The impact of cash transfers on subjective wellbeing and mental health in low- and middle-income countries: A systematic review and meta-analysis

December 2020

Joel McGuire[†] , Anders Malthe Bach-Mortensen[+], Caspar Kaiser[+,§]

Abstract

Background: A large body of evidence evaluates the impact of cash transfers (CTs) on physical health and economic indicators in low- and middle-income countries (LMICs). A growing amount of research on CTs contains measures of subjective wellbeing (SWB) and mental health (MH) but no attempt has been made to systematically synthesize this work.

Objective: To evaluate whether CTs improve the SWB and MH of recipients in LMICs.

Methods/design: We undertook a systematic review and meta-analysis of randomised controlled trials (RCTs) and quasi-experimental studies, including peer-reviewed publications and grey literature (e.g. reports, pre-prints, and working papers), conducted over the period 2000-2020, examining the impact of CTs on self-reported SWB and MH outcomes. A protocol for this review was prospectively registered with Prospero (CRD42020175464).

Results: Thirty-seven studies were included in our meta-analysis, covering 100 outcomes, and a total sample of 112,245 individuals. After an average follow-up time of two years, the average effect size on MH and SWB is estimated to be 0.10 standard deviations (SDs). CT value, both in absolute terms ($\hat{\beta}$=0.08 SDs per $100 PPP) and relative to previous income ($\hat{\beta}$=0.10 SDs for each doubling), are strong predictors of the effect size. Moreover, unconditional CTs have a larger impact than conditional CTs ($\hat{\beta}$=0.04). The impact of CTs diminishes marginally over time ($\hat{\beta}$=-0.02 SDs per year). We find no significant evidence of negative spillover effects to non-recipients.

Discussion: Cash transfers significantly increase MH and SWB in LMICs. More research on longitudinal (5+ years) and spillover effects is needed. Future impact evaluations should collect data on MH and SWB to enable comparisons of the relative cost-effectiveness of development interventions at improving people's wellbeing.

† Corresponding author. Email: joel@happierlivesinstitute.org. Happier Lives Institute.
+ Department of Social Policy and Intervention, University of Oxford.
§ Wellbeing Research Centre, University of Oxford.

1 Introduction

Cash transfers (CTs) - commonly understood as direct payments made to people in poverty - are among the most extensively studied and implemented interventions in low- and middle-income countries (LMICs) (Vivalt, 2015). Previous systematic reviews and meta-analyses of CTs found improvements on several outcomes. These outcomes include material poverty (Kabeer & Waddington, 2015), human capital (Baird et al., 2013b; Millán et al., 2019), social capital (Owusu-Addo et al., 2018), health (Lagarde et al., 2007; Behrman & Parker, 2010; Crea et al., 2015), intimate partner violence (Baranov et al., 2020; Buller et al., 2018), child labor (Kabeer & Waddington, 2015), the spread of HIV (Pettifor et al., 2013), spending on tobacco and alcohol (Evans & Poponova, 2014; Handa et al, 2018), and labor supply (Baird et al., 2018; Banerjee et al., 2017).

Although these factors are relevant to wellbeing, measures of mental health (MH) and subjective wellbeing (SWB), which probe how individuals themselves assess the quality of their lives, are often thought to track wellbeing more accurately. Indeed, measures of SWB are increasingly considered to be essential components in applied policy analyses (Benjamin et al., 2020; Frijters et al., 2020). It therefore seems pertinent to evaluate the effectiveness of CTs with respect to these measures.

Individual income and SWB are known to be positively associated (Powdthavee, 2010; Stevenson & Wolfers, 2013; Jebb et al., 2018), especially for those at low income levels (Clark, 2017; Deaton, 2008). A similar relationship is observed in the MH literature (Karimli et al., 2019; Tampubolon & Hanandita, 2014; Schilbach et al., 2016; Ridley et al., 2020). Moreover, mental health problems may engender and perpetuate poverty (Haushofer & Fehr, 2014). Unfortunately, the literature on the link between income and SWB and MH in LMICs has long lacked *causal* evidence, which the growing body of primary research on CTs may address.

While CTs may improve the SWB and MH of recipients, these interventions could also have negative psychological consequences on non-recipients. Qualitative research suggests the presence of negative psychological spillovers (Fisher et al., 2017; MacAuslan & Riemenschneider, 2011), and some recent quantitative work echo this worry (Haushofer et al., 2019). For example, envy among non-recipients may be a concern (Ellis, 2012). Community disruptions and crime rates may also increase if CTs are mistargeting to formally ineligible recipients (Agbenyo et al., 2017; Fisher et al., 2017). However, there is also some evidence of positive spillovers. For example, CTs have been found to decrease the intergenerational transmission of depression (Eyal & Burns, 2019) and to lead to decreased suicide rates in the areas they are implemented (Alves et al, 2018).

We know of no previous systematic reviews on this subject. A non-systematic meta-analysis by Ridley et al. (2020), which evaluates the impact of CTs on MH, is closest to our work.[1] We build on their work in four directions. First, we conducted a full systematic review and search of the existing literature in accordance with the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidance (Moher, Liberati, Tetzlaff, & Altman, 2010). Second, we consider SWB measures alongside MH measures[2]. Third, we consider quasi-experimental designs (in addition to randomised controlled trials (RCTs)). Fourth, we evaluate the quality of included studies, assess publication bias, and perform a moderator analyses across (1) outcome type (MH and SWB), (2) CT value, and (3) duration of the transfer.

---

[1] Also see the systematic review by Owusu-Addo et al. (2018). They focus on determinants of health inequalities in sub-Saharan Africa and include a descriptive section on MH.

[2] Unlike Ridley et al. (2020), we focus on measures of affective or mood disorders and exclude measures of stress or other psychological disorders. An affective or mood disorder refers to depression or anxiety. Mental health issues we do not consider are disorders relating to addiction or personality.

2 Methods

2.1 Eligibility criteria

For a study to be included it must satisfy four criteria: First, the study must investigate the effect of an unbundled cash transfer (defined below). Second, the study must include a measure of self-reported affective mental health or subjective wellbeing, but these need not be the primary focus of the study. Third, the study context must not be a high-income country.[3] Fourth, the study design must be experimental or quasi-experimental[4] and afford standardizing the mean difference between treatment and control groups.

Regarding our first criterion, we distinguish between unconditional cash transfers (UCTs) and conditional cash transfers (CCTs). Conditional cash transfers formally require adherence to certain actions, such as school enrollment or vaccination. The strictness of conditions varies widely, and conditions are sometimes left unmonitored due to high administrative costs (Davis et al., 2016). UCTs have no requirements, although they are often targeted to a vulnerable subset of the population, commonly defined by a combination of regional statistics, means tests and selection by prominent members of the community. We consider noncontributory social pensions and enterprise grants to be UCTs. CTs are typically paid out in lump-sums or streams (monthly installments). Some stream or multi-installment CTs have graduation mechanisms where individuals stop receiving transfers once they meet certain conditions (Villa & Niño-Zarazúa, 2019). All included CTs must be "unbundled", i.e. implemented and tested independently of other services such as asset transfers, training, or therapy.

Concerning our second criterion, we note that SWB measures tend to assess overall wellbeing (Diener, 2009; Diener et al., 2018), which sometimes include separate measures of positive and negative mental states (Busseri & Sadava, 2011). By contrast, affective MH questionnaires tend (1) to only measure the negative components of SWB, i.e., how badly someone is doing and, (2) to also capture information on an individual's behaviors and habits (in addition to their thoughts and feelings). In our analyses, we include measures of valenced mental states, but no measures of behavior or habits. See the "Measures" column of Table A3 in the appendix for a list of all included measures.

2.2 Data

We searched studies using academic search engines and databases. These included: EBSCO: MEDLINE, PsycINFO, PubMed, Business Source Complete, EconLit, Social Sciences Full Text (H.W. Wilson), APA PsycARTICLES, Psychology and Behavioral Sciences Collection, Academic OneFile, Academic Search Premier, CINAHL, Open Dissertations, Web of Science, Science Direct, JSTOR, ECON PAPERS, 3ie, IDEAS/REPEC, and Google scholar. These efforts were complemented by a forward and backward citation search of eligible studies, contacting authors, and through Google Scholar notifications. Our search string can be found in Appendix A.

We stored all retrieved records in the reference management system Zotero. Double-blind screening of the titles and abstracts was done using the software Rayyan by JM and CK. Any disagreements were discussed until consensus was reached. Studies that passed the double-screening were reviewed in full text by JM.

---

[3] We use the World Bank's thresholds (as of 2019) for high-income countries as having a GNI of more than $12,375. See: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.

[4] Common quasi-experimental designs employ a natural random assignment into control or treatment groups. Relevant identification strategies include regression discontinuity, difference-in-differences, instrumental variables or propensity score matching.

We extracted study details such as author name, CT program, number of participants, MH and SWB outcomes, and effect sizes. We also collected information on the size of the cash transfer, time between start of intervention and follow-up, and whether it was a CCT or UCT, paid out in a stream or lump sum, or directed towards adolescents, prime age adults or elders. All data were extracted by one author (JM) and the full extraction results were checked for accuracy by CK and ABM.

2.3 Quality

To assess the quality of included research, we evaluated the following domains: causal identification strategy, pre-registration, balance between treatment and control groups, attrition, sample size, contamination, treatment compliance, and whether intention-to-treat (as opposed to a complete case) analyses were performed.

2.4 Statistical Methods

We used the statistical programming language R for data analysis. Since most RCTs and quasi-experimental designs are based on mean differences,[5] we standardized these using Cohen's d. We used the independent t-statistic from a test of the mean difference to calculate Cohen's d in nearly all cases. We use $d = t\sqrt{1/n_t + 1/n_c}$ where n_t = treatment sample size and n_c = control sample size (Goulet-Pelletier & Cousineau, 2018). If the effect size of a study was expressed via odds ratios ($n = 2$), we converted from odds ratios to Cohen's d using $d = \ln(OR)\sqrt{3}/\pi$.[6]

If a study contained multiple outcome measures, we coded each as MH or SWB. To achieve a single effect size for each study-follow-up combination, we combined outcomes using the method of Borenstein et al., (2009), specifying a correlation of 0.7 for within construct aggregations, 0.5 for between constructs and 0.6 for both within and between aggregations. Specifying different correlations changes only the aggregate standard error, not the mean of effect sizes.

We used random effects (RE) models for our meta-analysis, which assume that true effects of each included study are drawn from a distribution of true effects (Borenstein et al, 2010). Each study in our model was weighted by the inverse of the standard error of the study's estimated effect size. Since there are sometimes multiple follow-ups in a study and multiple studies in a sample or program, we clustered standard errors at the level of the study and program. We assessed evidence of publication bias and p-hacking by using a funnel plot, the Egger regression test (Borenstein et al, 2011), and a "p-curve" (Simonsohn et al., 2014).

We conducted meta-regressions to test if certain study characteristics moderated estimated effect sizes. We focused on three potential moderating variables: years since CT began, size of CT, and whether CTs had conditionality requirements.

Concerning size of CT, we considered both the absolute and relative CT size. We operationalized absolute size as the average monthly value of a CT in purchasing power parity (PPP) adjusted US 2010 dollars, with lump sum CTs (comprising about 25% of our sample) divided by 24 months, which is the mean follow-up time.[7] For relative size, we used monthly CT value as a proportion of previous

[5] There is a concern that differences in subjective Likert scales are not meaningful (Bond & Lang, 2019). However, Bond and Lang's arguments require that individuals use Likert scales in a highly non-linear fashion (Kaiser & Vendrik, 2020). See Plant (2020) for arguments against such non-linear scale use.

[6] We do not use Hedge's-g as a small sample correction for Cohen's d because the two measures are identical to at least three decimal places for $n > 500$, the lower bound of the samples included in our study.

[7] We also test whether the results are sensitive to using 12, 36, 48, or 60 months instead. Results are qualitatively unchanged when doing so.

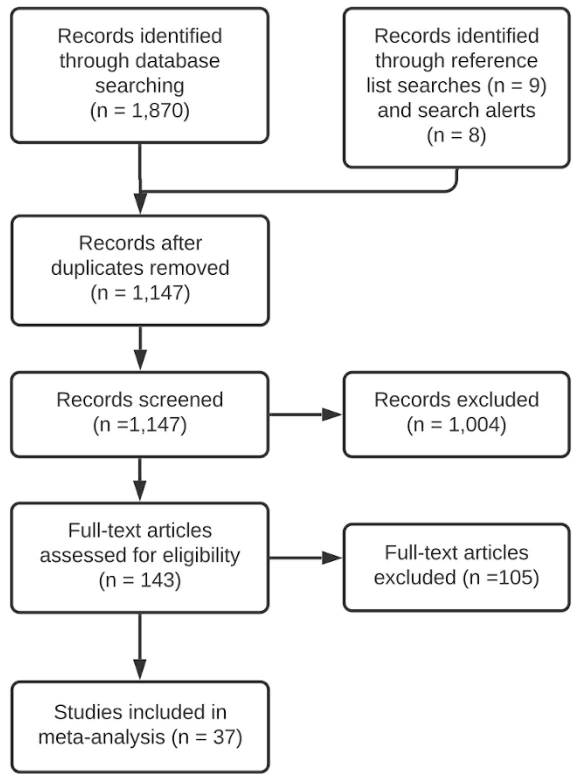
household monthly income. This was either directly reported or easily derived in many studies (21 out of 37 studies). If a study did not report sample information on income, we used consumption (10 studies) or expenditure (3 studies) information as a proxy. To convert between individual income and household income (8 studies) we assumed that $household\ income = individual\ income * \sqrt{household\ size}$ (see Chanfreau & Burchardt, 2008). If there was insufficient information to impute average household income (4 studies), we used regional statistics. Finally, as a robustness test, we also computed yearly CT value as a proportion of annual gross domestic product per capita (GDPpc).

3 Results

3.1 Description of Studies and Quality

We retrieved 1,870 records from implementing our search string. After removing duplicates, we were left with 1,147 records. After an initial round of double screening titles and abstracts by JM and CK, 143 met the eligibility requirements (see Figure 1 for a diagram of selection flow). After JM performed the final round of screening, there were 32 unique studies drawn from the initial search and five from Google Scholar alerts and citation searches. We thus included a total of 37 studies[8] reporting on 100 outcomes. Table A3 in the appendix summarizes the key characteristics of the included studies. Of the outcomes, 46 measured depression or general psychological distress, 21 measured happiness or positive feelings, 18 measured life satisfaction and two measured anxiety. The remaining 13 were summary indices of MH, SWB, or both.

Figure 1. Prisma Flow Diagram



Note: The flow chart shows the records screened at each stage of the systematic review.

Most of the studies were conducted in Africa (23), followed by Latin America (10) and Asia (4). The most commonly investigated CT type was UCT (26; 19 plain, 6 pensions and 1 enterprise grant) followed by CCTs (10) and one study that contained both a CT and UCT (Baird et al., 2013a). Country context was relatively evenly divided into low, low-middle, and upper-middle income countries (see Figure A2 in the appendix). Over half of the included studies included random assignment (22), while the rest were quasi-experimental (15).[9] The average time from the start of the CT to follow-up was two years. The average monthly payment was $38 PPP. A quarter of the studies were implemented as predominantly lump sum (10). All other studies (27) were paid out on a monthly basis.

In Table 1, we list the results of our quality assessments. While blinding of participants is impossible for CTs, blinding personnel and outcome assessment was mentioned (but not performed) in only one study (McIntosh & Zeitlin, 2020). Overall, few studies (9/37) referred to pre-registered protocols. The adherence to pre-specified statistical procedures and outcomes was generally unclear, thus making it

[8] One study breaks each follow-up into a separate paper (Haushofer et al., 2016; 2018).

[9] We labeled studies as "random assignment" if researchers did not have a role in the randomization process.

Table 1. Components of Quality

Subject	Question	Studies by Category
Design	What is the design of the study: cluster randomized control trial (cRCT), random assignment (RA), or quasi-experimental (QE)?	cRCT=13, RCT=5, RA =4, QE=15
Balance	Are there differences at baseline?	Yes=10, No=27
Balanced	Are baseline differences controlled for?	Yes=33, No=4
Attrition	Is there attrition or a low response rate?	Yes=24, No=13
Differential Attrition	Is the attrition differential, i.e., are there significant differences in response rates between treated and control groups?	Yes=19, No=18
Sample	How large is the sample? We operationalize this as a sample large enough to identify an effect size of 0.10=large (>3142), 0.15=medium (>1398), 0.20=small (>788), assuming a power level of 0.8 and significance level of 0.05.	Large=10, Medium=18 Small=9
Pre-registered	Is the study pre-registered?	Yes=9, No=28
Causal Identification Strategy Described	Is the randomization process or causal identification strategy described in detail?	Yes=33, No=4
Compliance	Is compliance with the treatment reported?	Yes=20, No=17
Contamination Proxy	Are treatment and control groups geographically separate? This is a proxy for contamination.	Yes=17, Unclear=20
ITT	Is an intention to treat analysis performed, i.e., do they use a complete case analysis (excluding noncompliant observations)?	Yes=28, Unclear=9
Blinding	Were surveyors and analysts blinded?	Yes=0, Unclear=37

impossible to assess whether outcomes were 'cherry-picked' post treatment. Moreover, about half of the included studies (17/37) did not assess treatment compliance. Therefore, aspects relating to implementation (e.g. intervention fidelity and adaptation) could not be assessed (Moore et al., 2015). Furthermore, contamination by the CT on control groups was rarely discussed or addressed. Only 13 out of 37 studies were geographically-clustered RCTs (cRCTs), which are more robust to possible contamination effects. Of the 15 quasi-experimental studies, one used a natural experiment (Powell-Jackson et al., 2016), two used instrumental variables (Ohrnberger et al., 2020a; Chen et al., 2019), and four used a regression discontinuity approach (based on a means test). The eight remaining studies used a propensity score matching approach. Of those using propensity score matching, six also employed a difference-in-difference estimator.

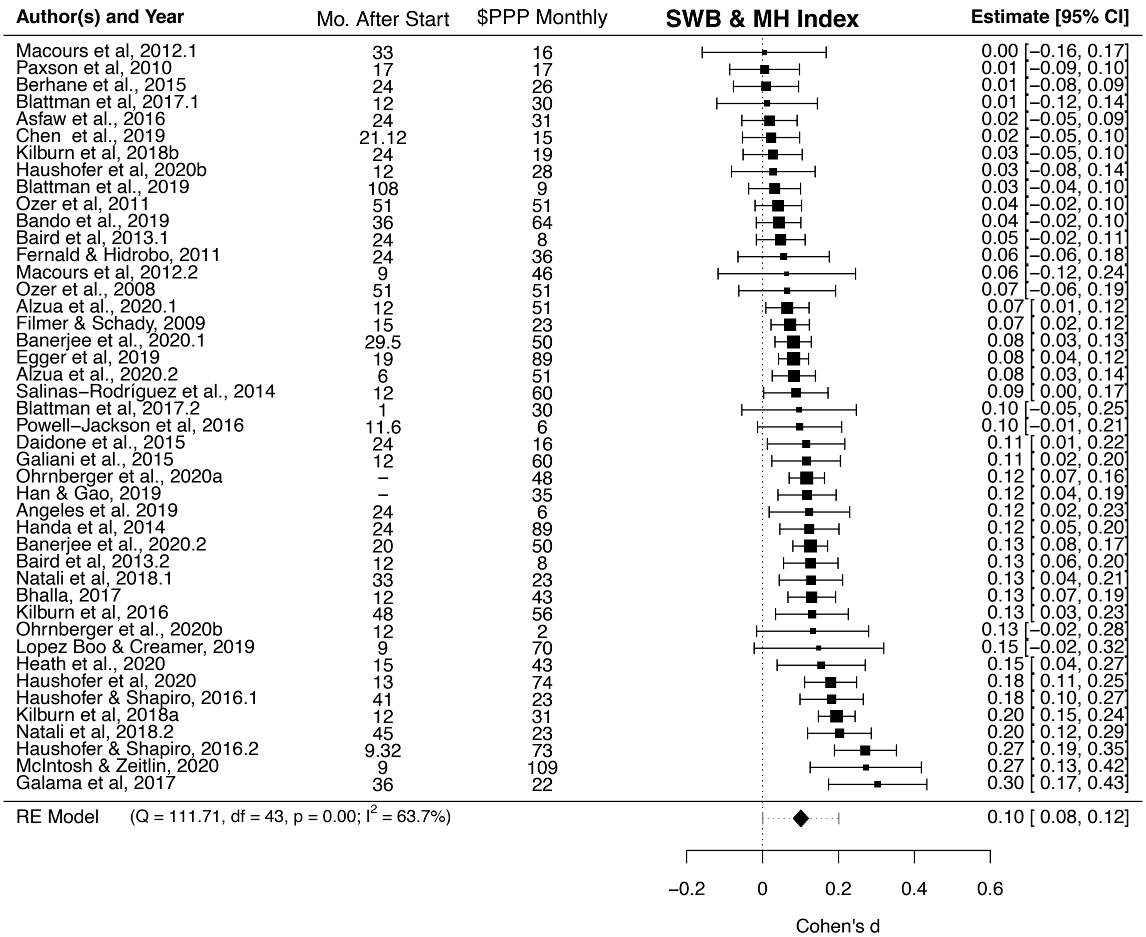
Despite the aforementioned concerns, we assess the synthesized evidence to be fairly reliable. Importantly, most studies clearly explained their causal identification strategy, were well balanced, performed intention-to-treat analyses, and controlled for differential attrition when present. Sample sizes were generally large compared to common sample sizes in clinical or psychological studies (n<500; Billingham et al., 2013; Kühberger et al., 2014; Sassenberg & Ditrich 2019).

3.2 Baseline results

For our baseline results, we aggregated effect sizes across studies using a random effects model. Throughout our analyses, we omitted measures of stress, optimism, and hope, and one outcome reported from Galama et al. (2017), which was a clear outlier.[10] The average overall effect size, as indicated by a black diamond at the bottom of Figure 2, is 0.10 SDs in the composite of SWB & MH measures (95% CI: 0.08, 0.12; given by the width of the diamond). The overall effect size does not

[10] In that study, Cohen's d for life satisfaction was 0.10 and for happiness it was 0.05. However, for an aggregation of 10 domains of satisfaction it was 0.76. The effect size was unusually high due to a very small standard error. This result could be due to chance as they ran and presented a very high number of specifications (~50). Results are qualitatively similar when the outlier is included.

Figure 2. Forest Plot



Note: Forest plot of the 37 included studies. Subjective wellbeing (SWB) and mental health (MH) outcomes in each study are aggregated with equal weight. Mo. after start is the average number of months since the cash transfer began. $PPP Monthly is the average monthly value of a CT in purchasing power parity adjusted US 2010 dollars. Lump sum cash transfers were converted to monthly value by dividing by 24 months, the mean follow-up time.

change substantially when accounting for dependency between multiple follow-ups, and multiple studies in a program in a multilevel model (ES: 0.095, 95% CI: 0.071, 0.118, or if we combine all the outcomes, without first averaging at the study-follow-up level (ES: 0.091, 95% CI: 0.066, 0.116.

Heterogeneity, as calculated by the I^2index, is substantial; 63.7% of the total variation in outcomes is due to variation between studies.[11] In other words, 63.7% of total variability can be explained by variability between studies instead of sampling error. To account for the impact of this substantial heterogeneity, we calculate a 95% predicted interval.[12] The estimated 95% prediction interval, given by the dashed line bisecting the black diamond in Figure 2, suggests that 95% of similar future studies would be expected to fall between 0.001 and 0.201 SDs in our composite of MH and SWB.

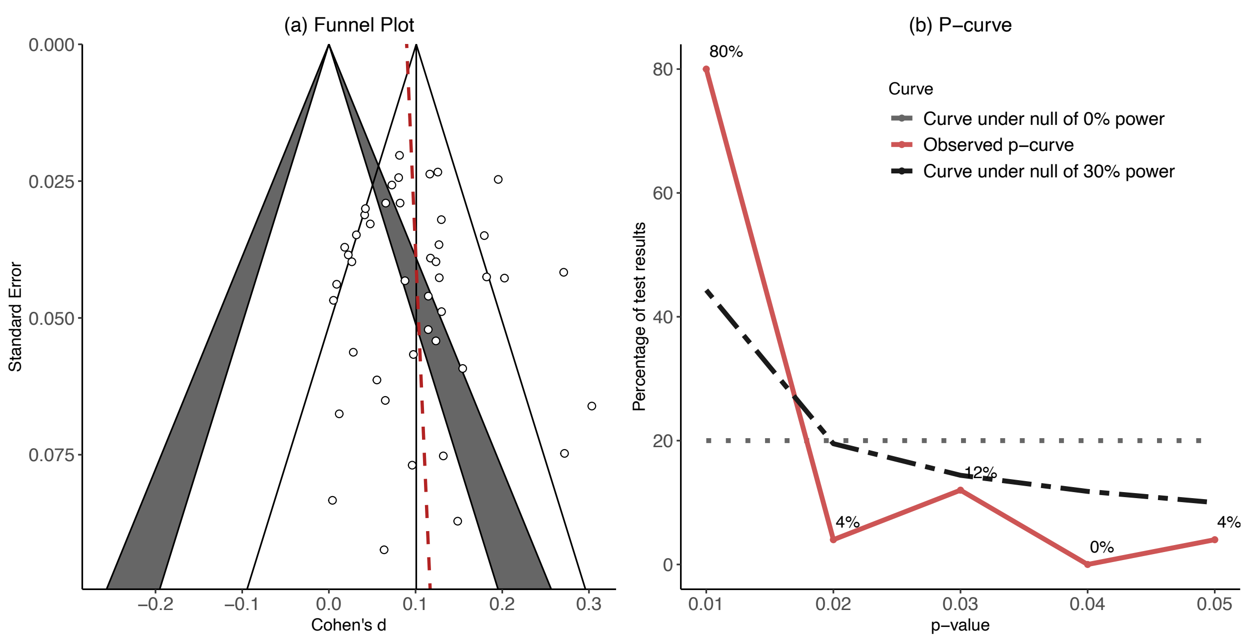
[11] 50-70% for I^2 is considered substantial (Higgins et al., 2019).

[12] See Riley et al., (2011) for further details on the calculation of prediction intervals. Note that prediction intervals are always larger than confidence interval in the presence of heterogeneity (IntHout et al., 2016).

Figure 3 displays the risk of publication bias and "p-hacking" (researchers testing a high number of outcomes and cherry-picking the coefficients that fall below a threshold p-value). In Figure 3a, we show a funnel plot, with standard error plotted against effect size, and the mean effect shown as a black vertical line.[13] If there are significantly more studies to the right than the left of the mean effect size, this would suggest that studies on the left may be missing, possibly indicating publication bias. This is known as asymmetry. Figure 3a shows little asymmetry, indicating that studies with more positive effects appear no more likely to be published. We use Egger's regression test to check this quantitatively by regressing the standard error on the effect size. The test does not reject the null of funnel plot symmetry (p=0.549), supporting our reading of the plot.

Figure 3b shows the percentage of results with different p-values. If "p-hacking" were an issue, we would expect that the distribution of p-values is left-skewed (an upward slope in the figure). The p-curve is downwardly sloped, which suggests no widespread p-hacking. However, it is possible that regression specifications with insignificant dependent variables were not reported at all. P-curves are unable to address such scenarios (Bishop & Thompson, 2016).

Figure 3. Funnel Plot and P-curve for evidence of potential bias



Note: Forest plot of the 37 included studies. Subjective well-being (SWB) and mental health (MH) outcomes in each study are aggregated with equal weight. Mo. after start is the average number of months since the cash transfer began. $PPP Monthly is the average monthly value of a CT in purchasing power parity adjusted US 2010 dollars. Lump sum cash transfers were converted to monthly value by dividing by 24 months, the mean follow-up time.

[13] It is expected that larger studies fall both nearer the mean effect size and have a smaller standard error and would therefore form the top of the funnel.

3.3 Meta Regression and Moderator Analysis

We focus on three types of variables that we expect to moderate the observed effects: (1) Whether a CT had conditionality requirements or not. (2) Value of CT (in absolute terms and relative to previous income). (3) Years since the transfer began, allowing us to assess whether effects dissipate over time. Throughout, we use multi-level models that account for multiple outcomes in a follow-up, multiple follow-ups in a study and multiple studies in a sample or program. Standard errors are clustered at the study and program level.[14] In every specification presented, the dependent variables are the study's estimated effect on MH or SWB. We standardized the effect sizes into Cohen's d.

In Figure 4, we present six plots that illustrate the bivariate moderating relationship of our variables of interest. Panel (a) shows the distribution and average effect size for UCTs and CCTS. Panels (b) through (f) show effect size on the y-axis and the time or size on the x-axis. Plots (b) through (f) are simple scatter plots meant to illustrate the raw correlation between two variables.

In Table 2, we present our main results. All models include a measure of CT size and years since the CT began. Model 1 includes a dummy indicating whether the CT had conditionality requirements. Models 1, 2 and 3 estimate the effect of relative CT size. Models 4 and 5 estimate the effect of absolute CT size (using $PPP monthly value). Models 3 and 4 include an interaction term between payment mechanism and "years since CT began" to identify the effect of decay conditional on whether a CT was paid out in a lump sum or stream.

In Model 1 we find that conditionality requirements reduce estimated effect sizes by almost 50%. In so far as UCTs are less costly to administer than CCTs, this suggests that UCTs are likely to be more efficient in promoting recipients' wellbeing.

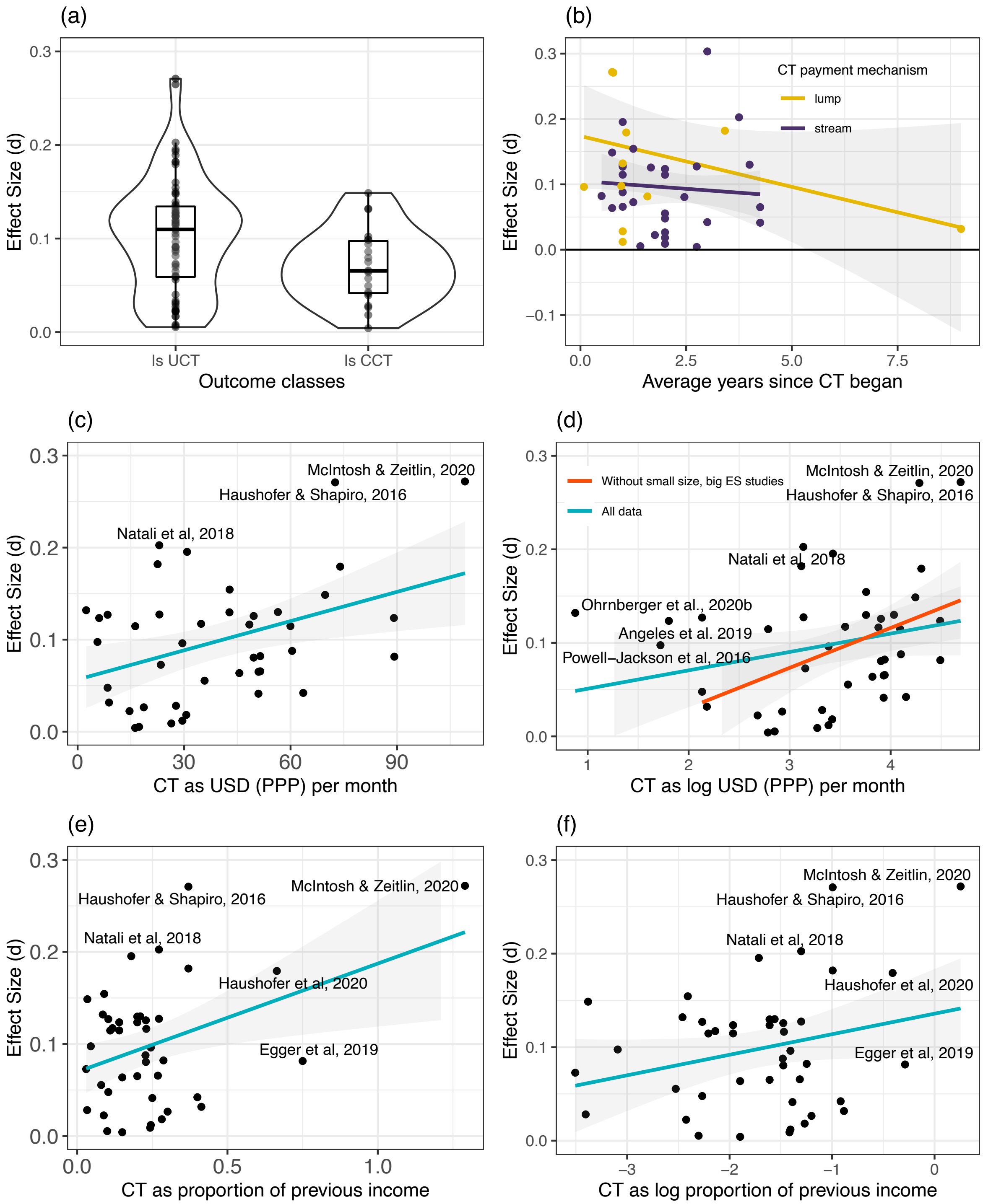
Table 2. Main results

	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	0.106***	0.091***	0.104**	0.097**	0.089***
	(0.016)	(0.071)	(0.028)	(0.031)	(0.021)
CT is CCT	-0.041**				
	(0.014)				
CT as Proportion of previous income	0.088***	0.099***	0.112***		
	(0.012)	(0.011)	(0.011)		
Years since CT began	-0.015*	-0.015**	-0.019	-0.017	-0.016*
	(0.004)	(0.005)	(0.013)	(0.013)	(0.007)
CT is lump sum			-0.051+	-0.024	
			(0.028)	(0.029)	
Years since * lump sum			0.006	0.001	
			(0.014)	(0.015)	
Monthly value in 100$ PPP				0.071*	0.080*
				(0.034)	(0.032)
Number of outcomes	97	97	97	97	97
Number of studies	35	35	35	35	35

Note: ***p < 0.001; **p < 0.01; *p < 0.05; +p < 0.1. "Time since CT began" is in years. "CT is lump sum" is an indicator for whether CTs were paid out in a lump sum. Otherwise CTs were paid out in (bi)monthly streams. Robust standard errors are clustered at the level of the program.

[14] We use rma.mv() and robust() from the metafor package in R (Viechtbauer, 2010).

Figure 4. Bivariate Moderator Relationships



Note: Panel (a) shows violin-box plots of effect size by outcome class. Panel (b) illustrates differences in decay of effect size between CTs paid in lumps (colored yellow) and streams (colored purple). Although there appears a decay amongst the studies paid out in lump sums, this may largely be driven by the study of Blattman et al. (2019), which follows-up eight years after the CT began. Panel (c) illustrates a positive relationship between absolute CT value and effect size. Panel (d) illustrates the increase in the slope of the regression line when very small (and surprisingly effective) transfers are omitted. Panel (e) illustrates a positive relationship between size of the transfer as a proportion of previous income and effect size. Panel (f) illustrates a positive relationship between size of the transfer as the *log* proportion of previous income and effect size.

In Model 2 we omit the indicator of whether CTs where CCTs or UCTs. Based on this specification, one can expect that doubling a recipient's consumption (by receiving a CT 100% of previous consumption) to roughly lead to a 0.10 SD increase in MH/SWB at the average follow-up time. Results in Models 1 and 3 are similar. See panels (e) and (f) of Figure 4 for the correlational relationship between relative size of a CT and magnitude of effect.

Models 4 and 5 shows our results for absolute CT value, yielding a significant and positive coefficient in both specifications. These results indicate that a CT with a monthly value of $100 PPP leads to an approximately 0.07 to 0.08 SD increase in SWB and MH outcomes. See Figure 4, panel (c) for the bivariate relationship. Increases in income are typically assumed to yield diminishing gains in wellbeing. To test if that is the case in our sample of studies, we log transformed our measures of relative and absolute CT size. We find a significant effect for log-relative value but no significant effect of log-absolute value (see Table A2 in the appendix).[15]

Taken together, models 1, 2 and 4 provide evidence that the effect of CTs on wellbeing decays over time. Using the coefficient from Model 2, each year the effect is estimated to decline by 0.015 SDs. With that estimate, a CT which doubles household income would take almost two decades to decay.[16] However, the effects of "years since CT began" could differ depending on whether the recipient was given the CT in a lump sum or still receives monthly transfers. Our bivariate plot (Figure 4, panel (b)) suggests a difference in decay between the two payment mechanisms. Lump CTs appear to decay over time while stream CTs (which are nearly all ongoing at the time of the last follow-up) show a flat trend. In Models 3 and 4 we formally test for differences in decay between lump and stream CTs. The interaction, "years since * CT is lump sum" gives the difference in decay between lump and stream CTs. Since stream CTs are ongoing, we expected lump CTs to exhibit a larger decay in effect size than streams. Surprisingly, this is not the case in models 3 and 4. These display a positive, albeit insignificant interaction term. Thus, although there is a significant overall decay in effect size (as indicated by Models 1, 2, and 5), we are unable to precisely estimate the effect over time for a specific payment type.

Finally, we note that seven studies in our study include multiple follow-ups. As shown in Figure A1 in the appendix, six of these show a decline in effects size across follow-ups. A repeated t-test of whether mean effect size is different between first and second follow-up yields a p-value of 0.007, indicating that this decline is statistically significant.

The relatively large and significant intercepts in Table 2 suggest that CTs could have an effect independent of the size of the cash transfer (i.e., an effect from being enrolled). An enrolment effect, however unintuitive, is not implausible. Being awarded an amount of cash might boost someone's sense of good fortune, which could explain the intercept. Another explanation for the intercepts is that they are an artifact of a concave relationship between CT size and effect. A linear model will generally overestimate the intercept on data that contains a true concave relationship. However, the insignificance of the log-transformed absolute CT value is evidence against a clear concave relationship (see appendix Table A2, Model 2).

---

[15] The latter result may be due to the studies by Ohrnberger et al., (2020b), Powell-Jackson et al., (2016) and Angeles et al., (2019). These all have relatively small transfer values (the smallest in our sample: less than $7 PPP monthly value) but relatively large effect sizes (0.10 - 0.25 d). See Figure 4 panel (d) for an illustration of the change in slope when omitting these high leverage low-value high-effect studies.

[16] This follows from setting d equal to zero where $d = 0.091 + 0.099 * proportion\ of\ previous\ consumption - 0.015 * Years\ Since\ CT\ began$. This calculation yields that d would become zero after 19 years.

In addition to these analyses, we also tested whether RCT design, type of measure, or the study context moderated the effect size (see Table A1 in the appendix). Whether a study uses a RCT design does not affect the magnitudes of the estimated effects of CTs. This suggests that studies which rely on natural experiments or other causal identification strategies are reasonably robust. However, we do find that, compared to pure MH measures, effects of CTs on measures of SWB are significantly larger. Moreover, the largest effect sizes occur for studies in which a compound index of both MH and SWB was used.[17] Notably, CTs conducted in Latin America have a near zero estimated effect. This appears to be primarily driven by the fact that many CTs in Latin America have conditionality requirements. When including both a dummy for conditionality and for the CT being conducted in Latin America, we find that the coefficient on Latin America is roughly halved and significant at the 10% level only.

As discussed in section 2, we ran alternative specifications of our size variables (see appendix Table A2). In particular, we checked if using CT value relative to GDP per capita changes our results. Although the coefficient is somewhat larger compared to results presented in Table 2 (with $p<0.05$), our conclusions remain unaffected.

Finally, in appendix D we consider how our type of results could potentially be used in policy analyses to study cost-effectiveness. Specifically, we calculate how many "wellbeing-adjusted life years" (see De Neve et al. 2020, Frijters et al. 2020), a given type of cash-transfer could buy for a given transfer size. We find that 1000$ lump-sum payment may be expected to buy roughly 0.330 "wellbeing-adjusted life years".

3.4 Spillovers

Four RCTs (two with multiple follow-ups) in our sample enabled assessment of spillover effects on non-recipients of CTs by including two control groups in a geographically-clustered RCT design: a spillover control made up of non-recipients living near recipients, and a "pure" control comprising non-recipients living spatially separate from the treatment locations.[18]

This design allowed comparison of wellbeing across (a) non-recipients who are "treated" to a spillover effect by living near recipients to (b) recipients living further away (who form the "pure" control). To ascertain the average effect of spillovers we performed a meta-analysis of the observed effects, using a multilevel random effects model, inverse-weighted by study standard error, and errors clustered at the level of the sample. Our results are illustrated in Figure 5.

The average effect of CTs on non-recipients' MH and SWB (represented by the diamond), is close to zero and is not significant at the 95% level, suggesting no significant spillover effects on average.

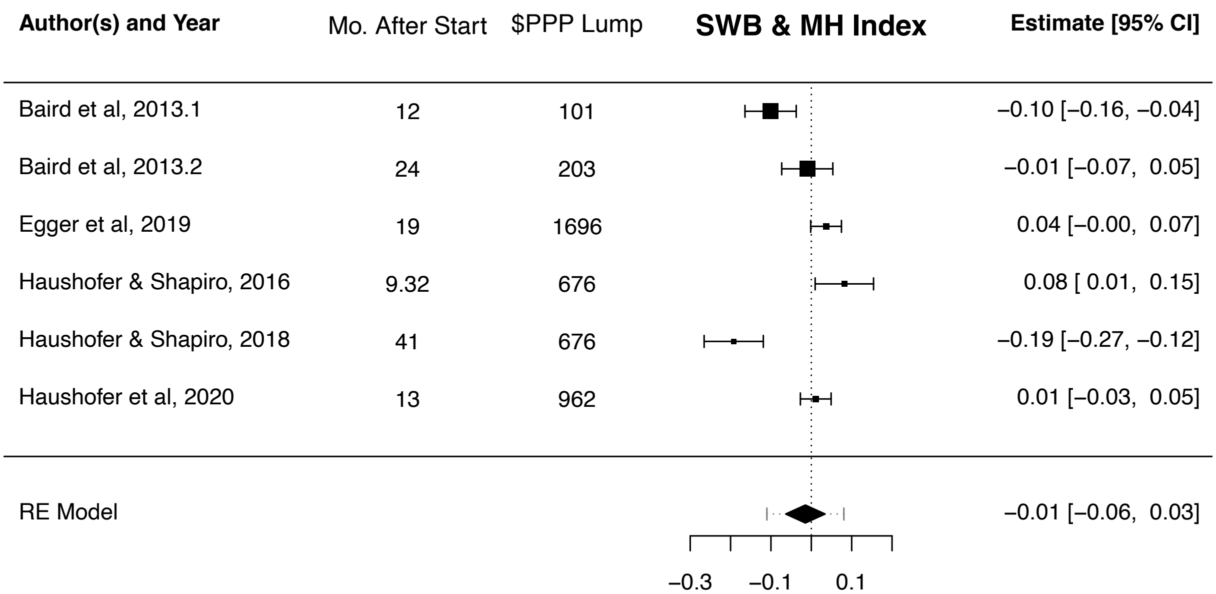
4 Discussion

Our results represent a systematic synthesis and meta-analysis of all the available causal evidence of the impact of CTs on mental health and subjective wellbeing in low- and middle-income contexts. In sum, we find that CTs, on average, have a positive effect on MH and SWB indicators among recipients. More precisely, we find an average impact of about 0.10 SDs. Additionally, we observe that the effects

[17] Studies in which this is the case are Egger et al. (2019), Haushofer & Shapiro (2016), Haushofer & Shapiro (2018), Haushofer et al. (2020a), and Haushofer et al. (2020b).

[18] There is some further variation in how spillovers are accounted for. Most spillovers are from *within* the (treated) village. An exception is Egger et al. (2019), who look at spillovers *across* treated and untreated villages. Most studies identify the spillover treatment categorically with geographic proximity of a non-recipient to a recipient (usually in the same village). An exception is Haushofer, Reisinger and Shapiro (2019) where the spillover is formulated as how many recipients live near a non-recipient (proxied by increases in average wealth of the village). Thus, it is the only study that looks at the degree of spillover intensity.

Figure 5. Forest Plot of Spillover Effects



Note: A forest plot of the studies in our sample that include MH and SWB spillovers. A random effects multilevel model (with levels for study and sample) with robust standard errors (clustered at the level of the program) shows an effect of -0.01. The 95% confidence interval overlaps with zero. All of the CTs except Baird et al., (2013a) were implemented by GiveDirectly, an NGO.

of CTs appear to only dissipate slowly over time. The estimated effects were substantially larger for unconditional CTs. Our results were consistent across a battery of robustness tests and the observed effects did not vary according to study design (RCT and quasi-experimental). Notably, our results indicate that CTs are less efficacious in Latin America, which may be explained by the prevalence of CCTs (as opposed to UCTs) in that region. We find no significant evidence of negative spillover effects on non-recipients. However, spillover effects were rarely reported upon (n=4). We therefore encourage more research on this aspect going forward.[19]

4.1 Limitations

Like most meta-analyses, using study averages for moderator variables means that we do not capture within-study variation, which limits the precision of our estimates. Some of our insignificant results may be due to low power. This could be remedied if we had access to the data at the level of the individual. Some of the studies we include have open access data policies (Haushofer et al., 2016; Paxson & Schady, 2010; Ohrnberger et al., 2020a). An individual level analysis may therefore be possible but was outside the scope of this paper. Another limitation arises from the paucity of longitudinal follow-ups. There was only one study in our sample that followed up more than five years after the cash transfer began (Blattman et al., 2020). This limits what we can say about the long run effects of CTs on SWB and MH. There is also only one study that discusses effects of CTs on the SWB and MH of individuals who share a household with recipients.[20] Unfortunately, our evidence was limited to spillovers relating to non-recipients in the geographic proximity of recipients.

An important feature of this meta-analysis is that it does not offer evidence on the *mechanisms* by which CTs improve SWB and MH. One possible mechanism worth investigating is whether the effect on

[19] Baird et al. (2014) make some useful recommendations concerning this research direction.
[20] Baird et al., (2013a) finds positive albeit insignificant effects of a CT on recipients' siblings.

SWB or MH stems from increased consumption relative to one's peers or from previous levels of consumption. Indeed, there is a rich set of possible mediators and moderators, and we have only analyzed a small subset of them.

Finally, we know of no other systematic review and meta-analysis which estimates the total effect of an intervention on SWB and MH. This limits our capacity to compare the cost-effectiveness of CTs to other poverty alleviation or health interventions.

4.2 Implications and suggestions for future research

Although there is some preliminary evidence that CTs are cost-effective interventions in LMICs compared to a USAID workforce readiness program (McIntosh & Zeitlin, 2020) and psychotherapy (Haushofer, Shapiro & Mudida, 2020), the work done to compare the cost-effectiveness of interventions in terms of SWB and MH is scarce, especially in LMICs. Our meta-analysis contributes to this literature by providing a comprehensive empirical foundation to compare the cost-effectiveness of cash transfers to interventions aimed at improving MH or SWB. Although limited, the practical implications of our meta-analysis are clear: direct cash transfers improve the wellbeing of poor recipients in LMICs.

There are several research questions to be pursued in future work on subjective wellbeing and mental health. What are the long run (5+ years) effects of CTs? What are the effects on a recipient's household and community? Relevant spillover data should be collected in RCTs or evaluated in quasi-experiments. The costs of CTs and other poverty alleviation interventions should be published. For instance, since a UCT requires less administration (as there are no conditions to monitor), it seems likely that UCTs are cheaper and, based on our results, more effective than CCTs. However, there appears to be no available evidence to answer this question. More broadly, we recommend a greater inclusion of SWB and MH data in intervention evidence collection efforts such as Aid Grade.[21]

5 Conclusion

Cash transfers have a small[22] ($d<0.2$) but significant and lasting effect on wellbeing with only mild adaptation effects. Although modest in size, if SWB and MH measure wellbeing more directly than other indicators, these reported improvements are an indicator of genuine success. How important CTs are as a means of improving wellbeing depends on their cost-effectiveness relative to the alternatives. Even if effect sizes are small, CTs may nevertheless be among the most efficient ways of improving lives. There is no evidence that CTs have, on average, significant negative spillover effects within the community they are implemented in. However, the evidence on this is scarce, meriting further research on the topic.

[21] Aid Grade synthesizes research from international development. http://www.aidgrade.org.

[22] With medium = 0.4 and large = 0.8 as established by Cohen (1992) in the context of psychological effects.

Data availability: As this is a systematic review and meta-analysis, all data is already available in published and unpublished manuscripts. The extracted data used to produce our results are available upon reasonable request.

Code availability: The statistical code used to create the results and figures in the manuscript and appendices will be made available upon reasonable request.

Author contributions: JM screened abstracts, performed the primary data extraction, contributed to drafting the manuscript, and contributed to the data analysis. ABM double-checked a subset of the data extraction, contributed to drafting the manuscript and provided expertise on systematic reviewing and meta-analyses. CK screened abstracts, double-checked a subset of the data extraction, contributed to drafting the manuscript, and contributed to performing the data analysis.

References

Agbenyo, F., Galaa, S. Z., & Abiiro, G. A. (2017). Challenges of the targeting approach to social protection: An assessment of the Ghana Livelihood Empowerment against Poverty Programme in the Wa Municipality of Ghana. *Ghana Journal of Development Studies*, 14(1), 19-38.

Alves, F. J. O., Machado, D. B., & Barreto, M. L. (2018). Effect of the Brazilian cash transfer programme on suicide rates: a longitudinal analysis of the Brazilian municipalities. *Social Psychiatry and Psychiatric Epidemiology, 54(5), 599-606.*

Angeles, G., de Hoop, J., Handa, S., Kilburn, K., Milazzo, A., Peterman, A., and Team, M. S. C. T. E. (2019). Government of Malawi's unconditional cash transfer improves youth mental health. *Social Science & Medicine*, 225:108–119.

Baird, S., Bohren, J. A., McIntosh, C., and Özler, B. (2014). Designing experiments to measure spillover effects. *The World Bank Policy Research Working Paper.*

Baird, S., De Hoop, J., and Özler, B. (2013a). Income shocks and adolescent mental health. *Journal of Human Resources*, 48(2):370–403.

Baird, S., Ferreira, F. H., Özler, B., and Woolcock, M. (2013b). Relative effectiveness of conditional and unconditional cash transfers for schooling outcomes in developing countries: a systematic review. *Campbell Systematic Reviews*, 9(1):1–124.

Baird, S., McKenzie, D., & Özler, B. (2018). The effects of cash transfers on adult labor market outcomes. *IZA Journal of Development and Migration*, 8(1), 22.

Banerjee, A. V., Hanna, R., Kreindler, G. E., and Olken, B. A. (2017). Debunking the stereotype of the lazy welfare recipient: Evidence from cash transfer programs. *The World Bank Research Observer*, 32(2):155– 184.

Baranov, V., Cameron, L., Contreras Suarez, D., & Thibout, C. (2020). Theoretical Underpinnings and Meta-analysis of the Effects of Cash Transfers on Intimate Partner Violence in Low-and Middle-Income Countries. *The Journal of Development Studies*, 1-25.

Behrman, J. R. and Parker, S. W. (2010). Impacts of conditional cash transfer programs in education. *Conditional cash transfers in Latin America*, 191–211.

Benjamin, D. J., Cooper, K. B., Heffetz, O., and Kimball, M. S. (2020). Self-reported wellbeing indicators are a valuable complement to traditional economic indicators but aren't yet ready to compete with them. *Behavioural Public Policy*. 4(2):198-209

Billingham, S. A., Whitehead, A. L., and Julious, S. A. (2013). An audit of sample sizes for pilot and feasibility trials being undertaken in the United Kingdom registered in the United Kingdom clinical research network database. *BMC Medical Research Methodology*, 13(1):104.

Bishop, D. V. and Thompson, P. A. (2016). Problems in using p-curve analysis and text-mining to detect rate of p-hacking and evidential value. *PeerJ*, 4:e1715.

Blattman, C., Fiala, N., and Martinez, S. (2020). The long-term impacts of grants on poverty: Nine-year evidence from Uganda's youth opportunities program. *American Economic Review: Insights*, 2(3):287–304.

Bond, T. N. and Lang, K. (2019). The sad truth about happiness scales. *Journal of Political Economy*, 127(4):1629–1640.

Borenstein, M., Cooper, H., Hedges, L., and Valentine, J. (2009). Effect sizes for continuous data. *The Handbook of Research Synthesis and Meta-analysis*, 2:221–235.

Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2):97–111.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

Buller, A. M., Peterman, A., Ranganathan, M., Bleile, A., Hidrobo, M., and Heise, L. (2018). A mixed method review of cash transfers and intimate partner violence in low-and middle-income countries. *The World Bank Research Observer*, 33(2):218–258.

Busseri, M. A. and Sadava, S. W. (2011). A review of the tripartite structure of subjective wellbeing: Implications for conceptualization, operationalization, analysis, and synthesis. *Personality and Social Psychology Review*, 15(3):290–314.

Chanfreau, J. and Burchardt, T. (2008). Equivalence scales: rationales, uses and assumptions. *Scottish Government, Edinburgh*.

Chen, X., Wang, T., and Busch, S. H. (2019). Does money relieve depression? Evidence from social pension expansions in China. *Social Science & Medicine*, 220:411–420.

Clark, A. E. (2017). Happiness, income and poverty. *International Review of Economics*, 64(2):145–158.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, 1(3):98–101.

Crea, T. M., Reynolds, A. D., Sinha, A., Eaton, J. W., Robertson, L. A., Mushati, P., Dumba, L., Mavise, G., Makoni, J., Schumacher, C. M., et al. (2015). Effects of cash transfers on children's health and social protection in sub-Saharan Africa: differences in outcomes based on orphan status and household assets. *BMC Public Health*, 15(1):511.

Davis, B., Handa, S., Hypher, N., Rossi, N. W., Winters, P., and Yablonski, J. (2016). *From evidence to action: the story of cash transfers and impact evaluation in sub Saharan Africa*. Oxford University Press.

Deaton, A. (2008). Income, health, and wellbeing around the world: Evidence from the Gallup world poll. *Journal of Economic Perspectives*, 22(2):53–72.

De Neve, Jan-Emmanuel, Andrew E. Clark, Christian Krekel, Richard Layard, and Gus O'Donnell. 2020. Taking a Wellbeing Years Approach to Policy Choice. *BMJ* 371.

Diener, E. (2009). Subjective wellbeing. In *The Science of Wellbeing*, pages 11–58. Springer.

Diener, E., Lucas, R. E., and Oishi, S. (2018). Advances and open questions in the science of subjective wellbeing. *Collabra. Psychology*, 4(1).

Egger, D., Haushofer, J., Miguel, E., Niehaus, P., and Walker, M. W. (2019). General equilibrium effects of cash transfers: experimental evidence from Kenya. *Technical report, National Bureau of Economic Research*.

Ellis, F. (2012). 'We are all poor here': Economic difference, social divisiveness and targeting cash transfers in sub-Saharan Africa. *Journal of Development Studies*, 48(2):201–214.

Evans, D. K. and Popova, A. (2014). Cash transfers and temptation goods: A review of global evidence. *The World Bank Policy Research Working Paper*.

Eyal, K., & Burns, J. (2019). The parent trap: cash transfers and the intergenerational transmission of depressive symptoms in South Africa. *World Development,* 117, 211-229.

Fisher, E., Attah, R., Barca, V., O'Brien, C., Brook, S., Holland, J., Kardan, A., Pavanello, S., and Pozarny, P. (2017). The livelihood impacts of cash transfers in sub-Saharan Africa: Beneficiary perspectives from six countries. *World Development*, 99:299–319.

Frijters, P., Clark, A. E., Krekel, C., and Layard, R. (2020). A happy choice: Wellbeing as the goal of government. *Behavioural Public Policy*, 4(2):126–165.

Galama, T. J., Morgan, R., and Saavedra, J. E. (2017). Wealthier, happier and more self-sufficient: When anti-poverty programs improve economic and subjective wellbeing at a reduced cost to taxpayers. *Technical report, National Bureau of Economic Research*.

Goulet-Pelletier, J.-C. and Cousineau, D. (2018). A review of effect sizes and their confidence intervals, part i: The Cohen's d family. *The Quantitative Methods for Psychology*, 14(4):242–265.

Handa, S., Daidone, S., Peterman, A., Davis, B., Pereira, A., Palermo, T., and Yablonski, J. (2018). Mythbusting? Confronting six common perceptions about unconditional cash transfers as a poverty reduction strategy in Africa. *The World Bank Research Observer*, 33(2):259–298.

Haushofer, J., Chemin, M., Jang, C., & Abraham, J. (2020b). Economic and psychological effects of health insurance and cash transfers: Evidence from a randomized experiment in Kenya. *Journal of Development Economics*, 144, 102416.

Haushofer, J. and Fehr, E. (2014). On the psychology of poverty. *Science*, 344(6186):862–867.

Haushofer, J., Mudida, R., and Shapiro, J. (2020a). The comparative impact of cash transfers and psychotherapy on psychological and economic well-being. *NBER Working Paper*

Haushofer, J., Reisinger, J., and Shapiro, J. (2019). Is your gain my pain? Effects of relative income and inequality on psychological well-being. *Technical report, Working Paper*.

Haushofer, J. and Shapiro, J. (2016). The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya. *The Quarterly Journal of Economics*, 131(4):1973–2042.

Haushofer, J. and Shapiro, J. (2018). The long-term impact of unconditional cash transfers: Experimental evidence from Kenya. *Busara Center for Behavioral Economics, Nairobi, Kenya*.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., and Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

IntHout, J., Ioannidis, J. P., Rovers, M. M., and Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ open*, 6(7).

Jebb, A. T., Tay, L., Diener, E., and Oishi, S. (2018). Happiness, income satiation and turning points around the world. *Nature Human Behaviour*, 2(1):33–38.

Kabeer, N. and Waddington, H. (2015). Economic impacts of conditional cash transfer programmes: A systematic review and meta-analysis. *Journal of Development Effectiveness*, 7(3):290–303.

Kaiser, C. and Vendrik, M. C. M. (2020). How threatening are transformations of happiness scales to subjective wellbeing research? *INET Oxford Working Paper*, No. 2020-19.

Karimli, L., Ssewamala, F. M., Neilands, T. B., Wells, C. R., and Bermudez, L. G. (2019). Poverty, economic strengthening, and mental health among aids orphaned children in Uganda: Mediation model in a randomized clinical trial. *Social Science & Medicine*, 228:17–24.

Kühberger, A., Fritz, A., and Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PloS one*, 9(9):e105825.

Lagarde, M., Haines, A., and Palmer, N. (2007). Conditional cash transfers for improving uptake of health interventions in low-and middle-income countries: A systematic review. *Jama*, 298(16):1900–1910.

MacAuslan, I. and Riemenschneider, N. (2011). Richer but resented: What do cash transfers do to social relations? *IDS Bulletin*, 42(6):60–66.

McIntosh, C. and Zeitlin, A. (2020). Using household grants to benchmark the cost effectiveness of a usaid workforce readiness program. *arXiv preprint. arXiv:2009.01749*.

Millán, T. M., Barham, T., Macours, K., Maluccio, J. A., and Stampini, M. (2019). Long-term impacts of conditional cash transfers: review of the evidence. *The World Bank Research Observer*, 34(1):119–159.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. International Journal of Surgery, 8(5), 336–341.

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., et al. (2015). Process evaluation of complex interventions: Medical research council guidance. *BMJ*, 350.

Ohrnberger, J., Fichera, E., Sutton, M., and Anselmi, L. (2020a). The effect of cash transfers on mental health–new evidence from South Africa. *BMC Public Health*, 20:1–13.

Ohrnberger, J., Fichera, E., Sutton, M., & Anselmi, L. (2020b). The worse the better? Quantile treatment effects of a conditional cash transfer programme on mental health. *Health Policy and Planning.*

Owusu-Addo, E., Renzaho, A. M., and Smith, B. J. (2018). The impact of cash transfers on social determinants of health and health inequalities in sub-Saharan Africa: a systematic review. *Health Policy and Planning*, 33(5):675–696.

Paxson, C. and Schady, N. (2010). Does money matter? The effects of cash transfers on child development in rural Ecuador. *Economic Development and Cultural Change*, 59(1):187–229.

Pettifor, A., Bekker, L. G., Hosek, S., DiClemente, R., Rosenberg, M., Bull, S., ... & Cowan, F. (2013). Preventing HIV among young people: research priorities for the future. *Journal of acquired immune deficiency syndromes (1999),* 63(0 2), S155.

Plant, M. (2020). A Happy Possibility About Happiness (And Other Subjective) Scales: An Investigation and Tentative Defence of the Cardinality Thesis. *Happier Lives Institute working paper.*

Powdthavee, N. (2010). How much does money really matter? Estimating the causal effects of income on happiness. *Empirical Economics*, 39(1):77–92.

Powell-Jackson, T., Pereira, S. K., Dutt, V., Tougher, S., Haldar, K., and Kumar, P. (2016). Cash transfers, maternal depression and emotional wellbeing: Quasi-experimental evidence from India's janani suraksha yojana programme. *Social Science & Medicine*, 162:210–218.

Ridley, M. W., Rao, G., Schilbach, F., and Patel, V. H. (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. *Technical report, National Bureau of Economic Research.*

Riley, R. D., Higgins, J. P., and Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *BMJ* 342.

Sassenberg, K. and Ditrich, L. (2019). Research in social psychology changed between 2011 and 2016: Larger sample sizes, more self-report measures, and more online studies. *Advances in Methods and Practices in Psychological Science*, 2(2):107–114.

Schilbach, F., Schofield, H., and Mullainathan, S. (2016). The psychological lives of the poor. *American Economic Review*, 106(5):435–40.

Simonsohn, U., Nelson, L. D., and Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of experimental psychology: General*, 143(2):534.

Stevenson, B. and Wolfers, J. (2013). Subjective wellbeing and income: Is there any evidence of satiation? *American Economic Review*, 103(3):598–604.

Tampubolon, G. and Hanandita, W. (2014). Poverty and mental health in Indonesia. *Social Science & Medicine*, 106:20–27.

Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software*, 36(3):1–48.

Villa, J. M. and Niño-Zarazúa, M. (2019). Poverty dynamics and graduation from conditional cash transfers: a transition model for Mexico's progresa-oportunidades-prospera program. *The Journal of Economic Inequality*, 17(2):219–251.

Vivalt, E. (2015). How much can we generalize from impact evaluations? *Journal of the European Economic Association*.

Appendix A Search string

Our boolean search string was as follows:

(Cash transfer OR "non-contributory pension*" OR "enterprise grant*") AND (satisfaction OR depression OR happiness OR "mental health" OR mental OR happy OR "subjective wellbeing" OR eudai* OR "subjective well*" OR subjective OR "self report*" OR SWB OR emotion* OR "positive emotion*" OR "negative emotion*" OR anxiety OR stress OR "positive affect" OR affective OR "negative affect" OR PHQ OR PHQ-9 OR SWLS OR GHQ OR GHQ-12 OR CES-D OR PERMA OR K10 OR trust OR "social cohesion" OR "social bonds" OR "interpersonal trust" OR "social capital" OR "community building")*

Appendix B Further tables

Table A1. Additional moderators of CTs' effects on MH and SWB

	Model 1	Model 2	Model 3	Model 4
Intercept	0.046* (0.019)	0.066** (0.021)	0.084*** (0.020)	0.052** (0.018)
Measure of SWB	0.042* (0.016)			
Compound measure of SWB & MH	0.070*** (0.009)			
Monthly value in 100$ PPP	0.063+ (0.036)	0.100** (0.034)	0.070+ (0.036)	0.071* (0.033)
CT deployed in Asia		-0.010 (0.023)	0.001 (0.027)	
CT deployed in Latin America		-0.061** (0.020)	-0.045+ (0.022)	
CT is CCT			-0.040* (0.016)	
CT is RCT				-0.015 (0.018)
Number of outcomes	99	99	99	99
Number of studies	37	37	37	37

Note: ***p < 0.001; **p < 0.01; *p < 0.05; +p < 0.1. Robust standard errors are clustered at the level of the program.

Table A2. Alternative specifications for CT size

	Model 1	Model 2	Model 3
Intercept	0.080***	0.065+	0.182***
	(0.013)	(0.034)	(0.026)
Years since CT began	-0.016*	-0.016*	-0.018**
	(0.005)	(0.007)	(0.006)
Monthly value relative to GDPpc	0.288**		
	(0.087)		
Log monthly value in $PPP		0.016	
		(0.011)	
Log monthly value relative to previous income			0.034**
			(0.009)
Number of outcomes	97	97	97
Number of studies	35	35	35

Note: ***p < 0.001; **p < 0.01; *p < 0.05; +p < 0.1. Robust standard errors are clustered at the level of the program.

Table A3. Summary of Included Studies

Citation	Title	Program	Country	Payment frequency	Design	Type	Scale	Measures	N	Mo. since start	Abs. mo. value	Total value	Rel. value	Baseline year	HH size
Natali et al, 2018	Does money buy happiness? Evidence from an unconditional cash transfer in Zambia	Zambian Child Grant (ZCG)	Zambia	Bi-monthly	cRCT	UCT	SWB	Happy	2203	33; 45	$24	$760; $1035	27%	2010	5.75
Kilburn et al, 2018	Paying for Happiness: Experimental Results from a Large Cash Transfer Program in Malawi	Malawi Social Cash Transfer (SCTP)	Malawi	Bi-monthly	cRCT	UCT	SWB	QoL, LS, Happy	3365	12	$33	$396	18%	2013	4.6
Kilburn et al, 2019	Cash Transfers, Young Women's Economic Well-Being, and HIV Risk: Evidence from HPTN 068	HIV Prevention Trials Network study number 068 (HPTN 068)	South Africa	Monthly	RCT	CCT	SWB; MH	CESD20	2533	24	$20	$469	22%	2012	6.15
Kilburn et al, 2016	Effects of a large-scale unconditional cash transfer program on mental health outcomes of young people in Kenya	Orphans & Vulnerable Children (CT-OVC)	Kenya	Monthly	cRCT	UCT	SWB, MH	Optimism, CESD10	2006	48	$54	$2576	21%	2007	5.5
Baird et al, 2013	Income Shocks and Adolescent Mental Health	(Nearly) Unique to Study	Malawi	Monthly	cRCT	UCT & CCT	MH	GHQ-12, MHI-5	2066	12; 24	$8	$100; $200	10%	2008	-
Paxson et al, 2010	Does Money Matter? The Effects of Cash Transfers on Child Development in Rural Ecuador	Bono de Desarrollo Humano	Ecuador	Monthly	RA	UCT (28% thought CCT)	MH	CESD	1430	17	$15	$126	10%	2004	4.78
Handa et al, 2014	Subjective Well-being, Risk Perceptions and Time Discounting: Evidence from a large-scale cash transfer programme	Orphans & Vulnerable Children (CT-OVC)	Kenya	Monthly	cRCT	UCT	SWB	Enjoyment, LS, enjoyment + positive feelings	1805	24	$85	$2034	14%	2007	5.5
Angeles et al. 2019	Government of Malawi's unconditional cash transfer improves youth mental health	Malawi Social Cash Transfer Program (SCTP)	Malawi	Bi-monthly	cRCT	UCT	MH	CESD20, CESDbinary	1366	24	$7	$156	18%-23%	2013	5.7
Haushofer & Shapiro, 2016 & 2018	The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya; The long-term impact of unconditional cash transfers: experimental evidence from Kenya	GiveDirectly	Kenya	Monthly (9 or 7) or lump	cRCT	UCT	MH, SWB	PWB, WVS Happy, WVS LS, CESD10	1474	9.32; 41	$118; $23.63	$709	37%	2012	5.14
Haushofer et al, 2020a	The Comparative Impact of Cash Transfers and Psychotherapy on Psychological and Economic Wellbeing	GiveDirectly	Kenya	Weekly (5) or lump	cRCT	UCT	MH, SWB	PWB, WVS Happy, WVS LS, GHQ12	5309	14 (3-28)	$83	$1076	66%	2017	4
Egger et al, 2019	General equilibrium effects of cash transfers: experimental evidence from Kenya	GiveDirectly	Kenya	3 payments over 12 months	cRCT	UCT	MH, SWB	PWB	5432	19 (9-31)	$98	$1871	75%	2015	4.3
Haushofer et al, 2020b	Economic and psychological effects of health insurance and cash transfers: Evidence from a randomized experiment in Kenya	GiveDirectly	Kenya	Lump	RCT	UCT	MH, SWB	Happy, LS, CESD20	690	12 (SD ~1)	$22	$564	3%	2011	-
Blattman et al, 2017	Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia	Unique to Study	Liberia	Lump	RCT	UCT	MH	Positive MH, Depression, anxiety and distress, LS, Happy	470	1;12	$30	$360	25%	2011	3.8
Blattman et al., 2020	The Long-Term Impacts of Grants on Poverty: 9-Year Evidence from Uganda's Youth Opportunities Program	Ugandan Govt. Skills Grant	Uganda	Lump	cRCT	UCT: Enterprise Grant	MH	Depression, Distress	1981	108	$9	$944	41%	2008	5.86
Powell-Jackson et al, 2016	Cash transfers, maternal depression and emotional wellbeing: Quasi-experimental evidence from India's Janani Suraksha Yojana programme	Janani Suraskha Yojana (JSY)	India	Lump	ED	CCT	SWB, MH	Happy, K10, Worried	1695	11.6 (SD 6.5)	$6	$74	~5%	2015	5.7

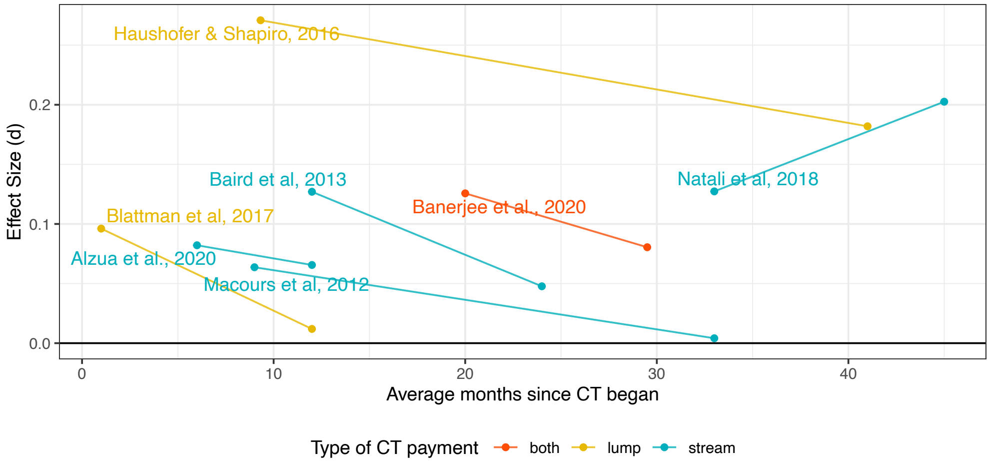
Macours et al, 2012	Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment	Atencion a Crisis Pilot	Nicaragua	Bi-monthly	RA	CCT	MH	CESD20	469 & 576	9; 33	$45; $16	$145-$385	~15%-26%.	2008	6.05
Galama et al, 2017	Wealthier, Happier and More Self-Sufficient: When Anti-Poverty Programs Improve Economic and Subjective Wellbeing at a Reduced Cost to Taxpayers	Familias en Accion Urbano	Colombia	Monthly	RD	CCT	SWB	LS, Happy, LS10domains	563	~36	$22	$338	10%	2010	3.95
Salinas-Rodríguez et al., 2014	Impact of the Non-Contributory Social Pension Program 70 y más on Older Adults' Mental Wellbeing	70 y más	Mexico	Bi-monthly	Match & DD	UCT: Pension	MH	GDS-15	2241	12	$57	$690	4%	2007	5.16
Fernald & Hidrobo, 2011	Effect of Ecuador's cash transfer program (Bono de Desarrollo Humano) on child development in infants and toddlers: A randomized effectiveness trial	Bono de Desarrollo Humano	Ecuador	Monthly	RA	UCT	MH	CESD20	1196	24	$31	$744	8% (6%-10%)	2004	5
Lopez Boo & Creamer, 2019	Cash, Conditions, and Child Development: Experimental Evidence from a Cash Transfer Program in Honduras	Bono 10,000	Honduras	Lump	RA	CCT	SWB	LS (RSE-10)	791	9	$73	$658	3%	2012	5.2
Ozer et al, 2011	Does alleviating poverty affect mothers' depressive symptoms? A quasi-experimental investigation of Mexico's Oportunidades programme	Oportunidades	Mexico	Bi-monthly	Match	CCT	MH	CESD20	6343	51 (42-60)	$43	$2193	~20%-25%	2003	4.32 (2.0)
Ozer et al., 2008	Effects of a Conditional Cash Transfer Program on Children's Behavior Problems	Oportunidades	Mexico	Bi-monthly	Match	CCT	MH	BPI-sub	945	51 (42-60)	$43	$2193	~20%-25%	2003	4.32 (2.0)
Han & Gao, 2020	Does Welfare Participation Improve Life Satisfaction? Evidence from Panel Data in Rural China	Rural Dibao	China	Monthly	Match & DD	UCT	SWB	LS	12761	-	$36	-	12%	2012	4.7
Bando et al., 2017	The Effects of Non-Contributory Pensions on Material and Subjective Well Being	Pension 65	Peru	Bi-monthly	RD	UCT: Pension	SWB, MH	Self-worth, Empowerment, SWB index 8, GDS-15	3342	36	$70	$2526	40%	2015	2.84 (AE)
Galiani et al., 2016	Non-contributory pensions	Adultos Mayores	Mexico	Bi-monthly	DD	UCT: Pension	MH	GDS-15	1950	12	$59	$708	14%	2009	5.6 (AE)
Chen et al., 2019	Does money relieve depression? Evidence from social pension expansion in China	China's New Rural Pension Scheme (NRPS)	China	Monthly	IV	UCT: Pension	MH	CESD20	2701	21.12 (SD 11.5)	$59	$ 708	9%	2011	2.87
Heath et al., 2020	Cash transfers, polygamy, and intimate partner violence: Experimental evidence from Mali	Programme de Filets Sociaux	Mali	Quarterly	cRCT	UCT	MH	Anxiety	1143	15	$47	$698	9%	2014	8.32
Ohrnberger et al., 2020	The effect of cash transfers on mental health – new evidence from South Africa	Child Support Grant	South Africa	Monthly	IV: Age eligibility	UCT	MH	CESD10	10925	-	$48	-	20%-25%	2008	6.43
Filmer & Schady, 2009	School Enrollment, Selection and Test Scores	CESSP Scholarship Program (CSP)	Cambodia	Quarterly	RD	CCT	MH	GHQ	3225	15	$22	$325	3%	2006	5
Bhalla, 2017	Chapter 3: Mediation Analysis of The Impact of An Unconditional Cash Transfer on Subjective Wellbeing	Harmonized Social Cash Transfer (HSCT)	Zimbabwe	Monthly	Match & DD	UCT	SWB	SWLS, Happy, Positive	2630	12	$46	$549	20%	2013	5.18
Ohrnberger et al., 2020b	The worse the better? Quantile treatment effects of a conditional cash transfer programme on mental health. Health Policy and Planning.	Malawi Incentive Program	Malawi	Lump	RCT	CCT	MH	SF-12	790	12	$2	$27	9%	2006	6.5
Berhane et al., 2015	Evaluation of The Social Cash Transfer Pilot Programme, Tigray Region, Ethiopia	Social Cash Transfer Pilot Programme	Ethiopia	Monthly	Match & DD	CCT	MH	SRQ-20	2080	24	$28	$665	24%	2012	2.42

Asfaw et al., 2016	Productive Impact of Ethiopia's Social Cash Transfer Pilot Programme (also Tigray). P.133	Social Cash Transfer Pilot Programme	Ethiopia	Monthly	Match & DD	CCT	SWB	LS (how things have been going)	2908	24	$32	$770	29%	2012	2.55
Daidone et al., 2015	Social Networks and Risk Management in Ghana's Livelihood Empowerment against Poverty Programme	Livelihood empowerment against poverty (LEAP)	Ghana	Bi-monthly	Match & DD	UCT	SWB	Happy	1504	24	$16	$390	11%	2010	3.86
Alzua et al., 2020	Mental Health Effects of an Old Age Pension: Experimental Evidence for Ekiti State in Nigeria	Ekiti Pilot Old Age Pension	Nigeria	Monthly	cRCT	UCT	SWB & MH	LS (index), GDS-15, MH (index)	3286	12	$55	$330; $661	29%	2013	3.03
McIntosh & Zeitlin, 2020	Using Household Grants to Benchmark the Cost Effectiveness of a USAID Workforce Readiness Program	GiveDirectly	Rwanda	Lump	RCT	UCT	SWB & MH	LS (index), MH (index)	1160	9	$96; $125; $153; $228	$866; $1122, $1374; $2048	99%; 129%; 158%; 235%	2018	5
Banerjee et al., 2020	Effects of a Universal Basic Income during the pandemic	GiveDirectly	Kenya	Monthly or Lump	cRCT	UCT	MH	CES-D	8330	20; 29.5	$57; $45, $52	$1673; $1381; $1260	30%, 34%, 37%	2018	4.9

Note: Cells with multiple values represent values for the first and second follow-ups or multiple treatment arms. cRCT = cluster randomized control trial, UCT = unconditional cash transfer, CCT = conditional cash transfer, MH = mental health, SWB = subjective wellbeing, PWB = psychological wellbeing, CESD = center for epidemiological studies depression inventory, LS = life satisfaction, SF-12 = short form (mental health), SWLS = satisfaction with life scale, GHQ = general health questionnaire, MHI = mental health inventory, GDS = geriatric depression scale, BPI = behavioral problems inventory (anxiety and depression subscale), RSE = Rosenberg self-esteem scale (first question which was used is a life satisfaction question), K10 = Kessler depression scale, WVS = world values survey, QoL = quality of life, AE = adult equivalent individuals, Happy = self-reported happiness, Match = propensity score matching, DD = difference-in-difference estimation.

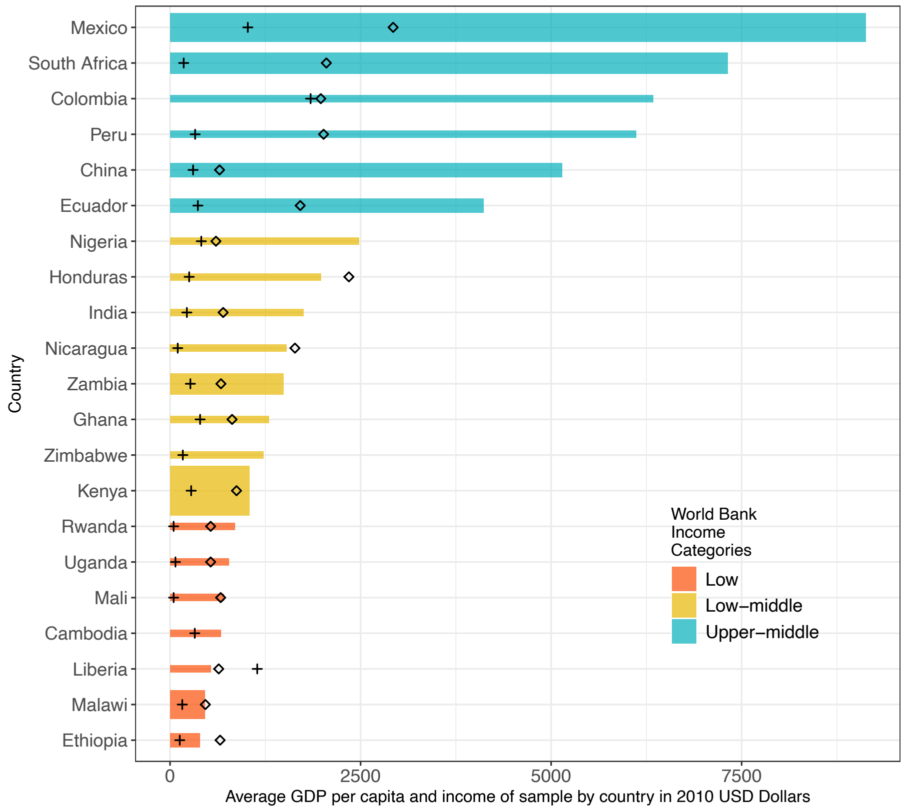
Appendix C Figures

Figure A1. Effect sizes for studies with multiple follow-ups



Note: Six out of seven studies with multiple follow-ups show a decline in effect size except Natali et al., (2018)

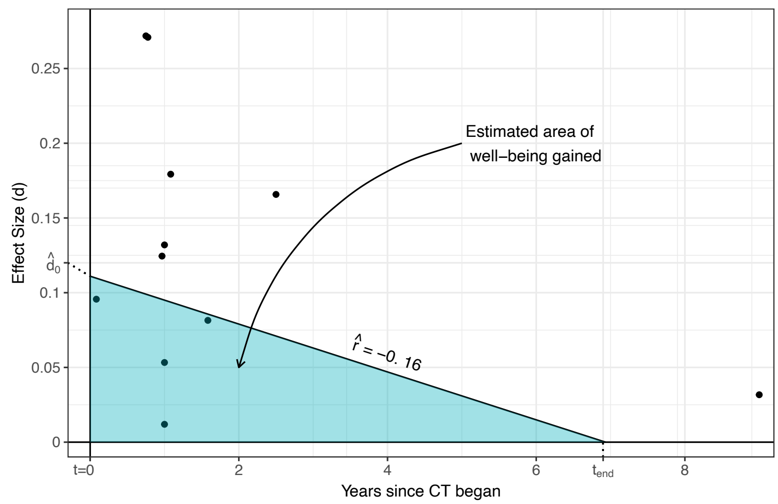
Figure A2. GDP per capita in the countries the studies took place in



Note: Width of bar plot is proportional to the number of studies that were conducted in that country (the most studies were conducted in Kenya). Diamonds indicate the poverty line. Crosses indicate the average income of the sample. Both indicate less variation in income of the extreme poor than variation in GDPpc alone would suggest.

Appendix D Wellbeing-Adjusted Life Years Analyses

Figure A3. Estimated total effect of $1,000 PPP Lump sum CT on well-being.



Note: The slope of the hypotenuse of the triangle is the same as the decay effect depicted by Model 4 in Table 2. The area of the triangle is equivalent to the definite integral. This graph differs from Figure 4.b because it does not include studies with stream payments and the slope is lower.

To further aid in the interpretation of our results, we illustrate how our estimate could potentially be used in a cost-effectiveness analysis to calculate "wellbeing-adjusted life years". First, we define a $\Delta WELLBY$ to denote a one SD change in wellbeing lasting for one year (see Frijters et al. 2020 for a similar definition).[23]

How many $\Delta WELLBY$ is a lump-sum payment of $1,000 estimated to buy? Assume, as in Model 4 of Table 2, that the instantaneous effect of a lump-sum CT linearly decreases over time. Further assume that after the time at which the effect is estimated to become zero, the effect will not further decrease (and thereby become negative). Call this time t_{end} . Let $t = 0$ at the start of the CT. For a lump-sum payment of $1,000, the estimated effect at $t = 0$ is given by $\hat{d}_0 = \hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_4 0.42$.[24] Here, $\hat{\beta}_0$, $\hat{\beta}_2$, and $\hat{\beta}_4$ respectively denote the estimated intercept and coefficients on "CT is Lump" and "Monthly Value in $100 PPP" from Model 4 in Table 2. Finally, the rate at which the effect decays over time is given by $\hat{r} = \hat{\beta}_1 + \hat{\beta}_3$, where $\hat{\beta}_1$ and $\hat{\beta}_3$ denote "Years since CT began" and "Years Since * CT is Lump", respectively.

We can then calculate the total effect as $\Delta WELLBY = \int_0^{t_{end}} \hat{d}_0 + \hat{r}t \, dt = d_0 t_{end} + \frac{1}{2}\hat{r}t_{end}^2$. Notice that in the present case $t_{end} = -\frac{\hat{d}_0}{\hat{r}}$. Thus, $\Delta WELLBY = -\frac{\hat{d}_0^2}{\hat{r}} + \frac{\hat{r}\hat{d}_0^2}{2\hat{r}^2} = -\frac{\hat{d}_0^2}{2\hat{r}} = -\frac{(\widehat{\beta_0} + \widehat{\beta}_2 + \widehat{\beta}_4 0.420)^2}{2(\widehat{\beta_1} + \widehat{\beta}_3)}$. Using estimates from Model 4 we get $\Delta WELLBY = \frac{(0.097 - 0.024 + 0.071 * 0.420)^2}{2(0.017 - 0.001)} = 0.330$.

An intuitive expression of $-\frac{\hat{d}_0^2}{2\hat{r}}$ in our special case is given by $\frac{\hat{d}_0 \, t_{end}}{2}$. Respectively interpreting $\hat{d}_0$ and t_{end} as the height and base of the triangle shown in Figure A3, that expression gives the area of such a triangle. Of course, Figure A3 shows that such calculations are somewhat imprecise. They should therefore be seen as an illustrative exercise, rather than as definite judgment on the total $\Delta WELLBY$ effects of CTs.

With this in mind, we nevertheless perform an analogous calculation for the total effect using relative instead of absolute size. A $1,000 lump sum would be 17% of previous income if spent in two years

[23] Frijters et al., (2020) define a WELLBY as a one-point change in life-satisfaction per year.
[24] The value 0.420 comes from assuming a $1,000 lump sum is consumed in 24 months, which is $42 dollars a month. The coefficients in Table 2 are expressed in $100s of dollars. We must thus divide by 100, yielding 42/100=0.420.

for the average household[25]. Using the estimates of Model 3 in Table 2 in such a case, we find $\Delta WELLBY = 0.197$.

A CT paid out in monthly increments requires a slightly different interpretation, given that nearly all CTs were still being paid at the time of the last follow-up. Therefore, our analysis does not afford a prediction of effects after the payments end. Instead, we calculate the effects for a two-year time period, the time in which we assume a lump cash transfer is consumed. For a monthly value of $42 PPP (yielding a total of $1,000 when paid out for two years), the effect at $t = 0$ is estimated to be $\widehat{d}_0 = \hat{\beta}_0 + \hat{\beta}_4 0.420 = 0.127$. Its yearly decay rate is given by $\hat{r} = -0.017$ (see the coefficient "Years since CT began" in Model 4 of Table 2). Thus, we estimate a total effect after two years of $\Delta WELLBY = d_0 t_{end} + \frac{1}{2}\hat{r}t_{end}^2 = 0.127 * 2 - 0.5 * 0.017 * 2^2 = 0.220$. Finally, using an analogous calculation on the basis of Model 3, we find that a stream cash transfer with a size equal to 17% of average household income in the sample would generate an estimated 0.207 $\Delta WELLBY$.

[25] The average yearly household income in our sample is $2,994 PPP. If the cash transfer is spent in two years, then it is $500 per year, which is 500/2,994=0.167≈17%. The annual individual income in USD is $378 at market exchange rates, which means many individuals in our sample live off less than a dollar a day.