Supplementary online appendix – Quantitative synthesis of included meta-analyses

This online appendix provides a quantitative synthesis of the included meta-analyses to supplement the results set out in the main paper.

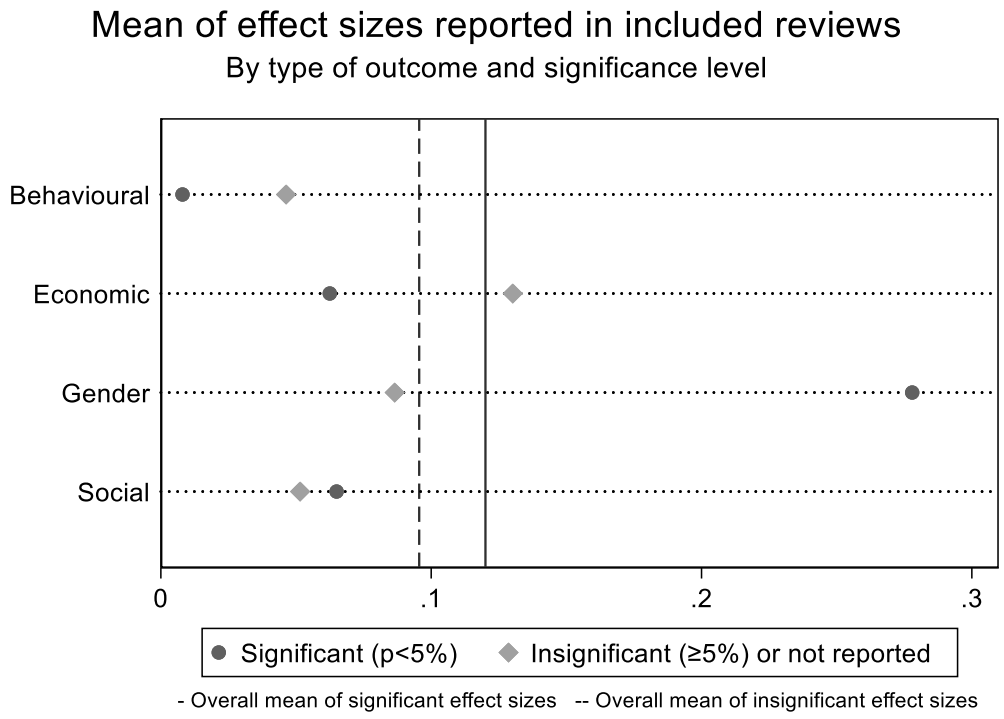
Due to high levels of heterogeneity within our evidence base and the small sample of studies, performing network meta-analysis or drawing on a meta-regression approach was not possible; nevertheless, we adopt the reporting guidelines created by Stanley et al (2013) where appropriate[1]. As discussed in section 5 of the main paper, a simplistic reading of the results summaries in Table A3 and Figure 3 suggests an overall positive, if mixed, set of findings. However, we argue that the positive evidence needs to be considered with caution in light of the low quality of the underlying primary evidence base. Furthermore, and more notably, the effect sizes overall are quite small and do not plausibly indicate transformative changes. We explore this notion further in this appendix and present the mean effect sizes reported across the different outcome categories in five of the 11 meta-studies that adopted a quantitative meta-analytical approach.

Figure A1 is a good starting point for allowing us to investigate the mean effect sizes by broad outcome category and level of significance. 76 effect sizes have been extracted from five meta-analyses, covering 15 different outcomes across a wide range of intervention types (see notes section below Figure A1 for more details). We can see that, for each of the broad outcome categories, the effect sizes are positive but very small, ranging from 0.040 for behavioural effects that are statistically not significant to 0.140 for economic effects that are also statistically not significant. They range from 0.010 for behavioural effects that are statistically significant to 0.280 for

[1] We adopt all points of the guidelines except the ones that relate to meta-regression modelling, as we did not conduct a meta-regression for reasons set out above. The guidelines were subsequently updated in 2019 (see here: https://www.maer-net.org/post/revision-of-reporting-guidelines) to include 12 recommendations on weights, outliers, reconstructed standard errors, etc. Given our unique position of reporting a meta-meta-analysis, we felt that many of these issues would not apply, as we were mainly concerned with resolving more fundamental issues such as dealing with the high levels of heterogeneity and the small sample of studies. As we are now moving into novel methodological territory in relation to meta-analyses, we refer the interested reader to recent prominent examples pushing methodological frontiers in meta-analysis research, e.g. Stanley (2017), Stanley and Doucouliagos (2017), Carter et al (2019).

gender effects that are statistically significant. The overall mean of significant effect sizes is 0.120, with the value for the insignificant effects being slightly lower. This means that, on average, an individual or household reached by financial inclusion initiatives will be slightly advantaged compared to one that is not participating. These figures, however, should also be treated with some caution, as they are highly heterogeneous and compare and aggregate effects on very different measures, giving rise to the 'apples and oranges' problem (Lipsey and Wilson 2001).

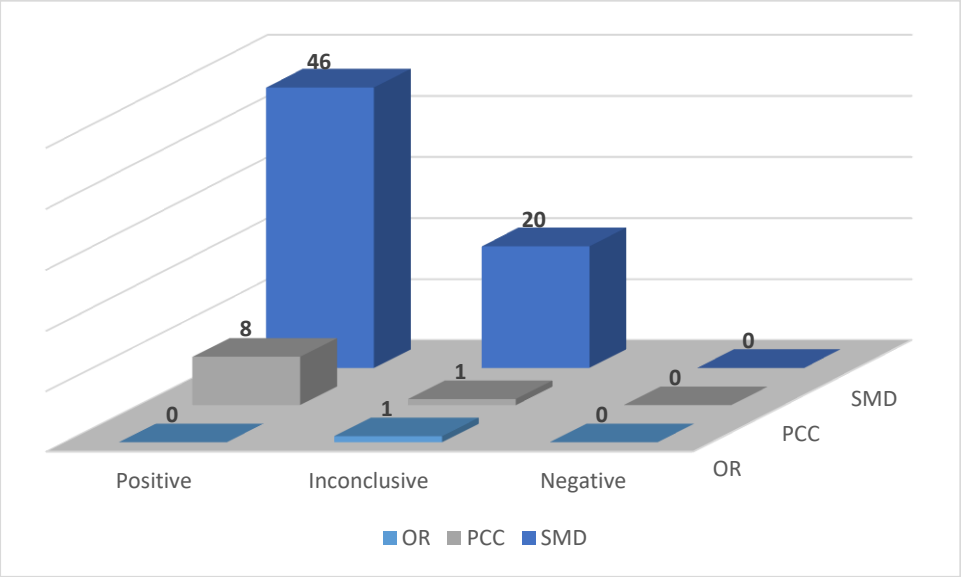
Figure A1: Mean effect sizes reported in 5 medium-/high-confidence meta-analyses by broad outcome category and level of significance



Note: The figure above is based on 76 effect sizes (66 standardised mean differences, 9 partial correlation coefficients and 1 odds ratio) measured from 15 different outcomes categorised into 4 broad outcome themes. 29 of the effect sizes relate to gender outcomes, followed by 21 for economic outcomes (11 for consumption and 10 for savings amount). We have taken the average of the raw effect sizes as observed from the systematic reviews. These come from 5 meta-analyses (Brody et al. 2015 Vaessen et al. 2014, Steinert et al. 2018, Chliova et al. 2015, and Gopalaswamy et al. 2016) covering a wide range of intervention types: microcredit, microsavings, CBSGs and microfinance more broadly.

A related figure is Figure A2, where we break down the number of effect sizes by direction and type of effect size, further supporting the notion that positive effects dominate.

Figure A2: Breakdown of the number of effect sizes by direction and type of effect size



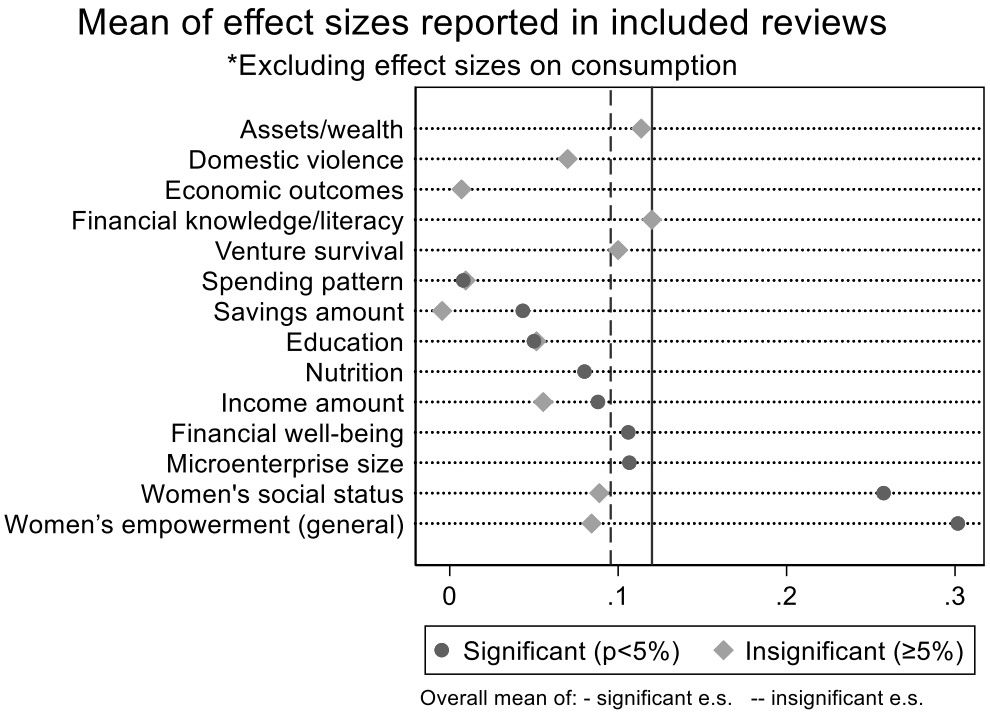
Notes: OR = Odds ratio, PCC = Partial correlation coefficient, SMD = Standardised mean difference. The 46 positive SMDs are reported by 3 of the meta-analyses (Brody et al. 2015, Steinert et al. 2018, Gopalaswamy et al. 2016), the 20 inconclusive SMDs are reported by 4 of the studies (Brody et al. 2015, Steinert et al. 2018, Gopalaswamy et al. 2016, and with Vaessen et al. 2014 accounting for 9 of these effect sizes). The 8 positive and the 1 inconclusive PCCs are reported by 1 study (Chliova et al. 2015), the 1 inconclusive OR is reported by 1 study (Steinert et al. 2018).

We should note, however, that many of the effects we found are strongly heterogeneous, both across studies and over time, places, populations, and between interventions. There may also be an issue with small sample bias. Slavin and Smith (2009) and others (e.g. Kjaergard et al. 2001) suggest that reviews with small sample sizes ($n < 100$) tend to report larger, more positive effect sizes than reviews with larger sample sizes ($n > 100$), and that they are often of lower methodological quality. In the case of our medium- and high-confidence reviews, the sample of primary studies they included range from 12 to 90, positioning our reviews in the small-sample category. 36% of the 11 medium- and high-confidence studies also voice concerns about the limited quantity of evidence they included.

In addition, we should flag that positive findings tend not to repeat from one context to another. With reference to the financial inclusion theory of change (presented in section 2 'theoretical background'), most of the positive impact estimates are for outcomes that are early along the causal

chain, such as in health-focused meta-studies which find changes in health knowledge, but not in health outcomes, or meta-studies looking at enterprise activity which find growth in business ventures run by households but not in household incomes. An exception appears to be for savings, where both immediate outcomes *and* wider poverty measures are affected in a positive but relatively small way, as suggested by Figure A3, which breaks down Figure A1 by sub-type of outcome – note the positive and significant effect for savings amount; the second-smallest significant finding.

Figure A3: Mean effect sizes reported in 5 medium/high confidence meta-analyses by sub-type of outcome and level of significance



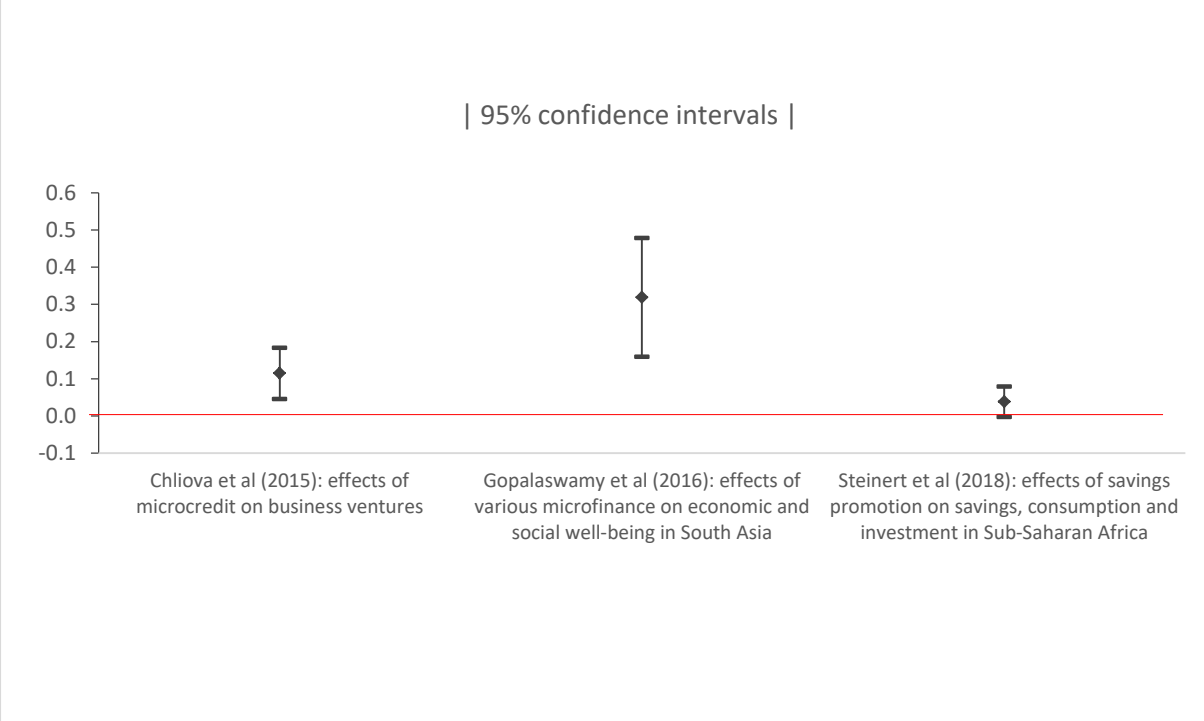
Note: The figure excludes the mean effect size for consumption, as it is an (insignificant) outlier.

In the following, more detailed, discussion of results, we cluster the findings of studies for four outcome categories – economic, social, behavioural and gender outcomes – and relate the findings to different financial intervention-types (as applicable) with reference to the theory of change presented in section 2.

Economic outcomes

Figure A4 summarises the quantitative evidence on economic outcomes complementing the analysis presented in section 5 'results'. Overall, the effects of financial inclusion interventions, particularly microcredit and combined/mixed microcredit-microsavings initiatives, on economic outcomes such as income or assets are positive but inconsistent and not particularly large, also judging by the few effect sizes presented in Figure A4, which range from 0.0387 to 0.3185 (mean values, mean CI 95% -0.0028 to 0.4783). We should note that the figure below is based on 35 effect sizes from 3 studies across 8 sub-types of outcomes and all broad intervention types, indicating high levels of heterogeneity.

Figure A4: Estimates of the impact of financial services on economic outcomes by meta-studies adopting a meta-analytical approach



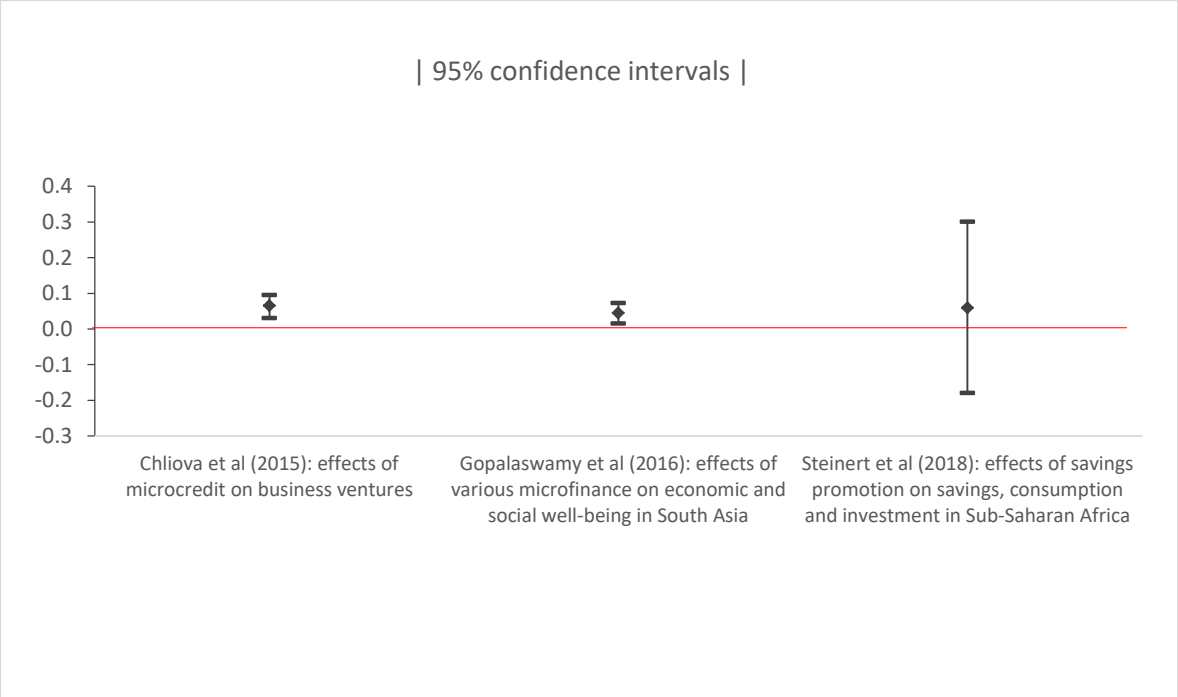
Notes: This figure is based on 35 number of effect sizes for 8 sub-types of outcomes (e.g. assets/wealth, consumption, financial well-being, income, savings amount, broad economic outcomes, microenterprise size and venture survival). The average effect size over the review is taken including both statistically significant and not significant values (due to the small number of highly heterogeneous effect sizes we decided not to disaggregate them further by statistical significance).

Credit and other financial services delivered through microfinance programming appear to have overall positive but decidedly mixed impacts, in terms of both lower- and higher-order outcomes. The picture for microsavings looks more hopeful, suggesting small but more consistently positive effects, especially on savings accumulation and incomes (and not on non-financial asset accumulation), and with fewer downsides for clients compared to credit. Having said that, Stewart et al. (2012) indicate that microsavings access does not enable the poor to engage in economic opportunities, but they also support the view that in some cases an increase in income, savings, expenditures and the accumulation of non-financial assets is observable.

Social outcomes

Summarising the effects for social outcomes and comparing them to those for economic outcomes, it appears that the effects for social outcomes are even smaller, and even more mixed, see Figure A5 below.

Figure A5: Estimates of the impact of financial services on social outcomes by meta-studies adopting a meta-analytical approach



Notes: This figure is based on 4 effect sizes for 2 sub-types of outcomes (e.g. nutrition and education). The average effect size over the review is taken including both statistically significant and not significant values

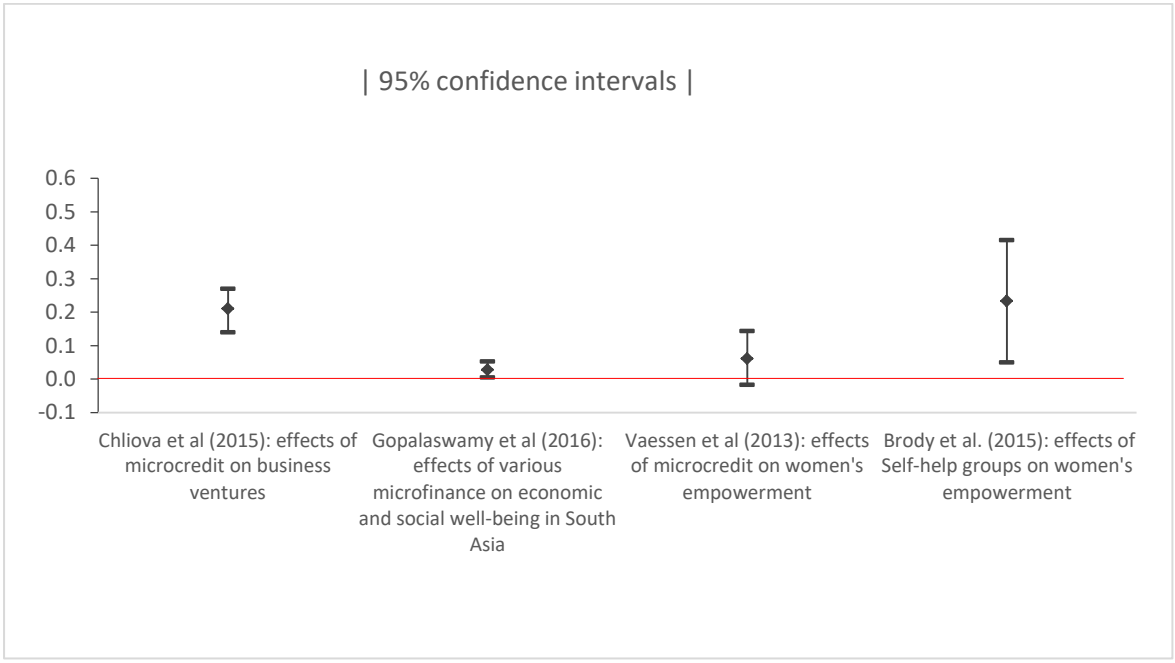
(due to the small number of highly heterogeneous effect sizes we decided not to disaggregate them further by statistical significance).

This finding is supported by the limited quantitative evidence we found, which suggests positive but very small effects (Figure A5), ranging from 0.0440 to 0.0650 (mean values, mean CI 95% -0.1800 to 0.3000). We should note that the figure is based on only 4 effect sizes from 3 studies across 2 sub-types of outcomes. These findings hold across all financial inclusion intervention types and across all geographical focal areas.

Gender outcomes

Summarising the effects of financial inclusion interventions on women's empowerment, they appear to be positive on the whole, albeit relatively small; a view which is also supported by the quantitative evidence (Figure A6) with mean effect sizes ranging from 0.0280 to 0.2338 (mean CI 95% from -0.0170 to 0.4156), but these effects are based on only 4 studies reporting a total of 29 effect sizes from 3 sub-types of outcomes.

Figure A6: Estimates of the impact of financial services on gender outcomes by meta-studies adopting a meta-analytical approach



Notes: This figure is based on 29 number of effect sizes for 3 sub-types of outcomes (e.g. women's empowerment (general), women's social status and domestic violence). The average effect size over the review is taken including both statistically significant and not significant values (due to the small number of highly heterogeneous effect sizes we decided not to disaggregate them further by statistical significance).

The effects heavily depend on how gender outcomes are conceptualised and operationalised in the underlying primary evidence base, where studies often use different indictors or index variables. The effects also depend on programmatic features of the interventions, with several meta-studies raising the question to what extent financial services themselves, rather than other programme elements, such as exposure to women's rights, awareness-raising, or efforts at group-building and social networking (which may also be delivered independently from any financial intervention) explain the effects. The effects of specifically gender-targeted programme elements were larger than those of the actual financial service (Chliova et al. 2015; Peters et al. 2016). The main enablers of empowerment effects appear to be group interactions, opportunities to leave the house, and exposure to additional rights-related training, rather than financial services.

Behavioural outcomes

Behaviour-changes could be enablers of more transformative changes. However, the meta-analyses in our study sample synthesised too few of the behaviour-related effects to warrant a quantitative analysis of these effects.

In summary, looking across the meta-analytical studies, almost all effect sizes are quite small – based on a small sample of meta-analyses (n=5) capturing only 76 effect sizes across 15 very diverse outcomes – and are hardly indicative of transformative changes from financial inclusion, as dominantly lower-order outcomes are affected. Many effects are strongly heterogeneous, both across studies and over time, places, populations, gender, and ethnicity as well as between interventions[2]; this suggests them to be unreliable and/or context-dependent. Positive findings tend

[2] We used an adapted PROGRESS checklist (O'Neill et al. 2014) to identify these factors that seem to drive heterogeneity in the financial inclusion context. However, it was very difficult to further unpack these drivers of heterogeneity as the reviews we included did not provide further disaggregated information, they rather

not to repeat from one context, intervention-type or study to another, and at least as many findings are mixed or inconclusive as are positive. Consequently, the positive results found for financial inclusion are fragile, and need to be treated with caution. Given the small sample of high confidence studies included in our review, the next section examines whether any meaningful patterns in outcome reporting can be detected by comparing them to the low confidence evidence we excluded.

Differences between high- and low-quality systematic review studies in terms of outcome reporting

We initially identified 32 eligible meta-studies (systematic reviews and meta-analyses) examining the impact of financial inclusion interventions on a range of economic, social, gender and behavioural outcomes. After subjecting these to a quality appraisal process (discussed in section 4 in the main paper), we excluded 21 reviews due to quality concerns, leaving a core sample of 11 medium- and high-confidence meta-studies. Our analysis focused on the medium- and high quality studies but we wish to note a few key differences between the studies we included and those we excluded.

For the 21 excluded studies, due to the low confidence in their findings, we do not include the directions of reported outcomes in our synthesis. However, we would suggest that knowing the patterns of outcome reporting in these other studies can be useful for the design of future, higher-confidence meta-studies that complement the existing medium- and high-quality evidence base. We note that the picture regarding types of outcomes reported is not very different for these low confidence studies, as we can see from Table A1; again there is an emphasis on economic outcomes and a relative paucity of reporting of social and behavioural outcomes. However, insurance and community-based saving groups (CBSGs) feature more strongly as modes of service delivery among

created broad categories lumping together a range of diverse outcomes and intervention types as a way to deal with high levels of heterogeneity.

the low confidence studies.[3] A similar share of the effects were reported for "mixed" microfinance as among the included medium- and high-confidence studies.

[3] These contain four studies of insurance, two of which we were surprised to have had to exclude following our formal quality assessment criteria.

Table A1: Number of outcomes reported and interventions for 11 medium and high confidence meta-studies and 21 low confidence meta-studies

Type of outcome	Microcredit	Microinsurance	Micro savings	CBSGs	Mixed microfinance
11 medium and high confidence meta-studies:					
Economic	32	9	33	9	19
Social	15	3	9	6	7
Gender	10	3	4	12	6
Behavioural	4	0	4	4	4
21 low confidence meta-studies:					
Economic	9	32	8	32	22
Social	1	13	7	10	9
Gender	7	14	13	5	11
Behavioural	4	8	4	4	4

It is important to note, however, that the evidence base for both the low and medium/high confidence studies is highly heterogeneous in terms of focusing on different intervention types, outcomes and geographies. As with the 11 medium and high confidence studies, many of the effects we find being reported are positive, but often are very small and occurring early on in the causal chain, which, if these meta-studies had a higher confidence level, would similarly suggest a lack of long-lasting and transformative changes.

Conclusion

The quantitative synthesis supports the findings presented in the narrative synthesis in the main paper, i.e. the results reported across the medium- and high- quality meta-studies raise the question whether financial inclusion interventions are supported by sufficiently strong evidence for having transformative impacts. We have raised quality concerns in relation to the meta-study evidence base, and provided a brief comparison of low- and high- quality studies in terms of their outcome reporting arguing that detecting patterns in this regard may prove useful for the design of future high-quality studies. We also discussed the implications of small sample bias, which further caution

the reader to place too much faith in the small and positive effects that we can report from reviewing these meta-studies.

References

Brody, C., De Hoop, T., Vojtkova, M., Warnock, R., Dunbar, M., Murthy, P., and Dworkin, S. L., 2015. Economic self-help group programmes for improving women's empowerment: a systematic review, 3ie Systematic Review 23. London: International Initiative for Impact Evaluation (3ie).

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science, 2*(2), 115-144.

Chliova, M., Brinckmann, J., and Rosenbusch, N., 2015. Is microcredit a blessing for the poor? A meta-analysis examining development outcomes and contextual considerations. *Journal of Business Venturing*, *30*(3), 467-487.

Gopalaswamy, A. K., Babu, M. S., and Dash, U., 2016. Systematic review of quantitative evidence on the impact of microfinance on the poor in South Asia. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University College London. Retrieved from http://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/Microfiannce-ST%20-%20IITM2.pdf?ver=2016-05-04-143152-407.

Lipsey, M. W., and Wilson, D. B., 2001. Practical Meta-Analysis, Applied Social Research Methods, Sage Publications, London.

Kjaergard, L. L., Villumsen, J., and Gluud, C., 2001. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Annals of internal medicine*, *135*(11), 982-989.

O'Neill, J., Tabish, H., Welch, V., Petticrew, M., Pottie, K., Clarke, M., ... and Tugwell, P., 2014. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. *Journal of clinical epidemiology, 67(1),* 56-64.

Peters, C. J., Picardy, J. A., Darrouzet-Nardi, A., and Griffin, T. S., 2014. Feed conversions, ration compositions, and land use efficiencies of major livestock products in US agricultural systems. *Agricultural Systems*, *130*, 35-43.

Slavin, R., and Smith, D., 2009. The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational evaluation and policy analysis*, *31*(4), 500-506.

Stanley, T. D. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychological and Personality Science, 8*(5), 581-591.

Stanley, T. D., & Doucouliagos, H. (2017). Neither fixed nor random: Weighted least squares meta-regression. *Research synthesis methods, 8*(1), 19-42.

Stanley, T. D., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., ... and Rosenberger, R. S., 2013. Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys*, *27*(2), 390-394.

Steinert, J. I., Zenker, J., Filipiak, U., Movsisyan, A., Cluver, L. D., and Shenderovich, Y., 2018. Do saving promotion interventions increase household savings, consumption, and investments in Sub-Saharan Africa? A systematic review and meta-analysis. *World Development*, *104*, 238-256.

Stewart R., van Rooyen, C., Korth, M., Chereni, A., Rebelo Da Silva, N., and de Wet, T., 2012. Do micro-credit, micro-savings and micro-leasing serve as effective financial inclusion interventions enabling poor people, and especially women, to engage in meaningful economic opportunities in low- and middle-income countries? A systematic review of the evidence. Technical report, EPPI-Centre, Social Science Research Unit, University of London.

Vaessen, J., Rivas, A., Duvendack, M., Palmer Jones, R., Leeuw, F., Van Gils, G., Lukach, R., Holvoet, N., and Bastiaensen, J., 2014. The Effects of microcredit on women's control over household

spending in developing countries. Campbell Systematic Review. Retrieved from http://www.campbellcollaboration.org/lib/project/178/.