

This article was downloaded by: [Institute of Education]

On: 05 November 2012, At: 01:43

Publisher: Routledge

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



## Journal of Development Effectiveness

Publication details, including instructions for authors and subscription information:

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

### Purity or pragmatism? Reflecting on the use of systematic review methodology in development

Ruth Stewart <sup>a b</sup>, Carina van Rooyen <sup>b c</sup> & Thea de Wet <sup>b c</sup>

<sup>a</sup> EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, London, UK

<sup>b</sup> Centre for Anthropological Research, University of Johannesburg, Johannesburg, South Africa

<sup>c</sup> Department of Anthropology and Development Studies, University of Johannesburg, Johannesburg, South Africa

Version of record first published: 18 Sep 2012.

To cite this article: Ruth Stewart, Carina van Rooyen & Thea de Wet (2012): Purity or pragmatism? Reflecting on the use of systematic review methodology in development, *Journal of Development Effectiveness*, 4:3, 430-444

To link to this article: <http://dx.doi.org/10.1080/19439342.2012.711341>

PLEASE SCROLL DOWN FOR ARTICLE

For full terms and conditions of use, see: <http://www.tandfonline.com/page/terms-and-conditions>

esp. Part II. Intellectual property and access and license types, § 11. (c) Open Access Content

The use of Taylor & Francis Open articles and Taylor & Francis Open Select articles for commercial purposes is strictly prohibited.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## Purity or pragmatism? Reflecting on the use of systematic review methodology in development

Ruth Stewart<sup>a,b\*</sup>, Carina van Rooyen<sup>b,c</sup> and Thea de Wet<sup>b,c</sup>

<sup>a</sup>EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, London, UK; <sup>b</sup>Centre for Anthropological Research, University of Johannesburg, Johannesburg, South Africa; <sup>c</sup>Department of Anthropology and Development Studies, University of Johannesburg, Johannesburg, South Africa

Systematic review methodology pioneered in health care has been increasingly applied to development questions of importance in lower- and middle-income countries. This paper reports one such review on the topic of microfinance in sub-Saharan Africa and reflects on the number of pragmatic methodological compromises made when applying the method to a new field. These compromises relate to multidisciplinary teamwork, application of regional filters, drawing on evidence from additional study types and exploring mechanisms for change through the development and testing of a causal pathway. The paper concludes that a pragmatic rigorous approach to systematically reviewing evidence of effectiveness is needed for international development.

**Keywords:** systematic review; microfinance; development; multidisciplinary teamwork; sub-Saharan Africa

### 1. Background

There is increasing pressure for development policy and practice to be based on evidence of what works (Sutcliffe and Court 2005, Deaton 2009, Clemens and Demombynes 2010). This pressure led to initiatives such as the Abdul Latif Jameel Poverty Action Lab (J-PAL), Innovations for Poverty Action and the International Initiative for Impact Evaluations (3ie) which advocate better evidence for more effective poverty alleviation, rather than relying on theoretical models (Deaton 2009, Banerjee and Duflo 2011, Karlan and Apple 2011). It has resulted in the conduct of randomised controlled trials (RCTs) to assess impact in development and, more recently, systematic reviews. Systematic reviews evolved in the 1980s in health care as a means to collate and synthesise the findings of RCTs. Under the auspices of the Cochrane Collaboration an international network of clinicians and academics now undertake and update systematic reviews annually, with over 4500 now available and a further 2000 underway (Cochrane Collaboration 2011). National and international bodies such as the National Institute for Health and Clinical Excellence in the UK and the World Health Organization routinely draw on these Cochrane reviews to inform their guidelines and policies. Since the late 1990s this methodology has been extended beyond health care into education and social policy through networks such as the Campbell Collaboration (2011) and the EPPI-Centre (2011), and has most recently been

---

\*Corresponding author. Email: [r.stewart@ioe.ac.uk](mailto:r.stewart@ioe.ac.uk)

applied in the field of development, advocated by bodies such as 3ie and the International Development Coordinating Group set up within the Campbell Collaboration.

However, the evidence-based approach does not necessarily translate smoothly to the context of development (Court *et al.* 2005, Sutcliffe and Court 2005). There is both a theoretical and an instrumental gap between the traditional worlds of systematic reviews of effectiveness and traditional impact evaluation in development. Systematic review methodology is based on the gold standard of RCTs which is now widely accepted in health care, but such a narrow approach to evidence is not so easily accepted in a field where trials are costly, have ethical dilemmas and are often lacking; where solutions are urgently required; and where heterogeneity raises serious concerns about the external validity of such trials (Bhargava 2008, Chambers *et al.* 2009, Deaton 2009, Jones 2009, Barrett and Carter 2010, Algoso 2011).<sup>1</sup> Esther Duflo though, one of the leading lights on the application of RCTs for poverty reduction, is quoted as saying, 'Creating a culture in which rigorous randomised evaluations are promoted, encouraged, and financed has the potential to revolutionise social policy during the 21st century, just as randomised trials revolutionised medicine during the 20th' (Duflo *et al.* 2004).

Part of the recent debate regarding the use of RCTs in the field of development is about the appropriateness of trials for development; for the debates in the media and on blogs, see Bendavid (2011), Blattman (2011), Buckley (2010), Devarajan (2011), Glennerster and Kremer (2011), Goldacre (2011), Haddad (2011), Kristof (2011), Lindley (2011) and Subramanian (2011). Other considerations are what Clemens and Demombynes (2010, p. 1) call luxury versus necessity; achieving rigour in research is costly, both financially and time-wise, leading practitioners to accept 'less' rigorous research wherein less is made of quantifiable outcomes. Hulme (2000) identifies four categories of rigorous research: monitoring and validation of impact which does not actually assess impact itself; a simple approach to assessing impact that seeks timely information at low cost; a moderate approach that is more costly and yields more reliable findings; and a complex approach with high cost and high reliability of causality. Hulme (2000) then argues for achieving 'fit' as an acceptable level of rigour. The question remaining is whether we have to choose between technical quality and policy influence (White 2011a).

Since 2010 there have been a number of calls for academics to undertake systematic reviews in development funded by the UK's Department for International Development (DFID), AusAID and 3ie. Our team successfully bid for one of the pilot reviews on the topic of microfinance (Stewart *et al.* 2010, van Rooyen *et al.* 2012). This paper reports the methodological approach which we employed and reflects on the implications of the pragmatic compromises we made when applying these methods to a new field. We consider each of the following aspects of our approach and draw out considerations for the future: the multidisciplinary nature of our team, our regional scope, the use of a wider range of sources for identifying relevant literature, our inclusion of a wider range of study designs than usual, the development and testing of a causal pathway to understand the mechanisms of change and the tight time frame within which the review was conducted. In presenting our methodology as a case study, we intend to prompt discussion and debate about the possible implications of our pragmatic approach.

## 2. Our methodological approaches and their implications

### 2.1. A multidisciplinary team

In traditional systematic reviews about the impact of health care interventions, review teams typically include those with methodological expertise and clinical experience. The Cochrane Collaboration offers all review teams training, and expect all groups to be

informed by, if not include, those with experience of delivering the health care intervention. The Cochrane Consumers Network (homepage) advocates the inclusion of service users' views when a review is designed and disseminated. The EPPI-Centre has employed two models. Initially all EPPI-Reviews were conducted by core staff with methodological expertise, advised by a multidisciplinary group with experience of the intervention or topic of interest. This approach had extended over the last 10 years to include a wider number of novice review groups supported by training and mentoring from EPPI-Centre staff (homepage).

*2.1.1. Our approach.* Our team was both multidisciplinary and international, including individuals with expertise in systematic reviewing and development, and spanning the UK and South Africa. We also drew on wider expertise through virtual networks. The use of email and Twitter enabled us to also draw on a wider range of people with direct experience of systematic reviews, impact evaluation in development and microfinance specifically. This wider group of people essentially took the role more traditionally taken by a project advisory group.

*2.1.2. Implications and reflections.* The variety within our team was not necessarily unique amongst systematic review teams; however, it was none-the-less significant for the conduct of the review. Each member of our team had a sharp learning curve in getting to grips with either a new topic area or a new methodology. Nevertheless, we were able to rely on one another at key stages in the review enabling us to overcome hurdles as they came along. Trusting one another when working outside of our own areas of expertise was essential to making the most of this multidisciplinary team.

As well as simply preventing us from getting stuck and allowing us to deliver our systematic review on time, the multidisciplinary nature of our team also had positive implications for our output, which benefitted further from our wider virtual network of advisors. Whilst many reviews do attempt to search a range of sources, our use of new media enabled us to conduct more innovative searching, identifying the most up-to-date research in the field. We have also been able to access more varied opportunities for dissemination than experienced on previous reviews. By using Twitter and drawing on a wider network of expertise this provided, our review was blogged by one of the leading academics in the field of microfinance which elicited some interesting discussion about the quality of the evidence available (Roodman 2011).

Lastly, the mixed team has allowed for considerable capacity building, both in systematic review methodology amongst the team in South Africa, and also in development for those in the UK. This training extended to postgraduate students at the University of Johannesburg for whom we ran a half-day workshop to introduce them to systematic reviews and the evidence-based approach to decision-making.

## *2.2. Scope of the review restricted by region*

Systematic reviews are (almost) always worldwide in scope, the aim being to identify and synthesise the findings of *all* available evidence on a subject. However, development is fundamentally a spatial field of work, with great emphasis placed on context and locale raising questions over the appropriateness of synthesising evidence from around the world to inform decisions in one region, and therefore leading to doubts over the generalisability of review findings.

*2.2.1. Our approach.* We set out to conduct a review of impact evaluations conducted only within sub-Saharan Africa. Our logic for this regional scope related to the nature and history of microfinance in the region compared to elsewhere in the world. Much has been written about the practices and impacts of microcredit in Asia, where the microcredit movement has its origin. In contrast, there is relatively little known about microfinance in sub-Saharan Africa to where the microcredit movement spread in the 1980s, and where it became stronger in the 1990s.<sup>2</sup> Sub-Saharan Africa therefore typically ‘disappears’ in the wealth of data on microfinance from Asia and Latin America. Further, sub-Saharan Africa is the poorest region in the world, according to the new multidimensional poverty index developed by Oxford University (Alkire and Santos 2010). With microfinance aiming to serve the poor, this region should then reveal much when reviewing the impact of microfinance. Sub-Saharan Africa is also the region in the world with the least access to formal financial services,<sup>3</sup> and the only region in the world where donor funding outstrips private portfolio funding (Honohan and Beck 2007). Our arguments were shared by our funder DFID, who agreed to fund this sub-Saharan African review, maybe partly because DFID – together with the World Bank – is developing a new capacity building fund for microfinance in Africa, called MICFAC. This regional scope affected our searching and criteria for including studies in the review.

*2.2.2. Implications and reflections.* Our team included experts in development in sub-Saharan Africa, enabling us to make the most of our regionally focused review with better understanding of the microfinance interventions reviewed and more sophisticated interpretation of our results than might otherwise have been possible. It has also proved beneficial in disseminating the findings of our work, as Africa-based academics and aid agencies have identified with the scope of the review. This was illustrated by the attention given to the review on CNN’s Market Place Africa website (Kermeliotis 2011).

On a practical level, limiting this review to one region considerably reduced the workload, making it possible to deliver the project in a very short time frame (more on this below). This increased efficiency was possible because of the methodology of systematic reviews which allowed studies from outside the region to be excluded very early in the review process (when searching for relevant literature), meaning far fewer studies to screen and process. This allowed us more time to explore studies with various comparative research designs, and not only RCTs.

Despite the apparent benefits of this regionally focused review, it looks unlikely that systematic review funders will commission similarly scoped reviews. This may be because the scope of such funders extends beyond regional boundaries and therefore such products do not meet their needs for global solutions, especially in the light of Sumner’s (2010) finding that three-quarters of the world’s poor live in middle-income countries. However, this may be short-sighted given the acknowledged limitations of one-size-fits-all development initiatives, and the need to consider context in development initiatives. One solution may be regional subgroup analysis within worldwide reviews.

### *2.3. Searching using range of sources*

Systematic reviews draw on a range of sources to identify potentially relevant literature (Harden 2001); however, they are traditionally heavily reliant on searching electronic databases. This may not be so appropriate in the field of development, where the publication and cataloguing of research is less standardised amongst a wide array of role-players, with

no equivalent of the large freely accessible health care library, PubMed. Broader searching is therefore required.

**2.3.1. Our approach.** Our search strategy included traditional database searching, but rather than searching one main database and a few additional specialist sources, we searched 18 different databases, as well as the websites of 24 organisations, and an online directory of books. We also contacted 23 key microfinance networks, organisations and individuals requesting relevant evidence, conducted citation searches for two key publications and searched the reference lists of initially included papers. Whilst our searching was all conducted in English, we did not exclude studies based on language, but worked with native speakers to assess foreign language papers for relevance and obtain translations when appropriate. Lastly, we identified a number of relevant research papers through our participation in informal microfinance networks by way of Twitter.

**2.3.2. Implications and reflections.** Reviewing the sources of the 15 studies included in our final review, we see that they were identified from a wide range of different searches including online searches of databases (IDEAS, ELDIS, Psychinfo, EconLit, SSCI, CSA and the Cochrane Library), searching organisational websites, checking the reference lists of relevant papers, contacting authors and citation searching (Stewart *et al.* 2010). Nine of the included 15 studies were identified from non-bibliographic databases, and five were not published in peer-reviewed journals. This suggests that our broad approach to searching and use of a range of different sources was valuable in ensuring identification of relevant studies for inclusion in the review. In particular, some studies were only found from searching reference lists of other relevant papers, highlighting the importance of investing time in this method, even though it often occurs later in the systematic review process than is ideal for collecting and including these additional papers.

Although only searching in English limited our review and may have introduced some publication bias, our scope to include papers in languages other than English also appears to have been vital as two of our 15 studies were in French and Spanish. This presents a challenge for review teams who may not have the scope to include non-English papers, but may in fact be particularly important in development where the language of publication may be shaped by the source of aid funding rather than academic culture. The Spanish paper in our review would support this argument as the language was determined by the funder (the Spanish Red Cross) and not the country studied (Rwanda). We do acknowledge that translation is not always available to review teams. However, as online resources such as Google Translate improve in quality and become easier to use, we envisage greater scope to include foreign language papers in development systematic reviews.

#### **2.4. Included a wider range of study designs than conventional systematic reviews of impact**

Systematic reviews of evidence of effectiveness employ a strict hierarchy of evidence with RCTs upheld as the gold standard. In such a counterfactual causal approach to impact assessment, steps are taken to remove potential biases and isolate the true impact of the specific intervention. These primarily include randomisation to intervention (that is, those who receive the service) and control (that is, comparison) groups and the collection of data before and after the intervention is implemented (White 2011b), and careful consideration of sample size and selection method to ensure sufficient evidence to conclude on impact

(Abadie and Imbens 2009). Whilst some systematic review groups will include studies with proxy randomisation, and even trials without randomisation, most insist on measurement of double-difference as a minimum standard (Campbell Collaboration 2011). Some advocate only *ex ante* experimental designs whilst others accept the value of impact evaluations using quasi-experimental approaches so long as sufficiently rigorous techniques are used to reduce bias (Bamberger *et al.* 2006). Whilst in health care debates around how to measure impact are largely a thing of the past, in development there is far less consensus than in health care, and the jury is still out regarding the necessity, appropriateness, ethics and limitations of RCT study designs (Cartwright 2007, Deaton 2009, Jones 2009, Ravallion 2009, Bamberger *et al.* 2010, Algoso 2011, White 2011b). Odell (2010) and El-Zoghbi and Martinez (2011) discuss the various approaches (including RCTs), and their appropriateness, to measure the impact of microfinance. Copestake *et al.* (2009) argue that RCTs are the best way to measure the impact and improve product design. Morduch (2011), on the other hand, raises some concerns of their use as impact assessment for microfinance includes issues of external validity and replication.

Hulme (2000), whilst acknowledging what he calls the scientific method (in which control groups are used during surveys to produce statistically valid results, such as RCTs and quasi-experimental research designs), identifies two further methodological approaches to study the impact of microfinance (and development), namely the humanities tradition (which makes use of mainly qualitative methods) and participatory learning and action (which use various participatory qualitative research tools). Except for proving impact, Hulme (2000) argues that many assessments done in the microfinances field (and we argue also in the wider development field) are about improving practices (which typically make more use of the latter research designs).<sup>4</sup> The translation of systematic review methodology from health, with its integral assumptions regarding the status of RCTs, thus has to contend with different histories and approaches to evidence in the development field, as well as the complexities of simultaneous causality, and the realities of development as political in nature.

**2.4.1. Our approach.** Anticipating that there would not be many RCTs on microfinance in sub-Saharan Africa, and that there would be many studies by the microfinance industry itself with varying rigour and more qualitative in nature (possibly relying more on anecdotal evidence), we took the decision to include in our review all good quality impact evaluations which included a comparison group, incorporating RCTs, non-randomised trials (including quasi-experimental designs such as pipeline studies and panel data) and simple comparison studies comparing those with and without microfinance, but not necessarily having before and after data. We then applied quality criteria to all these studies and excluded some for poor quality. Our criteria for judging quality, reported in detail elsewhere (Stewart *et al.* 2010, van Rooyen *et al.* 2012), included assessments of how studies had addressed placement bias, selection bias, attrition bias and reporting bias. We did not then use study design as a measure of quality beyond our requirement for a comparison group, but instead considered the implementation of those designs and the validity of the findings to further sort the available studies.

We initially reported all our findings from the four RCTs, two non-randomised trials and nine comparison studies all together (Stewart *et al.* 2010). Since publication, we have reanalysed our findings to explore the implications of including all comparison studies, only trials or only RCTs. With both our initial analysis, and these further data, we have taken the approach that the debate around study design in development impact evaluation is ongoing, and the best approach in these early ‘pilot’ phases of systematic reviews in this

field is transparency. We therefore included additional detailed reporting in appendices of our full report to allow the reader to distinguish between these study designs should they wish.

**2.4.2. Implications and reflections.** The first implication of including a wider range of studies is simply that we have been able to draw on a wider pool of good evidence as illustrated in Table 1.

Including a wider range of studies has enabled us to include evidence from more countries: the included RCTs were based in South Africa, Kenya and Uganda; the non-randomised trials in Zimbabwe and Uganda; and the simple comparison studies provided further data from Ethiopia, Tanzania (Zanzibar), Ghana, Malawi, Madagascar and Rwanda. It also resulted in evidence from a wider range of interventions, including two studies which assessed the impacts of combined savings and credit (one controlled trial and one comparison study), and additional evidence on microcredit (one controlled trial and eight comparison studies). For microsavings, on the other hand, the evidence base did not change with the inclusion of study designs wider than RCTs, as both identified studies employed RCT designs.

The impact of including additional study designs on our findings is summarised below, with the directions of effect for each outcome and intervention presented from all the included evidence, trials only or RCTs only. The sign (+, -, mixed) indicates the direction of effect, or the identification of 'no effect', and the number in brackets represents the number of studies with that result.

Table 2 shows that for savings, the inclusion of broader study designs has no impact as the only two identified studies were RCTs. For both credit and combined interventions, however, the strength of evidence in terms of number of studies reduces, and in many

Table 1. The number of included studies by intervention and study design.

|                              | RCTs | Non-randomised trials | Simple comparison studies |
|------------------------------|------|-----------------------|---------------------------|
| Microsavings                 | 2    | –                     | –                         |
| Microcredit                  | 2    | 1                     | 8                         |
| Microcredit and microsavings | –    | 1                     | 1                         |

Table 2. A summary of our synthesis results addressing the question 'Is microfinance effective in impacting on poor people's savings, expenditure and assets, or incomes?'.

| Outcomes               | Source of evidence | Savings          | Credit      | Combined credit and savings |
|------------------------|--------------------|------------------|-------------|-----------------------------|
| Savings                | All evidence       | YES + (2)        | YES + (1)   | YES + (1)                   |
|                        | Trials only        | YES + (2)        | No evidence | YES + (1)                   |
|                        | RCTs only          | YES + (2)        | No evidence | No evidence                 |
| Expenditure and assets | All evidence       | MAYBE mixed (1)  | YES + (4)   | MAYBE mixed (2)             |
|                        | Trials only        | MAYBE mixed (1)  | YES + (1)   | MAYBE mixed (1)             |
|                        | RCTs only          | MAYBE mixed (1)  | No evidence | No evidence                 |
| Incomes                | All evidence       | YES + (2)        | YES + (1)   | YES + (1)                   |
|                        | Trials only        | NO no effect (1) | YES + (2)   | No evidence                 |
|                        | RCTs only          | NO no effect (1) | YES + (1)   | No evidence                 |

cases, when including only trials or only RCTs no evidence is available. The overall direction of effect does not change though, although some may consider this to be a fortunate co-incidence.

Our synthesis results for non-financial impacts such as health and education are summarised in Table 3 below, which similarly presents those findings from all included studies, trials only and RCTs only.

As with financial outcomes, Table 3 shows that for savings, the inclusion of broader study designs has no impact as the only two identified studies were RCTs. For both credit and combined interventions, however, the strength of evidence in terms of number of studies reduces, and in many cases, when including only trials or only RCTs no evidence is available. The overall direction of effect does change in a few instances (represented in bold in the table). The impact of microcredit on health becomes more positive when only trials or only RCTs are included, and the same is true for food security and nutrition. The evidence of the impact of microcredit on job creation, on the other hand, is less positive when only trials are included. Despite these few cases, it is noteworthy that the major impact of including wider study designs in our review is to provide evidence on more outcomes, rather than to provide contradictory evidence.

Table 3. A summary of our synthesis results addressing the question 'Is microfinance effective in impacting on poor people's lives?'.

| Outcomes                    | Source of evidence | Savings     | Credit                                       | Combined credit and savings |
|-----------------------------|--------------------|-------------|--|-----------------------------|
| Education                   | All evidence       | YES + (1)   | MAYBE + (2), mixed (2), – (1), no effect (1) | NO – (1), no effect (1)     |
|                             | Trials only        | YES + (1)   | MAYBE mixed (1)                              | NO – (1)                    |
|                             | RCTs only          | YES + (1)   | No evidence                                  | No evidence                 |
| Women's empowerment         | All evidence       | No evidence | MAYBE + (1), mixed (2), no effect (1)        | No evidence                 |
|                             | Trials only        | No evidence | MAYBE mixed (2)                              | No evidence                 |
|                             | RCTs only          | No evidence | MAYBE mixed (1)                              | No evidence                 |
| Health                      | All evidence       | YES + (2)   | MAYBE + (4), mixed (1)                       | YES + (1)                   |
|                             | Trials only        | YES + (2)   | <b>YES + (2)</b>                             | No evidence                 |
|                             | RCTs only          | YES + (2)   | <b>YES + (1)</b>                             | No evidence                 |
| Food security and nutrition | All evidence       | YES + (1)   | MAYBE + (2), mixed (2), no effect (1)        | YES + (1)                   |
|                             | Trials only        | YES + (1)   | <b>YES + (1)</b>                             | No evidence                 |
|                             | RCTs only          | YES + (1)   | No evidence                                  | No evidence                 |
| Housing education           | All evidence       | No evidence | YES + (1)                                    | YES + (2)                   |
|                             | Trials only        | No evidence | No evidence                                  | YES + (1)                   |
|                             | RCTs only          | No evidence | No evidence                                  | No evidence                 |
| Job creation                | All evidence       | No evidence | MAYBE + (1), no effect (1)                   | No evidence                 |
|                             | Trials only        | No evidence | <b>NO no effect (1)</b>                      | No evidence                 |
|                             | RCTs only          | No evidence | No evidence                                  | No evidence                 |
| Social cohesion             | All evidence       | No evidence | No evidence                                  | No evidence                 |
|                             | Trials only        | No evidence | No evidence                                  | No evidence                 |
|                             | RCTs only          | No evidence | No evidence                                  | No evidence                 |

These analyses, although crude, would suggest that we were able to increase the policy relevance of our review by expanding the range of study designs we include, without considerably reducing rigour. Having said this, it is based on only directions of effect and not full statistical meta-analysis, and assessing the validity of our approach is not currently possible. The debates as to the importance of randomisation and the appropriateness of trial methodologies in development are still ongoing and whilst these questions remain, we advocate the explicit consideration of different 'cut-off points' for study designs and the inclusion of explicit and transparent synthesis of studies from different study designs within systematic reviews to inform these debates and provide information for decision-making. Others may argue for exclusion of non-experimental (and/or non-quasi-experimental) data, or greater weighting of findings from different study designs. Another approach would be to report findings of a wider range of study designs, as we have in this review, but to make policy recommendations based on only the highest quality studies. In recognition of decisions being made in the absence of evidence, we provide insights from the good quality evidence available to help inform decision-makers, rather than leaving them in a vacuum when the evidence is not of the very highest quality. We therefore argue for the inclusion of this wider range of study designs – an approach of transparent pragmatism rather than purity, or what Clemens and Demombynes (2010) call necessity versus luxury.

### 2.5. *Causal pathway development and testing*

Systematic reviews have tended to report evidence of effectiveness in terms of effect sizes, traditionally summarised in the form of forest plots, so-called black box impact evaluations (White 2011b). There has been less emphasis on the mechanisms of effectiveness, largely due to the consistent nature of the interventions reviewed. As reviews have moved beyond the relatively straightforward assessment of the effectiveness of drugs, to tackle broader and more complex interventions in education and social policy, questions have arisen about *how* interventions work, rather than just *whether* they work. This has led to the development of broader review methodology to include process evaluations and qualitative research. The EPPI-Centre has led these innovations combining evidence about 'what works' with the 'views studies' exploring not only effectiveness but also process (Brunton *et al.* 2005). If these innovations can be translated to development, they should go some way to tackling the central issues of process and context in development interventions.

**2.5.1. *Our approach.*** Aware of the limitations of simply reporting evidence of effectiveness for our review, we decided to take our analysis one stage further and explore the causal pathway of how microfinance works. The importance of this became particularly apparent given our finding that microfinance interventions can do both good and harm to the poor clients they purport to serve (Stewart *et al.* 2010). We therefore reflected upon the theories of change within the literature we had reviewed. We began by mapping out a simple theory of change for microfinance. We then undertook a continuous process of reflection and adaptation, taking into account the theories tested within the included studies, the process data reported (for example, on how people spent their money) and the evidence of effectiveness. As a result we were able to develop a complex flow chart of microfinance, its outcomes and impacts (Stewart *et al.* 2010).

**2.5.2. *Implications and reflections.*** The methodology we employed to develop and test our causal pathway was experimental. The process of trial and error was evidence-informed

but was potentially open to bias. However, through teamwork and persistence, we developed an evidence-informed causal chain which has helped us to move forward in our understanding of how microfinance works to improve the lives of the poor in sub-Saharan Africa. In particular it has helped us to understand and explain both the positive and negative outcomes of microfinance and how they relate to one another. It has proved an effective means to illustrate and explain our findings, and has led to more sophisticated conclusions and recommendations than we would otherwise have been able to make (Korth *et al.* 2012).

There is still scope to develop this methodology. We acknowledge the slightly different approaches taken by 3ie (Waddington *et al.* 2009, King *et al.* 2010) which focus more closely on the process evaluations linked to the trials of effectiveness included in their review, and welcome further opportunities to conduct, discuss and debate approaches to causal pathway development and analysis as more systematic reviews are conducted in development.

### *2.6. Tight deadline*

Last but not least of our pragmatic compromises is the tight time frame allowed for these systematic reviews. It is not uncommon for systematic reviews to take 12 months or more (EPPI-Centre 2011); indeed reviews of 6 months are usually described as Rapid Evidence Assessments rather than considered to be full systematic reviews, and come with an understanding that there will be methodological compromises in order to deliver within the required time frame (GSR 2009). If systematic reviews are to inform policy and practice in development, there is a need for reviewers to be willing to deliver quick reviews for decision-making. Having said this, there is also a need for decision-makers and funders to commission longer-term reviews with scope for broader focus and more in-depth understanding to build up a foundation of evidence to inform longer-term effective decision-making.

*2.6.1. Our approach.* Due to a delay in the funding decision, we were faced with delivering our systematic review in 5–6 months. Others within the same programme of work had similarly tight deadlines of 6 or 7 months. We took the approach that we would deliver as good a product as possible within the allocated timeline, whilst acknowledging that this would be a learning process for us, as well as our funders.

We achieved our goal through a tight timetable. The necessity to plan travel in advance and work around other commitments helped motivate our work. As already highlighted above, some of the compromises we made in adapting systematic review methodology for development enabled us to save time (for example, employing a regional focus), whilst others took more time (for example, in searching a wide range of sources). We reflect below on the impacts of conducting our review in such a short period of time.

*2.6.2. Implications and reflections.* Whilst we were determined to meet our deadline, as much to demonstrate what can be achieved in a short period of time (and what cannot), we did however put in additional hours during the 5 months and which did have implications for other work.

We also took some methodological decisions to allow us to complete the review quickly. We did not have time to collect full texts of all papers, or to chase missing results, leaving some gaps in our review (Stewart 2011). However, it is not unusual for systematic reviews to be unable to collect all relevant papers or not have time to contact authors for missing information.

In addition, rather than have two researchers independently read and extract data from each paper and then compare notes (a process known as double coding), we adopted a pragmatic approach: two researchers started double coding papers and discussing differences until we had achieved high level of inter-reviewer correlation. We then divided the remaining papers between us and only one researcher coded each separately, thus saving time. Had we had more reviewers involved in this task, we could have completed it still quicker. Whilst including more people and sharing the task, rather than all coding all the papers, may introduce errors, we maintained accuracy by working simultaneously and in the same physical space so we could discuss any uncertainties and agree definitions as needed. Furthermore, both researchers read the final smaller pool of included studies. By not fully double coding all papers we did run the risk of inaccuracies but balanced this by working very closely for dedicated periods of time which meant we could maintain the quality of our coding. This close working was enabled by use of technologies such as email, Skype and Dropbox, as well as the specialist systematic review software, EPPI-Reviewer. Also essential was our ability to travel to enable the team to work together in London and Johannesburg.

Delivering our review on time was further enabled by the reduced volume of papers to scan initially due to our regional focus. Worldwide reviews in the same round of DFID-funding have taken longer and have had to limit their scope in other ways, for example, focusing only on microcredit.

Reflecting on whether our rapid approach has reduced technical rigour, we would acknowledge that this might have been the case but only to a minor extent. Instead quick delivery of the evidence on microfinance in sub-Saharan Africa has enabled dissemination to decision-makers, and enabled us to apply for a second development review to further establish the evidence base in this important area of financial inclusion (Stewart *et al.* 2011).

### **3. Lessons for those commissioning, undertaking and using systematic reviews for development**

Reflecting on the pragmatic approaches we have employed in our review, we have identified a number of lessons for those commissioning, undertaking and using systematic reviews for development. Perhaps most importantly, we argue for pragmatism over purity, but with specific steps to ensure that rigour is not overly compromised. If asked to choose technical rigour or policy relevance we choose both.

We have learnt the importance of understanding that systematic reviews are not merely literature reviews, but a specific methodology which requires training and experience to employ. The novices on our team could not have conducted this review without the input of the experienced reviewers. Ad hoc support would not have been sufficient.

We acknowledge that many of the pragmatic compromises we made to ensure delivery of our review are not so different from approaches used for Rapid Evidence Assessments (GSR 2009), and would argue that any reviews delivered in such tight time frames should be renamed Rapid Systematic Reviews. To not do so would mislead the commissioners, reviewers or readers about what full reviews entail. Expanding the time frame of reviews, but not the budget, does not necessarily enable the team to input more time on the review, and perhaps should also be re-labelled in some way to avoid assumptions about the product.

We believe that there are benefits to development in including comparison studies with non-randomised designs. However, we follow the debates on this topic with interest, and are open to learning and developing our ideas. Rather than advocate purity, we recommend

reviewers conduct subgroup analysis by study design to allow the reader to reflect on the differences in the results and to learn from them.

We are convinced that developing causal pathways for the evidence of effectiveness should be encouraged for systematic reviews in the development field. This will contribute not only to theory but also to unpacking complexities and contextualising evidence.

In acknowledgement that development interventions are highly context specific and, where a policy area and regional context demands, we argue for commissioning regionally specific reviews and/or requiring worldwide reviews to conduct subgroup analysis by region or country. We believe this will have a greater impact on policy-makers and funders within regions who seek context-specific evidence to inform their decisions.

Finally, we advocate further expansion of systematic reviews in the field of development. They are feasible and can be useful. Through capacity building, and partnership working with reviewers, decision-makers and review-users in the South, reviews can provide valuable syntheses of research evidence to further our understanding of what works.

### Acknowledgements

The authors thank their host institutions, the University of London and the University of Johannesburg, and their funder, the UK's Department for International Development.

### Notes

1. In this paper we accept the use of RCTs to assess impact, and do not engage here the ontological and epistemological debates regarding the usefulness and relevance of this method of development impact assessment – for such debates see Bamberger *et al.* (2010), Cartwright (2007), Deaton (2009), Ravallion (2009) and White (2011b), as well as the debates in the spring 2010 issue of the *Journal of Economic Perspectives* and the June 2010 issue of the *Journal of Economic Literature*.
2. While the microfinance movement spread late to sub-Saharan Africa, mutual models of monetary help have a long history in Africa; for example, the Susu system originates in the 1900s, and the first credit union in the region was formed in Ghana by Catholic missionaries in 1955 (Nanor 2008).
3. Only around 20 per cent of adults in the region have an account at a formal or semi-formal financial institution (Honohan and Beck 2007). The ratio of private credit to GDP is 18 per cent, while it is 30 per cent in South Asia. For low-income countries in the region it is 11 per cent compared to 21 per cent for low-income countries in the rest of the world (Honohan and Beck 2007). And the diversity of microfinance types – in terms of technology applied, organisational structure, degree of formality and regulation, and clientele – seems to be wider than in other regions (Honohan and Beck 2007).
4. Makina and Malobola (2004) highlight that new developments in impact assessments of microfinance have fostered a greater emphasis on improving practice by monitoring and learning from impact to improve management and design better-fit products, that is, organisational learning and social performance management.

### References

Abadie, A. and Imbens, G., 2009. *Matching on the estimated propensity score*. NBER Working Paper No. w15301. Cambridge, MA: National Bureau of Economic Research.

Algoso, D., 2011. *Limitations of RCTs* [online]. Blog on 1, 3, 13 and 24 April at Find what works. Available from: <http://findwhatworks.wordpress.com/2011/04/01/limitations-of-rcts-politics-and-context/> [Accessed 24 May 2011].

Alkire, S. and Santos, M.E., 2010. *Multidimensional poverty index: 2010 data* [online]. Oxford: Oxford Poverty and Human Development Initiative. Available from: [www.ophi.org.uk/policy/multidimensional-poverty-index/](http://www.ophi.org.uk/policy/multidimensional-poverty-index/) [Accessed 24 May 2011].

Bamberger, M., Rao, W., and Woolcock, M., 2010. *Using mixed methods in monitoring and evaluation: experiences from international development*. Policy Research Working Paper No. 5245. Washington, DC: Development Research Group, World Bank.

Bamberger, M., Rugh, J., and Mabry, L., 2006. *Realworld evaluation: working under budget, time, data and political constraints*. Thousand Oaks, CA: Sage Publications.

Banerjee, A. and Duflo, E., 2011. *Poor economics: a radical rethinking of the way we fight global poverty*. New York, NY: Public Affairs.

Barrett, C. and Carter, M., 2010. The power and pitfalls of experiments in development economics: some non-random reflections. *Applied economic perspectives and policy*, 32 (4), 515–548.

Bendavid, E., 2011. *Those who fail to learn from medical history. Article for small changes, big results forum of the Boston review, March/April* [online]. Available from: [http://www.bostonreview.net/BR36.2/eran\\_bendavid\\_behavioral\\_economics\\_global\\_development.php](http://www.bostonreview.net/BR36.2/eran_bendavid_behavioral_economics_global_development.php) [Accessed 18 June 2011].

Bhargava, A., 2008. Randomised controlled experiments in health and social sciences: some conceptual issues. *Economics and human biology*, 6, 293–298.

Blattman, C., 2011. *Behavioural economics and randomised trials: trumpeted, attacked and parried* [online]. Blog on 24 March at Chris Blattman. Available from: <http://chrisblattman.com/2011/03/24/behavioral-economics-and-randomized-trials-trumpeted-attacked-and-parried/> [Accessed 24 May 2011].

Brunton, V., et al., 2005. Promoting physical activity amongst children outside of physical education classes: a systematic review integrating intervention studies and qualitative studies. *Health education journal*, 64, 323–338.

Buckley, C., 2010. To test housing program, some are denied aid [online]. *The New York Times*, 8 December. Available from: [http://www.nytimes.com/2010/12/09/nyregion/09placebo.html?\\_r=2&hpAlways=&pagewanted=all](http://www.nytimes.com/2010/12/09/nyregion/09placebo.html?_r=2&hpAlways=&pagewanted=all) [Accessed 24 May 2011].

Campbell Collaboration, 2011. [online]. Available from: <http://www.campbellcollaboration.org/> [Accessed 24 May 2011].

Cartwright, N., 2007. Are RCTs the gold standard? *BioSocieties*, 2, 11–20.

Chambers, R., et al., 2009. *Designing impact evaluations: different perspectives*. Working Paper 4. New Delhi: 3ie.

Clemens, M.A. and Demombynes, G., 2010. *When does rigorous impact evaluation make a difference? The case of the Millennium Villages*. CGD Working Paper No. 225. Washington, DC: Centre for Global Development.

Cochrane Collaboration, 2011. [online]. Available from: <http://www.cochrane.org/cochrane-reviews> [Accessed 24 May 2011].

Cochrane Consumer Network, homepage. [online]. Available from: <http://consumers.cochrane.org/> [Accessed 18 June 2011].

Copestake, J., Golberg, N., and Karlan, D., 2009. Randomised control trials are the best way to measure impact of microfinance programmes and improve microfinance product designs. *Enterprise development and microfinance*, 20 (3), 167–176.

Court, J., Hovland, I., and Young, J., eds., 2005. *Bridging research and policy in development: evidence and the change process*. London: ITDG Publishing.

Deaton, A., 2009. Instruments of development: randomisation in the tropics, and the search for the elusive keys to economic development. *Proceedings of the British Academy*, 162, 123–160.

Devarajan, S., 2011. *Can randomised control trials reduce poverty?* [online]. Blog on 23 March at Africa can . . . end poverty. Available from: <http://blogs.worldbank.org/africacan/can-randomized-control-trials-reduce-poverty> [Accessed 24 May 2011].

Duflo, E., Glennerster, R., and Kremer, M., 2004. Randomised evaluations of interventions in social service delivery. *Development Outreach*, March (special report) [online]. Available from: <http://stuff.mit.edu:8001/afs/athena/course/other/d-lab/DlabIII06/random-eval.pdf> [Accessed 18 June 2011].

El-Zoghbi, M. and Martinez, M., 2011. *Measuring changes in clients lives through microfinance: contributions of different approaches*. CGAP Brief, May. Washington, DC: Consultative Group to Assist the Poor.

EPPI-Centre, 2011. [online]. Available from: <http://eppi.ioe.ac.uk/> [Accessed 24 May 2011].

Glennerster, R. and Kremer, M., 2011. Behavioural economics at work in poor countries. Article for Small Changes, Big Results, a forum of the Boston Review, March/April [online]. Available from: [http://www.bostonreview.net/BR36.2/glennerster\\_kremer\\_behavioral\\_economics\\_global\\_development.php](http://www.bostonreview.net/BR36.2/glennerster_kremer_behavioral_economics_global_development.php) [Accessed 18 June 2011].

Goldacre, B., 2011. How can you tell if a policy is working? Run a trial [online]. *The Guardian*, 14 May. Available from: <http://www.guardian.co.uk/commentisfree/2011/may/14/bad-science-ben-goldacre-randomised-trials> [Accessed 18 June 2011].

GSR, 2009. *Rapid evidence assessment toolkit* [online]. London: Government Social Research Service. Available from: <http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment/how-to-do-a-reo> [Accessed 24 May 2011].

Haddad, L., 2011. *A new harvest of RCTs?* [online]. Blog on 15 May at Development horizons. Available from: <http://www.developmenthorizons.com/2011/05/new-harvest-of-rcts.html> [Accessed 24 May 2011].

Harden, A., 2001. Finding research evidence: systematic searching. In: S. Oliver and G. Peersman, eds. *Using research for effective health promotion*. Buckingham: Open University Press, 47–68.

Honohan, P. and Beck, T., 2007. *Making finance work for Africa*. Washington, DC: World Bank.

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

Jones, H., 2009. *The 'gold standard' is not a silver bullet for evaluation*. ODI Opinion 127. London: Overseas Development Institute.

Karlan, D. and Apple, J., 2011. *More than good intentions: how a new economics is helping to solve global poverty*. New York: Penguin.

Kermeliotis, T., 2011. *Experts warn Africa must learn from India's microfinance problems* [online]. CNN Marketplace Africa. Available from: <http://edition.cnn.com/2011/BUSINESS/03/23/microfinance.africa.lessons/index.html> [Accessed 23 March 2011].

King, E., Samii, C., and Snistveit, B., 2010. *Interventions to promote social cohesion in sub-Saharan Africa* (Synthetic review 002). New Delhi: 3ie.

Korth, M., et al., 2012. Microfinance: development intervention or just another bank? *Journal of agrarian change*, forthcoming.

Kristof, N.D., 2011. Getting smart on aid [online]. *The New York Times*, 18 May. Available from: [http://www.nytimes.com/2011/05/19/opinion/19kristof.html?\\_r=1&src=tptw](http://www.nytimes.com/2011/05/19/opinion/19kristof.html?_r=1&src=tptw) [Accessed 18 June 2011].

Lindley, H., 2011. *Monique and the mango rains: qualitative research in action* [online]. Blog on 10 May at Pas à pas. Available from: <http://pasapasblog.wordpress.com/2011/05/10/monique-and-the-mango-rains-qualitative-research-in-action/#more-189> [Accessed 24 May 2011].

Makina, D. and Malobola, L.M., 2004. Impact assessment of microfinance programmes, including lessons from Khula Enterprise Finance. *Development Southern Africa*, 21 (5), 799–814.

Morduch, J., 2011. *Is microfinance research as bad as medical research?* [online]. Blog on 5 January at CGAP Microfinance Blog. Available from: <http://microfinance.cgap.org/2011/01/05/is-microfinance-research-as-bad-as-medical-research/> [Accessed 24 May 2011].

Nanor, M.A., 2008. *Microfinance and its impact on selected districts in eastern region of Ghana*. Master of Philosophy Dissertation. Kwame Nkrumah University of Science and Technology.

Odell, K., 2010. *Measuring the impact of microfinance: taking another look*. Washington, DC: Grameen Foundation.

Ravallion, M., 2009. Should the randomistas rule? *The economists' voice*, 6 (2), Article 6.

Roodman, D., 2011. *Literature review on microfinance impacts in Africa* [online]. Blog on 9 February at David Roodman's microfinance open book blog. Available from: [http://blogs.cgdev.org/open\\_book/2011/02/literature-review-on-microfinance-impacts-in-africa.php](http://blogs.cgdev.org/open_book/2011/02/literature-review-on-microfinance-impacts-in-africa.php) [Accessed 24 May 2011].

Stewart, R., 2011. *What is being excluded in systematic reviews?* [online]. Blog on 1 April at Ruth Stewart. Available from: <http://www.ruthstewartresearch.org/what-is-being-excluded-in-systematic-reviews/> [Accessed 1 February 2012].

Stewart, R., et al., 2010. *What is the impact of microfinance on poor people? A systematic review of evidence from sub-Saharan Africa*. Technical Report. London: EPPI-Centre, Social Science Research Unit, University of London.

Subramanian, A., 2011. *Nicholas Kristof and aid* [online]. Blog post on 20 May at Global Development: Views from the Center. Available from: <http://blogs.cgdev.org/globaldevelopment/2011/05/nicholas-kristof-and-aid.php> [Accessed 24 May 2011].

Sumner, A., 2010. *Global poverty and the new bottom billion: what if three-quarters of the world's poor live in middle-income countries?* Working Paper. Brighton: Institute for Development Studies.

Sutcliffe, S. and Court, J., 2005. *Evidence-based policymaking: what is it? How does it work? What relevance for developing countries?* London: Overseas Development Institute.

van Rooyen, C., Stewart, R., and de Wet, T., 2012. *The impact of microfinance in sub-Saharan Africa: a systematic review of the evidence* [online]. World Development. Available from: <http://www.sciencedirect.com/science/article/pii/S0305750X12000496> [Accessed 12 March 2012].

Waddington, H., et al., 2009. *Water, sanitation and hygiene interventions to combat childhood diarrhoea in developing countries* (Synthetic review 001). New Delhi: 3ie.

White, H., 2011a. *The 3ie policy window* [online]. Presented at the 5th NONIE meeting, Paris, March 2011. Available from: <http://nonie2011.org/?q=content/conference-programme> [Accessed 24 May 2011].

White, H., 2011b. *An introduction to the use of randomised control trials to evaluate development interventions*. Working Paper No. 9. New Delhi: 3ie.