the treatment and comparison group (p 31)

which makes it comparable to a randomised approach. As a result, independence between potential outcomes and treatment assignment is accomplished. Assuming the following holds

(6) $$E(Y^0 \mid X, D = 1) = E(Y^0 \mid X, D = 0) = E(Y^0 \mid X)$$

and

(7) $$E(Y^1 \mid X, D = 1) = E(Y^1 \mid X, D = 0) = E(Y^1 \mid X)$$

then the counterfactual outcomes can be deduced from the outcomes obtained from participants and non-participants.

In addition, the assumption of common support or overlap will have to be met and applies to all *X* (Caliendo 2006, Caliendo and Hujer 2005):

(8) $$0 < \Pr(D = 1 \mid X) < 1$$ (Overlap)

According to Caliendo (2006), this assumption of overlap indicates that treatment and control groups provide equal support of *X*. It further ensures that *X* is not a perfect predictor that identifies a corresponding match for each participant from the pool of non-participants and the other way round. The literature encourages matching over the common support region only when

[57] The remaining notations in this section follow Caliendo and Hujer (2005).

> *... there are regions where the support of X does not overlap for the treated and non-treated individuals ... (Caliendo 2006, p 31).*

Rosenbaum and Rubin (1983) introduce the term 'strong ignorability' which applies when CIA can be maintained and when there is in fact overlap between treatment and control groups. If 'strong ignorability' is the case, then the average treatment effect (ATE) and the ATT can provide valid estimates for all *X*. However, the notion of 'strong ignorability' is often difficult to observe in practice and can be relaxed to a certain degree when the focus is on estimating ATT only. In this case, it is sufficient to assume $Y^{\uparrow}0 \amalg D \mid X$ and *P* $(D = 1 \mid X) < 1$ and hence ATT is denoted as follows (Caliendo and Hujer 2005):

(9) $$\Delta_{\downarrow}ATT^{\uparrow}MAT = E(Y^{\uparrow}1 \dashv |X, D = 1) - E_{\downarrow}x\,[E\,(Y^{\uparrow}0 \dashv |X, D = 0)|\, D = 1]$$

Where $E(Y^{\uparrow}1 \dashv |X, D = 1)$ calculates the mean outcomes of treated individuals and $E_x[E(Y^0|X, D = 0)|\, D = 1]$ provides the calculation for the matches from the control group (Caliendo and Hujer 2005). Treatment effects can be estimated by comparing mean outcomes of the matches; the differences obtained are estimates of the programme impact for these particular observations (Ravallion 2001).

PSM was not designed to control for selection on unobservables (Smith and Todd 2005). In fact, the technique is heavily dependent on the CIA or unconfoundedness assumption, i.e. selection on observable characteristics (Caliendo and Kopeinig 2005, 2008). This assumption must be maintained in order to produce unbiased PSM estimates. Selection on unobservables, or 'hidden bias' described by Rosenbaum (2002), exist without a doubt. They are driven by unobserved variables that influence treatment decisions as well as potential outcomes (Becker and Caliendo 2007). Matching estimators are commonly not robust enough to deal with selection on unobservables.

Therefore, to test the likelihood that one or more unobservables could play a role in selection, which would explain unobserved differences, sensitivity analysis has become increasingly important. Sensitivity analysis attempts to gauge the vulnerability of the assignment process into treatment to unobservables, and hence assess the quality of the matching estimates (Becker and Caliendo 2007). Few approaches for sensitivity analyses have been developed, the most well-known method is the bounding approach developed by Rosenbaum (2002).

Rosenbaum (2002) developed the 'conceptual advance' (ibid p106) of Cornfield et al. (1959) that the robustness of the estimate of difference in outcome between treatment and control groups (the impact estimate) could be assessed by asking what magnitude of selection on unobservables (hidden bias) one would need in order to explain away the observed impact. For example, in the context of death from lung cancer for smokers and non-smokers, Cornfield et al. (1959) suggested that if the ratio of likelihood of death from lung cancer for smokers to likelihood of death from lung cancer for non-smokers was high, then a similar high ratio for unobserved characteristic(s) would be required to make the unobserved characteristic(s) the true cause of higher prevalence of death from lung cancer by smokers.

Rosenbaum (2010) explained that

> *a sensitivity analysis in an observational study asks how the conclusions of the study might change if people who looked comparable were actually somewhat different ... (p367).*

In other words, the objective of sensitivity analysis is to explore whether the matching estimates are robust to selection on unobservables (Rosenbaum 2002).

Furthermore, Rosenbaum's (2002) approach did not directly assess CIA itself but tested the sensitivity of impact estimates in view of a possible violation of this identifying assumption. In the case that matching results are indeed sensitive to possible violations of CIA, alternative estimation strategies will have to be considered (Becker and Caliendo 2007). Few studies have extensively dealt with sensitivity analysis; Rosenbaum (2002) provided an in-depth discussion on the topic. Becker and Caliendo (2007), Ichino et al. (2006) and Nannicini (2007) also provided further insights[58].

Matching is a good choice when high quality datasets are available but might not be an appropriate evaluation strategy if that is not the case (Smith and Todd 2005). Dehejia (2005) concluded that PSM is indeed not the panacea for solving the evaluation problem pointing out that the correct specification of the propensity score is crucial, i.e. the balancing properties of the propensity score should be satisfied – as emphasized by Caliendo and Kopeinig (2005, 2008) and Smith and Todd (2005) – and that the sensitivity of the results require testing – as advocated by Rosenbaum (2002), Becker and Caliendo (2007), Ichino et al. (2006) and Nannicini (2007).

6.7.2.2 Instrumental variables

The IV approach is widely used in the evaluation arena and claims to control for selection on observables as well as unobservables (Heckman and Vytlacil 2007b, Basu et al. 2007). This contrasts with PSM which tries to construct an appropriate set of counterfactual cases to counteract selection on observables only. The main goal of the IV method is to identify a variable or a set of variables, i.e. instruments, that influence the decision to participate in a programme but at the same time do not have an effect on the outcome equation. Only when adequate instruments can be identified, then the IV approach is an effective strategy for estimating causal effects (Morgan and Winship 2007). In the words of Caliendo (2006),

> *the instrumental variable affects the observed outcome only indirectly through the participation decision and hence causal effects can be identified through a variation in this instrumental variable (p25).*

Imbens and Angrist (1994), Angrist et al. (1996), Heckman (1997), Angrist and Krueger (2001), Basu et al. (2007), Heckman and Vytlacil (2007b), Heckman and Urzua (2009) and Imbens (2009) have discussed IV approaches in depth.

A regressor qualifies as an instrument for Z*, which represents programme participation when uncorrelated with the error terms and it is not entirely influenced by *X*, the set or other explanatory variables (Caliendo 2006, Caliendo and Hujer 2005). The IV estimator for a binary instrument Z* $\in \{0,1\}$ can thus be denoted as follows[59]:

(10) $$\Delta^{IV} = \frac{E(Y|X,Z*=1) - E(Y|X,Z*=0)}{P(D=1|X,Z*=1) - P(D=1|X,Z*=0)}$$

[58] Various STATA commands were developed to implement sensitivity analyses such as mhbounds (developed by Becker and Caliendo 2007) which executes the bounding approach developed by Rosenbaum (2002) or sensatt (Nannicini 2007) which executes the approach developed by Ichino et al. (2006).

[59] Notation for equation (10) follows Caliendo and Hujer (2005).

The main challenge of the IV method is to identify an adequate instrument which influences programme participation but at the same time does not influence the outcome equation. In the words of Imbens and Angrist (1994), the variable Z* should be

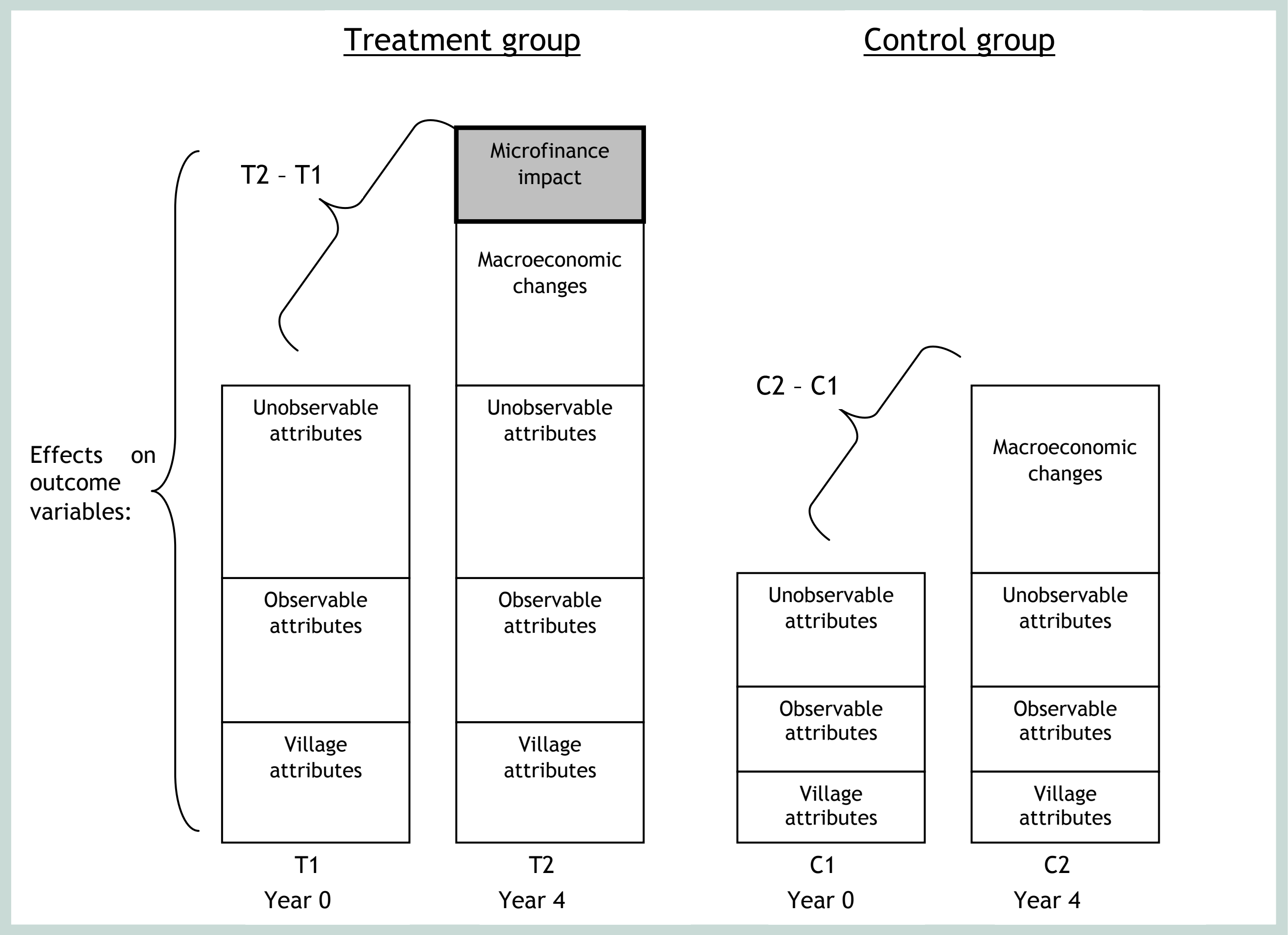
> …independent of the responses Y_i^0 and Y_i^1, and correlated with the participation indicator D_i (p468).

In many cases weak instruments are employed which can have adverse effects on the accuracy of IV estimates (Caliendo 2006, Caliendo and Hujer 2005). Overall, it can be concluded that IV estimates are only as good as the underlying instruments they employ. Heckman and Vytlacil (2007b) argued that IV estimates are not necessarily better than simple ordinary least square (OLS) estimates and might even be more biased.

6.7.2.3 Differences-in-differences

The DID approach can be explained with the help of an example from the area of microfinance. Isolating the effects of microfinance participation from all changes or events that occur during participation is a challenge (Johnson and Rogaly 1997, Armendáriz de Aghion and Morduch 2005). Furthermore, microfinance impact evaluations are confounded by placement and selection bias. Assuming an adequate dataset is available, DID can help to isolate microfinance treatment effects by eliminating certain observed and unobserved attributes. Armendáriz de Aghion and Morduch (2005) argued that there were village attributes, observable and unobservable attributes and macroeconomic changes. The authors further explained that village attributes refer to the particulars of where a person lives, e.g. access to markets which affects the likely returns to microfinance. For individuals, observable attributes can be age, education and experience, while unobservable attributes are for example entrepreneurial skills, organisational abilities, motivation, etc., all of which play a role when assessing the impact of microfinance. Hence, the aim is to isolate the microfinance impact controlling for all such attributes and measuring what is seen in the shaded box in Figure 5. However, things are more complicated because attributes also influence the decision of microfinance participants to initially join the programme. Armendáriz de Aghion and Morduch (2005) suggested a high correlation between entrepreneurial skills, age and microfinance participation. Also, if microfinance participants are wealthier than their non-participating peers before joining the programme, as suggested in studies by Coleman (2006) and Alexander (2001), then they will have more potential for income growth.

Figure 5: Illustration of DID approach, microfinance context



Source: Adapted from Armendáriz de Aghion and Morduch (2005, p204).

In summary, it can be expected that participants differ from non-participants due to unobservable characteristics. These differences can lead to contrasting reactions in the event of macroeconomic changes, therefore macroeconomic effects can have diverse impacts across both groups (Blundell and Costa Dias 2002). The authors further claimed that for a DID estimator to be unbiased, the decision to self-select into treatment must be independent from any temporary individual-level effects. It was further argued that any fixed individual-level and macroeconomic effects will eventually even out during the differencing procedure (Blundell and Costa Dias 2002).

Returning to Figure 5, Armendáriz de Aghion and Morduch (2005) suggested comparing T1 with T2. By doing this, village attributes, observable and unobservable attributes that are assumed to be time-invariant are netted out, thus enabling the microfinance impact to be captured. However, the macroeconomic changes occurring between the years of observation and which are independent of the microfinance impact are also displayed but not yet controlled for. Thus, attributing the entire difference of T2 - T1 to the impact of microfinance would be misleading. This problem cannot be solved without the introduction of a control group (Armendáriz de Aghion and Morduch 2005).

Hence, Figure 5 identifies a control group consisting of individuals who never had access to microfinance. This control group, however, is clearly not identical to the treatment group because of observable and unobservable differences. In a next step, Armendáriz de Aghion and Morduch (2005) suggested comparing T2 with C2 as this will address the biases arising from macroeconomic changes. This comparison appears to be adequate as these economic changes are felt in the same way by control and treatment groups. To isolate the true impact of microfinance, however, the single difference of T2 – T1 must be compared to the difference of C2 – C1; this is the DID approach. This approach would work well in terms of accurately measuring the causal impact of microfinance if only the underlying assumptions would hold. As mentioned earlier, it is assumed that village attributes, observable and unobservable attributes in treatment and control group are time-invariant. As a result, their effects net out when analysing T2 – T1 and C2 – C1. This assumption, however, does not hold in practice since attributes are bound to change over time and thus negatively affect the quality of DID estimates (Armendáriz de Aghion and Morduch 2005).

Overall, the DID approach is immensely popular but not without flaws, therefore it should be combined with other techniques. For example, Heckman et al. (1997) and Khandker et al. (2010) advocated combining DID with matching methods as this would account for selection on both observables and unobservables by comparing outcomes of participants before and after an intervention with before and after outcomes of matched non-participants (Caliendo 2006, Caliendo and Hujer 2005). This, however, assumes that observable and unobservable attributes in treatment and control groups are time-invariant, which is often not the case.

6.8 Appendix 8: Papers included in synthesis

Number	Study	Score	Paper
1	Abera 2010	1.099	Abera H 2010. Can microfinance help to reduce poverty? With reference to Tigrai, Northern Ethiopia. *Economics*. Mekele.
2	Abou-Ali et al. 2010	1.386	Abou-Ali H, El-Azony H, El-Laithy H, Haughton J, Khandker S (2010) Evaluating the impact of Egyptian social fund for development programmes. *Journal of Development Effectiveness*, 2 (4): 521 - 555.
3	Ahmed et al. 2000	2.073	Ahmed SM, Adams AM, Chowdhury M, Bhuiya A (2000) gender, socioeconomic development and health-seeking behaviour in Bangladesh. *Social Science & Medicine*, 51 (3): 361-371.
4	Aideyan 2009	2.485	Aideyan O (2009) Microfinance and poverty reduction in rural Nigeria. *Savings and Development*, 33 (3): 293-317.
5	USAID	1.253	Augsburg B (2006) Econometric evaluation of the SEWA bank in India: applying matching techniques based on the propensity score. Working. Paper MGSoG/2006/WP003, Maastricht University, October.
6	Banerjee et al. 2009	1.099	Banerjee A, Duflo E, Glennerster R, Kinnan C (2009) The miracle of microfinance? Evidence from a randomised evaluation. Available at: http://econ-www.mit.edu/files/4162.
7	USAID	1.946	Barnes C (2001) Microfinance programme clients and impact: an assessment of Zambuko Trust, Zimbabwe. Report submitted to USAID assessing the impact of microenterprise services (AIMS), October.
8	Bhuiya and Chowdhury 2002	1.792	Bhuiya A, Chowdhury M (2002) Beneficial effects of a woman-focused development programme on child survival: evidence from rural Bangladesh. *Social Science & Medicine*, 55 (9): 1553-1560.
9	PnK 1998	1.386	Chemin M (2008) The benefits and costs of microfinance: evidence from Bangladesh. *Journal of Development Studies*, 44 (4): 463-484.
10	USAID	2.073	Chen MA, Snodgrass D (1999) An assessment of the impact of SEWA bank in India: baseline findings. Report submitted to USAID assessing the impact of microenterprise services (AIMS), August.
11	USAID	1.946	Chen MA, Snodgrass D (2001) Managing resources, activities, and risk in urban India: the impact of SEWA bank. Report submitted to USAID assessing the impact of microenterprise services (AIMS), September.
12	Coleman 1999	0.693	Coleman BE (1999) The impact of group lending in northeast Thailand. *Journal of Development Economics*, 60 (1): 105-141.
13	Coleman 2006	0.693	Coleman BE (2006) Microfinance in northeast Thailand: who benefits and how much? *World Development*, 34 (9): 1612-1638.
14	Copestake 2002	0.693	Copestake J (2002) Inequality and the polarizing impact of microcredit: evidence from Zambia's copperbelt. *Journal of International Development*, 14: 743-755.
15	Copestake 2001	0.693	Copestake J, Bhalotra S, Johnson S (2001) Assessing the impact of microcredit: a Zambian case study. *Journal of Development Studies*, 37 (4): 81-100.
16	Copestake 2005	0.693	Copestake J, Dawson P, Fanning JP, McKay A, Wright-Revolledo K (2005) Monitoring the diversity of the poverty outreach and impact of microfinance: a comparison of methods using data from Peru. *Development Policy Review*, 23 (6): 703-723.
17	Cotler and Woodruff 2008	0.693	Cotler P, Woodruff C (2008) The impact of short-term credit on microenterprises: evidence from the Fincomun-Bimbo Program in Mexico. *Economic Development and Cultural Change*, 56 (4): 829-849.
18	Cuong 2008	1.099	Cuong NV 2008. Is a governmental microcredit Program for the poor really pro-poor? Evidence from Vietnam. *Developing Economies*, 46 (2): 151-187.

19	Deininger and Liu 2009	0.693	Deininger K, Liu Y (2009) Economic and social impacts of self-help groups in India. The World Bank, Policy Research Working Paper Series: 4884.
20	Diagne and Zeller 2001	1.386	Diagne A, Zeller M (2001) Access to credit and its impact on welfare in Malawi. Research Report 116. Washington, DC: International Food Policy Research Institute.
21	Doocy et al 2005	2.079	Doocy S, Teferra S, Norell D, Burnham G (2005) Credit Program outcomes: coping capacity and nutritional status in the food insecure context of Ethiopia. *Social Science and Medicine*, 60 (10): 2371-2382.
22	Duflo et al 2008	0.693	Duflo E, Crépon B, Parienté W, Devoto F (2008) Poverty, access to credit and the determinants of participation in a new microcredit Program in Rural Areas of Morocco. J-PAL Impact Analyses Series, No 2.
23	USAID	2.079	Dunn, E (1999) Microfinance clients in Lima, Peru: baseline report for AIMS core impact assessment. Report submitted to USAID assessing the impact of microenterprise services (AIMS), June.
24	USAID	1.946	Dunn E, Arbuckle JG (2001) The impacts of microcredit: a case study From Peru. Report submitted to USAID assessing the impact of microenterprise services (AIMS), September.
25	USAID	1.253	Duvendack M (2010a) Smoke and mirrors: evidence of microfinance impact from an evaluation of SEWA bank in India. Working Paper 24, DEV Working Paper Series, The School of International Development, University of East Anglia, UK.
26	USAID, PnK	1.253	Duvendack M (2010b) Smoke and mirrors: evidence from microfinance impact evaluations in India and Bangladesh. *Unpublished PhD Thesis. School of International Development.* Norwich: University of East Anglia.
27	PnK	1.386	Duvendack M, Palmer-Jones R (2011) High noon for microfinance impact evaluations: re-investigating the evidence from Bangladesh. Working Paper 27, DEV Working Paper Series, The School of International Development, University of East Anglia, UK.
28	Hadi 2001	2.079	Hadi A (2001) Promoting health knowledge through microcredit programmes: experience of BRAC in Bangladesh. *Health Promotion International*, 16 (3): 219-227.
29	Hashemi et al 1996	2.079	Hashemi SM, Schuler SR, Riley AP (1996) Rural credit Programs and women's empowerment in Bangladesh. *World Development*, 24 (4): 635-653.
30	Hoque 2004	2.079	Hoque S (2004) Microcredit and the reduction of poverty in Bangladesh. *Journal of Contemporary Asia*, 34 (1): 21 - 32.
31	Imai et al.2010	1.386	Imai KS Arun T, Annim SK, 2010. Microfinance and Household Poverty Reduction: New Evidence from India. World Development.
32	Imai and Azam 2010	1.099	Imai KS, Azam MS (2010) Does microfinance reduce poverty in Bangladesh? New evidence from household panel data. Discussion Paper, DP2010-24, Kobe University, September.
33	Kaboski and Townsend 2005	1.609	Kaboski JP, Townsend RM (2005) Policies and impact: an analysis of village-level microfinance institutions. *Journal of the European Economic Association*, 3 (1): 1-50.
34	Kaboski and Townsend 2009	1.609	Kaboski JP, Townsend RM (2009) The Impacts of Credit on Village Economies. SSRN eLibrary.
35	Karlan and Zinman 2010	0.693	Karlan D, Zinman J (2010) Expanding Credit Access: Using randomised supply decisions to estimate the impacts. *Review of Financial Studies*, 23 (1): 433-464.
36	PnK	2.079	Khandker SR (1996) Role of targeted credit in rural non-farm

			growth. *Bangladesh Development Studies*, 24 (3&4).
37	PnK	1.386	Khandker SR (2000) Savings, informal borrowing, and microfinance. *Bangladesh Development Studies*, 26 (2-3): 49-78.
38	PnK	1.099	Khandker SR (2005) Microfinance and poverty: evidence using panel data from Bangladesh. *The World Bank Economic Review*, 19 (2): 263-286.
39	PnK	1.386	Khandker SR, Latif MA (1996) The role of family planning and targeted credit programs in demographic change in Bangladesh. *World Bank Discussion Papers*, 337.
40	PnK	1.386	Khandker SR, Samad HA, Khan ZH (1998) Income and employment effects of microcredit programmes: village-level evidence from Bangladesh. *Journal of Development Studies,* 35 (2): 96-124.
41	Kondo et al 2008	0.693	Kondo T, Orbeta A, Dingcong C, Infantado C (2008) Impact of microfinance on rural households in the Philippines. *Ids Bulletin-Institute of Development Studies,* 39 (1): 51-70.
42	PnK	2.079	Latif MA (1994) Programme impact on current contraception in Bangladesh. *Bangladesh Development Studies,* 22 (1): 27-61.
43	PnK	1.386	McKernan S-M (2002) The impact of microcredit programmes on self-employment profits: do non-credit programme aspects matter? *Review of Economics and Statistics,* 84 (1): 93-115.
44	PnK	1.386	Menon N (2006) Non-linearities in returns to participation in Grameen bank programmes. *Journal of Development Studies,* 42 (8): 1379-1400.
45	Mohindra et al. 2008	2.079	Mohindra K, Haddad S, Narayana D (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: 2.
46	Montgomery 2005/ Setboonsarng and Parpiev 2008	0.693	Montgomery H (2005) Serving the poorest of the poor: the poverty impact of the Khushhali bank's microfinance lending in Pakistan. *Poverty Reduction Strategies in Asia: Asian Development Bank Institute (ADBI) Annual Conference.* Tokyo, 9 December 2005
47	PnK	1.386	Morduch J (1998) Does microfinance really help the poor? New evidence from flagship programmes in Bangladesh. Unpublished mimeo.
48	PnK	1.386	Nanda P (1999) Women's participation in rural credit programmes in Bangladesh and their demand for formal health care: is there a positive impact? *Health Economics,* 8 (5): 415-428.
49	Pisani and Yoskowitz 2010	2.485	Pisani MJ, Yoskowitz DW (2010) The efficacy of microfinance at the sectoral level: urban pulperias in Matagalpa, Nicaragua. *Perspectives on Global Development and Technology,* 9 (3-4): 418-448.
50	PnK	1.386	Pitt MM (1999) Reply to Morduch's 'Does microfinance really help the poor? New evidence from flagship programmes in Bangladesh'. Unpublished mimeo.
51	PnK	1.386	Pitt MM (2000) The effect of non-agricultural self-employment credit on contractual relations and employment in agriculture: the case of microcredit programs in Bangladesh. *Bangladesh Development Studies,* 26 (2 & 3): 15-48.
52	PnK	1.386	Pitt M, Khandker, SR, Chowdhury OH, Millimet DL (2003) Credit programs for the poor and the health status of children in rural Bangladesh. *International Economic Review,* 44 (1): 87-118.

53	PnK	1.386	Pitt MM Khandker SR (1998) The impact of group-based credit programs on poor households in Bangladesh: does the gender of participants matter? *Journal of Political Economy,* 106 (5): 958-996.
54	PnK	1.386	Pitt M, Khandker SR, Cartwright J (2006) Empowering women with microfinance: evidence from Bangladesh. *Economic Development and Cultural Change* 54(4) 791-831.
55	PnK	1.386	Pitt MM, Khandker SR, McKernan S-M, Latif MA (1999) Credit programmes for the poor and reproductive behavior of low-income countries: are the reported causal relationships the result of heterogeneity bias? *Demography,* 36 (1): 1-21.
56	Rafiq et al. 2009	2.079	Rafiq RB, Chowdhury JA, Cheshier PA (2009) Microcredit, financial improvement and poverty alleviation of the poor in developing countries: evidence from Bangladesh. *Journal of Emerging Markets,* 14 (1): 24-37.
57	Rahman et al. 1996	2.079	Rahman M, Davanzo J, Sutradhar SC (1996) Impact of the Grameen bank on childhood mortality in Bangladesh. *Glimpse,* 18 (1): 8.
58	Rahman 2010	2.079	Rahman S (2010) Consumption difference between microcredit borrowers and non-borrowers: a Bangladesh experience. *Journal of Developing Areas,* 43 (2): 313-326.
59	PnK	1.253	Roodman D, Morduch J (2009) The impact of microcredit on the poor in Bangladesh: revisiting the evidence. Center for Global Development, Working Paper No. 174, June.
60	Seiber and Robinson 2007	2.079	Seiber EE, Robinson AL (2007) Microfinance investments in quality at private clinics in Uganda: a case-control study. *BMC Health Services Research,* 7: 168.
61	Montgomery 2005/ Setboonsarng and Parpiev 2008	0.693	Setboonsarng S, Parpiev Z (2008) Microfinance and the millennium development goals in Pakistan: impact assessment using propensity score matching. Asian Development Bank Institute (ADBI) Discussion Paper No. 104, March.
62	Shimamura and Lastarria-Cornhiel 2010	1.386	Shimamura Y, Lastarria-Cornhiel S (2010) Credit program participation and child schooling in rural Malawi. *World Development,* 38 (4): 567-580.
63	Shirazi and Khan 2009	1.253	Shirazi NS, Khan AU (2009) Role of Pakistan poverty alleviation fund's microcredit in poverty alleviation: a case of Pakistan. *Pakistan Economic and Social Review,* 47 (2): 215-228.
64	Smith 2002	2.079	Smith SC (2002) Village banking and maternal and child health: evidence from Ecuador and Honduras. *World Development,* 30 (4): 707-723.
65	Steele et al. 2001	0.693	Steele F, Amin S, Naved RT (2001) Savings/credit group formation and change in contraception. *Demography,* 38 (2): 267-282.
66	Swain et al. 2008	2.485	Swain RB, Van Sanh N, Van Tuan V (2008) Microfinance and poverty reduction in the Mekong delta in Vietnam. *African and Asian Studies,* 7 (2-3): 191-215.
67	Swain and Wallentin 2009	1.386	Swain RB, Wallentin FY (2009) Does microfinance empower women? Evidence from self-help groups in India. *International Review of Applied Economics,* 23 (5): 541-556.
68	Takahashi et al. 2010	1.099	Takahashi K, Higashikata T, Tsukada K (2010) The short-term poverty impact of small-scale, collateral-free microcredit in Indonesia: a matching estimator approach. *The Developing Economies,* 48 (1): 128-155.
69	USAID	1.253	Tedeschi GA (2008) Overcoming selection bias in microcredit impact assessments: a case study in Peru. *Journal of Development Studies,* 44 (4): 504-518.

70	USAID	1.253	Tedeschi GA, Karlan D (2010) Cross-sectional impact analysis: bias from drop-outs. *Perspectives on Global Development and Technology,* 9 (3-4): 270-291.
71	Tesfay 2009	1.099	Tesfay GB (2009) Econometric analyses of microfinance credit group formation, contractual risks and welfare impacts in northern Ethiopia. *Agricultural Economics and Rural Policy.* Wageningen: Wageningen University.
72	Van der Weele and Van der Weele 2007	2.079	Van der Weele KD, Van der Weele TJ (2007) Microfinance impact assessment: evidence from a development program in Honduras. *Savings and Development,* 31 (2): 161-192.
73	Zaman 1999	1.386	Zaman H (1999) Assessing the impact of microcredit on poverty and vulnerability in Bangladesh. The World Bank, Policy Research Working Paper Series: 2145.
74	Zeller et al. 2001	1.386	Zeller M, Sharma M, Ahmed AU, Rashid S (2001) Group-based financial institutions for the rural poor in Bangladesh: an institutional- and household-level analysis. *Research Report of the International Food Policy Research Institute,* (120): 97-100..

6.9 Appendix 9: Papers excluded after scoring

1	Ahmed SM, Adams AM, Chowdhury M, Bhuiya A (2000) Gender, socioeconomic development and health-seeking behaviour in Bangladesh. *Social Science & Medicine*, 51 (3): 361-371.
2	Aideyan O (2009) Microfinance and poverty reduction in rural Nigeria. *Savings and Development*. 33 (3): 293-317.
3	Doocy, S., Teferra, S., Norell, D. & Burnham, G., 2005. Credit Program outcomes: coping capacity and nutritional status in the food insecure context of Ethiopia. *Social Science and Medicine*, 60 (10):.2371-2382.
4	Duflo E, Crépon B, Parienté W, Devoto F (2008) Poverty, access to credit and the determinants of participation in a new microcredit program in rural areas of Morocco. J-PAL Impact Analyses Series, No 2.
5	Hadi A, (2001) Promoting health knowledge through microcredit programmes: experience of BRAC in Bangladesh. *Health Promotion International,* 16 (3): 219-227.
6	Hashemi SM, Schuler SR, Riley AP (1996) Rural credit programs and women's empowerment in Bangladesh. *World Development,* 24 (4): 635-653.
7	Hoque S (2004) Microcredit and the reduction of poverty in Bangladesh. *Journal of Contemporary Asia,* 34 (1): 21 - 32.
8	Mohindra K, Haddad S, Narayana D (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: 2.
9	Pisani MJ, Yoskowitz DW (2010) The efficacy of microfinance at the sectoral level: urban pulperias in Matagalpa, Nicaragua. *Perspectives on Global Development and Technology,* 9 (3-4): 418-448.
10	Rafiq RB, Chowdhury JA, Cheshier PA (2009) Microcredit, financial improvement and poverty alleviation of the poor in developing countries: evidence from Bangladesh. *Journal of Emerging Markets,* 14 (1): 24-37.
11	Rahman M, Davanzo J, Sutradhar SC (1996) Impact of the Grameen bank on childhood mortality in Bangladesh. *Glimpse,* 18 (1): 8.
12	Rahman S (2010) Consumption difference between microcredit borrowers and non-borrowers: a Bangladesh experience. *Journal of Developing Areas,* 43 (2): 313-326.
13	Seiber EE, Robinson AL (2007) Microfinance investments in quality at private clinics in Uganda: a case-control study. *BMC Health Services Research,* 7: 168.
14	Smith SC (2002) Village banking and maternal and child health: evidence from Ecuador and Honduras. *World Development,* 30 (4): 707-723.
15	Swain RB, Van Sanh N, Van Tuan V (2008) Microfinance and poverty reduction in the Mekong delta in Vietnam. *African and Asian Studies,* 7 (2-3): 191-215.
16	Van der Weele KD, Van der Weele TJ (2007) Microfinance impact assessment: evidence from a development programme in Honduras. *Savings and Development,* 31 (2): 161-192.

6.10 Appendix 10: RCT Checklist

Study	Outcomes	Sign	Risk of bias
Banerjee et al	**Table 3a**		Low-moderate - ideally would have analysed a panel data set, but baseline proved unusable. Possible bias due to unobserved attrition in treatment locations., and behavioural changes in control areas
	Business Creation	**All**	
	New Business	+ive (10%)	
	Stopped business	+ive (ns)	
	Existing busineses	Profit/Inputs/Revenues/Employees/Wages/Value of assets +ive (ns)/+ive (ns)/+ive (ns)/-ive (ns)/-ive (ns)/+ive (ns)	
	New businesses		
	+ selection effect	-ive (ns)/-ive (ns)/-ive (ns)/-ive (10%)/-ive (ns)/-ive (ns)/	
	Table 3c		
	Industries of business	Old/new	
	Fd&agric	-ive(ns)/+ive(10%)	
	Clothing/sewing	+ive (ns)/-ive(ns)/	
	Rickshaw driving	-ive (ns)/-ive(10%)	
	Repair./constr	-ive (ns)/-ive(ns)	
	Crafts vendor	-ive (ns)/-ive(ns)	
	other	+ive (ns)/+ive(ns)	
	Monthly hh expenditure	Total pce/Nondurable pce/Food pce/Durable pce/Durables for business/Temptation goods/Festivals – not weddings/Any home repair/75th pctile home repair +ive (ns)/-ive (ns)/-ive (ns)/+ive (10%)/+ive (10%)/-ive (10%)/-ive(5%)/+ive (ns)/-ive (ns)	
	Women Empowerment		
	All hh		
	Primary decision maker	+ive (ns)	
	Primary non-food spend.	-ive (ns)	
	Health expend	+ive (ns)	
	Index of social outcomes	+ive (ns)	
	Women with loans		
	Primary decision maker	+ive (ns)	
	Child major illness	+ive (ns)	
	Incidence of Shocks	Health shock/Property loss/Job losss/Death +ive (ns)/-ive (ns)/-ive (ns)/-ive (ns)	
	Borrowing to deal with shocks	Borrowed/Amount/Borrowed from MFI/Amount from MFI/Borrowed from Spandana -ive (ns)/-ive (ns)/+ive (5%)/+ive (ns)/+ive (1%)	
	Conditional on shock		
	Borrowed from	Spandana/Relatives/friends/Money lender/Other source/Received gift/Other financing/Missed	

		any work/Days missed +ive (1%)/-ive (ns)/+ive (ns)/-ive (10%)/-ive (ns)/-ive (ns)/+ive (ns)/-ive (ns)	
	Table 8		
	Effects by business status		
	Borrowing	Main New/any old business/w interaction no old/new/any old	
	Any MFI	+ive (ns)/+ive (1%)/+ive (5%)/-ive(ns)/+ive (10%)	
	Non-MFI	-ive(5%)/-ive(5%)/-ive(ns)/-ive(ns)/-ive(ns)	
	Pmce		
	Durable	+ive(ns)/+ive(5%)/-ive(5%)/+ive(5%)/+ive(5%)	
	Business durable	-ive(ns)/+ive(ns)/-ive(ns)/+ive(ns)/+ive(5%)	
	Nonodurable	+ive(1%)/+ive(1%)/+ive(1%)/-ive(1%)/+ive(ns)	
	Temptation	-ive(5%)/-ive(ns)/+ive(10%)/-ive(1%)/-ive(10%)	
	Business outcomes		
	Started new business	+ive(5%)/+ive(5%)/-ive(ns)/+ive(10%)/+ive(ns)	
	Stopped business	-ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)	
	Social index	+ive(1%)/+ive(1%)/+ive(ns)/-ive(ns)/+ive(ns)	
	Table 9		
	Existing business owners		
	OLS		
	Profits	+ive (ns)	
	Drop business with zero	+ive (ns)	
	95th percentile quantile		
	Drop business with zero	+ive (ns)	
	Median regression		
	Drop business with zero	+ive (ns)	
Karlan & Zinman	Borrowing[a], business outcomes[b] and other outcomes[c]	Intention to treat basis	Moderate mainly due to high attrition (response rate) and randomisation effects (loan officers perceiving loanees with low credit scores)
	Effects on borrowing		
	From lender of close subs	All/male/female/>median/<mdian	
	Any loan outst < 50k		
	Loan size >= 50k	+ive(1%)/+ive(1%)/+ive(1%)/+ive(1%)/+ive(1%)	
	Number of loans <=50k	+ive(1%)/+ive(1%)/+ive(1%)/+ive(1%)/+ive(1%)	
	All formal	+ive(1%)/+ive(1%)/+ive(1%)/+ive(1%)/+ive(1%)	
	Any loan outst < 50k	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)	
	Loan size >= 50k	+ive(5%)/+ive(5%)/+ive(5%)/+ive(ns)/+ive(1%)	
	Number of loans <=50k	+ive(ns)/+ive(ns)/+ive(10%)/+ive(ns)/+ive(ns)	
	All informal		
	Any loan outst < 50k	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)	
	Loan size >= 50k	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)	
	Number of loans <=50k	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)	
	All loans	+ive(ns)/-ive(ns)/+ive(ns)/-ive(ns)/+ive(ns)	
	Any loan outst < 50k	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)/+ive(5%)	

Loan size >= 50k	+ive(ns)/+ive(ns)/+ive(ns)/-ive(ns)/+ive(10%)
Number of loans <=50k	
Business outcomes	+ive(ns)/+ive(ns)/+ive(10%)/+ive(ns)/+ive(ns)
Table 5	+ive(ns)/+ive(ns)/+ive(10%)/+ive(5%)/+ive(ns)
profit	+ive(ns)/+ive(ns)/+ive(10%)/+ive(ns)/+ive(ns)
profit trimmed	-ive(ns)/+ive(ns)-ive(ns)/-ive(ns)/-ive(ns)
log profit	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
total sales	-ive(ns)/+ive(ns)/-ive(ns)/-ive(ns)/+ive(ns)
total sales trimmed	
log total sales	-ive(ns)/-ive(ns)/+ive(ns)/-ive(ns)/+ive(ns)
Business inputs	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)/-ive(ns)
Value of inventory	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)/-ive(ns)
” “ “ trimmed	-ive(10%)/-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
Lovg inventory	-ive(10%)/-ive(ns)/-ive(ns)/-ive(5%)/-ive(ns)
Number of businesses	-ive(5%)/-ive(10%)/-ive(ns)/-ive(10%)/-ive(ns)
Number of helpers	+ive(ns)/+ive(10%)/-ive(ns)/-ive(ns)/-ive(ns)
Number of paid helpers	
Number unpaid helpers	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)/+ive(ns)
Table 6	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
Second job	-ive(ns)/-ive(ns)/-ive(5%)/-ive(ns)/-ive(ns)
Any member helping	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
“ “ empl outside	-ive(ns)/-ive(ns)/+ive(10)/-ive(ns)/-ive(ns)
Any overseas wkr	+ive(ns)/+ive(ns)/-ive(ns)/+ive(10%)/-ive(ns)
Any students	-ive(ns)/-ive(ns)/+ive(ns)/-ive(ns)/-ive(ns)
Table 7	+ive(ns)/+ive(ns)/-ive(10%)/+ive(ns)/-ive(ns)
Non-inventory fixed	-ive(ns)/+ive(ns)/-ive(1%)/+ive(ns)/-ive(ns)
Purchased any assets	-ive(ns)/-ive(ns)/-ive(1%s)/+ive(ns)/-ive(5%)
Sold any assets	-ive(ns)/-ive(ns)/+ive(ns)/-ive(ns)/-ive(5%)
wall_concrete	
Floor_concrete	-ive(ns)/-ive(ns)/-ive(5%)/-ive(5%)/+ive(ns)
roof_concrete_metal	-ive(5%)/-ive(ns)/-ive(ns)/-ive(1%)/-ive(ns)
Phone	+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)/-ive(ns)
Table 8	+ive(ns)/-ive(ns)/+ive(ns)/-ive(ns)/+ive(ns)
health_insurance	ordered probit
other_insurance	-ive(ns/-ive(ns)/+ive(ns)/+ive(ns)/-ive(ns)
any_savings	+ive(5%)/+ive(10%)/+ive(10%)/-ive(ns)/-ive(ns)
any_remittances_out	+ive(ns)/-ive(ns)/+ive(ns)/ive(10%)/-ive(ns)
Table 9	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)
trust_1	OLS
trust_2	+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)
trust_3	+ive(1%)/+ive(1%)/+ive(1%)/+ive(1%)/+ive(ns)
trust_4	+ive(5%)/+ive(1%)/+ive(1%)/+ive(5%)/+ive(ns)
get_emergency_friend	
get_emergency_family	
get_emergency_famfriend	
	-ive(ns)/-ive(ns)/-ive(ns)/+ive(ns)/-ive(ns)
Table 10	-ive(ns)/+ive(ns)/-ive(ns)/+ive(ns)/-ive(ns)
Household income	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
income	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
income_trimmed	-ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)/-ive(ns)

log_income	+ive(ns)/-ive(ns)/+ive(ns)/-ive(ns)/+ive(ns)
not_poor	-ive(ns)/+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)
any_remittances_recieved	
food_qual_improved	
could_visit_doctor	
	-ive(1%)/-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
Table 11 Aggregate index of subjective well-being[60]	

Notes: a. 3 variables – any outstanding loans > 50k pesos, <= 50k pesos; loan size, for three loan sources – Lender or close substitutes; all formal loan, all informal loans; and all loans.

b. Business profits (profits and sales - 6 variables); business inputs (7 variables);

c. Human capital and occupational choice (5 variables)
non-inventory fixed assets (6 variables)
household investments and risk management (4 variables)
trust and informal access (4 variables – estimated by probit and OLS)
household income and consumption (7 variables)
subjective measures of well-being (9 variables and an aggregate index)

[60] Scales of optimism, calmness, (lack of) worry, life satisfaction, work satisfaction, job stress, decision making power, and socioeconomic status" (p. 17).

6.11 Appendix 11: Pipeline checklist

Study	Design & Method	Outcomes(Table 2)	Sign & significance	Further details	Risk of Bias and Comments
Coleman	Classic pipeline with non-participants in both	**Table 2**	Fe/nonfe/naïve/super_naive	Except super-naive model	High No coefficients significant except in super-naive model
		Months as VB member	+ive(ns)/ +ive(ns)/ +ive(ns)/ +ive(ns)		
		Sex of household head (females1)	+ive(1%)/+ive(1%)/ +ive(1%)/ +ive(1%)		
		highest educated female years.	-ive(ns)/-ive(ns)/-ive(ns)/+ive(1%)		
		highest educated male years.	+ive(ns)/ +ive(ns)/ +ive(ns)/ +ive(ns)		
		Number generations family in	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)		
		village	+ive(ns)/ +ive(ns)/ +ive(ns)/ +ive(ns)		
		Number relatives in village	+ive(10)/ +ive(10)/+ive(10)/-ive(ns)		
		member village chief or assistant?	+ive(10)/ +ive(ns)/ +ive(ns)/ +ive(ns)		
		Is female in hh a civil servant?	-ive(ns)/-ive(ns)/-ive(ns)/-ive(1%)		
		Is male in hh a civil servant?	+ive(1%)/+ive(1%)/+ive(1%)/+ive(1%)		
		Female-owned land value 5 years ago baht.	-ive(10)/-ive(ns)/-ive(ns)		
		Male-owned land value 5 years ago baht.	+ive(10)/ +ive(ns)/		
		Does hh have a village bank member?			
		Number females age 5–15	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)		
		Number females age 16–21	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)		
		Number females age 22–39	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)		
		Number females age 40–59	-ive(ns)/-ive(ns)/-ive(ns)/+ive(5%)		
		Number females age 60 and over	+ive(ns)/+ive(ns)/+ive(ns)/+ive(1%)		
		Number males age 5–15	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)		
		Number males age 16–21	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)		
		Number males age 22–39	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)		
		Number males age 40–59	+ive(ns)/+ive(ns)/+ive(ns)/-ive(ns)		
		Number males age 60 and over	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)		
		Table 3			
		Physical assets			
		Household wealth	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)		
		Women's wealth	+ive(ns)/+ive(ns)/+ive(ns)/+ive(5%)		
		Men's wealth	+ive(ns)/+ive(ns)/-ive(ns)/-ive(5%)		
		Household land value _	-ive(ns)/-ive(ns)/-ive(ns)/+ive(ns)		
		Women's land value *T*	+ive(ns)/+ive(ns)/+ive(5%)/+ive(1%)		
		.Men's land value *T*	-ive(ns)/-ive(5)/-ive(1%)/-ive(1%)		
		.Household nonland assets	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)		
		Women's nonland assets	+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)		
		Men's nonland assets	+ive(ns)/+ive(ns)/+ive(ns)/-ive(ns)		
		.Household productive assets	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)		
		Women's productive assets.	+ive(ns)/+ive(ns)/+ive(5)/+ive(ns)5		
		Men's productive assets	-ive(ns)/+ive(ns)/-ive(ns)/-ive(ns)		
		.Household nonland farm assets	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)		
		Women's nonland farm assets	+ive(ns)/+ive(ns)/+ive(ns)/+ive(5)		
		Men's nonland farm assets	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)		
		Household livestock	+ive(ns)/+ive(ns)/+ive(ns)/-ive(ns)		
		Women's livestock	+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)		
		Men's livestock	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)		
		Household business assets	+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)		
		Women's business assets	+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)		
		.Men's business assets	+ive(ns)/+ive(ns)/+ive(ns)/-ive(ns)		
		Household consumer durables	-ive(ns)/+ive(ns)/-ive(ns)/-ive(ns)		
		Women's consumer durables	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)		
		Men's consumer durables	-ive(ns)/-ive(ns)/-ive(ns)/-ive(10)		
		House value	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)		

Savings, debt, lending	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
Household savings cash	
bank deposits, etc	-ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)
Women's savings	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
Men's savings	
Household low interest debt	+ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
(interest rateF2%rmonth)	
SaÍings, debt, lending	+ive(ns)/-ive(5)/-ive(ns)/-ive(ns)
Women's low interest debt	+ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
Men's low interest debt	+ive(ns)/+ive(ns)/-ive(ns)/-ive(ns)
Household high interest debt	
interest rate)2%rmonth	+ive(5)/+ive(ns)/+ive(ns)/+ive(ns)
Women's high interest debt	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
.Men's high interest debt	
Household lending out at positive interest	+ive(5)/+ive(ns)/+ive(ns)/+ive(ns)
Women's lending out at positive interest	+ive(5)/+ive(ns)/+ive(ns)/+ive(ns)
Production, sales, expenses, and labor	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)
Household self-employment	
production sales and own	
consumption	-ive(ns)/+ive(ns)/+ive(10)/+ive(10)
Women's self-employment sales	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
Men's self-employment sales	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)
Household agricultural production	+ive(ns)/+ive(ns)/+ive(10)/+ive(10)
Women's agricultural sales	-ive(ns)/-ive(ns)/-ive(5)/-ive(5)
Men's agricultural sales	
Household animal production sales	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
and own consumption.	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
Women's animal sales	-ive(ns)/-ive(ns)/-ive(10)/-ive(5)
Men's animal sales	-ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)
Household business sales	+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)
Women's business sales	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)
Men's business sales	-ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)
Household self-employment	-ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)
expenses	+ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
Women's self-employment expenses	-ive(ns)/-ive(ns)/-ive(5)/-ive(5)
Men's self-employment expenses	+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)
Household farming expenses	-ive(ns)/-ive(ns)/-ive(1)/-ive(1)
Women's farming expenses	-ive(ns)/-ive(1)/+ive(ns)/+ive(ns)
Men's farming expenses	
Household animal-raising expenses	
	-ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)
Production, sales, expenses, and	-ive(ns)/-ive(1)/-ive(5)/-ive(5)
labor	-ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)
Women's animal raising expenses	+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)
Men's animal-raising expenses	+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)
Household business expenses	
Women's business expenses	+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)
Men's business expenses	
Household self-employment labor	+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)
hours	-ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)
Women's self-employment labor	
hours	-ive(5)/-ive(ns)/-ive(ns)/-ive(ns)
Men's self-employment labor hours	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
Health care, education	-ive(10)/-ive(ns)/-ive(ns)/-ive(ns)
Household medical expenses	+ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
Medical expenses made for women	
Medical expenses made for men	+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)
Medical expenses made for _.children	-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)

		Medical expenses made for girls
		Medical expenses made for boys
		School expenses for children in household
		School expenses made for girls
		School expenses made for boys
			+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)
			+ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
			+ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)	

		Many economic and social variables: Coleman, 1999, Table 3 lists 25 outcome variables under "assets", 12 variables under "Savings, debt lending"; 27 under "Production, Sales, expenses, and labour"; & 9 under "Health care , education"[61]	6 out of 292 possible impacts were significant at p>0.10 or better	
COleman, 2006	Pipeline t-tests	**Table 11** Household wealth Women's wealth Men's wealth Household land value Women's land value (T) Men's land value (T) Household nonland assets Women's nonland assets Men's nonland assets Household productive assets Women's productive assets (T) Men's productive assets (T) Household nonland farm assets Women's nonland farm assets (T) Men's nonland farm assets (T) Household livestock Women's livestock (T) Men's livestock (T) Household business assets (T) Women's business assets (T) Men's business assets (T) Household consumer durables Women's consumer durables (T) Men's consumer durables (T) House value Savings, debt, lending Household savings (cash, bank deposits, etc.) Women's savings (T) Men's savings (T) Household low-interest debt (interest rate 6 2%/month) (T) Women's low-interest debt (T) Men's low-interest debt (T) Household high-interest debt (interest rate > 2%/month) (T) Women's high-interest debt (T) Men's high-interest debt (T) Household lending out at positive interest (T) Women's lending out at	Ols with controls mnths rf/mnths comm. Memb -ive(ns)/+ive(1) -ive(ns)/+ive(1) -ive(ns)/+ive(ns) -ive(ns)/+ive(1)	high

[61] E.g. Women's wealth; household wealth; women's wealth; men's wealth; household land; women's land.

		positive interest (T) Production, sales, expenses, labor Household self-employment production (sales and own consumption) Women's self-employment sales (T) Men's self employment sales (T) Household agricultural production Women's agricultural sales (T) Men's agricultural sales (T) Household animal production (sales and own consumption) Women's animal sales (T) Men's animal sales (T) Household business sales (T) Women's business sales (T) Men's business sales (T) Household self-employment expenses (purchase of inputs) Women's self-employment expenses (T) Men's self-employment expenses (T) Household farming expenses (purchase of inputs) Women's farming expenses (T) Men's farming expenses (T) Household animal-raising expenses (purchase of inputs) Women's animal-raising expenses (T) Men's animal-raising expenses (T) Household business expenses (purchase of inputs) (T) Women's business expenses (T) Men's business expenses (T) Household self-employment labor hours Women's self-employment labor hours Men's self-employment labor hours Health care, education Household medical expenses (T) Medical expenses made for women (T) Medical expenses made for men (T) Medical expenses made for children (T)			

Copestake 2001	Pipeline & DID on growth of outcomes	Growthrate of profits
		Growth rate of profits
		first loan
		second loan
		
		Business diversification
		first loan
		second loan
		
		Household income growth
		first loan
		second loan
			-ive(ns)
			
			+ive(ns)
			+ive(sig)
			
			-ive(ns)
			+ive(sig)
			
			-ive(ns)
			+ive(ns)
				Control function analysis	high

Copestake 200	Pipeline & DID	**Table 2** Household income	Value of loans * loans * poverty loans - pooled/loans pooled/pooled loans*p/>pl/<pl +ive(5%)/+ive(ns)/+ive(1%)+ive(ns)/+ive(1%)	Controls	High

2			ess -ive in real terms for those with larger loans		
Copestake et al 2005	Non-clients & DID control function	Poverty status	+ive	Effect sizes not reported	
		Table 6 DiD			
		Business activities			
		change in sales	+ive(ns)		
		change in profits	-ive(ns)		
		change in family income	+ive (1%)		
		change in monthly income	+ive (1%)		
		Table 7 & 8	(model 1/model 2/model 3/poorer/richer		
		change monthly per capita income	+ive(1%)/+ive(1%)/+ive(1%)/+ive(1%)/+ive(1%)		
Cotler & Woodruff	Pipeline, panel with non-comparable control	**Table 2**		for overall results, but some differences by loan size	Medium Short term – 3rd wave of data dropped
		Business performance	Recieving a Loan/sales		
		profits	-ive (sig)/+ive(ns)		
		revenues	+ive (ns)/-ive(ns)		
		inventories	+ive (sig)/-ive(ns)		
		fixed assets	+ive(sig)/-ive(ns)		
		Table 3	Recieving loan/loan*assets/sales		
		profits	+ive(ns)/ive(sig)/+ive(ns)		
		revenues	+ive(ns)/-ive(ns)/-ive(ns)		
		inventories	+ive(sig)/-ive(sig)/+ive(ns)		
		fixed assets	+ive(ns)/-ive(ns)/-ive(ns)		
		Table 4	Amount of loan/amount * size/sales		
		profits	+ive(ns)/-ive(sig)/+ive(ns)		
		revenues	+ive(sig)/-ive(ns)/-ive(ns)		
		inventories	+ive(ns)/-ive(sig)/-ive(ns)		
		fixed assets	+ive(sig)/-ive(ns)/-ive(ns)		
		Table 5	Recieving a loan/loan * assets		
		profits	+ive(sig)/+ive(ns)		
		revenues	+ive(sig)/-ive(sig		
		inventories	+ive(ns)/-ive(sig)		
		fixed assets	+ive(sig)/-ive(ns))		
Deininger & Liu	Pipeline & PSM	**Table 5**		PSM kernel matching ATT estimates	Basically similar results for two sub-groups of participants (new, and converted)
		Female empowerment;	(trimmed ps weight/kernel)		
		social capital[a]	+ive (1%)/+ive(1%)		
		economic empowerment[b]	+ive (1%)/+ive(1%)		
		political participation[c]	+ive (1%)/+ive(1%)		
		nutritional status;			
		energy	+ive ns/+ive(ns)		
		protein	+ive (1%)/+ive(1%)		
		per capita income, consumption, and	-ive (ns) /-ive(ns)		
		assets	-ive (ns)/-ive(ns)		
			New ps/new kernel/conv ps/conv conv/non ps/non kern		
		Table 7	+ive(1)/+ive(1)/+ive(1/+ive(1)/+ive(1)/+ive(1)		
		Change	+ive(1)/+ive(1)/+ive(1/+ive(1)/+ive(1)/+ive(1)		
		female social capital	+ive(1)/+ive(1)/+ive(1/+ive(1)/+ive(1)/+ive(1)		
		female econ empowerment	+ive(5)/+ive(1)/+ive(1/+ive(ns)/+ive(ns)/+ive(ns)		
		female polit representation	+ive(1)/+ive(1)/+ive(5)/+ive(5)/+ive(ns)/+ive(ns)		
		energy intake	+ive(1)/+ive(5)/+ive(ns/+ive(ns)/+ive(ns)/+ive(ns)		
		protein intake	+ive(ns)/+ive(ns)/+ive(ns/+ive(ns)/+ive(ns)/+ive(ns)		
		consumption	+ive(ns)/+ive(ns)/+ive(ns/+ive(ns)/+ive(ns)/+ive(ns)		
		income pc			

		asset pc
		
		Table 8 Change
		female social capital
		female econ empowerment
		female polit representation
		energy intake
		protein intake
		consumption
		income pc
		asset pc
			Diff new-conv/diff new-non-part
			-ive(ns/+ive(ns)
			-ive(ns)/+ive(ns)
			0ive(ns)/+ive(5%)
			+ive(10%)/+ive(10%)
			+ive(ns)/+ive(5%)
			+ive(ns)/+ive(5%)
			-ive(ns)/-ive(ns)
			+ive(ns)/+ive(ns)		

Kondo – Philippines	Pipeline w/wo; matched baranguays; control function DID	**Table 6** simple t-tests
		Economic
		Income per capita
		expenditure per capita
		savings 1
		savings 2
		food expenditure
		poor[62]
		subsistence poor
			Participation vs non-participation
			existing/extension
			+ive(10%)/-ive(ns)
			+ive(ns)/-ive(ns)
			+ive(5%)/+ive(ns)
			+ive(5%)/-ive(ns)
			+ive(ns)/-ive(ns)
			-ive(1%)/-ive(ns
			-ive(ns)/-ive(ns)
				Expansion areas +ive (ns) all
				
				Participants likely to be poor
					moderate

		Table 7 Program loans Economic Income per capita expenditure per capita savings 1 savings 2 food expenditure	Existing areas (psm)/expansion area +ive (10%) +ive (10%) +ive (ns) +ive (ns) +ive(10%)		
		Table 8 non-GBA loans amont other loans number other loans	Existing areas/expansion areas -ive(5%)/0ive(ns) -ive(ns)/-ive(ns) +ive(1%)/+ive(ns)		
		Table 9 took up non-GBA loans Amount non-GBA loans No. non-GBA loans	-ive(10%) 0ive(ns) 0ive(ns)		

		Table 10
		has savings account
		savings < 5k
		savings 5-10k
		savings >10k
			Existing areas/expansion areas
			+ive(1%)/+ive(1%)
			+ive(ns)/+ive(ns)
			+ive(ns)/-ive(ns)
			-ive(ns)/-ive(1%)		

		Table 11 total savings
		has savings account
		savings < 5k
		savings 5-10k
		savings >10k
			+ive(1%)
			-ive(1%)
			+ive(1%)
			+ive(1*)		

		Table 12
		enterprisese & employment
		with hh enterprise
		no enterprise
		empl fam
		empl non-fam
		tot empl
			Existing areas/expansion areas
			+ive(1%)/+ive(1%)
			+ive(1%)/+ive(1%)
			+ive(1%)/+ive(ns)
			+ive(ns)/+ive(ns)
			+ive(5%)/+ive(ns)		

		Table 13 no enterprises	+ive(1%) +ive(1%)		

[62] Difference between published paper and original report – latter is consistent with text in paper

no employees

Table 14 Household Assets
agric&communal land
agr&commercial land
farm equipment
livestock & poultry
hh appliances
value hh appliances

Table 15 education outcomes
with children 6-12
% enrolled
with children 13-16
% enroled
with children 17-24
among children 17-24
years per school age child
year eper attening child

Table 16
% ill/injured
with ill/inj members
% seeking treatment
with 0-5
prop immunised
percap medical expend.

Table 17 Hunger and reduced food
hunger incidence
reduced food intake

Table 18 per capita income
per capita income
per capita expenditure
per capita savings 1
percapita savings 2
per capita food expend

Table 21 by education status
per capita income
per capita expenditure
per capita savings 1
percapita savings 2
per capita food expend

existing areas/expansion areas
+ive(ns)/0ive(ns)
-ive(ns)/-ive(ns)
-ive(ns)/+ive(ns)
+ive(1%)/+ive(1%)
+ive(ns)/+ive(ns)
-ive(ns)/+ive(ns)

+ive(ns)/+ive(1%)
+ive(ns)/+ive(ns)
+ive(5%)/+ive(5%)
-ive(ns)/+ive(ns)
+ive(1%)/-ive(ns)
+ive(ns)/+ive(ns)
+ive(ns)/+ive(ns)
+ive(ns)/+ive(ns)

+ive(ns)/+ive(ns)
+ive(1%)/+ive(5%)
+ive(ns)/+ive(ns)
-ive(1%)/+ive(ns)
+ive(ns)/+ive(ns)
-ive(ns)/-ive(ns)

-ive(ns)/+ive(ns)
-ive(ns)/+ive(ns)

q1/g2/g3/g4
-ive(1%)/-ive(1%)/-ive(ns)/-iove(1%)
-ive(1%)/-ive(5%)/+ive(ns)/+ive(1%)
-ive(1%)/-ive(1%)/-ive(ns)/+ive(1%)
-ive(1%)/-ive(1%)/-ive(ns)/+ive(1%)
-ive(1%)/-ive(ns)/+ive(ns)/+ive(1%)

primary/secondary/tertiary
-ive(ns)/+ive(1%)/+ive(5%)
-ive(ns)/+ive(1%)/+ive(ns)
-ive(10%)/+ive(ns)/+ive(5%)
-ive(10%)/+ive(ns)/+ive(5%)
-ive(ns)/+ive(10%)/+ive(5%)

Montgomery – Pakistan	Pipeline (OLS DID)	**Table 9.2** Household expenditure	(OLS) Client/corepoor/duration/duration*/poor*duration/amnt loans/core poor*amount/cycles/core poor*cycles	Logit & OLS coefficients	Results also given for interaction terms with core poor & amount borrowed And core poor * number of loan cycles Most values non-significant
		p. c: Food	-(ns)/-(1)/-(ns)/0(ns)/0)ns)/+)(ns)/+(ns)/-(ns)		
		p. C.: Non-Food	+(ns)/-(5)/+(ns)/-(ns)/0(ns)/0(ns)/-(ns)/-(ns)		
		per child medical expenditure.	-(ns)/+(ns)/+(ns)/+(ns)/0(ns)/0(ns)/+(ns)/+(ns)		
		per child education	+(ns)/-(1)/+(ns)/+(ns)/0(10)/0(10)/-(5)/+(5)		
		Table 9.3 Health and education	(logit)		
		prob attending school	+(5)/-(1)/-(1)/+(5)/0(50)/0(ns)/-(1)/+(5)		
		days absent	+(ns)/+(ns)/-(ns)/+(ns)/0(10)/0(ns)/-(ns)/-(ns)		
		seeking medical treatment	-(ns)/-(ns)/+(5)/+(ns)/0(1)/0(ns)/+(1)/-(ns)		
		treatment from trained practitioner	-(ns)/-(ns)/+(10)/+(5)/+(ns)/+(ns)/+(5)/-(ns)		
		prob. take medicine	+(ns)/-(ns)/0(ns)/+(ns)/0(ns)/0(ns)/+(ns)/+(ns)		
		prob. vaccinated	-(ns)/-(5)/+(ns)/+(1)/0(ns)/0(ns)/+(ns)/+(ns)		
		Table 904 Income and assets	(ols) client/core poor/mnths/urban*mnths/core poor*mnths/amnt/urban*amnt/core poor*amnt/cycles/urban*cycles/core poor*cycles		
		livestock (sales)	-(ns)/+(ns)/+(ns)/ /-(ns)/+(ns)//-(ns)/+(ns)//-(ns)		
		livestock (profits)	-(ns)/+(ns)/+(ns)/ /-(ns)/+(ns)//-(ns)/+(ns)//-(ns)		
		microenterprise (sales)	+(ns)/+(1)/-(ns)/+(1)/-(10)/+(ns)/+(1)/-(10)/+(ns)/+(1)/-(10)		
		microenterprise (profits)	+(ns)/-(1)/-(ns)/+(1)/+(ns)/-(ns)/+(1)/-(ns)/-(ns)/+(1)/-(ns)		
		agriculture (sales)	+(1)/-(ns)/+(ns)/ /+(5)/+(1)//+(5)/+(1)//-(5)		
Setboonsarng and Parpiev – Pakistan	Pipeline with PSM & DID (same data as Montgomery)	**Table 13** Consumption and expenditure	**Nearest neighbour** mpce_total/mpce/mpce_food/mpce_nonfood **-ive(ns)/+ive(ns)/-ive(ns)/+ive(ns)**		Impact of membership on MDG variables. Second set of results of lending on MDG variables – similar mixed results of low statistical significance)
		Agricultural inputs sales and profits	ag sales/pesticides/ farm equipment -value/rent income **+ive(ns)/+ive(ns)/+ive(sig)/+ive(sig)**		
		Animal Raising	livestock value/sales/annual inputs/agricultural profits **+ive(ns)/+ive(ns)/+ive(sig)/+ive(sig)**		
		Household outside income	**-ive(ns)**		
		Quantity and value of consumer durables	**quant/value** **-ive(ns)/-ive(ns)**		
		Non-ag enterprise assets	capital assets/net assets/monthly inputs/sales/profits **+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)**		
		Household savings	**-ive (ns)**		
		School expense per child (by sex) \ monthly expend per child	school exp per child/ per girl / per boy **-ive(ns)/-ive(ns)/-ive(ns)/+ive (ns)**		
		Health seeking	medical expenditure per capita/when ill seeks treatment/can pay for medic expend/ORS/under 5 vaccinated **+ive (ns)/+ive (ns)/+ive (ns)/+ive (ns)/+ive (ns)**		
		Women empowerment	have say in schooling/have say in health care/use contraception/incidence of domestic violence **-ive(ns)-ive(ns)/-ive(ns)/+ive(ns)**		
		Working hours for adults and children by activity	working hours adult on crops/animals/non-agric business/ total/child working hours on farm crops/animal raising/ non-agric bs/ total **+ive(ns)/+ive(ns)/-ive(ns)/+ive(ns)/+ive(ns)/+ive(sig)/-ive(sig)/+ive(ns)**		
		Table 14 effects on poor Consumption and expenditure	mpce_total/mpce/mpce_food/mpce_nonfood **+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)**		
		Agricultural inputs sales and profits	ag sales/pesticides/value of farm equipment/rental income from farm equipment		

			+ive(ns)/+ive(ns)/+ive(sig)/+ive(ns)
		Household outside income	**-ive(ns)**
		Quantity and value of consumer durables	**Quant/val** **-ive(ns)/+ive(ns)**
		Non-ag enterprise assets	gross capital assets/net capital assets/monthly inputs/sales/profits **+ive(ns)/+ive(ns)/-ive(ns)/-ive(ns)/+ive(ns)**
		Household savings	**-ive(ns)**
		School expense per child (by sex)	monthly expend per child/school exp per child/school exp per girl/school exp per boy **+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)**
		Health seeking	medical expenditure per capita/when ill seeks treatment/can pay for medic expend/ORS/under 5 vaccinated **+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)**
		Women empowerment	have say in schooling/have say in health care/use contraception/incidence of domestic violence **+ive(ns)/-ive(ns)/-ive(ns)/+ive(ns)**
		Working hours for adults and children by activity	working hours adult on crops/animals/non-agric business/ total/child working hours on farm crops/animal raising/ non-agric bs/ total **+ive(sig)/+ive(sig)/-ive(ns)/+ive(sig)/+ive(ns)/+ive(sig)/-ive(sig)/+ive(ns)**
Steele, Amin and Naved, 2001	Pipeline panel fixed/random effects models	**Table 7**Use of modern contraceptives	
		Basic (naive) estimate use all & modern methods all ASA & SC members	+ive (0.1%)
		Random effects use of modern contraception	old area member/new area non-ASA SC member/new area SC member /new area-ASA ASA member +ive (1 %)/+ive ns/+ive (ns)/+ive (5%)

Notes: a. self-reported level of trust in individuals of the same or different caste or religion from within or outside the village as well as in government officials and police, all on a 1-5 scale.
b. woman can set aside money for her own use, go to the market, to the clinic or the community centre, visit friends, or work on fields outsides the village, without asking permission from her husband or other males in the family. In either case, we use principal component method to generate an index based on a single factor.
c. frequency of their attendance at village meetings.

6.12 Appendix 12: With/without 2SLS Checklist

Study	Identification/specification tests performed?[a]	Identification problem addressed?	Method, tests	Outcome variables[63]	Sign and significance	Comments	Risk of Bias[b]
Cuong 2008	Y	Y	IV & LIML	log pcexpend () logpcincome	loan size - 2sls/gmm/liml/2sls/gmm/liml participation - 2sls/gmm/liml/2sls/gmm/liml +ive(1)/+ive(1)/+ive(1)/ +ive(1)/+ive(1)/+ive(1) +ive(1)/+ive(1)/+ive(1)/ +ive(1)/+ive(1)/+ive(1)	Various estimations: IV with FE and LIML Various estimations: IV with FE, LIMIL	Moderate
				logpcexpend (loan size) logpcincome (participation)	+ive(1) +ive(1)		
				Table 6 poverty indexes hcr pg pg2	(size of loans – not clear) 2sls/gmm/liml -ive(10)/-ive(5)/-ive(5) -ive(ns)/-ive(5)/-ive(5) -ive(10)/-ive(5)/-ive(10)		

				Appendix Table 4 2SLS
				log pc expend
				log income pc
					loan size/participation
					+ive(1)/+ive(1)
					+ive(5)/+ive(5)		

				Table 5 2sls/gmm/liml/2sls fe log pc expend log income pc	+ive(1)/+ive(1)/+ive(1)/+ive(1) +ive(1)/+ive(1)/+ive(1)/+ive(1)		
				Table 6 poverty index hcr pg pg2	2sls/gmm/liml -ive(10)/-ive(ns)/-ive(10) -ive(5)/-ive(5)/-ive(5) -ive(5)/-ive(5)/-ive(10		
				Appendix Table 7 interaction prog * pov log pc expend	Lnsize/loansize*poor/poor2002/part /part*poor/poor2002		
				2sls	+ive(ns)/-ive(ns)/-ive(1)/+ive(ns)/-ive(ns)/-ive(1)		
				gmm	+ive(ns)/-ive(ns)/-ive(1)/+ive(ns)/-ive(ns)/-ive(1)		
				liml	+ive(ns)/-ive(ns)/-ive(5)/+ive(ns)/-ive(ns)/-ive(ns)		
				log income pc 2sls	+ive(ns)/-ive(ns)/-ive(1)/+ive(ns)/-ive(ns)/-ive(5)		
				gmm	+ive(ns)/-ive(ns)/-ive(1)/+ive(ns)/-ive(ns)/-ive(5)		
				liml	+ive(ns)/-ive(ns)/-ive(ns)/+ive(ns)/-ive(ns)/-ive(ns)		
Diagne and Zeller 2001	Y	Possibly	LIML model	**Table 25** Income per capita	**All results insignificant** **+ive(ns)**	Broken down by credit membership	High
				Table 26 Crop income	**+ive(ns)**		

[63] As noted above, many papers do not set out a clear theory of change elaborating pathways between microfinance and desirable outcomes. Hence, many papers list and report impact estimates on very large numbers of outcome variables. They also use different significance levels. The variables reported are listed more fully in an Excel spreadsheet which can be made available on request. The original papers are the most useful source of both.

				Table 27 nonfarm seasonal income	**+ive(ns)**		
				Table 28 Food expenditure 5 dummies	**-ive(ns)**		
				Table 29 calorie intake	**+ive(ns)**		
				Table 30 Protein 5 dummies	**+ive(ns)**		
				Table 31 waz	**+ive(ns)**		
				Table 32 haz	**-ive(ns)**		

Imai et al. 2010 – discussed in section 5.3.1.2	Y	Possibly	Treatment effects model, essentially a Heckman procedure	**Table 2** 2stage treatment effects
				IBR (Index based Ranking)
				Income
				Food Security
					Access - Total/urban/rural/Use-Total/urban/rural
					+ive(1)/+ive(1)/+ive(1)/+ive(1)/+ive(1)/+ive(1)
					+ive(1)/+ive(1)/+ive(1)/-ive(1)/-ive(1)/-ive(1)
					+ive(1)/+ive(1)/+ive(1)/+ive(1)/+ive(1)/+ive(1)
						Results for total sample also presented but the heterogeneity renders them uninteresting.	High

				Table 3 tobit IBR Income Food Security	Amount of prod loan Total/urban/rural/Use +ive(1)/+ive(1)/+ive(10) -ive(1)/-ive(ns)/-ive(1) +ive(1)/+ive(ns)/+ive(1)		

				Table 4 tobit
				IBR
				Income
				Food Security
					Total amount of loan
					+ive(1)/+ive(1)/+ive(1)
					+ive(1)/+ive(1)/+ive(5)
					+ive(ns)/-ive(ns)/+ive(ns)		

				Appendix 2 - PSM poverty reducing effect whole sample access	Nn total/u/r/kern t/r.u +ive(1)/+ive(1)+ive(ns)/+ive(1)/+ive(1)/+ive(1)		
				Productive loan	+ive(1)/+ive(ns)+ive(1)/+ive(1)/+ive(1)/+ive(1)		

				Appendix 3-psm
				access
				poor
				mod por
					+ive(5)/+ive(ns)+ive(5)/+ive(1)/+ive(ns)/+ive(1)
					+ive(1)/+ive(ns)+ive(1)/+ive(1)/+ive(1)/+ive(1)		

				productive loan
				poor
				mod por
					+ive(1)/+ive(ns)+ive(1)/+ive(1)/-ive(ns)/+ive(1)
					+ive(1)/+ive(5)+ive(1)/+ive(1)/+ive(1)/+ive(1)		

PnK (all except those applying PSM on PnK data) (SEE APPENDIX 17 FOR FURTHER DETAILS)	Y - RnM N other studies	N - issue of mistargeting remains	LIML model	Main PnK paper consumption female non-landed assets male & female labour supply boys & girls school enrolment	Main PnK paper: positive & significant for most outcome variables for other outcome variables see Appendix 17	Morduch, Chemin, RnM, Duvendack and Duvendack and Palmer-Jones find different results see Table 10b	Main PnK paper: High
Shimamura and Lastarria-Cornhiel 2010	N	Unclear	IV	**Table 8** school attendance all 6-14 15-18	OLS/2SLS -ive (ns)/-ive (10%) -ive (ns)/+ive (ns)		High

				Table 9 + 10 Girl			
				6-14	-ive (10%)/-ive (10%)		
				15-18	+ive (ns)/+ive (ns)		
				Boy			
				6-14	+ive (ns)/-ive (ns)		
				15-18	+ive (ns)/+ive (ns)		
				Table 12 Child work Crop farming			
				6-14	-ive (5%)/+ive(ns)		
				15-18	+ive (ns)/+ive (ns)		
				Table 13 Household chores			
				6-14	-ive (5%)/-ive (1%)		
				15-18	-ive (ns)/-ive (ns)		
Zaman 1999	Y	Possibly	Heckman	**Table 4.0** poverty	Equ3.1/eq3.2/equ3.3/equ3.4: identificationwith no elig hh/no elig ff/ff/ols vill fe -ive(ns)/-ive(ns)/-ive(5%)/-ive(ns)	Paper has many problems, for example including (presumably wrongly) of BRVO as the dependent variable in the first stage and also on the RHS in the second stage of the Heckit estimation of equations	High
				Table 5.0 membership 6 land ownership * loan amount variables	-ive(10%) ml1/ml2/ml3/ul1/ul2/ul3 +ive(ns)/-ive(ns)/+ive(10)/+ive(ns)/+ive(ns)/+ive(ns)		
				Table 11 Aware that dowry is illegal	BRVO/LOADUM1/LOADUM2/LOADUM3 0.033(ns)/-0.001(ns)/ 0.047(ns)/.077(ns)		
				Aware of method of divorce	-0.006(ns)/-0.01(ns)/0.007(ns)/0.045(5%)		
				Aware of minimum marriage age	0.036(ns)/-0.076(5%)/-0.064(5%)/0.008(ns)		
				Aware of local chairman’s name	0.021(ns)/0.160(5%)/0.089(ns)/0.123(5%)		
				Owns land	0.112(5%)/0.036(ns)/0.106(10%)/0.058(ns)		
				Owns poultry	0.121(1%)/-0.094(ns)/-0.133(5%)/-0.144(5%)		
				If owns poultry % that can sell poultry independently (N = 980)	-0.103(10%)/0.048(ns)/-0.007(ns)/0.245(1%)		
				Owns livestock	-0.046(10%)/0.058(10%)/0.036(ns)/0.046(ns)		
				If owns livestock % that can sell livestock independently (N = 103)	-0.178(ns)/-0.021(ns)/0.094(ns)/-0.265(ns)		
				Owns jewelry	0.08(10%)(ns)/-0.014(ns)/-0.093(ns)/-0.089(ns)		
				If owns jewelry % that can sell jewelry independently (N = 694)	0.017(ns)/0.032(ns)/0.011(ns)/0.079(10%)		
				Has savings	0.473(1%)/0.086(10%)/0.110(5%)/0.118(1%)		
				If has savings % can use savingsindependently (N = 379)	-0.345(1%)/0.085(ns)/0.064(ns)/0.151(10%)		
				Forced pregnancy			
				Visits local market -	0.004(ns)/-0.035(10%)/-		

				Visits Matlab market -	0.006(ns)/-0.001(ns)
					0.037(ns)/0.084(ns)/0.097(5%)/0.029(ns)
					0.038(ns)/0.037(ns)/0.026(ns)/0.007(ns)		

Zeller et al. 2001	N	Unclear	IV	**Table 5.4 & 5.5** (BRAC&ASA vs RDRS) **Table 5.6** Mpce food expenditure **Table 5.7** Calorie cons. all seasons Income	Credit limit HYV cultivation/ mpce - aus/aman/boro -ive((10%)/+ive(sig)/+ive(sig)/+ive(ns) Aus/Aman/Boro +ive (5%)/+ive (5%)/+ive (ns) aus/aman/boro +ive sig)/+ive (sig)/+ive (sig +ive sig)/+ive (sig)/+ive (sig		High

Notes: a. Identification and/or specification tests can include Sargan-Hansen, Durbin-Wu-Hausman tests, and/or others.

b. For reasons explained elsewhere in this report we have included these studies which have low inherent credibility because of their with/without design, but among these we provide an essentially subjective assessment of risk of bias based on criteria in data extraction tables. It is, in our view, better to provide this judgement which influences our overall judgements rather than leave this unreported. Readers are, of course, free to disagree on the informational basis of our account.

6.13 Appendix 13: With/without PSM Checklist

Study	Balancing tests performed?	More controls than treated?	Matching quality assessed through tests?	Sensitivity analysis?	Outcome variables	Sign & Significance	Comments	Risk of bias

Abera 2010)	Reported that balancing properties are satisfied but no evidence provided.	Unclear	N	N	**Table 4.3**
					Household medical expenditures
					Expenditure on children's education
					Expenditures on social occasions
					Expenditure on clothing and personal items
					Table 4.4
					Household fixed assets (house)
					Household fixed assets(without house)
					Household productive assets
					Table 4.5
					Household expenditure on food
					Household expenditure on food & non-food
					Household poverty gap squared
						Strat/radius/nn/kernel[c]
						-ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)
						+ive(1%)/+ive(1%)/+ive(ns)/=ive(5%)
						+ive(ns)/+ive(ns)/++ive(ns)/+ive(ns)
						+ive(ns)/+ive(ns)/+-ive(ns)/-ive(ns)
						+ive(ns)/+ive(ns)/++ive(10%)/+ive(ns)
						+ive(ns)/+ive(ns)/++ive(10%)/+ive(ns)
						+ive(10%)/+ive(ns)/+ive(5%)/+ive(ns)
						-ive(5%)/-ive(ns)/-ive(ns)/-ive(ns)
						-ive(5%)/-ive(ns)/-ive(ns)/-ive(ns)
						+ive(5%)/-ive(ns)/-ive(ns)/-ive(ns)
						-ive(ns)/-ive(ns)
						+ive(5%)
						+ive(ns)
						olsfe /ols
						+ive(ns)/+ive(ns)
						re/ols
						-ive(ns)/-ive(ns)
						+ive(ns)/+ive(ns)
							PSM results only	Moderate

Study	Balancing tests performed?	More controls than treated?	Matching quality assessed through tests?	Sensitivity analysis?	Outcome variables	Sign & Significance	Comments	Risk of bias
					Table 4.7 (2step) Household total and food expenditure **Table 4.8** (2step fe) Productive assets Fixed assets **Table 4.11** hhpercapita monthly expenditure **Table 4.12** pov gap ratio **Table 4.13** pove gap ratio			

Abou-Ali et al (2010)	N
	Y in working paper version
		Unclear	N	Y – but not for the microcredit estimates of their study	(MF results only) PSM kernel matching
					Farm income per capita
					Non-farm income per capita
					Ln(expenditure per capita)
					Ln(income per capita)
					Ln(food expenditur
						Met/LEU/LER/UEU/UER[a]
						+ive(10%)/-ive(10%)/-ive(10%)/+ive(5%)/-ive(10%)
						+ive(10%)/+ive(10%)/+ive(10%)/+ive(10%)/+ive(10%)
						+ive(10%)/+ive(10%)/-ive(ns)/-ive(10%)/-ive(10%)
						+ive(10%)/-ive(ns)/-ive(10%)/-ive(10%)/-ive(10%)
						+ive(10%)/+ive(10%)/-ive(ns)/-ive(10%)/-ive(10%)
						+ive(10%)/+ive(10%)/+ive(ns)/+ive (10%)/+ive(10%)
						+ive(5%)/+ive(ns)/+ive(10%)/-ive(ns)/+ive(5%)
						+ive(ns)/+ive(10%)/+ive(10%)/-ive(ns)/+ive(10%)
						+ive(10%)/+ive(10%)/+ive(10%)/+i
							Lower & Upper Egypt rural mainly -ive	High

Study	Balancing tests performed?	More controls than treated?	Matching quality assessed through tests?	Sensitivity analysis?	Outcome variables	Sign & Significance	Comments	Risk of bias
					e per capita)	ve(10%)/+ive(10%)		
					Food as share ofexpenditure	-ive(10%)/-ive(ns)/-ive(10%)/-ive(ns)/-ive(10%) -ive(10%)/-ive(10%)/+ive(10%)/+ive(10%)/+ive(10%)		
					Unemployment rate (in area)	-ive(10%)/-ive(ns)/+ive(10%)/+ive(10%)/+ive(10%)		
					Percentage in area working for awage			
					Percentage in area self-employed			
					Illiteracyrate (in area)			
					Povertygap rate (P1)			
					Headcount poverty rate (P0)			
					Percentage in area			

Study	Balancing tests performed?	More controls than treated?	Matching quality assessed through tests?	Sensitivity analysis?	Outcome variables	Sign & Significance	Comments	Risk of bias
Imai, Arun and Annim (2010)[1]	N	N	N	N	See Table in appendix 12	See Table in appendix 12		High
Imai and Azam (2010)	Y	N	N	N	Income per capita panel wave 1 wave 2 wave 3 wave 4	 -ive (ns) TE/NN/Knl[b] +iv(1)/+ive(5)/+ive(5%) -ive(ns)/+ive(ns)/+ive(ns) +ive(ns)/-ive(ns)/+ive(ns) -ive(1)/-ive(ns)/-ive(1%)		Moderate
PnK (Chemin, Duvendack)	Y	N	Y	Y	See Table in appendix 17	See Table in appendix 17		High

Takahashi, Higashikata and Tsukuda (2010)	Y	Y	Y	N	**Table 5**
					Income/profits
					income
					Profits
					self-empl business
					Profits
					non-fm enterprise
					Profits
					aqua/farming
					Sales
					Self-empl business
					non-fm enterprise
					Profits
					aqua/farming
					Assets
					savings
					durables
					livestock
					Expenditures
					Schololing per attend
					Schooling Per child
					Medical
					Female clothing
						Ols/did
						
						+ive(ns)/+ive(ns)
						+ive(ns)/+ive(ns)
						+ive(ns)/+ive(ns)
						-ive(ns)/-ive(ns)
						
						+ive(ns)/+ive(10%)
						+ive(ns)/+ive(10%)
						-ive(ns)/-ive(ns)
						
						-ive(ns)/-ive(ns)
						-ive(ns)/-ive(ns)
						-ive(ns)/-ive(ns)
						
						+ive(ns)/+ive(ns)
						+ive(10%)/+ive(ns)
						+ive(ns)/-ive(ns)
						-ive(ns)/-ive(ns)
						
						Att/slope poor
						+ive(ns)/-ive(ns)
						+ive(ns)/+ive(ns)
						+ive(ns)/+ive(ns)
						+ive(ns)/-ive(ns)
						
						+ive(5%)/-ive(5%)
						+ive(5%)/-ive(5%)
						-ive(ns)/+ive(ns)
						
						-ive(ns)/+ive(ns)
						+ive(ns)/+ive(ns)
						-ive(ns)/-ive(ns)
							PSM
							DID
								High

Study	Balancing tests performed?	More controls than treated?	Matching quality assessed through tests?	Sensitivity analysis?	Outcome variables	Sign & Significance	Comments	Risk of bias
					Table 6 Income/profits income Profits self-empl business Profits non-fm enterprise Profits aqua/farming **Sales** Self-empl business non-fm enterprise Profits aqua/farming **Assets** savings durables livestock **Expenditures** Schololing per attend Schooling Per child Medical Female clothing	+ive(ns)/+ive(5%) +ive(ns)/+ive(5%) -ive(ns)/-ive(ns) -ive(ns)/-ive(ns)		

USAID (Augsburg,	Y	N	Y	N	Main USAID papers:
					Table 3
					Hh yearly income
					Hh income per capita
					Total income last year
					Table 4
					Hh yearly income
						Main USAID papers:
						Nn1/nn2/kernel1/kernel2
						+ive(ns)/+ive(ns)/+ive(sig)/+ive(sig)
						+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)
						+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)
						Nm31/nmsav35/nm260/nm+sav475
						+ive(sig)/+ive(ns)/+ive(sig)/+ive(sig)
						+ive(sig)/+ive(sig)/+ive(ns)/+ive(sig)
						+ive(ns)/-ive(ns)/+ive(ns)/-ive(ns)
						First diff
						+ive(ns)/+ive(ns)
								Main USAID papers:
								High

Study	Balancing tests performed?	More controls than treated?	Matching quality assessed through tests?	Sensitivity analysis?	Outcome variables	Sign & Significance	Comments	Risk of bias
					Hh income per capita Total income last year **No table** Hh yearly income Hh income per capita Total income last year	+ive(ns)/+ive(ns) -ive(ns)/-ive(ns)		

Study	Balancing tests performed?	More controls than treated?	Matching quality assessed through tests?	Sensitivity analysis?	Outcome variables	Sign & Significance	Comments	Risk of bias
Usaid Duvendack				Y		Round 1 - USAID/PSM - 5 nearest neighbour matching/PSM – kernel matching/PSM - kernel matching bandwidth 0.01 ROUND 2		
					Table 5 Total household income per annum in Rupees R1 R2	+ive(1%)/+ive (1%)/+ive (1%) +ive (1%)/+ive (1%)/+ive (1%) +ive (1%)/+ive (1%)/+ive (1%) +ive (1%)/+ive (1%)/+ive (1%)		
					Total household income per annum per capita in Rupees R1 R2	+ive(1%)/+ive (1%)/+ive (1%) +ive /+ive (1%)/+ive (1%) -ive (ns)/+ive(ns) /+ive(ns) -+ive (ns)/+ive(ns) /+ive(ns)		
					Expenditure for housing improvements in Rupees R1 R2	+ive (ns)/-ive(ns) /-ive (ns) +ive(ns) /+ive(ns) /-ive(ns) +ive (ns)/+ive /0+ive -+ive /+ive /+ive		
					School enrolment for girls aged 5 to 10 R1 R2	-ive(ns)/- ive(ns) /- ive(ns) - ive (1%)/- ive(ns) /- ive(ns)		

					School enrolment for boys aged 5 to 10 R1
					R2
						+ive(1%)
						+ive(1%)
						+ive(ns)		

					School enrolment for girls aged 11 to 17 R1
					R2
						+ive(1%)
						+ive(5%)
						-ive(ns)		

					School enrolment	+ive(ns)		

Study	Balancing tests performed?	More controls than treated?	Matching quality assessed through tests?	Sensitivity analysis?	Outcome variables	Sign & Significance	Comments	Risk of bias

					for boys aged 11 to 17 R1
					R2
					Table7 PSM and DID
					Household level hypotheses
					Total household income per annum in Rupees
					Total household income per annum per capita in Rupees
					Inverse Simpson index
					Expenditure for housing improvements in Rupees
					Expenditure on household assets in Rupees
					School enrolment for girls aged 5 to 10
					School enrolment for boys aged 5 to 10
					School enrolment for girls
						+ive(ns)
						-ive(ns)
						+ive(ns)
						+ive(5%)
						+ive(5%)
						+ive(5%)
						+ve(5%)
						+ive(ns)
						+ive(ns)
						+ive(1%)
						+ive(1%)
						R1 5nn/Kern.01/R2 R1/5nn Kern.01
						+ive(1)/+ive (1)/+ive (1)/+ive (1)
						+ive (1)/+ive/(1)/+ive (1)/+ive (1)
						+ive (5)/ +ive (5)/+ive (1)/+ive (1)
						+ive (5)/+ive (1)/+ive (1)/+ive (1)
						+ive(1)/ +ive (1)/+ive(ns)/		

Study	Balancing tests performed?	More controls than treated?	Matching quality assessed through tests?	Sensitivity analysis?	Outcome variables	Sign & Significance	Comments	Risk of bias
					aged 11 to 17			

					School enrolment for boys aged 11 to 17	+ive(1)/ +ive (1)/ +ive (1)/ +ive (1)
						+ive(5)/+ive(10)/+ive (5)/+ive(5)
						+ive (5)/ +ive(5)/ +ive(1)/ +ive(1)
						+ive(5)/ +ive(5)/ +ive(1)/ +ive(1)
						+ive(1)/ +ive(1)/ +ive(ns)/		

					Food expenditure per day per capita in Rupees			
					Enterprise level hypotheses	+ive (1)/ +ive (1)/ +ive(5)/ +ive(5) +ive (5)/+ive(5)/ +ive(ns)/+ive(ns) +ive(5)/ +ive(10)/ +ive(1)/+ive(10) +ive(1)/+ive(1)/ +ive(10)/+ive(10) +ive(1)/ +ive(1)/- +ive(ns)/		
					Informal sector income of whole household - per month in Rupees			
					Informal sector income of respondent only - per month in Rupees			
					Microenterprise revenues of all enterprises in household - per month in Rupees			
					Microenterprise revenues of microenterprises for which respondent is			

Study	Balancing tests performed?	More controls than treated?	Matching quality assessed through tests?	Sensitivity analysis?	Outcome variables	Sign & Significance	Comments	Risk of bias
					primarily responsible - per month in Rupees Current value of fixed assets of all microenterprises in household in Rupees Current value of fixed assets of microenterprises for which respondent is primarily responsible in Rupees Hours worked in previous week in all microenterprises in household Days worked in previous month in all microenterprises in household **Table 8** PSM with sub-groups Total household income			

Study	Balancing tests performed?	More controls than treated?	Matching quality assessed through tests?	Sensitivity analysis?	Outcome variables	Sign & Significance	Comments	Risk of bias
					per annum borr vs control borvs saver saver vs control one time vs control repeat vs control			
					Total household income per annum per capita borr vs control borvs saver saver vs control one time vs control repeat vs control			
					Expenditure for housing improvements borr vs control borvs saver saver vs control one time vs control repeat vs control			

6.14 Appendix 14: Other with/without studies

Study	Method	Control Group	Control Variabl es	Outcome variables	Sign & Significance	Comments	Risk of bias
Barnes 2001, USAID Zimbabwe	Panel	Y	N	Multiple Economic Social Empowerment variable	Mainly +ive & significant	Discussed in depth in section 3.4.2	High
Chen 1999, USAID India	Panel	Y	N	Multiple Economic Social Empowerment variable	Mainly +ive & significant	Discussed in depth in section 3.4.2	High
Chen 2001, USAID India	Panel	Y	N	Multiple Economic Social Empowerment variable	Mainly +ive & significant	Discussed in depth in section 3.4.2	High
Dunn 1999, USAID Peru	Panel	Y	N	Multiple Economic Social Empowerment variable	Mainly +ive & significant	Discussed in depth in section 3.4.2	High
Dunn 2001, USAID Peru	Panel	Y	N	Multiple Economic Social Empowerment variable	Mainly +ive & significant	Discussed in depth in section 3.4.2	High
Tedeschi 2008, USAID Peru	Panel	Y	N	Re-analysis of USAID Peru study			
Shirazi and Khan 2009	DID	Yes	No	Poverty Extremely poor Ultra poor Vulnerable Quasi-non poor Non poor	-ive(ng) +ive(n) -ive(ng) -ive(ng) +ive(ng) -ive(ng +ive(ng)	Significance levels not given	High
Swain and Wallentin 2009	Robust Maximum Likelihood	Yes	Yes	Empowerment	+ive(sig)	Control group does not account for unobervables	High
Tesfay 2009	Panel	Yes	Yes	(Table 5.3) Per captia consumption Housing improvement (Table 5.4) Per captia consumption Housing improvement	 +ive(1%) +ive(1%) +ive(1%) +ive(1%)		Moderate

6.15 Appendix 15: Selected summaries of key studies

The studies summarised in Appendix 15 were selected in an arbitrary way. We could not possibly summarise all studies included in this review due to time and budget constraints, therefore we only provide a summary of those studies we felt appeared to be particularly rigorous, i.e. with the least amount of bias, or have received a lot of attention from researchers and deserved further attention.

6.15.1 RCTs

6.15.1.1 Banerjee, Duflo, Glennerster and Kinnan 2009/2010 (India)

This study claimed, apparently correctly, to be the 'first randomized experiment of the impact of microfinance in a new market' (abstract); it found that 15-18 months after introduction of loans:

> *no effect of access to microcredit on average monthly expenditure per capita, but expenditure on durable goods increased in treated areas and the number of new businesses increased by one third. The effects of microcredit access are heterogeneous: households with an existing business at the time of the programme invest more in durable goods, while their nondurable consumption does not change. Households with high propensity to become new business owners increase their durable goods spending and see a decrease in nondurable consumption, consistent with the need to pay a fixed cost to enter entrepreneurship. Households with low propensity to become business owners increase their nondurable spending. We find no impact on measures of health, education, or women's decision-making.*

The randomisation design was a subset of 104 'slums' of Hyderabad which the MFI, lending 'almost exclusively' to women in self-formed groups, was considering entering. Slums were paired on a minimum distance by a set of variables, one of which was randomly chosen for entry by the MFI; approximately 65 households (not the poorest of the poor) from each slum (treatment and control) were selected[64]. External design validity was subject not only to the location in a single city at a specific time, but also to exclusion of the largest slum areas where the MFI was 'keen to start operations', and to areas with a 'high proportion of migrant' or construction workers.

The randomisation of areas, if they are separated spatially sufficiently, allows accounting for spill-over effects within the location; no evidence on spatial separation was provided.

The study planned a panel starting with a pre-intervention baseline and an endline survey, but the baseline survey turned out to be 'non-random' and too small to detect effects, although some data are reported. The endline survey was independently sampled based on a census of the locations, except for some 500 households which reported borrowing (from any source) in the baseline

[64] 'The remaining 104 were assigned to pairs based on minimum distance according to per capita consumption, fraction of households with debt, and fraction of households who had a business, and one of each pair was assigned to the treatment group' (p5). We presume that the MFI was considering entry in the control locations at a later date making this an incipient pipeline design with randomised allocation of areas to treatment and control. This does not seem to be reported but would be required for ethical reasons to not discriminate against the control locations.

survey who were retained[65]. All areas had few MFI loans but most households reported borrowing from informal or formal sector sources. About a third reported having an existing (very small) business. Health shocks were common as was borrowing consequent on such shocks.

Samples from treatment and control locations seem to have been chosen subsequent to allocation to treatment arm[66], a possible threat to validity (Puffer et al. 2003), and were shown not differ on a number of relevant variables from baseline and endline surveys; samples were also subdivided by propensity to become entrepreneurs. Although other MFIs entered both treatment and control areas, entry of the partner MFI was associated with increased borrowing in treatment areas (27% vs. 18.7%), particularly from MFIs. A further threat to validity was that potential borrowers in the control areas may have postponed investments in the expectation of entry by the MFI.

Households can respond to availability of credit through three channels; by lowering interest rates and relaxations of present and possibly future credit constraints leading to changed investments in fixed and working capital. Secondly, by lending to women it can alter their specific credit constraints, and also alter intra-household dynamics leading to changed patterns of employment, time allocation (including school enrolment), and expenditure, often argued particularly on health and children. Thirdly, it can alter behaviour that may be constrained by lack of savings vehicles.

The analysis was conducted on an intention-to-treat basis without covariates, and with dummies for business status differentiating between those who already have businesses and those with a high propensity to form new businesses. Estimates took account of heteroscedasticity and clustering (by slum). Many outcome variables were tested, with no allowance for increased likelihood of a chance effect due to multiple outcome assessment, but '[W]hile microcredit succeeds in affecting household expenditure and creating and expanding businesses, it appears to have no discernible effect on education, health, or womens' empowerment' (p30) within the 15-18 month time period of the study. This left the study to conclude:

at least in the short-term (within 15-18 months), microcredit does not appear to be a recipe for changing education, health, or womens decision-making. Microcredit therefore may not be the 'miracle' that is sometimes claimed (p31).

Thus, this study did not find very convincing meaningful impacts on well-being, did find impacts on intervening variables, but in a short time period only, and may not have had adequate statistical power to identify impacts on the well-being characteristics that were the primary purpose of development intervention. Had a true panel been available as presumably originally intended, there may have been sufficient statistical power to identify effects on well-being. Thus, the door is left open to others to conclude that '[The] study's

[65] 'These problems were both corrected in the follow up survey, at the cost of not having a panel. The exception to the non-resurveying of baseline households is a small sample of households (about 500 households) who indicated they had loans at the baseline, who were surveyed with the goal of understanding the impact of an increase in credit availability for those households who were already borrowing (though not from MFIs). This analysis is ongoing' (p6).

[66] It is not clear when selection into the control sample was undertaken; the baseline survey was conducted in 2005 prior to entry. It appears that allocation of cluster to treatment and control took place prior to 'Spandana then progressively [beginning to operate] in the 52 treatment areas, between 2006 and 2007' (2009, p5). A sample frame was constructed from a 'comprehensive census' in 2007, and the endline survey was conducted between August 2007 and April 2008 (ibid, p5). Thus, while recruitment of individuals into treatment occurred after randomisation, selection into treatment and control samples occurred subsequent to randomisation. In addition 500 households with high propensity to form businesses were held over from the baseline, although this occurred 'prior to cluster randomisation' (p6).

relatively short time frame .... limits the scope of results and their implications to the short term. Social outcomes may take longer to emerge. In the short-run, at least, nothing big and positive leaps out from the evaluation' (A&M 2010, p299). This seems an interpretation aimed to minimise type 2 errors.

6.15.1.2 Karlan and Zinman (K&Z) 2009 (Philippines)

The other RCT study included in our stage three selection (Karlan and Zinman 2009) randomised selection of applicants for consumer credit who marginally failed the selection criteria of the MFI, in the outskirts of Manila in the Philippines; loan officers were instructed to offer individual liability loans to some randomly chosen people whose creditworthiness score (an aggregation of scoring by various characteristics) came below the usual cut-off. Their uptake of the loan offer and repayment were reported. A similar approach was taken in another study by the same authors in South Africa which we do not include because it pertains to non poor individuals (Karlan and Zinman 2005, 2010). The methodology was explained as follows. After canvassing applicants by 'normal marketing procedures' and interviewing and scoring applicants, the final scoring was produced by the MFIs adapted software to instruct loan officers to offer loans to both those who achieved a high enough score, and a selection of marginal failures. Loan officers did not see the score and so were presumed not to know who was a marginal failure.

> *'Our sample frame is comprised of 1,601 marginally creditworthy applicants ..... (1,583) of whom were first-time applicants to the Lender' (p6). '1,272 marginal applicants were assigned 'approve', and 329 applicants were assigned 'reject'. The software simply instructed loan officers to approve or reject —it did not display the application score or make any mention of the randomisation. Neither loan officers, branch managers, nor applicants were informed about the credit scoring algorithm or its random component (p7).*

Survey data were produced by researchers who sought out the selected 1,601 applicants achieving a 70% response rate some 22 months after application for a loan. The impacts were estimated using the intention-to-treat control function approach with all marginal clients offered loans, treated regardless of whether they took up the offer or not. Two categories of marginal clients (those just below and those further below the normal cut-off) were distinguished[67] and the actual risk score and date of application and of interview were covariates[68].

The authors reported finding that marginal applicants who were offered loans did borrow more, but appeared to shrink their businesses while increasing profits if they were male, increased their access to informal credit 'to absorb shocks' and substituted informal for formal insurance. Males seemed to increase enrolment and decrease family employment outside the family; no evidence of other increases in well-being were identified, but there was some evidence of 'a small decline in self-reported well-being'(p18).

The authors termed these results 'diffuse, heterogeneous and surprising', and some commentators 'surprisingly positive impacts ... and ... a creative way to apply randomization' (A&M 2010, p297).

[67] This was necessary because different probabilities of assignment to be offered a loan were used for these two groups (0.85 and 0.60).

[68] It is not clear how the specification of dates 'control(s) flexibly for the possibility that the lag between application and survey is correlated with both treatment status and outcomes' (p10), rather than just indexing these dates.

It is not clear however that these conclusions are warranted. Crucial issues in RCTs are the randomisation effects which derive from the possibility that knowledge of taking part in an experiment affects behaviour[69]. These are part of the broader categories of 'meaning' effects (Moerman 2002) of which placebo effects are most well known.

These effects are not mentioned in K&Z. It was claimed that this is a double blinded approach[70], but it takes little imagination to perceive that loan officers will readily grasp that quite a number of individuals they are instructed to accept have lower creditworthiness than others, and indeed those they are used to dealing with. This will surely affect the disposition of loan officers towards marginally accepted clients. In a companion paper Karlan and Zinman (2008) showed that 47% of the marginally accepted were in fact rejected by the loan officers (loan officer non-compliance) leading to the intention-to-treat analysis. But an intention-to-treat analysis will not address issues of selective loan officer behaviour[71]. It is not clear what rates of non-compliance are in the Philippines study of Karlan and Zinman, although they write that 'In all, there were 351 applications assigned out of the 1,272 assigned to treatment that did not ultimately result in a loan. Conversely, there were 5 applications assigned to the control (rejected) group that did receive a loan (presumably due to loan officer noncompliance or clerical errors)' (p9). It is not clear why allocating loans to those rejected is 'non-compliance' but at least some loans not received were among those allocated to treatment.

The high attrition rate in the survey with only 70% of those in the original randomisation interviewed is another cause for concern, notwithstanding that the rates of attrition are similar between treated and control subjects.

6.15.2 Pipelines/Control functions

Since Coleman's pioneering[72] study (1999, 2006) pipeline designs have become quite widely used (we have eight papers – seven studies). As mentioned earlier, they allow randomisation and appropriate control groups. However there are several less than ideal variants on the design used by Coleman, and, indeed, it may be that Coleman's design has some problems, particularly in relation to statistical power.

6.15.2.1 Coleman 1999 (Thailand)

Coleman's (1999) highly regarded study on the impact of group-lending in Northeastern Thailand controlled for self-selection and non-random programme placement bias using observable village characteristics, and village-level fixed-effects using data from a quasi-experimental design carried out in 1995 – 1996. The study conducted a survey on 455 households; in addition to selecting participating and non-participating households in villages where MFIs were already active (had disbursed loans), the innovation was to get MFIs to identify households which would participate in villages where they planned to operate, and to survey a sample of these future participating households and a sample of

[69] Two of these effects are known as Hawthorne (those being treated act differently because they know they are being treated) and John Henry effects (those in the control group behave differently because they are not being treated).

[70] 'Only the Lender's Executive committee was informed about the details of the algorithm and its random component, so the randomisation was 'double-blind' in the sense that neither loan officers (nor their direct supervisors) nor applicants knew about assignment to treatment versus control' (p8).

[71] This is, of course, similar to the effect that doctor knowledge has on patient outcome (Moerman, Chapter 4), and why (genuine) double blinding, which is possible in the case of indistinguishable pills but not when loans are offered to identifiably different clients.

[72] Steele et al. (2001), is also a pipeline study for which the fieldwork and indeed initial reporting (Steele et al. 1998) actually precedes Coleman's.

non-selected control households from these future villages into which MFIs would expand. The control group was surveyed one year before receiving its first loan.

Coleman's design consisted of 14 village clusters (8 and 6 per treatment) and large numbers of cases per cluster. The intra- and inter-cluster correlations of important variables were not given, but it is clear that there were too few clusters per treatment and too many observations per cluster to provide much statistical power.

A DID estimation with village fixed effects and household covariates were used to estimate difference of incomes between participants and non-participants in programme villages with difference of incomes between participants and non-participants in control villages. Coleman's (1999) study concluded that the microfinance programme in Northeastern Thailand had little impact although other studies which were ignorant of selection bias provided evidence to the contrary. More importantly, Coleman's (1999) study found that microfinance had positive impacts on increased money-lending activities and leads to an increase in debts.[73] However, as the author pointed out, the results should be read critically because Thailand is already fairly rich and less credit constrained compared to other developing nations.

6.15.2.2 Copestake et al. 2001 (Zambia)

Copestake et al. (2001) reported impact of microcredit accessed by individuals in a group liability context in Zambia using a cross-section sample of two groups of borrowers (1-2 years since first loan and 12-18 months since first loan) and a pipeline group. Estimates were made of impacts of growth rates and profits (recall) from borrowing status, and of growth of profits, business diversification, and household income growth by size of loan in a control function framework. A variety of impacts were reported, with some barely statistically significant (Table 1, p88). Larger second loans had largest and most statistically significant impacts. But this finding was vitiated by the high exit rate of clients between first and second loans. This means that the full sample of pipeline clients were compared with a subset of borrowers who survived to second loans and were likely to be quite unlike the intake into either the first loan or pipeline. Part of the pipeline sample is from a different geographical area. Footnote 10 notes possibility of bias due to unobservable notwithstanding inclusion of proxies such as 'receipt of training and subscription to government health services'; see also footnote 21 which argues 'that the problems of programme endogeneity and selection bias were [not] fully controlled but that they were sufficiently dealt with fort yield plausible results and hence reduce expectations about likely impacts among key stakeholders' – because estimated impacts were slight and data on comparability of areas were not provided. It is likely that loanees were not poorer individuals as an initial payment of 10% of agreed loan had to be paid into the Loan Insurance Fund.

6.15.2.3 Copestake 2002 (Zambia)

Copestake (2002), focused on inequality in impacts of microfinance in Zambia comparing ''one-year-old' clients with a comparison group of 'pipeline' clients' interviewed once using recalled profits etc. over the previous year to assess impact using a control function. Significant polarising impacts of borrowing on

[73] Coleman (1999) discovered that many borrowers joined the microfinance programme mainly for social reasons (e.g. peer pressure). They had no projects to invest in and solely borrowed for consumption purposes. Hence, they frequently did not have the funds to repay the microfinance loan at the end of the loan cycle. As a result, they borrowed from moneylenders to repay the microfinance loan. Then, in order to repay the moneylender they had to apply for another microfinance loan. This circle continued until they ended up in a downward spiral of bad debt.

nominal profits and transfers to household budgets were found. The first year loan members and pipeline were drawn from a single source and seemed to refer to the same population, probably the pipeline were drawn at later dates. The samples of both groups involved high replacement rates borrowers and low (46 and 24%) response rates, and it is questionable whether those joining later in the same area were equivalent to those joining earlier; these differences may involve unobservables as well as observables. This survey involved continuing borrowers so neglected exits (drop-outs and graduates), although the paper provided evidence from another survey on exit which is high (27.8%), higher among second loan takers and later entrants adding evidence that the pipeline groups may differ from earlier borrowers.

6.15.2.4 Copestake et al. 2005 (Peru)

Copestake et al. (2005) used panel data from a sample of microfinance clients of two MFIs in Peru to estimate impacts on changes in sales, profits, and family and monthly incomes in a basic DID model and in a multivariate (control function) model. They found significant impacts; more for richer than poorer individuals. The panel started with microfinance clients already having received loans, up to a year or more prior to the survey, and so could not be shown to be equivalent to the control group of non-members, the sampling method is not explained. As explained above, if non-members had the opportunity to become members but 'choose' not to, there must be some difference between them and members, and a control function can only account for observables. It is not clear that either group is poor, although another sample of clients from similar MFIs are found to be 'generally worse off- than other people living in the same locality'. Furthermore, several variables in the control function could be endogenous, and IV or panel methods which might reduce bias due to unobservables were not used, presumably because of lack of suitable instruments and time varying covariates.

6.15.2.5 Montgomery 2005 (Pakistan)

Montgomery (2005) reported estimations of impact using a pipeline design of Khushhali Bank clients in Pakistan, some of whom had already received loans and others who had been selected but not yet received loans; new clients were drawn from different villages from current borrowers. Randomly selected non-clients from the same populations comprise the control groups for both borrowers (treatment) and new (pipeline) clients and a DID control function model was estimated with interactions of treatment and poverty status to estimate whether there were different impacts on the poorest individuals. Although the estimating equation (equation 1[74]) specifies village and household controls, no results are reported for these variables, and their presence was not reported in footnotes to tables of results, although summary statistics for a number of household characteristics were reported in Appendix 9 A2; thus we assume that the estimation is a control function using household and village characteristics although unreported. Montgomery reported that coefficients on the dummy for Khushhali 'interest, the regression results indicate the program does not impact most consumption expenditure measures – almost all coefficient estimates .... are insignificantly different from zero. There is some evidence that participation in the program has a positive impact on educational

[74] $Y{\downarrow}ij = \beta{\downarrow}1\ X{\downarrow}ij + \beta{\downarrow}2\ V{\downarrow}j + \beta{\downarrow}3\ M{\downarrow}ij + \beta{\downarrow}4\ P{\downarrow}ij + \beta{\downarrow}5\ T{\downarrow}ij + \beta{\downarrow}6\ P{\downarrow}ij\ T{\downarrow}ij + ({\downarrow}ij$, where X_{ij} are characteristics of household *i* in village *j*, V_j are village characteristics, M_{ij} is membership status (existing or new member of Khushhali Bank), P_{ij} is poverty status (poorer = 1), T_{ij} is treatment status (existing borrower = 1). $□_5$ and $□_6$ are the coefficients of interest referring to the impact of borrowing and the additional impact of borrowing for a poorer person.

expenditures for the very poor, as indicated by the statistically significant positive' (p11). In terms of social development indicators (education and health), participation had mixed impacts with some positive and some negative effects, although in some cases the core poor individuals appeared to benefit more (p12). There were also significant positive impacts on households with microenterprises in urban areas and very poor borrowers involved in agriculture (p13).

This pipeline study assumed that later (pipeline) and earlier (treatment) borrowers were perfect substitutes; however, as noted, if earlier borrowers had different characteristics then these estimates were biased. This issue was addressed in another paper included in this review (Setboonsarng and Parpiev 2008), who used PSM to address this selection bias; Setboonsarng and Parpiev (2008) is discussed below (and to which the reader might now refer, although to preserve the logic of our structure we next address the panel studies that use a pipeline approach).

6.15.3 Pipeline and panel

Panel methods are widely claimed to address problem of selection bias and the robustness can be tested by estimating fixed and random effects models, and testing the differences between coefficients estimated by these two methods. With this in mind we discuss the two included panel pipeline studies.

6.15.3.1 Steele et al. 2001 (Bangladesh)

This study, which seems to have been largely overlooked in the literature[75], also innovated the pipeline design attributed to Coleman[76]. It used 'three areas: an area where SC (Save the Children, USA) had operated non-credit programs since the mid-l970s (the 'old' area); an area where SC was soon to begin new programme interventions (the 'new' area); and villages in the same area that were similar to those in the new area but where no SC intervention was planned (the 'control' area)' (p269), for which a panel data set for 1993[77] and 1995[78] was available. It aimed to assess the impact of participation in women's savings and credit organisation on contraceptive use. It founds that 'the analysis of program impact on the use of modem contraceptives reveals a positive effect of the credit program, after we adjust for this selectivity: we see no evidence of an effect of participation in a savings group' (p267).

The authors emphasised that the sample clearly suffers from selectivity and placement biases, which were addressed through the use of fixed and random effects estimations applied to the panel data, and applying Hausman type tests to the difference in coefficients between the two models. It found

> *no significant differences at the .05 level between the fixed-effects and random-effects estimates. The test statistic comparing the two*

[75] i.e. it does not appear in Armendáriz and Morduch (2010). Goldberg (2005) refers to the 1998 working paper version, but does not highlight the methodological innovation in the study (pipeline and IV in panel analysis).

[76] Coleman's study conducted later (1995-6) than the field work used in the Steele study, but the earliest publication of the latter is 1998, which is a year earlier than the year in which the Coleman study was published. Steele et al. 2001 is methodologically more sophisticated using a panel analysis, while Coleman is restricted to cross-sectional analysis.

[77] Women surveyed in 1993 may be divided into four categories tor our analysis: (1) members of savings groups in the old area; (2) poor women in the new area where SC had not yet introduced a program, but who would be eligible for membership when the savings and credit groups were formed; (3) women in the same area who did not fulfil SC's eligibility criteria for group membership; and (4) the control group (p269).

[78] By 1995, category 2 had been divided into five subcategories: (a) those who had chosen to join one of the newly formed SC savings groups in a village where ASA did not work; (b) non-members in non-ASA villages; (c) SC members in villages where ASA worked; (d) SC-ASA members in ASA villages; and (e) those in ASA villages who did not participate in either programme.

> *sets of estimates from an analysis of ASA villages is calculated as 8.24 on 4 degrees of freedom (P= .08) (p 278).*

It referred to plots of the residual errors and asserted that:

> *Although we find some suggestion of a difference in the distribution of u, between the old area and the other categories, there is little evidence of a difference between non-members in ASA villages (group 6) and SC-ASA members in ASA villages (group 8), the contrast found to be significant in the random-effects model. This provides further evidence of little correlation between program membership status and the individual-level, time invariant unobservables. Therefore we conclude that the estimates of program effects obtained from the random effects model are consistent (p278).*

Unfortunately neither interpretation is very convincing. It is not clear why a P-value of 0.08 is indicative of insignificance rather than of marginal significance, and the tables of results presented refer to P-values <0.10. Further, the plots for groups 6 and 8 (non-members in ASA villages, and SC-ASA members in ASA villages) are visibly different in location and dispersion, even if the difference with ‘old’ areas appears greater.

This apparently rigorous and methodologically sophisticated paper warrants further investigation; it has design limitations in that treatment areas have significant differences in characteristics, which may have been largely controlled by sub-group analysis and/or village/area fixed effects, there is also differential attrition, as described in an appendix where it is asserted that ‘because our analysis is based only on women who respond in both surveys, we must control for area of residence to adjust for the effect of attrition bias on the parameter estimates (Little and Rubin 1987:15)’ (p281). However, it is not evidence that attrition control is conducted since the analysis is restricted to the sample for which there are responses in both waves[79].

While this paper found evidence that

> *After controlling tor non-random program effects, allowing for exclusion criteria in identifying comparison groups, controlling for prior characteristics such as women s propensity to use contraception before they joined, and controlling for the effect of child health interventions, our analysis shows a substantial impact of membership in SC-ASA credit programs on contraceptive use. This increase in contraception is over and above the substantial rise that can be attributed to the introduction of new health measures for children, directed primarily to mothers, which also included some motivational messages to accept family planning. Those who joined SC-ASA credit programs showed a proportionately larger gain in contraceptive use; this finding suggests that something about membership in credit groups spurred further change in an environment that already was changing with regard to family planning (p280).*

The authors hypothesised that ‘uptake of modem contraception is explained by social networks associated with membership and by exposure to new ideas fostered by group dynamics. We lack the data necessary to demonstrate these mechanisms, however’. They also discussed the difference between these findings and those of Pitt et al. (1999), which used data from a different area of

[79] One standard method to address attrition is to use the data from the first wave model the attrition in a 2-stage process (Heckman, 1979); there is no evidence that this is done in this paper.

Bangladesh just before the period studied in Steele et al. (2001), and also estimates the effects of MFI membership on contraceptive use, finding 'clear negative effects for female participants relative to nonparticipants'. This difference is attributed in part to the different specification of participation used by Pitt (amount of loans taken), which was narrower, and may be less appropriate given the mechanism of transmission suggested. It is not clear how valid an objection this is since in the programmes addressed in Pitt et al. (1999), all members were also borrowers, although the amount they borrowed did vary. This difference in specification, the queries about the Steele et al. study raised here (failure to address attrition, doubts about the interpretation of the fixed vs random effects model), and the doubts raised about the method used in the PnK oeuvre suggest a replication of both these studies. Until this is done no clear conclusion can emerge from the contrasting findings of these two studies.

6.15.3.2 Cotler and Woodruff 2008 (Mexico)

This pipeline study of small-scale retail outlets of the largest snack food company in Mexico collected data on clients of an MFI, which rolled out its programme 'neighbourhood by neighbourhood'. In one neighbourhood where the MFI started its programme in summer 2004, and another neighbourhood where the MFI planned to roll out its programme the following year it agreed to screen and select retail outlets which applied to be prospective clients of its programme when the MFI proposed to roll out its programme the following using 'identical methods' of screening (a point made on pages 830 and 831). Of course, the exception was that in the second neighbourhood, prospective clients were aware that they could not receive loans until the MFI programme was rolled out there, i.e. for a year or more[80] (p818; see further discussion below).

Retail enterprises (n=216) were selected in the former and 188 in the later neighbourhoods, which would serve as 'a comparison group'. A three wave survey was conducted with waves separated by 4-6 months. Attrition between the second and third waves was so large that only the first two waves were analysed (allowing a period of only some 4-6 months for impacts to accrue). Because of the way in which samples were selected the authors claimed that 'the two groups should be comparable in terms of unmeasured characteristics related to demand for credit, entrepreneurial ability, and so on' (p831). The main difference between the samples collected from locales which were 37 km apart would relate to 'local shocks [that] might have affected the treatment and control groups differently'. But these could be controlled using monthly data on sales from the two localities which, though highly seasonal showed 'very similar' trends[81].

The delay in access to funds might well alter the profile of applicants in the control area compared to those who accessed loans immediately. Successful applicants might also alter their activities in the interim between being selected and, later survey waves before, receiving the loan; they might do this because the MFI's interest rates were lower than the marginal loans immediately available, with a resulting bias in favour of estimating a positive impact (p838), although the delay might select out more profitable potential loanees.

[80] This is discussed on page 838, where it is stated that 'In sum, since members of both the treatment group and the control group were selected through similar screens, we expect that they possess similar entrepreneurial spirit and face similar economic restrictions and opportunities.'

[81] But there were differences in growth of sales in the two neighbourhoods, in one period the control neighbourhood experienced more than twice the growth rate (-1./2 vs -3.7%) around the time of the baseline and first follow up; and a difference of nearly 3.8% (9.3 vs 13.1%) between the first and second follow up surveys (p839). Different figures are given on page 841, where is appears that total sales of all firms of the population from which the sample was drawn, grew at 5.6% between July 2003 and October 2004 in the treatment neighbourhood but only 1.7% in the control.

The data are seen as of better quality than usual, because interviews were conducted by loan officers, who were given training, included much information pertinent to selection of successful loanees and whose remuneration depended on loan performance. Although re-interviews would normally only be to loanees who applied for further loans the loan officers agreed to re-interview all who had applied for loans in the first round (when presumably the incentive effects claimed above would not have applied, a contention that may gain some support from the fact that those only receiving a first loan were not re-interviewed at the third round.

Treatment and control groups had some statistically significant differences in business characteristics (mean and median fixed assets at P<=0.01%, mean inventories as P<=0.1%)

> *We use differences in the phasing in of a new lending program designed to serve clients of the largest snack food company in Mexico to identify the impact of credit on outcomes of small retail enterprise in Mexico City. We find that the loans have positive impacts on the smallest firms but negative impacts on larger firms. These results are consistent with hypotheses that smaller firms have higher returns to capital and face greater credit constraints. Given that the program involved loans given for 4-month terms, we find surprisingly large effects on investment in fixed assets.*

This study uses a pipeline design and an unbalanced panel data analysis; while here are three rounds of data the attrition between rounds 2 and 3 was substantial so only the first two rounds are used.

6.15.4 Pipeline and PSM

Three papers use PSM to conduct their impact assessments of a pipeline design (Kondo et al. 2008, Setboonsarng and Parpiev 2008, Deininger and Liu 2009). As noted above, PSM cannot account for the effects of unobservable characteristics which affect both selection into treatment and the outcomes of treatments. Nevertheless, it may be a useful technique to construct the counterfactual provided there are a large number of potential matches, with a wide range of good quality covariates which are causally related to selection on observables and good proxies for unobservables. It is also important to report the common support and numbers of potential matches, balancing of covariates and to conduct sensitivity analysis (as discussed earlier).

6.15.4.1 Kondo et al. 2008 (Philippines)

Conducted in the Philippines with a selected treatment group and a pipeline drawn from an ‘expansion area’:

> *The comparison barangays, on the other hand, are expansion areas where programme clients have been identified and organised into groups but no loans have yet been released to them (p51).*

Statements that control villages were matched, with no quality evidence of the matching process, provide little assurance that this was in fact achieved.

The innovation of this study was to include drop-outs and graduates in the pipeline design (p51), thus mitigating biases due to sub-selection of borrowers that occurs in Coleman’s design. The estimating equation is standard control function with village (V) and household (X) characteristics, and membership (M=1 if member in treatment or pipeline villages, 0) and treatment (T=1 if ever borrowed, 0) dummies.

$$Y_{ij} = \beta_1 X_{ij} + \beta_2 V_j + \beta_3 M_{ij} + \beta_4 T_{ij} + \varepsilon_{ij}$$

Characteristics of members and non-members in existing and expansion (pipeline) areas were compared and some differences found in the treatment and pipeline areas. Outcome variables were per capita income and expenditure, savings (2 definitions) and food expenditures, poverty status and subsistence poverty, with some differences between members and non-members in treatment areas but none in pipeline areas. Impacts estimated from the control function showed positive impacts significant only at 10% level for per capita incomes, total expenditures and food expenditure. Quite a number of participants and non-participants took up non-MFI loans, but, while take up of loans (impact on financial transactions of households) was investigated, the impact of loans on outcome variables was seemingly not addressed. Impacts on numbers of enterprises and employment, assets, education and health were also explored. Impact heterogeneity by per capita income quartile and educational status of the reference person was also explored.

The results showed that the majority of respondents were non-poor by the official definition (p67), that impacts were only marginally statistically significant, and somewhat regressive with some negative (and insignificant or marginally significant) impacts on lower income quartiles and significant positive impacts among higher income quartiles (p68). This heterogeneity explained the marginally significant impact for the whole sample, and corresponded with Coleman's findings of impacts only among the better-off (among clients who were themselves not among the 'official poor').

The study found reduced reliance on ('presumably') higher priced loans, and some consumption smoothing, positive impacts on employment and number of enterprises, but no significant impacts on assets or human capital, although the length of time in which impact could occur on these variables might be considered short.

6.15.4.2 Setboonsarng and Parpiev (S&P) 2008 (Pakistan)

This paper used the same dataset used in the paper by Montgomery (2005) discussed above, which it criticised describing that: '[T]he first study, conducted by Montgomery in 2005, assumed no self-selection bias occurred, whereas this study adopted econometric methods [PSM] to address that issue' (p1). Strangely they did not mention the pipeline nature of the data and Montgomery's approach was not discussed although it was noted that 'Montgomery (2005) drew causal linkages in part based on the assumption that the survey design would minimize the selection bias' (p10). S&P do not seem to make use of the 'future client' indicator.[82] Using the classification of S&P, existing borrowers were significantly different (wealthier) to non-borrowers including future clients; they claimed that borrowers 'appear to be initially wealthier than the control group', although this comparison was made after borrowers may have already benefitted from loans.

Using PSM the 'researcher can match participants from the treatment group with participants from the control group, so that the treatment group and control group can be balanced. This approach can significantly reduce bias in observational study' (p10-11). The probit estimation of propensity score has a reasonably high pseudo R-squared (0.38, N=2881), with coefficients indicating

[82] Montgomery (2005) reported 1,454 Khushhali clients and future clients, and 1,427 non-clients (p9); S&P reported 1,204 KB borrowers and 1,677 non-borrowers (Table 8, p9). This implied that there were 250 future clients.

that wealthier females from landowning higher status households who already borrowed from other sources and were living in somewhat more remote locations were more likely to be Khushhali Bank borrowers (Table 10, p13). Restricting the sample to the range of common support drops the sample from 2,881 to 2,856 in 11 blocks but 'the distribution of Khushhali Bank borrowers and non-borrowers along the propensity score is not similar' (p14). In fact the distribution of propensity scores of Khushhali Bank borrowers was distinctly bimodal while that of non-borrowers was unimodal, located around the lowest P scores. This means that the majority of control households (non-borrowers) had low P scores while of course the majority of borrowers had high P scores; there were some borrowers with low P scores (30%) for whom there were many potential matches (1,542 non-borrowers to match with 363 borrowers with low P scores), while for the bulk of Khushhali Bank borrowers there were very few potential matches (83 non-borrowers to match with 861 borrowers with high P scores). This was clearly a ridiculous basis on which to proceed to undertake an impact analysis, if only because as we have noted it is recommended that there are more potential matches than treatment cases in the relevant ranges.

The paper discussed the use of several methods of matching, but used nearest neighbour and presented results of kernel and stratification methods (not described in detail) by way of robustness checks; this only checked robustness to matching methods rather than to unobservables which was the main issue. It did not use sensitivity analysis. It found a number of largely positive impacts[83] with 't' values indicating statistical significance of difference between matched treatment and control cases.

When discussing the impact on the poor[84] S&P used a poverty expenditure threshold of Rps 878.6 (Montgomery uses Rps 1,000 per capita per month), and match '749 poor households who borrowed from Khushhali Bank .... with 439 non-poor [sic], non-KB borrowers' (p18) and find impacts 'essentially similar to those for clients in general' (p18).

As noted, robustness of results is assessed by comparing the results of nearest neighbour with kernel and stratification matching However, while this comparison showed broadly similar results, it revealed nothing about robustness with regard to unobservables. This, as we have emphasised earlier, can be addressed by sensitivity analysis, which was not performed. Consequently we rate the findings reported in this paper as highly vulnerable to bias.

As an example of the need for sensitivity analysis, properly interpreted we refer to the one example we know of where sensitivity analysis was reported in a paper that addressed the impact of microfinance (although sensitivity analysis was not reported for microfinance impact). Abou-Ali et al. (2010) (see below), reported sensitivity analysis for their results to describe the impact of SFD intervention in roads on transportation spending to compare SFD intervention with no SFD intervention. They found that the gamma at which the estimated difference in spending became insignificant was 1.17, and interpreted this to mean that 'In this example, the results are thus relatively robust to hidden bias' (p542). Unfortunately this is the wrong interpretation, as such a value close to 1

[83] Some positive impacts such as on use of pesticides were qualified as having potentially negative environmental and or health impacts not accounted for in the analysis.

[84] That was apart from the earlier claim that 'KB borrowers have PRs 6,494.2 higher profit on livestock than that of non-borrowers. This shows the strong positive effect of KB borrowing on a farmer's poverty situation' (p15).

indicates that the results are highly vulnerable to unobservables (Rosenbaum 2002, Rosenbaum 2005, p1812[85]).

Given the parlous effective common support, the failure to report balancing of covariates, and the lack of sensitivity analysis, one can only conclude that this study provides no satisfactory evidence of impact of microfinance on poor people.

6.15.4.3 Deininger and Liu 2009 (India)

This paper employed a pipeline design based on two phases of a World Bank Funded development project in Andhra Pradesh (AP), India. The project combined predominantly female social and economic empowerment oriented self-help groups (SHGs) with access to microcredit, targeting especially the 'leftover poor'[86]. The evaluation used the first three years of Indhira Kranthi Patham (IKP), known as District Poverty Initiatives Program (DPIP) which started in 2000 and aimed to strengthen SHGs in 6 poorest districts of the State. A second phase – Rural Poverty Reduction Project (RPPP) – expanded the project to the states remaining 16 districts from July 2003, which supplied the pipeline sample. The survey was conducted in 2004 of some 6,000 households. Villages were randomly selected villages and then households with 'weights applied based on their poverty status according to the census of PIP [Participatory Identification of the Poor). The survey included a community questionnaire and an SHG questionnaire, administered to randomly selected SHGs in DPIP areas. As expected, pipeline areas were clearly and markedly different from DPIP areas in a wide range of variables including 'backwardness', infrastructure, levels of female economic activities (outside the household), caste panchayats, untouchability and other variables. Participants could be classified into 'converted' (from previous SHGs), new, and non-participants. Propensity Score (PS) matching was used to control for differences in observables; both cross sectional (programme effects) and DID estimates based on recall were computed. The idea was that since DPIP and RPRP were two phases of the same project conducted in distinct districts of the Indian State of AP, identification could accomplished 'by combining pipeline, propensity score (PS) matching and difference-in-difference estimation methods' (p9), since self-selection would be the same in both DPIP and RPPP areas (p10). However, as discussed above, it is simply not plausible that selection could have been similar because of differences in time and circumstances; indeed the paper itself noted that the survey took place immediately following a famine which would surely affect recruitment, as would the different contexts of DPIP (in the six poorest districts) and RPPP (the remaining districts). Plots of propensity scores for the two areas (p 30 and 31) confirm these doubts, with almost mirror image distributions between DPIP and RPPP areas for each of the three categories (New and converted participants and non-participants). Nevertheless, the authors argued 'that under fairly general assumptions, villages or households in RPRP areas can serve as a control for those in DPIP areas' (p9). ATET effects were estimated without covariates. Sensitivity analysis was not conducted, and there was only limited information about the quality of matches (e.g. the number of control households used after matching). While there was no evidence of impact on incomes or assets, there were significant impacts on consumption and access to food nutrients, which may have been due to the immediate transfers involved in loans to SHG members, or to consumption smoothing. The authors suggested that

[85] 'The study of the effects of diethylstilbestrol becomes sensitive at about gamma = 7, while the study of the effects of coffee becomes sensitive at gamma = 1.3. A small bias could explain away the effects of coffee, but only an enormous bias could explain away the effects of diethylstilbestrol.'

[86] Those who tend to get excluded from 'normal' SHGs.

income effects might materialise in the longer term; however, it is unlikely that this can be demonstrated since the initial differences in timing to access project inputs is small. The likelihood is small that differences, due to this variation in timing of access, would remain some years after the project began in the pipeline area.

6.15.5 With/without and PSM

Most microfinance evaluations included in this SR adopt a with/without approach and we merely provide summaries of a few selected key papers that are largely representative of the with/without studies in this report. We begin summarising the with/without studies employing PSM.

6.15.5.1 Abou-Ali et al. 2010 (Egypt)

The study by Abou-Ali et al. (2010) was an impact assessment of the Egyptian Social Fund for Development (SFD) that actively promoted community development, public works projects and microcredit. The paper evaluated various SFD activities by measuring its impact on a wide range of outcome indicators using PSM. We focus on the microcredit part of this paper which assessed the impact of microcredit on income, expenditure, employment, literacy rates, and poverty levels. The authors' headline findings for microcredit were that it generally increased income per capita but their results varied substantially by region. In the metropolitan areas as well as in urban Upper Egypt microcredit increased household expenditure and reduced poverty, but these results could not be confirmed in other regions investigated.

Abou-Ali et al.'s paper is one of the few PSM microfinance IEs that applied sensitivity analysis to estimate the vulnerability of their matching estimates to selection on unobservables or 'hidden bias' using Rosenbaum's (2002) term. However, sensitivity analysis was not applied to their microfinance estimates and interpretation of sensitivity analysis, i.e. their gamma values, raises doubts. For example, the authors argued that '[U]sing a significance cut-off of 10 per cent, we see that [gamma] could be as high as 1.17 before the results lose their statistical significance. ... In this example, the results are thus relatively robust to hidden bias' (p543). The opposite is in fact true, the gamma value at which the estimated impact becomes insignificant (i.e. gamma = 1.17) indicates that their results are highly vulnerable to selection on unobservables or 'hidden bias'. There are further concerns about this paper such as lack of evidence on common support or balancing (though these issues were raised and reported in the working paper version of their paper), as well as a lack of reporting the number of matched comparison groups for most outcome indicators, which is crucial for assessing the quality of the matches.

6.15.5.2 Imai et al. 2010 (India)

Imai et al. (2010) assessed the impact of microfinance on poverty reduction in India. Cross-sectional data on 5,260 client (from 20 different MFIs) and non-client households was collected across India. They used a treatment effects model (i.e. a variant of Heckman's sample selection model – see Heckman 1979) and PSM to account for selection bias. Impact was assessed on an index-based poverty ranking indicator that contains information on landholdings, income, assets, housing and sanitation. The authors found that microfinance had significantly positive impacts on poverty reduction. The focus of this study was on the treatment effects model which claims to account for the unobservables, and which is essentially a more sophisticated and robust method than IV. Puhani (2000), however, criticised the selection models and argued that they were driven by narrow assumptions about functional form and error distributions. In other words, impact estimates obtained from selection models are only as good as these assumptions on distributional and functional form (Vytlacil 2002). This last point is similar to IV estimates whose reliability heavily depends on the quality of the underlying instruments (Caliendo and Hujer 2005). Hence, the authors applied PSM to check the robustness of results obtained from the treatment effects model. The treatment effect model and PSM results, which are not reported, are claimed to be conclusive and confirm that microfinance has significantly positive effects on poverty reduction. Details such as balancing, common support or quality of the matches, of the PSM estimation were not provided, and there was no evidence that sensitivity analysis was done on the PSM estimation. Therefore it remains unclear whether PSM estimates are vulnerable to unobservables; the PSM adds little independent support to the 2-stage estimates.

This paper is confusingly written and hard to interpret; for example, Table 2 erroneously reports that the 'Case B dependent variable: Index Based Ranking (the first stage probit estimates whether a household has taken a loan for productive purposes ('MFI productive'))'. This should not include the words 'Index Based Ranking'. There are many other parts of this text which are unclear. An article with quite so much econometric elaboration should perhaps be reviewed and published in a specialist econometrics journal rather than a multi-disciplinary one such as World Development.

6.15.6 With/without and 2-Stage

6.15.6.1 Cuong 2008 (Vietnam)

The study by Cuong (2008) assessed how well poor people are targeted by the microfinance programme of Vietnam Bank for Social Policies and measures its impact on expenditure and income per capita. Panel data was used and two methods applied: first, IV on the cross-section and second IV in combination with a fixed effects panel model. Cuong found that the microfinance programme under investigation had a positive impact on expenditure and income per capita and thus concluded that microfinance reduces poverty. However, the programme mainly targets non-poor people who also receive larger amounts of credit than poor people. Cuong's findings are not surprising, as Coleman (1999) argued, more wealthy poor people and individuals who are more entrepreneurial are more likely to participate in microfinance; the observed impacts might not be due to microfinance but due to other unobserved characteristics. It can be hypothesised that other forms of finance might have been equally effective as microfinance in the context of non-poor individuals. The study by Cuong appears to be technically sound and rigorous but doubts remain about the relevance of its results since essentially non-poor people were targeted (67.1% programme participants were non-poor people, p171).

6.15.6.2 Shimamura and Lastarria-Cornhiel 2010 (Malawi)

The study by Shimamura and Lastarria-Cornhiel (2010) was motivated by the discussion on trade-offs between child labour and schooling. They investigated a microfinance programme in Malawi and its impact 'on children's school attendance and the likelihood of being involved in other productive activities' (p569). The authors conducted a paired-site sampling survey to account for sample site variations and apply an IV approach (p569). They found that microfinance participation decreased school attendance by girls in particular, and that the programme did not reach the poorest people. Doubts are raised about the validity of the instrument and hence their findings; no identification tests were run to assess the validity of the instrument and no specification tests, i.e. a Hausman test, was conducted to gauge whether OLS estimates would have been as useful as the IV estimates presented in this paper.

6.16 Appendix 16: Impacts assessed by study

			Economic											Social							Political
STUDY	TITLE	YEAR	Business profits	Business revenues	Sales	Income per capita	Consumption/Expenditure[87]	Assets	Employment	Savings	Debts	Poverty index/status	Other	Children's school enrolment	School attendance	Contraceptive use[88]	Nutritional status[89]	Vulnerability to shocks	Social capital	Other	Empowerment
Abera H	Can microfinance help to reduce poverty? With reference to Tigrai, northern Ethiopia?	2010					X	X				X									
Abou-Ali H, El-Azony H, El-Laithy H, Haughton J, Khandker SR	Evaluating the impact of Egyptian social fund for development programs	2009				X	X		X			X								X	
Banerjee A, Duflo E, Glennerster R, Kinnan C	The miracle of microfinance? Evidence from a randomised evaluation	2009	X	X			X	X	X				X							X	X
Bhuiya A, Chowdhury M	Beneficial effects of a woman-focused development programme on child survival: evidence from rural Bangladesh	2002																		X[90]	
Coleman BE, Thailand	2 papers				X	X	X	X	X	X	X		X							X	
Copestake J, Bhalotra S, Johnson S	Assessing the impact of microcredit: a Zambian case study	2001	X			X							X						X		
Copestake J.	Inequality and the polarizing impact of microcredit: evidence from Zambia's copperbelt	2002	X			X							X							X	

[87] Consumption/xpenditure per capita (food and non-food)
[88] Contraceptive use or maternal health
[89] Nutritional status/calorie intake/food security
[90] Infant mortality

			Economic											Social							Political
STUDY	**TITLE**	**YEAR**	**Business profits**	**Business revenues**	**Sales**	**Income per capita**	**Consumption/ Expenditure**[87]	**Assets**	**Employment**	**Savings**	**Debts**	**Poverty index/status**	**Other**	**Children's school enrolment**	**School attendance**	**Contraceptive use**[88]	**Nutritional status**[89]	**Vulnerability to shocks**	**Social capital**	**Other**	**Empowerment**
Copestake J, Dawson P, Fanning JP, A. McKayA, Wright-Revolledo K	Monitoring the diversity of the poverty outreach and impact of microfinance: a comparison of methods using data from Peru	2005	X		X	X						X									
Cotler P, Woodruff C	The impact of short-term Credit on microenterprises: evidence from the Fincomun-Bimbo programme in Mexico	2008	X	X				X					X								
Cuong NV	Is a governmental microcredit programme for the poor really pro-poor? Evidence from Vietnam	2008				X	X					X									
Deininger K and Liu Y	Economic and social impacts of self-help groups in India	2009				X	X	X									X		X		X
Diagne A and Zeller M	Access to credit and its impact on welfare in Malawi	2001				X	X										X				
Imai KS, Arun T and Annim SK	Microfinance and household poverty reduction: new evidence from India	2010				X						X					X				
Imai KS and Azam MS	Does microfinance reduce poverty in Bangladesh? New evidence from household panel data	2010				X															
Karlan D and Zinman J	Expanding credit access: using randomized supply decisions to estimate the impacts	2010	X		X	X	X		X		X		X							X	
PnK, Bangladesh	20 papers		X			X	X	X	X				X	X		X					X
Kondo T, Orbeta A, Dingcong C, Infantado C	Impact of microfinance on rural households in the Philippines	2008				X	X			X											

			Economic											Social							Political
STUDY	**TITLE**	**YEAR**	Business profits	Business revenues	Sales	Income per capita	Consumption/ Expenditure[87]	Assets	Employment	Savings	Debts	Poverty index/status	Other	Children's school enrolment	School attendance	Contraceptive use[88]	Nutritional status[89]	Vulnerability to shocks	Social capital	Other	Empowerment
Montgomery H/ Setboonsarng Sand Parpiev Z, Pakistan	2 papers		X		X	X	X	X	X	X					X					X	X
Shimamura Y and Lastarria-Cornhiel S	Credit Program participation and child schooling in rural Malawi	2010							X						X						
Shirazi NS and Khan AU	Role of Pakistan poverty alleviation fund's microcredit in poverty alleviation: a case of Pakistan	2009										X									
Steele F, Amin S and Naved RT	Savings/credit group formation and change in contraception	2001														X					
Swain RB and Wallentin FY	Does microfinance empower eomen? Evidence from self-help groups in India	2009																			X
Takahashi K, Higashikata T and Tsukada K	The short-term poverty impact of small-scale, collateral-free microcredit in Indonesia: a matching estimator approach	2010	X		X	X	X	X													
USAID, India, Peru, Zimbabwe	10 papers		X	X		X	X	X		X			X	X				X	X	X	X
Tesfay GB	Econometric analyses of microfinance credit group formation, contractual risks and welfare impacts in northern Ethiopia	2009					X						X[91]								
Zaman H	Assessing the impact of microcredit on poverty and vulnerability in Bangladesh	1999								X			X							X	X

[91] Housing Improvements

			Economic											Social							Political
STUDY	**TITLE**	**YEAR**	**Business profits**	**Business revenues**	**Sales**	**Income per capita**	**Consumption/ Expenditure**[87]	**Assets**	**Employment**	**Savings**	**Debts**	**Poverty index/status**	**Other**	**Children's school enrolment**	**School attendance**	**Contraceptive use**[88]	**Nutritional status**[89]	**Vulnerability to shocks**	**Social capital**	**Other**	**Empowerment**
Zeller M, Sharma M, Ahmed AU, Rashid S	Group-based financial institutions for the rural poor in Bangladesh: an institutional- and household-level analysis	2001					X										X				

6.17 Appendix 17: Results of other included papers based on PnK dataset

Paper	Method	Outcome variable	Headline findings	Risk of Bias
PnK 1998	WESLIML	Log pc expenditure	ols/wmlols/wml/wmlfe	High
		Female		
		brac	+ive (sig)/+ive (sig)/+ive (sig)/ +ive (sig)	
		brdb	+ive (ns)/+ive (ns)/+ive (sig)/ +ive (sig)	
		gb	+ive (ns)/+ive (sig)/+ive (sig)/ +ive (sig)	
		Male		
		brac	+ive (sig)/+ive (sig)/-ive (ns)/ +ive (ns)	
		brdb	+ive (sig)/+ive (sig)/-ive (sig)/ +ive (sig)	
		gb	+ive (ns)/+ive (ns)/-ive (sig)/ +ive (ns)	
		Ln women n-lnd assets	unwtTobit/ ols/wmlols/wml/wmlfe	
		Female		
		brac	+ive (sig)/+ive (sig)/+ive (sig)/ +ive (sig)/+ive (ns)	
		brdb	+ive (ns)/-ive (ns)/+ive (ns)/ +ive (sig)/+ive (ns)	
		gb	+ive (sig)/+ive (sig)/+ive (ns)/ +ive (sig)/+ive (ns)	
		Male		
		brac	+ive (ns)/+ive (ns)/+ive (sig)/ +ive (ns)/+ive (ns)	
		brdb	+ive (sig)/+ive (ns)/+ive (sig)/ +ive (ns)/+ive (ns)	
		gb	+ive (sig)/+ive (sig)/+ive (ns)/ -ive (ns)/-ive (ns)	
		Ln female lab supply	unwtTobit/wmltobit/wmlliml/wmlfe/wmllimlfe	
		Female		
		brac	+ive (ns)/+ive (sig)/+ive (ns)/ +ive (sig)/-ive (ns)	
		brdb	+ive (sig)/+ive (sig)/+ive (sig)/ +ive (sig)/-ive (ns)	
		gb	+ive (sig)/+ive (sig)/+ive (sig)/ +ive (sig)/+ive (ns)	
		Male		
		brac	-ive (sig)/-ive (ns)/-ive (ns)/ -ive (ns)/-ive (ns)	
		brdb	-ive (ns)/+ive (ns)/+ive (ns)/ +ive (ns)/-ive (ns)	
		gb	+ive (sig)/+ive (sig)/+ive (ns)/ -ive (nsig)/-ive (ns)	
		Male labour supply	unwtTobit/wmltobit/wmlliml/wmllimlfe	
		Female		
		brac	+ive (nsig)/+ive (nsig)/-ive (sig)/ -ive (sig)	
		brdb	-ive (nsig)/-ive (nsig)/-ive (sig)/ -ive (sig)	
		gb	+ive (sig)/+ive (sig)/-ive (sig)/ -ive (sig)	
		Male		
		brac	+ive (nsig)/+ive (nsig)/+ive (nsig)/ -ive (sig)	
		brdb	+ive (nsig)/+ive (nsig)/+ive (nsig)/ -ive (sig)	
		gb	-ive (nsig)/-ive (sig)/+ive (nsig)/ -ive (sig)	

Paper	Method	Outcome variable	Headline findings	Risk of Bias
		Girl school enrolment		
		Female	Uwprobit/wmlprobit/wmlliml/wmlfe/wmllimlfe	
		brac	+ive (nsig)/+ive (nsig)/+ive (sig)/ +ive (nsig)/-icve(nsig)	
		brdb	+ive (nsig)/+ive (nsig)/+ive (nsig)/ +ive (nsig)/--ive(nsig)	
		gb	+ive (sig)/+ive (nsig)/+ive (sig)/ +ive (sig)/+ive(nsig)	
		Male		
		brac	+ive (sig)/+ive (sig)/+ive (sig)/ +ive (nsig)/+ive(nsig)	
		brdb	-ive (nsig)/-ive (nsig)/+ive (nsig)/ +ive (nsig)/+ive(nsig)	
		gb	+ive (nsig)/+ive (nsig)/+ive (sig)/ +ive (nsig)/+ive(nsig)	
		Boy school enrolment		
		Female	Uwprobit/wmlprobit/wmlliml/wmlfe/wmllimlfe	
		brac	+ive (nsig)/+ive (nsig)/+ive (sig)/ -ive (nsig)/+ive(sig)	
		brdb	+ive (sig)/+ive (nsig)/+ive (sig)/+ive (sig)/+ive(sig)	
		gb	+ive (sig)/+ive (sig)/+ive (sig)/+-ive (sig)/+ive(sig)	
		Male		
		brac	-ive (nsig)/+ive (nsig)/+ive (nsig)/ -ive (nsig)/-ive(nsig)	
		brdb	+ive (nsig)/+ive (nsig)/+ive (nsig)/+ive (nsig)/+ive(nsig)	
		gb	+ive (sig)/+ive (sig)/+ive (nsig)/+ive (sig)/+ive(sig)	

Paper	Method	Outcome variable	Headline findings	Risk of Bias
Morduch, 1998	DID	Ln pc expend	Hhchar/hh&vill char/full sample	High
		GB	-ive(ns)/-ive(ns)/-ive(ns)	
		BRAC	-ive(ns)/-ive(ns)/+ive(ns)	
		BRDB	-ive(sig)/-ive(ns)/-ive(sig)	
		Var Ln pc expend		
		GB	-ive(ns)/-ive(ns)/-ive(sig)	
		BRAC	-ive(ns)/-ive(ns)/-ive(ns)	
		BRDB	-ive(ns)/-ive(ns)/-ive(sig)	
		Log women non-land assets	Problems reconstructing so not assessed	
		Log lab per adult pm		
		GB	+ive(ns)/+ive(sig)/-ive(ns)	
		BRAC	ive(ns)/-ive(sig)/-ive(ns)	
		BRDB	+ive(ns)/+ive(ns)/+ive(sig)	
		Var Log lab p adult pm		
		GB	-ive(ns)/-ive(ns)/-ive(sig)	
		BRAC	-ive(ns)/-ive(ns)/-ive(ns)	
		BRDB	-ive(ns)/-ive(ns)/-ive(sig)	
		Male lab supply		
		GB	+ive(sig)/+ive(sig)/+ive(ns)	
		BRAC	+ive(sig)/-ive(ns)/+ive(sig)	
		BRDB	+ive(sig)/+ive(ns)/+ive(sig)	
		Female labour supply		
		GB	+ive(ns)/+ive(sig)/-ive(sig)	
		BRAC	-ive(sig)/+ive(ns)/-ive(sig)	
		BRDB	+ive(ns)/+ive(ns)/+ive(ns)	
		Boy school enrolment		
		GB	+ive(ns)/+ive(ns)/+ive(ns)	
		BRAC	+ive(ns)/+ive(sig)/+ive(ns)	
		BRDB	-ive(ns)/+ive(ns)/-ive(ns)	
		Girl school enrolment		
		GB	-ive(ns)/-ive(sig)/-ive(ns)	
		BRAC	-ive(ns)/-ive(ns)/+ive(ns)	
		BRDB	-ive(sig)/-ive(sig)/-ive(ns)	
Pitt, 1999		Log pc expend	0.5/0.66/1.20/1.60/2.0	High
		Female bor BRAC	+ive(sig)/+ive(sig)/ +ive(sig)/ +ive(sig)/ +ive(sig)	
		Male bor BRAC	+ive(ns)/+ive(ns)/ +ive(ns)/ +ive(ns)/ +ive(ns)	
		Female bor BRDB	+ive(sig)/+ive(sig)/ +ive(sig)/ +ive(sig)/ +ive(sig)	
		Male bor BRDB	+ive(sig)/+ive(ns)/ +ive(ns)/ +ive(ns)/ +ive(ns)	
		Female bor GB	+ive(sig)/+ive(sig)/ +ive(sig)/ +ive(sig)/ +ive(sig)	
		Male bor GB	+ive(ns)/+ive(ns)/ +ive(ns)/ +ive(ns)/ +ive(ns)	
		All BRAC	+ive (sig)	
		All BRDB	+ive (sig)	
		All GB	+ive (sig)	
			With land interactions	
			None/0.5/0.66/1.20/1.60/2.0	
		Female bor BRAC	+ive(sig)/+ive(sig)/+ive(sig)/ +ive(sig)/ +ive(sig)/ +ive(sig)	

Paper	Method	Outcome variable	Headline findings	Risk of Bias
		Male bor BRAC	+ive(ns)/+ive(ns)/+ive(ns)/ +ive(ns)/ +ive(ns)/ +ive(ns)	
		Female bor BRDB	+ive(sig)/+ive(sig)/+ive(sig)/ +ive(sig)/ +ive(sig)/ +ive(sig)	
		Male bor BRDB	+ive(sig)/+ive(sig)/+ive(sig)/ +ive(sig)/ +ive(ns)/ +ive(ns)	
		Female bor GB	+ive(sig)/+ive(sig)/+ive(sig)/ +ive(sig)/ +ive(sig)/ +ive(sig)	
		Male bor GB	+ive(ns)/+ive(ns)/+ive(ns)/ +ive(ns)/ +ive(ns)/ +ive(ns)	
Chemin, 2008	PSM		Kernel0.05/0.02/0.01	High
		Var Log per capita expenditure	-ive (ns)/-ive(ns)/-ive(nss)	
		Log women non-land assets	+ive(ns)/ +ive(ns)/ +ive(ns)	
		Female lab supply	+ive(ns)/ +ive(ns)/ +ive(ns)	
		Male labour supply	+ive(sig)/ +ive(sig)/ +ive(sig)	
		Girl school enrolment	+ive(sig)/ +ive(sig)/ +ive(sig)	
		Boy school enrolment	+ive(sig)/ +ive(ns)/ +ive(ns)	
Roodman & Morduch, 2009	cmp	**Table 3** - PnK		High
		Log female borr	Tgt:LIMhh OLS/Vill/VillFE/ all: LIMhh OLS/Vill/VillFE/	
		BRAC	+ive(sig)/+ive(ns)/-ive(ns)/ +ive(ns)/ -ive(sig)/ -ive(sig)	
		BRDB	+ive(ns)/-ive(ns)/-ive(ns)/ -ive(ns)/ -ive(sig)/ -ive(sig)	
		GB	+ive(sig)/-ive(ns)/-ive(ns)/ +ive(ns)/ -ive(sig)/ -ive(sig)	
		Log Male borr		
		BRAC	+ive(sig)/-ive(ns)/-ive(ns)/ +ive(sig)/ -ive(ns)/ -ive(ns)	
		BRDB	+ive(sig)/-ive(ns)/+ive(ns)/ +ive(sig)/ -ive(ns)/ +ive(ns)	
		GB	+ive(ns)/-ive(ns)/-ive(ns)/ -ive(ns)/-ive(ns)/ -ive(ns)	
		Table 4 - PnK		
		Log per capita consum	Rounds 1-3/1/2/3/Srvy&vill dummies1-3/1/2/3	
		log female bor		
		BRAC	-ive(ns)/-ive(ns)/-ive(ns)/ -ive(ns)/-ive(ns)/-ive(ns)/ -ive(ns)/-ive(ns)	
		BRDB	-ive(sig)/-ive(ns)/-ive(sig)/ -ive(ns)/-ive(sig)/-ive(ns)/ -ive(ns)/-ive(sig)	
		GB	-ive(ns)/-ive(ns)/-ive(ns)/ -ive(ns)/+ive(ns)/+ive(sig)/+ive(ns)/+ive(ns)	
		Log male bor		
		BRAC	+ive(sig)/+ive(sig)/+ive(ns)/ +ive(ns)/+ive(ns)/+ive(ns)/-ive(ns)/+ive(ns)	
		BRDB	-ive(ns)/-ive(ns)/+ive(ns)/ -ive(ns)/-ive(sig)/-ive(ns)/ +ive(ns)/-ive(sig)	
		GB	-ive(ns)/-ive(sig)/-ive(ns)/ -ive(ns)/-ive(ns)/-ive(ns)/ +ive(ns)/+ive(ns)	
		Table 5 – PnK wwliml	Fem borr BRAC/BRDB/GM: / male borr BRAC/BRDB/GB	
		Ln fem n-land assets	+ive(sig)/+ive(sig)/+ive(sig): /+ive(ns)/-ive(ns)/-ive(ns)	
		ln fem hrs/mnth	-ive(ns)/-ive(ns)/+ive(ns): /-ive(sig)/-ive(sig)/-ive(sig)	
		ln male hrs./mnth	+ive(ns)/+ive(ns)/+ive(ns): /-ive(ns)/-ive(ns)/-ive(ns)	
		Girl school enrolment	-ive(ns)/-ive(ns)/-ive(ns): /-ive(ns)/-ive(ns)/-ive(ns)	
		Boy school enrolment	-ive(ns)/+ive(ns)/-ive(ns): /-ive(ns)/+ive(ns)/+ive(ns)	
		Table 6 - Morduch		
		Ln pc expend	targ hh: hh char/arg hh hh& vill char/ all hh, hh & vill char Vill FE	
		GB	-ive(ns)/-ive(ns)/-ive(ns)	
		BRAC	-ive(ns)/-ive(ns)/+ive(ns)	
		BRDB	-ive(sig)/-ive(ns)/-ive(sig)	

Paper	Method	Outcome variable	Headline findings	Risk of Bias
		Var Ln pc expend		
		GB	-ive(ns)/-ive(ns)/-ive(ns)	
		BRAC	-ive(ns)/-ive(ns)/-ive(ns)	
		BRDB	-ive(ns)/-ive(ns)/-ive(ns)	
		Log lab per adult pm		
		GB	+ive(ns)/+ive(ns)/-ive(ns)	
		BRAC	-ive(ns)/-ive(ns)/-ive(ns)	
		BRDB	+ive(ns)/+ive(ns)/+ive(ns)	
		Var Log adult lab pm		
		GB	-ive(ns)/-ive(sig)/-ive(ns)	
		BRAC	+ive(ns)/-ive(ns)/-ive(ns)	
		BRDB	-ive(ns)/-ive(ns)/-ive(ns)	
		Male lab supply		
		GB	-ive(sig)/+ive(ns)/-ive(ns)	
		BRAC	+ive(ns)/-ive(ns)/-ive(ns)	
		BRDB	+ive(ns)/+ive(ns)/+ive(ns)	
		Adult Fem lab supply		
		GB	+ive(ns)/+ive(ns)/-ive(ns)	
		BRAC	-ive(ns)/+ve(ns)/-ive(ns)	
		BRDB	-ive(ns)/+ive(ns)/+ive(ns)	
		Boy school enrolment		
		GB	+ive(ns)/+ive(ns)/-ive(ns)	
		BRAC	-ive(ns)/+ive(ns)/-ive(ns)	
		BRDB	-ive(ns)/-ive(ns)/-ive(ns)	
		Girl school enrolment		
		GB	-ive(ns)/-ive(sig)/+ive(ns)	
		BRAC	-ive(ns)/-ive(sig)/+ive(ns)	
		BRDB	-ive(sig)/-ive(sig)/-ive(ns)	
		Table 7 – Morduch DID		
		ln pc exp	GB/BRAC/BRDB	
		did	-ive(sig)/-ive(sig)-ive(sig)	
		Table 10 - Khandker,05	OLS no FE/FE/2sls noFE/FE 2sls Interaction Vill dummies no FE /FE	
		Ln fem current loans	+ive(sig)/+ive(ns)/+ive(sig)/+ive(ns)/-ive(ns)/-ive(ns)	
		Ln fem past loans	+ive(sig)/+ive(ns)/+ive(sig)/+ive(sig)/+ive(sig) /+ive(sig)	
		Ln male current loans	+ive(sig)/+ive(ns)/+ive(sig)/+ive(ns)/+ive(ns)+ive(ns)	
		Ln male past loans	+ive(ns)/-ive(ns)/+ive(ns)/-ive(ns)/+ive(ns)/-ive(ns)	

Paper	Method	Outcome variable	Headline findings	Risk of Bias
Duvendack, 2010b, 2011	PSM , DID	**Table 19 -** Chemin	Kernel0.05/0.02/0.01	High
		Var Log pc expend	-ive (ns)/-ive(ns)/-ive(ns)	
		Log pc expend	-ive (ns)/-ive(ns)/-ive(ns)	
		Log fem n-land assets	+ive(sig)/ +ive(sig)/ +ive(sig)	
		Female lab supply	+ive(ns)/ +ive(ns)/ +ive(ns)	
		Male labour supply	-ive(ns)/ +ive(ns)/ +ive(ns)	
		Girl school enrolment	+ive(ns)/ +ive(ns)/ +ive(ns)	
		Boy school enrolment	+ive(sig)/ +ive(sig)/ +ive(sig)	
		Table 25		
		All	NN comparisons 1/2/3/4/Kernel comparisons 1/2/3/4	
		Var Log pc expend	+ive(ns)/- ive(ns)/-ive(ns)/ +ive(sig)/-ive(ns)/-ive(ns)/-ive(ns)/ +ive(sig)	
		Log pc expend	+ive(ns)/+ive(ns)/+ive(ns)/+ive(sig)/-ive(ns)/-ive(ns)/+ive(ns)/ +ive(sig)	
		Log fem n-land assets	+ive(sig)/-+ive(sig)/-ive(ns)/+ive(ns)/+ive(sig)/+ive(sig)/+ive(ns)/+ive(ns)	
		Female lab supply	+ive(sig)/+ ive(sig)/+ive(sig)/-ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)/-ive(sig)	
		Male labour supply	-ive(ns)/-ive(sig)/+ive(sig)/+ive(sig)/-ive(ns)/-ive(sig)/+ive(sig)/+ive(sig)	
		Girl school enrolment	+ive(ns)/+ive(sig)/+ive(ns)/+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)/+ive(ns)	
		Boy school enrolment	+ive(ns)/+ ive(sig)/+ive(sig)/+ive(sig)/+ive(ns)/+ive(ns)/+ive(sig)/+ive(sig)	
		Female borr		
		Var Log pc expend	-ive(ns)/- ive(ns)/-ive(ns)/+ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)/+ive(ns)	
		Log pc expend	-ive(ns)/+ive(ns)/+ive(ns)/-ive(ns)/-ive(ns)/-ive(ns)/+ive(ns)/ +ive(ns)	
		Log fem n-land assets	+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)/+ive(ns)	
		Female lab supply	+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)	
		Male labour supply	-ive(sig)/-ive(sig)/-ive(ns)/-ive(sig)/-ive(sig)/-ive(sig)/-ive(sig)/-ive(sig)	
		Girl school enrolment	+ive(ns)/+ive(ns)/+ive(sig)/+ive(ns)/+ive(ns)/+ive(sig)/+ive(ns)/+ive(sig)	
		Boy school enrolment	+ive(ns)/+ive(sig)/+ive(ns)/+ive(sig)/+ive(ns)/+ive(ns)/+ive(ns)/+ive(ns)	
		Male borr		
		Var Log pc expend	-ive(ns)/-ive(ns)/-ive(ns)/+ive(sig)/-ive(ns)/-ive(sig)/-ive(ns)/ +ive(sig)	
		Log pc expend	+ive(ns)/+ive(ns)/+ive(ns)/+ive(sig)/-ive(ns)/-ive(ns)/+ive(ns)/+ive(sig)	
		Log fem n-land assets	+ive(sig)/+ive(ns)/-ive(sig)/+ive(ns)/+ive(sig)/+ive(ns)/+ive(ns)/-ive(ns)	
		Female lab supply	-ive(sig)/-ive(sig)/+ive(sig)/-ive(sig)/-ive(sig)/-ive(sig)/+ive(sig)/-ive(sig)	
		Male labour supply	+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)/+ive(sig)	
		Girl school enrolment	+ive(sig)/+ive(ns)/+ive(sig)/+ive(ns)/+ive(ns)/+ive(ns)/+ive(sig)/+ive(ns)	
		Boy school enrolment	+ive(ns)/+ive(ns)/+ive(sig)/+ive(sig)/-ive(ns)/+ive(ns)/+ive(sig)/+ive(sig)	
		Table 30 – Panel	RE / PSM DID	
		Var Log pc expend	-ive(sig)/-ive(sig)	
		Log pc expend	-ive(ns)/+ive(ns)	
		Log fem n-land assets	+ive(sig)/-ive(ns)	
		Female lab supply	+ive(sig)/+ive(sig)	
		Male labour supply	-ive(sig)/-ive(sig)	
		Girl school enrolment	+ive(sig)/+ive(sig)	
		Boy school enrolment	+ive(sig)/+ive(sig)	
		Table 39 – RnM		
		log pc expenditure	femBRAC/maleBRAC/femBRDB/MaleBRDB/FfemGB/maleGB	
		RnM	-ive(ns)/+ive(sig)/-ive(sig)/-ive(ns)/-ive(ns)/-ive(ns)	
		MD	+ive(ns)/+ive(ns)/-ive(ns)/-ive(sig)/-ive(sig)/-ive(ns)	

Paper	Method	Outcome variable	Headline findings	Risk of Bias
Khandker and Latif 1996 (Latif, 1994 excluded)	Multivariate	Contraceptive use	Presence of Programme	High
		BRAC	+ive (sig)	
		BRDB	ns	
		GB	+ive (sig)	
		Fertility		
		BRAC	ns	
		BRDB	ns	
		GB	ns	
		Infant mortality		
		BRAC	ns	
		BRDB	ns	
		GB	ns	
Khandker, Samad and Khan, 1998	LIML	Production		High
		Farm activities		
		GB	+ive (ns)	
		BRAC	+ive (ns)	
		BRDB	+ive (sig)	
		non-farm		
		GB	+ive (sig)	
		BRAC	+ive (sig)	
		BRDB	+ive (sig)	
		all		
		GB	+ive (sig)	
		BRAC	+ive (sig)	
		BRDB	+ive (sig)	
		Employment – in LF		
		Farm activities		
		GB	+ive (sig)	
		BRAC	-ive (ns)	
		BRDB	+ive (sig)	
		non-farm		
		GB	+ive (sig)	
		BRAC	+ive (sig)	
		BRDB	-ive (ns)	
		all		
		GB	+ive (sig)	
		BRAC	+ive (ns)	
		BRDB	+ive (sig)	
		Employment – hr/ month self/wage/total		
		Farm activities		
		GB	+ive (ns)/-ive (sig)/-ive (ns)	
		BRAC	-ive (sig)/-ive (sig)/-ive (sig)	
		BRDB	+ive (sig)/+ive (ns)/+ive (sig)	
		non-farm		
		GB	+ive (sig)/-ive (ns)/+ive (sig)	
		BRAC	+ive (ns)/+ive (sig)/+ive (ns)	
		BRDB	+ive (sig)/-ive (sig)/-ive (ns)	
		all		
		GB	+ive (sig)	
		BRAC	-ive (sig)	
		BRDB	-ive (sig)	
		Income self/wage/both/total		

Paper	Method	Outcome variable	Headline findings	Risk of Bias
		Farm activities		
		GB	+ive (ns)/-ive (sig)/+ive (ns)	
		BRAC	+ive (ns)/-ive (sig)/-ive (ns)	
		BRDB	+ive (sig)/+ive (ns)/+ive (sig)	
		non-farm		
		GB	+ive (sig)/+ive (ns)/+ive (sig)	
		BRAC	+ive (ns)/+ive (sig)/+ive (sig)	
		BRDB	+ive (ns)/-ive (ns)/-ive (ns)	
		all		
		GB	+ive (ns)	
		BRAC	+ive (sig)	
		BRDB	+ive (ns)	
		Wages		
		GB	+ive (sig)	
		BRAC	+ive (ns)	
		BRDB	-ive(ns)	
Nanda, 1999	IV probit	Use of formal health car		High
		Women's participation.	+ive (10%)	
		Men's participatione	ns	
Pitt, Khandker, McKernan and Latif, 1999	LIML	Modern contraceptive use	Ns	High
		females	-ive (?5%)	
		males	+ive (ns)	
		Fertility		
		Female participation	+-ive for female participation	
		male participation	-ive for male participation	
Pitt, 2000	LIML	Contractual relations		High
		Employment	+ive for sharecropping, agricultural self-employment – more so for women	
		Sharecropped land		
		Elig0.5		
		Female		
		Aman	+ive (ns)/+ive (ns)	
		boro	+ive (sig)/+ive (sig)	
		aus	+ive (sig)/+ive (sig)	
		Male		
		Aman	+ive (ns)/-ive (ns)	
		boro	+ive (ns)/-ive (ns)	
		aus	+ive (sig)/-ive (ns)	
		Elig 1.0		
		Female		
		Aman	+ive (ns)/-ive (ns)	
		boro	+ive (sig)/+ive (sig)	
		aus	+ive (sig)/+ive (sig)	
		Male		
		Aman	-ive (ns)/-ive (ns)	
		boro	-ive (ns)/+ive (ns)	
		aus	+ive (ns)/+ive (ns)	
		Fixed Rental land		

Paper	Method	Outcome variable	Headline findings	Risk of Bias
		Elig0.5		
		Female		
		Aman	+ive (ns)/+ive (ns)	
		boro	+ive (ns)/+ive (ns)	
		aus	-ive (ns)/+ive (ns)	
		Male		
		Aman	+ive (ns)/+ive (ns)	
		boro	+ive (ns)/+ive (sig)	
		aus	+ive (ns)/+ive (ns)	
		Elig 1.0		
		Female		
		Aman	+ive (ns)/+ive (ns)/+ive (ns)	
		boro	-ive (ns)/+ive (ns)/+ive (ns)	
		aus	-ive (sig)/-ive (sig)/-ive (ns)	
		Male		
		Aman	+ive (ns)/-ive (ns)/+ive (ns)	
		boro	+ive (ns)/+ive (sig)/+ive (sig)	
		aus	+ive (ns)/+ive (ns)/+ive (ns)	
		Male self-employment in agric		
		Elig0.5		
		Female		
		Aman	+ive (sig)/+ive (sig)	
		boro	+ive (ns)/+ive (sig)	
		aus	+ive (ns)/+ive (sig)	
		Male		
		Aman	+ive (ns)/+ive (sig)	
		boro	+ive (sig)/+ive (sig)	
		aus	+ive (sig)/+ive (sig)	
		Elig 1.0		
		Female		
		Aman	+ive (ns)/+ive (ns)/+ive (ns)	
		boro	+ive (ns)/+ive (ns)/+ive (ns)	
		aus	+ive (ns)/+ive (ns)/+ive (ns)	
		Male		
		Aman	+ive (ns)/+ive (sig)/+ive (sig)	
		boro	+ive (ns)/+ive (sig)/+ive (sig)	
		aus	+ive (sig)/+ive (sig)/+ive (sig)	
		Male wage employment in agric		
		Elig0.5		
		Female		
		Aman	-ive (ns)/-ive (sig)	
		boro	-ive (ns)/-ive (sig)	
		aus	-ive (ns)/-ive (sig)	
		Male		
		Aman	+ive (ns)/+ive (ns)	
		boro	+ive (ns)/-ive (sig)	
		aus	+ive (ns)/-ive (sig)	

Paper	Method	Outcome variable	Headline findings	Risk of Bias
		Elig 1.0		
		Female		
		Aman	-ive (ns)/-ive (sig)	
		boro	-ive (ns)/-ive (sig)	
		aus	-ive (ns)/-ive (sig)	
		Male		
		Aman	+ive (ns)/+ive (ns)	
		boro	+ive (ns)/-ive (sig)	
		aus	+ive (ns)/-ive (sig)	
Khandker, 2000	LIM L	Log male MFborrowing	**Table 9**	High
		Formal borrowing	-ive(ns)	
		informal borrowing	-ive(ns	
		total savings	+ive (sig)	
		prog savings	+ive (sig)	
		voluntary savings	+ive (ns)	
		Log fe m MF borrowing		
		Formal borrowing	-ive (sig)	
		informal borrowing	+ive (ns)	
		total savings	+ive (sig)	
		prog savings	+ive (sig)	
		voluntary savings	+ive (ns)	
		Log male MF borrowing	**Table1 10**	
		GB prog savings	+ive (sig)	
		GB vol savings	+ive (ns)	
		BRAC prog savings	+ive (sig)	
		BRAC vol savings	-ive (sig)	
		BRDB prog savings	+ive (sig)	
		BRDB vol savings	+ive (ns)	
		Log fe m MF borrowing		
		GB prog savings	+ive (sig)	
		GB vol savings	+ive (sig)	
		BRAC prog savings	+ive (sig)	
		BRAC vol savings	+ive (sig)	
		BRDB prog savings	+ive (sig)	
		BRDB vol savings	+ive (ns)	
		(Table 11)		
		Log MF savings	*Male borrowing in - aman/boro/aus/*	
		pooled	+ive (sig)/+ive (sig)/ +ive (sig)	
		aman	+ive (sig)	
		boro	+ive (sig)	
		aus	+ive (sig)	
			Female borrowing in - aman/boro/aus	
		pooled	ive (sig)/+ive (sig)/ +ive (sig)	
		aman	+ive (sig)	
		boro	+ive (sig)	
		aus	+ive (sig)	
McKern an, 2002	LIM L	**Table 2**	All three programmes	High
		Log monthly profits	wesl/weslfe/wesllimlfeexog3/wesmllimlfeexog1	
		brac	+ive (sig)/+ive(sig)/+ive(sig)/+ive(sig)	
		brdb	+ive (sig)/+ive(sig)/+ive(sig)/+ive(sig)	

Paper	Method	Outcome variable	Headline findings	Risk of Bias
		gb	+ive (sig)/+ive(sig)/+ive(sig)/+ive(sig)	
		Table 3		
		Log monthly profit	0.5/strict0.5/1.0/1.5/2.0gbexog/2.0/2.0all*logland	
		Brac	+ive (sig)/+ive(sig)/+ive(sig)/+ive(sig) /+ive(sig) /+ive(sig) /+ive(sig)	
		brdb	+ive (sig)/+ive(sig)/+ive(sig)/+ive(sig) /+ive(sig) /+ive(sig) /+ive(sig)	
		gb	+ive (sig)/+ive(sig)/+ive(ns)/+ive(ns) /+ive(ns) /+ive(sig) /+ive(sig)	
		Table 4		
		Non-credit effect		
		Log monthly profit		
		spec 1	wesl/weslfe/wesllimlfeexog3/wesmllimlfeexog1	
		Brac	+ive (sig)/+ive(sig)/+ive(sig)/+ive(sig)	
		brdb	+ive (sig)/+ive(sig)/+ive(sig)/+ive(sig)	
		gb	+ive (sig)/+ive(sig)/+ive(sig)/+ive(sig)	
		spec 2	weslfe1/weslfe2/weslfe3	
		Brac	+ive (sig)/+ive(sig)/+ive(ns)	
		brdb	+ive (sig)/+ive(sig)/+ive(Ns)	
		gb	+ive (sig)/+ive(sig)/+ive(sig)	
Pitt et al, 2003	LIML-FE	Female borrowers		High
		Arm circumference		
		Girls	+ive (5%)	
		boys	+ive (5%)	
		BMI		
		irls	ns	
		boys	ns	
		Height-for-age		
		girls	+ive (5%)	
		boys	+ive (5%)	
		Male borrowers		
		Arm circumference		
		Girls	+ive (ns)	
		boys	+ive (ns)	
		BMI		
		irls	0 (ns)	
		boys	0 (ns)	
		Height-for-age		
		girls	-ive (ns)	
		boys	-ive (ns)	

Paper	Method	Outcome variable	Headline findings	Risk of Bias
Menon, 2006	Multivariate	Consumption **Table 3**	Returns to MF membership vary by length of participation. Mixed results.	High
		log of pc food expenditure	(length of membership - linear coeficients only m&f1/m&f2/male1/male2/female1/female2 +ive (sig)/+ive(sig)/+ive (ns) /+ive (ns)/+ive (sig)/+ive(sig	
		Table 4 Dif season1-2 log of pc food expenditure	length of membership - m&f/male/female -ive (sig)/-ive(sig)/-ive (sig)	
		Table 5 Dif season2-3 log of pc food expenditure	-ive (sig)/-ive(sig)/-ive (ns)	
			-ive (sig)/-ive(sig)/-ive (sig)	
		Dif season2-3 **Table 6** log of pc food expenditure		
			-ive (sig)/-ive(sig)/-ive (sig)	
		Table 7 diff 1-2 & 2-3 log of pc food expenditure	length of membership - m&f/male/female/ m&f/male/female -ive (sig)/-ive(sig)/-ive (sig)/+ive(sig)/+ive (ns)/-ive(ns)	
Pitt, Khandker and Cartwright, 2006	LIML	Women's empowerment **Table 4** OLS	(latent outcome empowerment variables) Female credit/purchasing/resources/finance/transaction management/mobility and networks/activism/attitudes/husband behaviour/fertility and parenting ols +ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/ olsvfe +ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/ Male credit/purchasing/resources/finance/transaction management/mobility and networks/activism/attitudes/husband behaviour/fertility and parenting ols -ive (ns)/-ive (ns)/-ive (ns)/+ive (ns)/-ive (sig)/-ive (ns)/+ive (ns)/+ive (ns)/-ive (ns) olsvfe -ive (ns)/-ive (ns)/-ive (ns)/+ive (ns)/-ive (sig)/-ive (ns)/+ive (ns)/+ive (ns)/-ive (ns)	High
		Table 5	(latent outcome empowerment variables) Female credit/purchasing/resources/finance/transaction management/mobility and networks/activism/attitudes/husband behaviour/fertility and parenting ols +ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/ Olsvfe +ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/ Iv_vfe +ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/+ive (sig)/ Male credit/purchasing/resources/finance/transaction management/mobility and networks/activism/attitudes/husband behaviour/fertility and parenting Ols +ive (ns)/+ive (ns)/-ive (ns)/-ive (ns)/-ive (ns)/+ive(ns)/+ive (ns)/+ive (ns)/-ive (ns) Olsvfe -ive (ns)/+ive (ns)/+ive (ns)/+ive (ns)/-ive (sig)/-ive (ns)/+ive (ns)/+ive (ns)/-ive (ns) Iv_vfe -ive (ns)/-ive (ns)/-ive (ns)/+ive (ns)/-ive (sig)/-ive (ns)/+ive (ns)/+ive (ns)/-ive (ns)	

The authors of this report were supported by the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) and the International Initiative for Impact Evaluation (3ie).

The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) is part of the Social Science Research Unit (SSRU), Institute of Education, University of London.

Since 1993, we have been at the forefront of carrying out systematic reviews and developing review methods in social science and public policy. We are dedicated to making reliable research findings accessible to the people who need them, whether they are making policy, practice or personal decisions. We engage health and education policy makers, practitioners and service users in discussions about how researchers can make their work more relevant and how to use research findings.

Founded in 1990, the Social Science Research Unit (SSRU) is based at the Institute of Education, University of London. Our mission is to engage in and otherwise promote rigorous, ethical and participative social research as well as to support evidence-informed public policy and practice across a range of domains including education, health and welfare, guided by a concern for human rights, social justice and the development of human potential.

This material has been funded by the Department for International Development. However the views expressed do not necessarily reflect the department's official policies. The report was first published in 2011 by:

Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre)
Social Science Research Unit
Institute of Education, University of London
18 Woburn Square
London WC1H 0NR

Tel: +44 (0)20 7612 6397
http://eppi.ioe.ac.uk
http://www.ioe.ac.uk/ssru

ISBN: 978-1-907345-19-7

Cover image © 2004 Ingrid Hesling, Courtesy of Photoshare

The International Initiative for Impact Evaluation (3ie) works to improve the lives of people in the developing world by supporting the production and use of evidence on what works, when, why and for how much. 3ie is a new initiative that responds to demands for better evidence, and will enhance development effectiveness by promoting better informed policies. 3ie finances high-quality impact evaluations and campaigns to inform better program and policy design in developing countries. 3ie Systematic Reviews examine the range of available evidence regarding a particular intervention.

International Initiative for Impact Evaluations (3ie)
Tel: + 44(0)20 7958 8351/8350
http://www.3ieimpact.org

The views expressed in this work are those of the authors and do not necessarily reflect the views of the EPPI-Centre, 3ie or the funder. All errors and omissions remain those of the authors.

This document is available in a range of accessible formats including large print. Please contact the Institute of Education for assistance:
telephone: +44 (0)20 7947 9556 email: info@ioe.ac.uk