The Life You Can Save

Quality of evidence form

Date: July 1st, 2025

Intervention: Unconditional Cash Transfers

1. Studies

Please list all studies considered for this appraisal. Provide their reference in APA style. Create a folder 'Sources' in the prospective RNP evaluation folder and save a copy of each study. Include a link for each study file next to the reference in the field below. Include a justification for why a subset of studies are examined if applicable.

Please note that this is not intended as a systematic review; rather, it represents a targeted scoping of the evidence to highlight the most pertinent and methodologically robust studies for our purposes. We prioritise systematic reviews and meta-analysis, and studies directly evaluating GiveDirectly. Four studies are reviewed:

- Pega, F., Liu, S. Y., Walter, S., Pabayo, R., Saith, R., & Lhachimi, S. (2022). *Unconditional cash transfers for reducing poverty and vulnerabilities: Effect on use of health services and health outcomes in low- and middle-income countries*. Cochrane Database of Systematic Reviews, (11). https://doi.org/10.1002/14651858.CD011135.pub3
- Bastagli, F., Hagen-Zanker, J., Harman, L., Barca, V., Sturge, G., & Schmidt, T. (2016). *Cash transfers: What does the evidence say?* Overseas Development Institute. https://thedocs.worldbank.org/en/doc/111531529868058319-0160022017/original/Day39am10749.pdf
- Chong, Z. Z., & Lau, S. Y. (2025). Unconditional cash transfers and child schooling: A meta-analysis. *Empirical Economics, 68*(2), 639–666. https://link.springer.com/article/10.1007/s00181-024-02647-3
- Haushofer, Johannes, and Jeremy Shapiro. 2016. "The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." Quarterly Journal of Economics, vol. Forthcoming. https://pmc.ncbi.nlm.nih.gov/articles/PMC7575201/
    - We review this individual study because it is an RCT of UCTs implemented by GiveDirectly.

Excluded studies

Baird, S., Ferreira, F. H. G., Özler, B., & Woolcock, M. (2013). *Relative effectiveness of conditional and unconditional cash transfers for schooling outcomes in developing*

countries: A systematic review. *Campbell Systematic Reviews, 9*(8). https://onlinelibrary.wiley.com/doi/full/10.4073/csr.2013.8

Evans, D., & Popova, A. (2014). Cash transfers and temptation goods: a review of global evidence. *World Bank Policy Research Working Paper*, (6886). http://documents.worldbank.org/curated/en/617631468001808739

- We exclude the above two systematic reviews because we prioritize more recent ones.

Papineni, S., Gonzalez, P., Goldstein, M., & Friedman, J. (2025). Cash Is Queen. *Development Research*. https://documents1.worldbank.org/curated/en/099209105052527084/pdf/IDU-172cd481-ec32-47d0-a1d3-1a3f658485f8.pdf

- We do not review this study because it is a single RCT study, while recent, it was not implemented by GiveDirectly. We prioritize reviewing systematic reviews.

2. Appraisal

2.1) Study design

Please provide details on the design of the studies considered in this appraisal. At the end indicate what is the highest study design level across the studies considered for this intervention.

Pega et al. (2022) - systematic review

Bastagli et al. (2016) - "rigorous literature review, which complies with core systematic review principles – breadth, rigour and transparency – while allowing for a more flexible handling of retrieval and analysis with the objective of ensuring comprehensiveness and relevance"

Chong and Lau (2025) - meta-analysis

Haushofer et al. (2016) - RCT

Appraisal for study design

Level	Description
☑ High	Evidence comes from one or more systematic reviews, meta-analyses or experimental studies such as Randomised Controlled Trials (RCTs).
☐ Medium	Evidence comes from quasi-experimental designs. These are well

	designed but non-randomised studies.
☐ Low	Evidence comes from non-experimental studies (e.g., observational studies).
☐ Not considered	Evidence comes from case studies, anecdotal reports, expert opinion or narrative reviews.

2.2) Risk of bias

Please provide details around potential risk of bias across the studies considered in this appraisal. If needed refer to *risk of bias checklists* prepared by the Joanna Briggs Institute[1] to aid the assessment of limitations of systematic reviews,[2] RCTs[3] and quasi-experimental[4] studies. Indicate for which study this is the case with a brief explanation as to why. At the end provide an overall appraisal for risk of bias across the studies considered.

Across the reviewed studies, the overall risk of bias is low. Each review applied clear eligibility criteria, rigorous critical appraisal methods, and appropriate synthesis strategies. Primary studies, including Haushofer et al. (2016), demonstrated strong design quality with randomisation, reliable outcome measurement, and high follow-up rates, supporting confidence in the validity of the findings.

Pega et al. (2022)

1. *Is the review question clearly and explicitly stated?* Yes. The primary objective is clearly stated: to assess the effects of UCTs on health service use and outcomes in LMICs (p. 12).
2. *Were the inclusion criteria appropriate for the review question?* Yes. The review used rigorous inclusion criteria, including only studies that met EPOC standards or equivalent (p. 12).
3. *Was the search strategy appropriate?* Yes. The search strategy was comprehensive, including 15 academic databases and additional grey literature and organizational websites (pp. 13–14).
4. *Were the sources and resources used to search for studies adequate?* Yes. Sources included major databases (e.g., MEDLINE, PubMed, CENTRAL), grey literature databases, and organizational websites (pp. 13–14).

[1] https://jbi.global/critical-appraisal-tools
[2] Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Medical Research Methodology. 2003;3:25 doi:10.1186/1471-2288-3-25.
[3] Barker TH, Stone JC, Sears K, Klugar M, Tufanaru C, Leonardi-Bee J, Aromataris E, Munn Z. The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. JBI Evidence Synthesis. 2023;21(3):494-506
[4] Barker TH, Habibi N, Aromataris E, Stone JC, Leonardi-Bee J, Sears K, et al. The revised JBI critical appraisal tool for the assessment of risk of bias quasi-experimental studies. JBI Evid Synth. 2024;22(3):378-88.

5. *Were the criteria for appraising studies appropriate?* Yes. The review used Cochrane and EPOC tools with additional domain-specific criteria for risk of bias (pp. 18–19).
6. *Was critical appraisal conducted by two or more reviewers independently?* Yes. Two reviewers extracted information independently, with a third resolving disagreements (p. 18).
7. *Were there methods to minimize errors in data extraction?* Yes. Data extractors were trained, piloted a form, and worked in pairs independently. A third reviewer resolved discrepancies (p.18).
8. *Were the methods used to combine studies appropriate?* Yes. Meta-analysis was performed when studies were sufficiently homogeneous. Rules for combining outcomes were clearly defined (p. 21).
9. *Was the likelihood of publication bias assessed?* Yes. Although funnel plots weren't created due to <10 studies per outcome, the review took extensive steps to mitigate publication bias (p.21).
10. *Were recommendations for policy and/or practice supported by the reported data?* Yes. Recommendations were carefully made in line with the strength and certainty of the evidence (p. 45).
11. *Were the specific directives for new research appropriate?* Yes. The review called for improved pre-registration, better reporting, and exploration of effect modification (p. 45),

---

Bastagli et al. (2016)

1. *Is the review question clearly and explicitly stated?* Yes. The review outlines three explicit research questions related to the impact of cash transfers, the influence of design/implementation features, and effects on women and girls (p. 5, 17).
2. *Were the inclusion criteria appropriate for the review question?* Yes. Studies were included based on relevance, methodological rigor, and predefined criteria assessing risk of bias (p. 6, 63).
3. *Was the search strategy appropriate?* Yes. The strategy used five search tracks, covering bibliographic databases, expert recommendations, websites, previous reviews, and snowballing (p. 65).
4. *Were the sources and resources used to search for studies adequate?* Yes. Over 38,000 studies were screened, with extensive sources used (p. 6, 65).
5. *Were the criteria for appraising studies appropriate?* Yes. Studies underwent second-stage screening to assess risk of bias and methodological rigor (p. 6, 67).
6. *Was critical appraisal conducted by two or more reviewers independently?* While not explicitly stated, the language indicates a rigorous review process, but does not confirm dual independent review (p. 67).
7. *Were there methods to minimize errors in data extraction?* Yes. A standardized extraction approach was applied, with disaggregation by gender and aggregation levels (p. 6, 70).
8. *Were the methods used to combine studies appropriate?* Yes. Both vote counting and narrative synthesis were used, accounting for statistical significance and effect direction (p. 6).

9. *Was the likelihood of publication bias assessed?* No formal statistical test (e.g., funnel plots) is reported; however, the inclusion of grey literature helps address this concern (p. 6, 73).
10. *Were recommendations for policy and/or practice supported by the reported data?* Yes. Policy implications are directly drawn from outcome evidence in all sections (p. 99, 121, 143, 166, 201, 227).
11. *Were the specific directives for new research appropriate?* Yes. The report explicitly identifies gaps and calls for more disaggregated, long-term, and regionally diverse studies (p. 266–267).

Chong and Lau (2025) - https://mpra.ub.uni-muenchen.de/121778/1/MPRA_paper_121778.pdf

1. *Is the review question clearly and explicitly stated?* Yes. The paper clearly asks whether UCTs impact student enrolment and attendance, and what explains heterogeneity in those impacts (p. 1 Abstract; p. 4).
2. *Were the inclusion criteria appropriate for the review question?* Yes. The authors used explicit eligibility criteria: experimental or quasi-experimental studies reporting effect sizes with statistical precision (p. 11–12).
3. *Was the search strategy appropriate?* Yes. The search covered global databases and grey literature, with keywords and cross-validation with past reviews (p. 11).
4. *Were the sources and resources used to search for studies adequate?* Yes. 152 studies across 90 programmes were initially identified, refined to 38 studies in 18 countries for final analysis (p. 11).
5. *Were the criteria for appraising studies appropriate?* Yes. A five-category risk of bias framework from IDCG (via Baird et al. 2014) was used (p. 14).
6. *Was critical appraisal conducted by two or more reviewers independently?* Not explicitly stated. The methodology section doesn't clarify if multiple independent reviewers assessed bias (p. 14).
7. *Were there methods to minimize errors in data extraction?* Yes. A standardized procedure was followed for coding effect sizes and moderator variables (p. 12–13).
8. *Were the methods used to combine studies appropriate?* Yes. A random-effects meta-analysis was used due to expected heterogeneity (p. 17–19).
9. *Was the likelihood of publication bias assessed?* Yes. Funnel plots, Egger's test, and trim-and-fill methods were used to assess and correct for bias (p. 28–30).
10. *Were recommendations for policy and/or practice supported by the reported data?* Yes. Policy implications were drawn cautiously based on findings and robustness checks (p. 31–32).
11. *Were the specific directives for new research appropriate?* Yes. The study calls for more UCT evaluations, especially regarding long-term impacts and moderators (p. 32–33).

Haushofer et al. (2016)

Bias related to selection and allocation

1. *Was true randomization used for assignment of participants to treatment groups?* Yes. The study was a two-level cluster-randomized controlled trial. Villages were randomly selected for treatment, and within treatment villages, eligible households were randomly assigned to treatment or control (spillover) groups *(p. 7).*
2. *Was allocation to treatment groups concealed?* No explicit mention is made of allocation concealment procedures (e.g., sealed envelopes or central randomisation). Therefore, this is unclear from the text.
3. *Were treatment groups similar at the baseline?* Yes. Baseline balance tests show that treatment and control households in treatment villages were similar across most index variables. Only one variable (income from self-employment) was marginally significant at baseline, but this did not survive correction for multiple inference *(p. 12–13).*

Bias related to administration of intervention/exposure

4. *Were participants blind to treatment assignment?* No. Participants were informed about the transfer after baseline. There is no indication that participants were blind to their treatment status *(p. 7, 9).*
5. *Were those delivering the treatment blind to treatment assignment?* No. The GiveDirectly staff directly informed households of the transfers, suggesting they were not blinded *(p. 7).*
6. *Were treatment groups treated identically other than the intervention of interest?* Yes. Control (spillover) and treatment households in the same village were surveyed identically, and no intervention other than the cash transfer was administered *(p. 9–10).*

Bias related to assessment, detection and measurement of the outcome

7. *Were outcome assessors blind to treatment assignment?* The study states that the survey and intervention teams were kept distinct, and surveyors denied any association with the intervention. While this may reduce detection bias, true blinding is not confirmed *(p. 28).*
8. *Were outcomes measured in the same way for treatment groups?* Yes. Surveys and biomarker measurements (e.g., cortisol) were administered uniformly across groups *(p. 9–10).*
9. *Were outcomes measured in a reliable way?* Yes. The study used validated instruments (e.g., CES-D scale, stress scales, anthropometric measures) and objective biomarkers (e.g., cortisol) *(p. 10).*

Bias related to participant retention

10. *Was follow-up complete and if not, were differences between groups in terms of their follow-up adequately described and analysed?* Yes. Follow-up was high (93.3% overall), and analyses showed no differential attrition or bias introduced by attrition.

Multiple sensitivity analyses (including Lee bounds) confirmed this *(p. 15–16).*

Statistical Conclusion Validity

11. *Were participants analysed in the groups to which they were randomized?*
    Yes. An intent-to-treat approach was used, including all households assigned to receive a transfer regardless of whether they actually received it *(p. 15).*
12. *Was appropriate statistical analysis used?*
    Yes. The study applied regression analysis with clustering at the household level, corrected for multiple inference (FWER), and conducted robustness checks using seemingly unrelated regressions (SUR), Lee bounds, and MDEs *(p. 13–16).*
13. *Was the trial design appropriate and any deviations from the standard RCT design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial?* Yes. The cluster-randomised design was appropriate for this intervention. The authors justified within-village comparisons and addressed the lack of baseline data for pure control villages *(p. 12–14).*

Appraisal for risk of bias

Level	**Description**
☑ Low	The *majority* of RCTs, systematic review(s) or meta-analysis considered show a *low* risk of bias.
☐ Medium	Across *some* of the considered studies the risk of bias may be *moderate*, but not high.
☐ High	A *high* risk of bias is identified across studies.

2.3) Imprecision

Please provide an assessment on the potential of imprecision in the reviewed studies. Imprecision refers to the uncertainty in how close the estimated effect size (e.g., the impact of an intervention) is to the true effect. It is influenced by the sample size, which directly impacts the statistical power of a study. In your answer specify for which studies and to what extent imprecision is of concern. At the end provide an overall appraisal for imprecision across the studies considered.

Across the studies reviewed, imprecision is a moderate concern. While some outcomes, especially those with larger sample sizes or multiple trials, show acceptable precision, many others are affected by small sample sizes, wide confidence intervals, or limited statistical power. Subgroup analyses and less commonly studied outcomes (e.g., long-term or third-order effects) are particularly prone to imprecision. Overall, although some studies transparently report effect sizes and MDEs, precision remains limited across much of the evidence base.

Pega et al. (2022)
Moderate imprecision. Authors note: “Estimated treatment effects showed acceptable or even good precision for the small number of outcomes that had evidence from several C-RCTs, relatively common events and relatively large (effective) sample sizes. However, for most outcomes, treatment effects were still relatively imprecise.” p. 46.

Bastagli et al. (2016)
Moderate concern for imprecision, especially in less commonly studied outcomes or those requiring long-term observation.

- The review does not directly quantify precision through confidence intervals or sample sizes but highlights variation in statistical significance and effect sizes. Some areas like child anthropometry, cognitive development, and employment sector effects showed mixed or weak significance, indicating imprecision. This stems from differences in programme types, sample sizes, and methods.
- Small evidence base and weak effects for third-order outcomes like long-term empowerment, business investment, or fertility.
- Limited effects on child stunting and underweight despite improvements in dietary diversity (pp. 129–131, 177–179, 237–238).

Chong and Lau (2025)
Imprecision is a moderate concern, especially for attendance outcomes and smaller studies.

- The study finds smaller effect sizes than earlier meta-analyses, noting that many studies have large standard errors and report non-significant effects, especially for attendance.
- Studies with significant effects often had higher standard errors (i.e., lower precision) (p. 22–23, 31).

Haushofer et al. (2016)
Moderate concern about imprecision, particularly for subgroup analyses. The study reports Minimum Detectable Effect Sizes (MDEs) and acknowledges limitations in statistical power, especially in subgroup comparisons (e.g., female vs. male recipients). For most main outcomes (e.g., health, education, empowerment), MDEs ranged from 0.16 to 0.20 SD, allowing detection of modest effect sizes. However, imprecision is a concern in some comparisons, such as differences between transfer modalities or genders, due to smaller sample sizes. Overall, imprecision is moderate but well-addressed through transparent reporting (p. 17–18; Appendix Table A.1).

Appraisal for imprecision

Level	**Description**
☐ Low	The studies have *high* statistical power to detect significant differences.

	Confidence intervals are sufficiently narrow meaning that statistical imprecision has negligible impacts on decision making. This typically means that the average study has large sample sizes.
☑ Medium	The studies only have *moderate* statistical power to detect significant differences. Confidence intervals may introduce some uncertainty into decision making. Some of the considered studies have small sample sizes.
☐ High	The studies have *low* statistical power to detect significant differences with confidence intervals that introduce uncertainty into decision making. This typically means that the studies have small sample sizes.

2.4) Inconsistency

Please provide an assessment on the potential of inconsistency among the reviewed studies. Inconsistency refers to variation in the results across studies evaluating the same intervention. At then end provide an overall appraisal for inconsistency across the studies considered.

We consider overall a moderate concern for inconsistency. The assessed systematic reviews report moderate to high inconsistency, while the reviewed RCT finds consistent results across study arms.

Pega et al. (2022)

Moderate inconsistency. Authors report: “The existing evidence was relatively consistent for some outcomes, such as the likelihood of having had any illness and of having attended school, but it was highly inconsistent for many others, including the level of household dietary diversity. Primary studies and future updates of this review could formally explore or even quantify effect modification, as a means of better understanding inconsistency in highly inconsistent bodies of evidence.” p.45

Bastagli et al. (2016)

Inconsistency is a moderate concern, especially across regions and less direct outcome areas.

- Some outcomes, like school attendance, food expenditure, and health service use, show consistent positive effects.
- Others, including test scores, cognitive development, anthropometry, and employment, exhibit mixed or no consistent pattern, with contradictory findings across contexts.
- The evidence on conditionality versus unconditionality also varies across regions and outcomes (p. 107, 129, 151, 177, 213, 240–241)

Chong and Lau (2025)

High inconsistency exists among the studies. High heterogeneity (I^2 ~70%) in both enrolment and attendance effect sizes across studies. This heterogeneity persisted even after excluding high-bias studies. Programme and country characteristics explain only a small portion of this variation (p. 19–20, 24–27).

Haushofer et al. (2016)
Low concern about inconsistency. There is little evidence of inconsistency in effects across the study's arms. While some effects vary by treatment type (e.g., monthly vs. lump-sum), these are explained by design rather than contradictory results. The robustness checks and use of multiple specifications support consistency. Moreover, null effects are supported by tight confidence intervals and small MDEs.

Appraisal for inconsistency

Level	Description
☐ Low	The estimated effects of the intervention are broadly consistent across the studies (e.g., RCTs). There may be some minor variation.
☑ Medium	The estimated effects are broadly consistent across the studies, although there may be some moderate variation.
☐ High	There is inconsistency in the estimated effects across studies.

2.5) Indirectness

Please provide an assessment on the potential of indirectness among the reviewed studies. Indirectness refers to situations where the evidence does not directly apply to the intervention being evaluated or the specific context of interest. At the end provide an overall appraisal for indirectness across the studies considered.

There is low concern for indirectness. Studies evaluate interventions implemented in LMICs, including a program implemented by GiveDirectly.

Pega et al. (2022)
Low concern for indirectness, as evidence is largely applicable to the priority regions (Sub-Saharan Africa and South Asia) (p. 23–24).

Bastagli et al. (2016)
Low concern for indirectness, since many studies come from TLYCS priority countries.

- The review includes substantial evidence from Sub-Saharan Africa (38%), and also studies from Bangladesh, Morocco, Kenya, Nigeria, Malawi, and Uganda.
- However, it acknowledges overrepresentation of Latin American CCTs and underrepresentation of non-CCTs and newer interventions (p. 76, 238).

Chong and Lau (2025)
Low concern for indirectness. The review covers 18 countries, including Kenya, Morocco, Nigeria, Malawi, and South Asia (2 studies), aligning with TLYCS priorities and where GiveDirectly operates. UCTs were not all education-targeted, but all had student participation outcomes (p. 15, 16, 31).

Haushofer et al. (2016)
No concern about indirectness. The evidence is directly applicable to the intervention (it is implemented by GiveDirectly) and setting of interest.

Appraisal for indirectness

Level	Description
☑ Low	The RCTs or quasi-experimental studies study the intervention directly as it is implemented by the prospective RNP or in highly relevant contexts.
☐ Medium	The studies are only moderately relevant to the context in which the intervention is implemented.
☐ High	The studies have very low relevance to the context in which the intervention is implemented.

3. Summary

Criteria	Appraisal		
Study design	☑ High	☐ Medium	☐ Low
Risk of bias	☐ High	☐ Medium	☑ Low
Imprecision	☐ High	☑ Medium	☐ Low
Inconsistency	☐ High	☑ Medium	☐ Low
Indirectness	☐ High	☐ Medium	☑ Low

4. Quality of evidence final score

Please award the prospective RNP a final *quality of evidence score*:

Level	Description
☑ High	There is a wide range of well conducted experimental and quasi-experimental evidence showing the effectiveness of the intervention (i.e., the studies show a low risk of bias and the estimates are precisely estimated). This includes evidence in the geography where the organization operates. There is at least one study assessing the impact of the intervention as implemented by the RNP (direct evidence). There is broad consistency on the effect of the intervention across considered studies.
☐ Medium	There is a medium range of experimental, quasi-experimental and non-experimental evidence showing the effectiveness of the intervention. Some studies have a moderate risk of bias, but none show a high risk. Some studies show moderate imprecision. There is primarily indirect evidence for the intervention. There is moderate variation of the results across the studies.
☐ Low	There is a low range of studies showing the effectiveness of the intervention. There is a high risk of bias and imprecision across some studies. There is only indirect evidence on the effectiveness of the intervention（none coming from studies where intervention is implemented by prospective RNP). There is high variation of the results across studies.

Please provide any comments to back up the decision if considered necessary. Only one quality score should be awarded to the RNP even if it implements more than one intervention. Hence, to inform this final score, first award an intervention specific score. Take the average across the interventions' final scores.