The Effects of Universal Basic Income on Labour Force Participation - A Meta-Analysis

Jorge de la Cal Medina Philipp Pag Viviane M. Meyer

December 10, 2021

1 Introduction

Universal Basic Income (UBI) is a periodic cash payment that is unconditionally delivered to all citizens individually by the state (Gentilini et al., 2020). The debate about this policy has been around for some time (Friedman, 1967), but in view of many state's lockdown measures against the current coronavirus pandemic and the need for states to support its citizens financially, it may never have been as relevant as it is today.

Among the potential benefits are the reduction in income inequality and the promise to leverage many out of poverty as well as creating a more inclusive labour market and improving gender equity (Gentilini et al., 2020). However, a voiced criticism, which, if revealed to be true, could defeat UBI's whole purpose, is that it would disincentivise work (Gibson et al., 2020a). In order to conduct an evidence-based inquiry on this matter, we conduct a meta-analysis with existing experimental data of UBI-like policies on labour market participation (LMP).

Our meta-analysis indicates that an UBI might indeed have a negative impact on the labor supply of males and females.

The problem with experimental evidence on UBI is that, as of today, it has not been implemented yet, at least not long term.[1] Evidence is scarce and scattered. Although there have been several pilots which implement policies mimicking certain aspects of UBI, almost none satisfy the definition

[1]Mongolia and the Islamic Republic of Iran implemented what has come nearest to a UBI for short periods of time (Salehi-Isfahani and Mostafavi-Dehzooei, 2018; Doojav and Bayarjargal).

of UBI as given above. For this reason the result of the meta-analysis has to be enjoyed with caution. A fundamental assumption of meta-analyses is that the effects are random expressions of the same distribution. However, since the best available evidence shows a high degree of heterogeneity, due to the different types of interventions and general surroundings, this assumption is difficult to uphold. Nonetheless, we believe that it is possible to abstract from this heterogeneity across studies and make a cautious prediction. Our hope is that further research along these lines will be provided with more and more homogeneous studies to reach more conclusive and robust results.

The paper unfolds as follows. First we present the data consisting of the estimates gathered for the meta-analysis. We also rate its quality after the standardized framework GRADE. Second, we conduct the meta-analysis. Further, to find possible drivers of the heterogeneity in our analysis, we conduct two subgroup analyses. Moreover, we run a meta regression to investigate whether possible combinations of between-study differences could better explain the observed heterogeneity. We finally conclude discussing our results and putting them into perspective to our data.

2 Data

2.1 Search Strategy

We develop a systematic search procedure that can capture any relevant and workable estimates of the interventional effect of UBI on labour market participation that have been produced. For this purpose we first search for appropriate UBI interventions conducted. For these interventions we then search for studies analyzing the interventional effects (see Figure 1).

2.1.1 Inclusion criteria for interventions

First we identify any relevant interventional programs that involve policies similar to UBI. Here we follow the approach of Gentilini et al. (2020). If we were to impose all four criteria for UBI, namely universality, unconditionality, cash-basedness, and state-providedness, we would end up with only two interventions. For that reason we include other programs and pilots that feature intervention designs similar to UBI, only satisfying some of the criteria. This results in a sample of 25 different interventions.

2.1.2 Search and exclusion criteria for studies

Subsequently we search for all studies that analyze the effects of those interventions. For this we use the search engine Google Scholar, as well as literature databases such as EconLit. We also use existing scope reviews (Gentilini et al., 2020; Gibson et al., 2020b) to gather relevant studies on the above defined interventions. This results in a starting sample of 37 studies. Next, we do not consider studies that do not feature a quantitative analysis, and from the remaining sample we again exclude studies that do not analyze effects on individual (male and female) labour market participation. In total we exclude 22 studies, resulting in a final sample of 14 studies, which analyse the effects of 9 interventions in total (see Table 5).

2.2 Sample Overview

The final set of selected studies analyzes the interventions shown in Table 1. In the following, we briefly discuss those interventions on which the meta-analysis is ultimately based. It becomes apparent that the interventions, although having in common certain characteristics which they also share with UBI, in other respects are quite different.

2.2.1 Negative Income Tax

Negative income tax is equivalent to a subsidy for people under a certain income threshold. For this reason it does not satisfy unconditionality. The U.S. government conducted four important experiments between 1968 and 1982 in different parts of the USA (Institute for Research on Poverty, 1973, 1976; Kehrer et al., 1979; Christophersen et al., 1983; Philip K. Robins, 1985).

2.2.2 Social Wealth Fund

Alaska used revenues from the large oil reserves on state-owned land to finance annual cash-transfers for all its residents (Feinberg and Kuehn, 2018; Guettabi et al., 2019).

2.2.3 Cash Transfer

There exist different money transfer interventions which vary greatly in their design, the amount of the benefit and the target population characteristics, although usually they are designed as a safety-net against poverty.

Kenya's money transfer program targeted households with orphans and vulnerable children, providing monthly cash transfers to promote child care and support their human capital development (Asfaw et al., 2014). The Ecuadorian program - targeting low-income mothers - was originally conceived as a conditional cash transfer; preventive health checks or a minimum level of school attendance for children were requirements. Now, due to the lack of verification by the authorities, the program is more similar to an unconditional cash transfer (Mideros and O'Donoghue, 2015). Pakistan's cash transfer program aimed at poor families, where payments are made to the female head of the family, but with rather strict eligibility requirements (Ambler and Brauw, 2019).

2.2.4 Tribal Dividends

Native American "Nations" implemented cash-transfers financed by casino profits to tribe members. The casino dividend program began in 1996 and distributes the profits to all tribal members (Akee et al., 2010).

2.2.5 Heterogeneity

Our sample of interventions covers a fairly wide range of different experiments, varying in scope, implementation and purpose. In the selection of these we have aimed at capturing all workable effects of UBI on LMP. This, however, results in a strong heterogeneity among interventions. This heterogeneity might compromise the validity of the meta-analysis' result. However, the goal of it is to examine the entire available, workable research on the question at stake. A more restricting selection would have made any meta-analysis impossible as a whole, since the number of studies wold have been insufficient.

2.3 Assessment of Study Quality

In order to assess the quality of the studies taken into consideration, we utilised the GRADE framework (Guyatt et al., 2008) which evaluates studies in terms of certain criteria and defines an aggregating quality measure between 1 and 4 points (see Table 2). We adopted it to our purpose which is particular in that a meta-analysis already treats several of GRADE's evaluation criteria, namely publication bias, imprecision and inconsistency. Further, since in all interventions we suspect lurking confounders related to the choice of population or time (or both), we downgrade all studies by one point. To the confidence in the overall result of our meta-analysis we assign the lowest score of any included study, as advised by the authors of GRADE.

2.3.1 Risk of bias

Risk of bias refers to the internal validity of the studies estimates. We rate a study's confidence down if any of the following is the case (Guyatt et al., 2011).

- The study does not define and apply appropriate eligibility criteria, i.e. treatment and control group come from different populations.
- The outcome is measured differently in the treatment and the control group.
- The study does not adequately control for confounding, i.e. does not measure all known explanatory variables factors.

2.3.2 Indirectness

Evidence is most certain when studies directly compare the interventions of interest in the population of interest, and report the outcome critical for decision-making. We downgrade the quality of evidence according if the population, intervention, or outcomes from relevant studies differ from those in which we are interested in our meta-analysis (Guyatt et al., 2011).

2.3.3 Rating up the evidence

In the following cases we rate up the evidence (Guyatt et al., 2011):

- When a large magnitude of effect exists.
- When all plausible confounders or other biases increase our confidence in the estimated effect, i.e. possible, unaccounted confounders and biases would result in an underestimate of an apparent treatment effect.

3 Analysis

3.1 Methodology

As already mentioned, the interventions we compare in this meta-analysis are quite heterogeneous. This is also the case for the methodology and the measurement of effects of the studies. They all measure the effect of UBI on LMP, but they do this in different units. Some studies measure the effect on the labor force participation rate (i.e. the percentage point difference induced by the UBI), others measure the effect on hours or days worked (either weekly, monthly or annually). So, in order to be directly comparable with each other, the results of the different studies need to be standardized. To do so, we transform the different effects into their standardized mean difference (SMD), which is calculated by dividing the mean effect of the papers by their standard deviation.

However, for this transformation we need the standard deviations of all studies. Unfortunately, some of them are rather old and don't exhibit standard deviations, although some of them report significance levels. To impute standard deviations for the studies where they are missing, we proceed in the following way, implementing the appropriate measures advised by Wiebe et al. (2006) to impute missing standard deviations in meta analyses:

1. For studies, where a significance level is available (which incidentally is the case for the significant effects), we compute the lower bound of standard deviation imputing the t-value based on the mean effect of the study, the significance and the number of observations.

2. For studies, where 1. is not applicable, but for which we have a study in our sample which is comparable (e.g. investigates the same experiment/similar intervention) and exhibits standard deviations, we calculate the so called coefficient of variation (CV) of the comparison study as follows: $CV = \frac{SD}{TE} * 100$
The CV is then used to calculate an approximate standard deviation for the other study: $SD = \frac{TE_{missing.SD}}{100} * CV$, where TE is the treatment effect of a study with standard deviation and $TE_{missing.SD}$ is the treatment effect of the study with a missing standard deviation.

3. For the remaining studies, where there is no significance level, which means the effect is insignificant, and no from our point of view directly

comparable study, we assume a p-value of 0.11, which is the upper bound of insignificant effects, and calculate the approximate standard deviations in the same way as in 1.

After having imputed all the missing standard deviations and computed the SMDs of all the studies, we have all the necessary inputs to execute our meta analysis. As most of the studies report the effects of UBI on labor supply separately for males and females we carry out separate meta-analyses for their respective effects. Some studies additionally report the effect on the household level or specifically for female family heads. This sex-based differentiaton is sensible in our eyes, since the estimated effects for males and females in the various studies we considered often differ greatly, hinting that they are not directly comparable. Moreover, we did not aggregate the effects to the household level, as only few studies report effects on the household level, which would make the calculated weighted averages quite imprecise for most of the effects. For these reasons we only consider in our study sample studies which report effects for males and females separately.

Finally, expecting a relatively high heterogeneity of our results, due to the large heterogeneity of the studies as it is explained above, we further perform a subgroup analysis. Here we group by developed and developing countries, as well as by intervention types. Lastly, we run meta regressions with multiple regressors to test whether a combination of the differences between the studies could explain part of the overall heterogeneity.

3.2 Results

3.2.1 Summary Statistics

Table 3 summarizes the variables of the two study subsets for which we perform our meta analyses for male and female effects. Either subsets includes 10 studies (we exclude the Study by Philip K. Robins (1985) as this would lead to the NIT studies being over weighted, because each of the NIT study results would otherwise appear twice with same SMD). The oldest intervention which is investigated by these studies started in 1968 while the latest started only in 2012, which illustrates the large differences between the studies in regard to the time of the interventions.

For the male subset, the number of participants per study ranges across studies from 264 to 12’950’193 while for our female subset it ranges from 264 to 7'000'428. This reveals the large differences in the size of the interventions. Besides this, for the male subset the SMD ranges from -57.2 to 1.2, while for the female subset it ranges from -39.3 to 0.9. Thus, there is also a large difference with respect to the size and the sign of the analyzed effects among studies.

3.2.2 Meta Analysis

We conduct two separate meta analyses for effects on LMP for male and female, respectively. Figure 2 shows the forest plot for the male subsample while Figure 3 shows the forest plot for the female subsample. One can see that in both subsamples, two of the studies have the same SMD (-1.60). This is due to the method how we imputed the standard deviations of these studies (see 3.1). In both of these studies there was no significant effect and therefore no indicated significance level from which we could have computed a lower bound of the standard deviation. Also, unfortunately we did not have a directly comparable study, where we could draw a CV from, to calculate an approximate standard deviation. This is why, following step 3 in 3.1, we simply assumed a p-value of 0.11 for both these studies. Consequently, the studies have the same SMD, which we acknowledge is suboptimal. This drawback could be improved if better data is available, specifically data where the standard deviations are indicated.

Figure 2 displays that the I^2 of the male subsample is 98%. This shows again, that there is substantial heterogeneity between the results of the different studies, which does not come as a surprise as discussed above (see 2.2.5.). However, some of the studies in the sample have a very high sample size (N), which reduces the sampling error and increases the I^2. Therefore, the high I^2 could also be a result of a high N in some of the studies and the associated low sampling error. The τ^2 is the between study variance and is 10.3598.

As not all our studies come from the same population, violating the basic assumption in the fixed effect model, we focus on the results of the random effects model. This model accounts for this varying population in the sample (the same applies to the female subsample). The random effects model reports an overall SMD of -5.91 with a confidence interval of [-8.57,-3.24]. As our effect is given in SMD, we cannot easily interpret the magnitude. Still, as 0 is not in the confidence interval, this indicates that on the basis of the included studies, the overall effect of an UBI on labor supply is negative and significant for males.

Figure 3 on the other hand displays that also the I^2 of the female subsample is relatively high with 94% and indicates substantial heterogeneity. Again, this could partly be the result of the large sample size of some of the studies. The τ^2 is 2.8656. The random effects model reports an overall SMD of -2.68 with a confidence interval of [-4.19,-1.18]. As before, the results on the basis of this subsample indicate that the overall effect of an UBI on the female labor supply is negative and significant. The results from Figure 2 and 3 provide some evidence that the overall effect of an UBI on the female labor supply is smaller than the effect on the male labor supply.

Figure 4 shows the funnel plots of the male and female subsample. Both the plots show that there is no clear evidence of publication bias, as studies without significant effects are published.

3.2.3 Subgroup Analysis

As our two meta analyses show signs of high heterogeneity, we try two control for possible sources of it. For this, for each of the subsamples we run two variations of our meta analyses where we perform two different subgroup analyses. In the first modification, we want to explore whether the heterogeneity comes from differences between developed and developing countries. In the second modification, we examine whether the differences in the type of the interventions could be the driver of the heterogeneity.

Figures 5 and 6 show the resulting forest plots for the developed / not developed subgroup analysis. For both of the subsamples the heterogeneity is not reduced by separately analyzing developed and not developed countries, as the I^2 remains roughly the same. Still we can see, that the I^2 and the τ^2 (in both subsamples) are large for studies conducted in developed countries and small for studies conducted in not developed countries. Nevertheless the sample of not developed countries is smaller than the sample of developed countries, which might be the driver of these differing results.

Figures 7 and 8 show the forest plot for the intervention type subgroup analysis. For both the subsamples, the separate analysis by intervention type does not reduce the I^2 indicating that the heterogeneity is not mainly caused by the variation in intervention type. Nonetheless, we see some variation in heterogeneity depending on the intervention type. For Social Wealth Fund the I^2 and τ^2 are very large while for NIT and Cash Transfer they are small in both subsamples. I^2 and τ^2 are both 0 for Cash Transfer in both subsamples and 0 for NIT in the male subsample, but slightly higher for NIT in the female subsample.

3.2.4 Meta Regression

As the subgroup analysis did not clearly reduce the heterogeneity in the two meta analyses, we perform one last analysis to check whether a combination of variables could be responsible for part of the heterogeneity in our samples. For this we run two meta regressions for each of the subsamples: One regression where we only use intervention type as a regressor, analogous to the subset analysis conducted above, and another regression where we additionally use the starting decade of the interventions as a second regressor. This meta regressions will then show whether the combination of difference in intervention type and difference in the start decade of intervention can explain part of the heterogeneity in the two meta models.

As the first regression with one regressor is analogous to the subgroup analysis we performed above, and this subgroup analysis did not provide evidence for any reduction of overall heterogeneity, in Table 4, we focus on the results of the second regression and whether they show a different result than the subgroup analysis (SE shown in parentheses). It can be seen that the two models both have a lower I^2 than the original meta analyses (91.95% vs. 98% for male and 83.57% vs. 94% for female subsample). This is a first indicator that the combination of two regressors can explain part of the heterogeneity in the models. Further, the R^2 indicates that the meta regression models amount for a notable amount of the heterogeneity (39.64% for the male, 42.50% for the female subsample). These results suggest, that at least part of the heterogeneity in our meta analyses stems from the differences in intervention type in combination with the difference in starting decades of the different interventions.

Finally, we use ANOVA to compare the full meta regression models (two regressors) to the reduced meta regression models (only intervention type as regressor). We find a significant difference between the two models for the female subsample, while the difference between the models remains insignificant for the male subsample. This provides evidence for the conclusion that the meta regression for the female subsample in Table 4 indeed better explains the heterogeneity in our meta analysis, while it fails to provide evidence to draw the same conclusions for the male subsample.

4 Discussion

In this paper we conduct a meta-analysis looking into the effects of UBI on participation in the labour market. This question is crucial to decide whether we should endorse this policy, and all the more relevant since UBI is currently a debated topic.

Because UBI involves a considerable magnitude and rather pervasive changes in the socio-economic structure, it is not normally feasibility, for political and/or financiation issues. To this date it has not been implemented fully, as prescribed by theory. In order to conduct an evidence-based analysis, we thus draw evidence from programs which involve policies featuring similar aspects as UBI. However imperfect these proxies might be, they are the only source of experimental evidence available. Arguably, our results will still give us better a standpoint to (try to) answer the question at stake. How much better off we will be dependent on the quality of the gathered evidence and meaningfulness of the analysis.

We collect evidence by conducting a systematic search for interventional data. First we identify all conducted programs featuring different types of intervention designs with similarities to UBI. For each intervention we then search for studies that meet our pre-defined criteria of eligibility. This systematic approach gives us confidence that our resulting set of studies contain all the existing and relevant interventions. Subsequently we rate the quality of each study according to the GRADE framework (Guyatt et al., 2008). As stated, this will allow us to place a specific, though subjective degree of confidence in the results.

The results of our meta-analyses indicate a negative effect of UBI-like interventions on male and female labour supply, whereas the effect seems to be more negative for men. Therefore, our findings are in line with critical voices about UBI's benefits. Furthermore, we found no evidence of publication bias.

However, our analysis has several limitations. Therefore, all the results of our analysis should be interpreted with caution. First, since the different study results were measured in different units and hence are not directly comparable, a standardization of all effects was necessary. This makes a clear numeric prediction impossible, since the overall effect is unitless (on the basis of our results we can only assume that the effect is negative for both genders).

Second, we could only use evidence from UBI-related proxies, since a complete universal and unconditional basic income program has not been implemented so far. We are thus not able to make a prediction for the impact of a full-blown UBI.

Third, a large degree of heterogeneity among the data is predestined due to a large variety of different intervention types included in this study. A simple subgroup analysis could not further explain the heterogeneity in the model, but some of it could be identified by a meta regression. Results suggest that part of the heterogeneity likely stems from a combination of both the differences in type and implementation time of the interventions. Nevertheless, some unexplained heterogeneity remains. Some of this residual heterogeneity might be caused by the large sample sizes of some studies, which drives the sampling error to zero and in turn increases the I^2 (the variation that cannot be explained by the sampling error).

Forth, with regard to the the quality scores assigned to the studies, which vary between 2 and 3, we have only limited confidence in these results.

In order to counteract these limitations future research should take the following points into consideration. Future meta-analyses can investigate the source of heterogeneity between studies more profoundly looking at individual treatment data, which could allow to control for important covariates. Experimenters can reduce the heterogeneity of currently available evidence by conducting standardized UBI-like interventions, even on a medium-scale.

References

Randall K.Q. Akee, William E. Copeland, Gordon Keeler, Adrian Angold, and E. Jane Costello. Parents' incomes and children's outcomes: A quasi-experiment using transfer payments from casino profits. *American Economic Journal: Applied Economics*, 2(1):86–115, 1 2010. ISSN 19457782. doi: 10.1257/app.2.1.86. URL `http://www.aeaweb.org/articles.php?doi=10.1257/app.2.1.86`.

Kate Ambler and Alan De Brauw. Household Labor Supply and Social Protection Evidence from Pakistan ' s BISP Cash Transfer Program, 2019.

Solomon Asfaw, Benjamin Davis, Josh Dewbre, Sudhanshu Handa, and Paul Winters. Cash Transfer Programme, Productive Activities and Labour Supply: Evidence from a Randomised Experiment in Kenya, 2014. ISSN 17439140.

Gary. Christophersen, Inc. Mathematica Policy Research, and S R I International. Final report of the Seattle-Denver income maintenance experiment. Technical report, Washington, D.C., 1983. URL `file://catalog.hathitrust.org/Record/000666543 http://hdl.handle.net/2027/mdp.39015011328542 (v.1) http://hdl.handle.net/2027/uc1.c041066740 (v.2)`.

Gan-Ochir Doojav and Ariun-Erdene Bayarjargal. Implications of Economic Growth, Poverty and Inequality in Mongolia over the Period of 2008-2012. Technical report.

Robert M. Feinberg and Daniel Kuehn. Guaranteed Nonlabor Income and Labor Supply: The Effect of the Alaska Permanent Fund Dividend. *B.E. Journal of Economic Analysis and Policy*, 6 2018. ISSN 19351682. doi: 10.1515/bejeap-2018-0042.

Milton Friedman. The Case for a Negative Income Tax. *National Review*, 7: 239–241, 1967.

Ugo Gentilini, Margaret Grosh, Jamele Rigolini, and Ruslan Yemtsov. Exploring Universal Basic Income: A Guide to Navigating Concepts, Evidence, and Practices. Technical report, 2020.

Marcia Gibson, Wendy Hearty, and Peter Craig. The public health effects of interventions similar to basic income: a scoping review. *The Lancet Public Health*, 5(3):e165–e176, 2020a. ISSN 24682667. doi: 10.1016/S2468-2667(20)30005-0. URL `http://dx.doi.org/10.1016/S2468-2667(20)30005-0`.

Marcia Gibson, Wendy Hearty, and Peter Craig. The public health effects of interventions similar to basic income: a scoping review, 3 2020b. ISSN 24682667.

Mouhcine Guettabi, Matthew Reimer, and Andrew Bibler. Short-term Labor Responses to Unconditional Cash Transfers Short-term Labor Responses to Unconditional Cash Transfers *. Technical report, 2019.

Gordon H. Guyatt, Andrew D. Oxman, Gunn E. Vist, Regina Kunz, Yngve Falck-Ytter, Pablo Alonso-Coello, and Holger J. Schünemann. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, 336(7650):924–926, 4 2008. ISSN 09598146. doi: 10.1136/bmj.39489.470347.ad.

Gordon H. Guyatt, Andrew D. Oxman, Regina Kunz, David Atkins, Jan Brozek, Gunn Vist, Philip Alderson, Paul Glasziou, Yngve Falck-Ytter, and Holger J. Schünemann. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *Journal of Clinical Epidemiology*, 64 (4):395–400, 2011. ISSN 08954356. doi: 10.1016/j.jclinepi.2010.09.012.

Institute for Research on Poverty. New Jersey Graduated Work Incentive Experiment. Summary Report, 1973.

Institute for Research on Poverty. The Rural Income Maintenance Experiment. Summary Report., 1976.

Kenneth C. Kehrer, John F. McDonalds, and Robert A. Moffitt. Final Report of the Gary Income Maintenance Experiment: Labor Supply, 1979.

Andrés Mideros and Cathal O'Donoghue. The Effect of Unconditional Cash Transfers on Adult Labour Supply: A Unitary Discrete Choice Model for the Case of Ecuador. *Basic Income Studies*, 10(2):225–255, 2015. ISSN 19320183. doi: 10.1515/bis-2014-0016.

Philip K. Robins. A Comparison of the Labor Supply Findings from the Four Negative Income Tax Experiments. *The Journal of Human Resources*, 20 (4):567–582, 1985.

Djavad Salehi-Isfahani and Mohammad H. Mostafavi-Dehzooei. Cash transfers and labor supply: Evidence from a large-scale program in Iran. *Journal of Development Economics*, 135:349–367, 11 2018. ISSN 03043878. doi: 10.1016/j.jdeveco.2018.08.005. URL `https://doi.org/10.1016/j.jdeveco.2018.08.005`.

Natasha Wiebe, Ben Vandermeer, Robert W. Platt, Terry P. Klassen, David Moher, and Nicholas J. Barrowman. A systematic review identifies a lack of standardization in methods for handling missing variance data. *Journal of Clinical Epidemiology*, 59(4):342–353, 2006. ISSN 08954356. doi: 10.1016/j.jclinepi.2005.08.017.

Tables

Table 1: Interventions studied

Abbr.	Intervention	Country	Type of Intervention	Duration	Benefit
BDH	Bono de Desarrollo Humano	Ecuador	Cash Transfer	2003-present	credit of up to USD840
BISP	Pakistan's Benazir Income Support Program	Pakistan	Cash Transfer	2008-present	quarterly RS4834
CT-OVC	Kenya Cash Transfer Programme for Orphans and Vulnerable Children	Kenya	Cash Transfer	2012-present	monthly Ksh1500, 2011/2012: increased to Ksh200
GNIT	Gary Income Maintenance Experiment	USA	NIT	1971-74	USD3,300 - USD4,300 per year
NNIT	New Jersey Graduated Work Incentive Experiement	USA	NIT	1968-72	USD1,650 - USD5,000 per year
RNIT	Rural Income Maintenance Experiment	USA	NIT	1969-73	USD1,741 - USD3,482 per year
SNIT	Seattle/Denver Income Maintenance Experiment	USA	NIT	1970-76/80	USD 3,800 - USD5,600 per year
APF	Alaska Permanent Fund	USA	Social Wealth Fund	1982-present	up to USD2,072 per year
NATD	Native American Nations Tribal Dividend	USA	Tribal Dividend	1990's-present	not specified

Table 2: Overall Quality Rating

4	High	We are very confident that the effect of the study reflects the actual effect
3	Moderate	We are quite confident that the effect in the study is close to the true effect
2	Low	The true effect may differ significantly from the estimate
1	Very Low	The true effect is likely to be substantially different from the estimated effect

Table 3: Summary Statistics of the Male/Female Subset

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Male							
Effect	10	−23.3	56.2	−182.1	−13.1	−0.03	1.3
SE	10	6.3	8.5	0.1	0.4	8.0	24.9
n	10	1,305,221.0	4,091,683.0	264	969.5	16,080	12,950,193
p-value	1	0.8		0.8	0.8	0.8	0.8
Intervention Start	10			1968			2012
Intervention End	10			1972			2020
SMD	10	−6.5	17.9	−57.2	−1.6	−0.1	1.2
Female							
Effect	10	−24.2	46.7	−150.7	−18.6	−0.3	1.2
SE	10	6.2	9.4	0.1	0.4	4.5	27.8
n	10	709,993.2	2,210,328.0	264	938.8	16,080	7,000,428
p-value	1	0.01		0.01	0.01	0.01	0.01
Intervention Start	10			1968			2012
Intervention End	10			1972			2020
SMD	10	−5.5	12.0	−39.3	−2.7	−1.0	0.9

Table 4: Meta Regression for Male and Female Subset

	Male Model	Female Model
Intercept	320.43	-64.51
	(3605.90)	(2448.00)
Intervention Type: NIT	-8.76	1.60
	(67.94)	(43.67)
Intervention Type: Social Wealth Fund	-32.18	-16.62
	(41.88)	(28.01)
Intervention Type: Tribal Dividend	-2.24	2.13
	(29.34)	(19.12)
Intervention Decade (start)	-0.16	0.031
	(1.80)	(1.22)
I^2 (residual heterogeneity)	91.95%	83.57%
τ^2 (estimated amount of residual heterogeneity)	187.00	73.83
H^2 (unaccounted variability)	12.43	6.09
R^2 (amount of heterogeneity accounted for)	39.64%	42.50%

Table 5: Final Set of included Studies

Study	Intervention ID	Intervention Type	Treatment Sample Size	Comments on quality	Quality (GRADE)	Risk of bias	Indirect-ness	Rate up
Mideros, O'Donoghue 2015	BDH	Cash Transfer	19,840		3	0	0	0
Ambler, de Brauw 2019	BISP	Cash Transfer	4,976	only woman eligible, poor people adressed	2	0	-1	0
Asfaw et al 2014	CT-OVC	Cash Transfer	2,294	money given to HH with children, poor households	2	0	-1	0
Kehrer et al 1979	GNIT	NIT	1,799	poor people adressed, NIT	2	0	0	-1
Christophsen 1983	SNIT	NIT	4,800	poor people adressed, NIT	2	0	0	-1
Institute for Research on Poverty 1973	NNIT	NIT	1,216	poor people adressed, NIT	2	0	0	-1
Institute for Research on Poverty 1976	RNIT	NIT	809	poor people adressed, NIT	2	0	0	-1
Robins 1985	GNIT	NIT	1,780		3	0	0	0
Robins 1985	NNIT	NIT	1,357		3	0	0	0
Robins 1985	RNIT	NIT	264		3	0	0	0
Robins 1985	SNIT	NIT	4,800		3	0	0	0
Bibler et al 2019	APF	Social Wealth Fund	28,027		3	0	0	0
Feinberg and Kuehn 2018	APF	Social Wealth Fund	7,000,428		3	0	0	0
Akee et al 2010	NATD	Tribal Dividend	1,420		3	0	0	0

Figures

Figure 1: Flow Diagram of Search Procedure

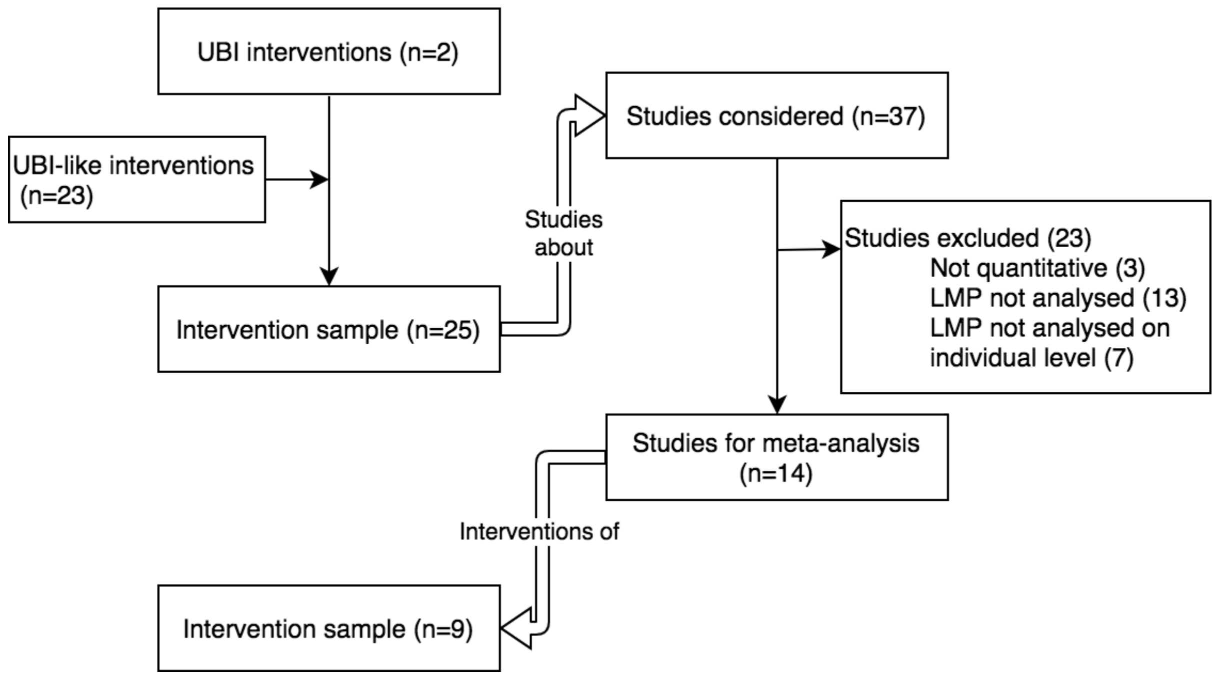

Figure 2: Forest Plot of Male Subsample

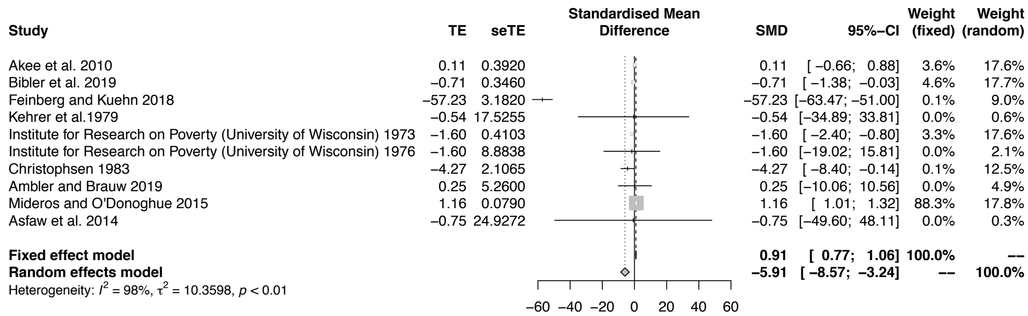

Figure 3: Forest Plot of Female Subsample

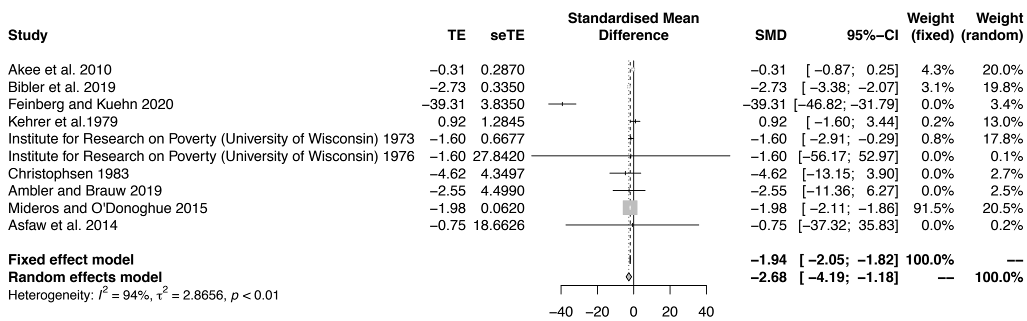

Figure 4: Funnel Plots

(a) Male Subsample (b) Female Subsample

Figure 5: Forest Plot of Male Developed/Not Developed Subgroups

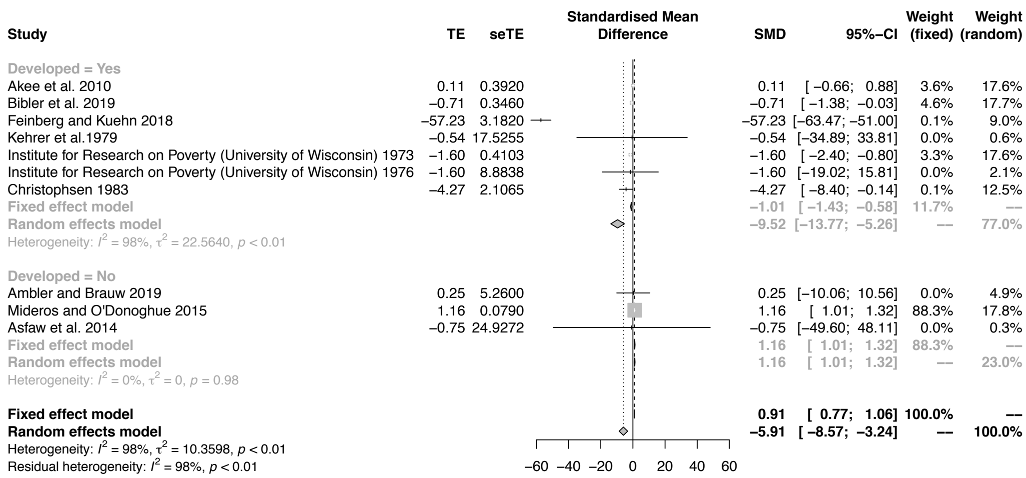

Figure 6: Forest Plot of Female Developed/Not Developed Subgroups

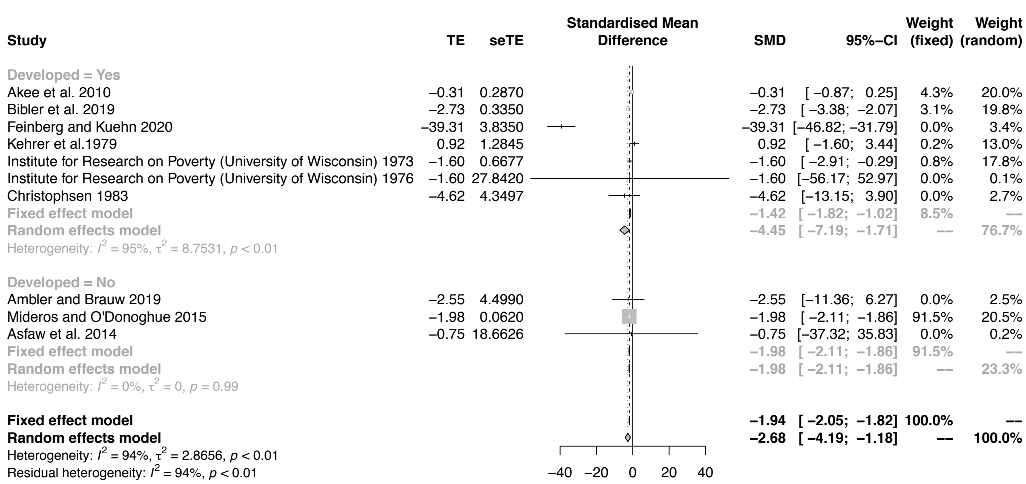

Figure 7: Forest Plot of Male Intervention Subgroups

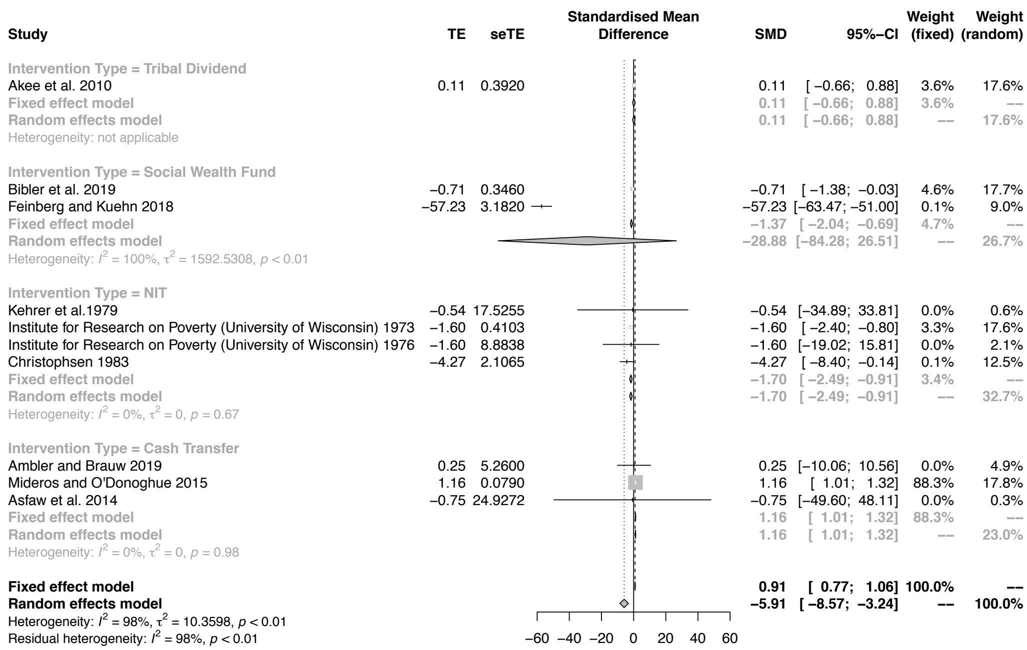

Figure 8: Forest Plot of Female Intervention Subgroups