THE SWIFT APPRAISAL OF HOUSEHOLD ECONOMIC STRENGTHENING STUDIES: A BRIEF GUIDE FOR PROGRAM STAFF

HES Research Dialogues: Methods Brief I

Marija Pantelic and Janina Steinert

Introduction

There is a growing focus on evidence-based decision making in international development. USAID, for example, calls for international development projects to be designed "based on evidence and supported by analytical rigour" (USAID 2015). Sackett and colleagues (1996, p.71) define evidence-based practice as:

> "The conscientious, explicit and judicious use of best currently available evidence, integrated with client values and professional expertise, in making decisions about the care of individuals."

Evidence-based programming is particularly important when resources are limited and when working with key populations such as orphans and vulnerable children (OVCs). Intervening in the lives of OVCs comes with an ethical obligation to guide program design with the best available evidence. Failure to do so may waste limited public funding and could even cause significant harm to intended beneficiaries (Chalmers 2003; Gibbs & Gambrill 2002).

This brief aims to aid program developers and implementers working on household economic strengthening (HES) interventions for OVCs to assess the quality of available evidence. It presents a *Scoring Sheet for the Swift Appraisal of Household Economic Strengthening Studies* (SSSA-HES), focusing on the most rigorous research designs. Ultimately, this tool should help HES program developers and implementers who encounter multiple studies with contradictory findings to decipher which studies were the most rigorously conducted and to become critical, objective consumers of evidence to guide their work.

Objectives

The Swift Appraisal of HES Studies draws on existing tools for the systematic assessment of evidence. They include – but are not limited to – the *Cochrane Collaboration's tool for assessing risk of bias for randomized controlled trials* (Higgins et al. 2011), the *Consolidated Standards of Reporting Trials (CONSORT) statemen*t (Schultz et al. 2010), the *Critical Appraisal Skills Programme (CASP) Randomized Controlled Trials Checklist* (CASP 2016a), and the *Grading of Recommendations Assessment, Development, and Evaluation (GRADE) Approach to Rating the Quality of Evidence* (Guyatt et al. 2011), all of which are tools for appraising the quality of randomized controlled trials (RCTs). The SSSA-HES also draws on the *Transparent reporting of evaluations with non-randomized designs (TREND) statement* (Des Jarlais et al. 2004), the *Cambridge Quality Checklists* (Murray et al. 2009) and the *CASP Cohort Study Checklist* (CASP 2016b), which can be adequately applied to selected observational research studies. Lastly, the *CASP Systematic Review Checklist* (CASP 2016c) for rating quality of systematic reviews also influenced the development of the Swift Appraisal of HES

This brief was produced under United States Agency for International Development (USAID) Cooperative Agreement No. AID-OAA-LA-13-00001 and was made possible by the generous support of the American people through USAID and the United States President's Emergency Plan for AIDS Relief. The contents are the responsibility of FHI 360 the SEEP Network and do not necessarily reflect the views of USAID or the United States Government.

Studies. These tools have been used both by researchers as well as policy makers for rating the quality of available evidence. Based on an in-depth review of these tools, this brief provides a condensed, single checklist applicable for appraising HES research covering various study designs. The language is straight-forward as it was adapted for non-academic consumers of research. The resulting scoring sheet is intended to help HES program developers and implementers answer three key questions to help guide their work and decision-making:

1) What are the results of available HES studies? What do they mean in practical terms?
2) To what extent are the results of a HES study accurate or biased? In other words, what is the quality of the evidence?
3) To what extent can the results be applied to the country or context of interest?

Scope

This brief will focus on the three most rigorous study designs according to the pyramid of evidence: systematic reviews, RCTs, cohort and case-control studies (Greenhalgh 1997). It is important to acknowledge that other relevant study designs exist (see *Figure 1*). However, whenever possible, evidence with the highest capacity to infer causality should be considered (i.e. systematic reviews, RCTs and cohort studies). Additional caution should be applied if less rigorous types of evidence are used to inform further research and funding.

DEFINITIONS

Quality of evidence: For the purpose of the SSSA-HES, quality of evidence is defined as the extent of confidence with which we can determine that the effect estimates presented in a study are correct (Balshem et al. 2011).

Systematic reviews: "A systematic review uses transparent procedures to find, evaluate and synthesize the results of relevant research. Procedures are explicitly defined in advance, in order to ensure that the exercise is transparent and can be replicated. This practice is also designed to minimize bias." (Campbell Collaboration 2016).

RCT: "A study in which a number of similar people are randomly assigned to 2 (or more) groups to test a specific drug, treatment or other intervention. One group (the experimental group) has the intervention being tested, the other (the comparison or control group) has an alternative intervention, a dummy intervention (placebo) or no intervention at all. The groups are followed up to see how effective the experimental intervention was. Outcomes are measured at specific times and any difference in response between the groups is assessed statistically." (National Institute for Health and Care Excellence 2016).

Observational studies: "A […] study in which the investigator observes the natural course of events with or without control groups (for example, cohort studies and case–control studies)." (National Institute for Health and Care Excellence 2016). Unlike RCTs, the investigator does not intervene or expose participants to any intervention in observational studies.

Cohort studies: "An observational study with 2 or more groups (cohorts) of people with similar characteristics. One group has a treatment, is exposed to a risk factor or has a particular symptom and the other group does not. The study follows their progress over time and

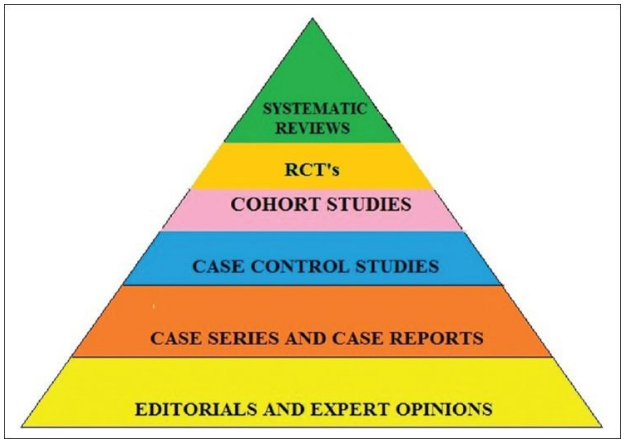
records what happens." (National Institute for Health and Care Excellence 2016).

Case-control studies: Another type of "observational study to find out the possible cause(s) of a disease or condition. This is done by comparing a group of patients who have the disease or condition (cases) with a group of people who do not have it (controls) but who are otherwise as similar as possible (in characteristics thought to be unrelated to the causes of the disease or condition). This means the researcher can look for aspects of their lives that differ to see if they may have caused the condition." (National Institute for Health and Care Excellence 2016).

Case series: "Reports of several patients with a given condition, usually covering the course of the condition and the response to treatment. There is no comparison (control) group of patients." (National Institute for Health and Care Excellence 2016).

THE HIERARCHY OF EVIDENCE

Figure 1. Hierarchy of Evidence (reprinted from Greenhalgh, 1997)



The three types of studies addressed in this brief rank differently on the hierarchy of evidence, with systematic reviews considered as the gold standard of research evidence, and observational studies such as cohort and case-control studies considered to be more prone to bias. However, in the field of HES research, observational studies are often the most, if not the only, feasible design due to ethical, practical and political constraints. For example, recent research showing the beneficial effects of the South African child-focused cash transfers (Cluver et al. 2013) could not have been carried out in the form of an RCT because there was a national policy guiding the implementation of the grant, therefore random allocation was not possible.

Scoring Sheet for the Swift Appraisal of HES Studies (SSSA-HES)

The scoring sheet provided below is a simplified short guide for *swiftly* appraising the quality of research evidence. It is not intended to be comprehensive, but instead is meant to provide non-academic practitioners with a basic and accessible tool for assessing quality of research. Therefore, it is recommended that program developers and implementers consult methodology and statistics experts prior to making decisions about developing and funding interventions or research.

Below are additional helpful resources for appraising the quality of research evidence.

Systematic reviews: *Cochrane Handbook for Systematic Reviews of Interventions*
http://training.cochrane.org/handbook

RCTs: *CASP Randomized Controlled Trial Checklist*
http://docs.wixstatic.com/ugd/dded87_4239299b39f647ca9961f30510f52920.pdf

Cohort studies: *CASP Cohort Study Checklist*
https://hhs.hud.ac.uk/lqsu/Useful/critap/Cohort%20Study%20Checklist/CASP-Cohort-Study-Checklist-31.05.13.pdf

Scoring Sheet for the Swift Appraisal of HES Studies (SSSA-HES) (adapted from Higgins et al. 2011; Schultz et al. 2010; CASP 2016a; CASP 2016b; Guyatt et al. 2011; Des Jarlais et al. 2004; Murray et al. 2009) First determine whether a systematic review is available. If yes, appraise it. If a systematic review is not available, or if you have assessed it to be of poor quality, you might want to appraise the primary studies yourself (RCTs and cohort studies).	
Study design score	
Yes = 3	Systematic Review
Yes = 2	Randomized control trial
Yes = 1	Cohort study
I. Quality Assessment of Systematic Reviews	
Yes = 1 No = 0	Has the systematic review been conducted fairly recently, i.e. in the last 5 years? Are the findings likely to be up to date?
Yes = 1 No = 0	Was the systematic review protocol published prior to the systematic review? This is important so as to rule out selective outcome reporting. The paper should refer to its protocol in the text and provide a link to where it can be accessed online.
Yes = 1 No = 0	Has the systematic review made sufficient efforts to identify evidence from grey literature? Is unpublished, in addition to published, research considered? Indication for a positive assessment would be if authors had conducted hand searches through online databases of relevant organizations/NGOs, contacted experts in the field to point to further relevant work on the topic, and if trial registries had been searched.
Yes = 1 No = 0	Are the results of the review objectively valid? Indication for a positive assessment would be if, for instance, screening of eligible studies, data extraction, as well as quality appraisal were carried out by more than one review author, if there is mention of discrepancies in quality ratings and decisions on how these were resolved (e.g. by involving a third review author), and if procedural steps of the review are outlined in a transparent way.
Yes = 1 No = 0	Was the quality of evidence of included studies assessed with a standardized tool?
Yes = 1 No = 0	Is the review likely unbiased in terms of geographic scope? Indication for a positive assessment would be if searches had been conducted in several languages and performing a search that was geographically inclusive.
II. Quality Assessment applicable to both RCTs and Cohort Studies	
Response rates – refusal to participate and drop out ideally would not be higher than 30%. Differential attrition refers to differences in drop out rates between the intervention and control/comparison group.	
Yes = 2	Refusal rates <30%
Yes = 1	Refusal rate >30%
Yes = 0	Not reported

Sample size score	
Yes = 1	Sample size ≥200
Yes = 0	Sample size <200
Follow-up score	
Yes = 1 No = 0	Lost to follow up <30%? Was the follow up complete enough? Participants who are lost to follow-up may have different outcomes than those available for assessment.
Yes = 1 No = 0	Was the follow up long enough to capture real effects of an intervention? The good or bad effects should have had long enough to reveal themselves. The appropriate length of follow up will depend on the HES intervention that you are studying.
II.A. Quality Assessment of RCTs only	
Yes = 1 No = 0	Was there a 'placebo' intervention such that participants would not expect one or the other to be effective. For example, testing a micro finance intervention against a health intervention rather than a waitlist control. If two interventions are delivered, participants are less likely to be aware of which intervention is being tested, thereby minimizing the likelihood of a placebo effect. However, if there is a waitlist control, participants are fully aware which intervention is being tested and might therefore be more susceptible to placebo effects.
Yes = 1 No = 0	Did the researchers use an intention-to-treat analysis? Intention-to-treat analysis analyses participants based on their assigned study arm, rather than being restricted to those who attended and completed the program. If only those who have completed the program are analysed, then we are not adequately accounting for drop out. Findings are likely to be biased and based only on a sub-sample of participants who completed the program and might not be representative of the whole sample.
Yes = 1 No = 0	Was contamination considered and ruled out? How likely were the control/comparison group participants to be exposed to parts or all of the intervention? For example, a school-based intervention that randomized individual children rather than schools, or a community-based intervention that randomized households rather than whole communities would run the risk of contamination.
Yes = 1 No = 0	Were baseline characteristics (including possible confounding factors that could influence the study outcomes) of the intervention and control groups assessed and found to be equal (differences not significant)? This would be indicative of successful randomization and adequate sample size.
Yes = 1 No = 0	Was differential attrition (difference in attrition between study arms) <10%?
Yes = 1 No = 0	Could selective outcome reporting be ruled out? Did the researchers register the trial protocol prior to publication of trial outcomes and do primary and secondary outcomes reported in the protocol correspond to what was reported as results? If this is not the case, reporting bias might exist in that authors only report significant findings and not insignificant findings and potentially harmful findings. A good place to search for trial registrations is http://apps.who.int/trialsearch/

II.B. Quality Assessment of cohort studies only	
Yes = 1 No = 0	Was total population or random sampling used?[1] If the participants were selected by convenience, or if the study does not explicitly report how the participants were selected, then findings could be more prone to bias. Therefore, if convenience or purposive sampling was used or the sampling method was not reported, please give a 0.
Yes = 1 No = 0	Did the study design and analysis take into account alternative explanations for the findings? For example, if researchers are interested in looking at how financial literacy training at baseline predicts poverty levels at follow up, they need to measure and statistically account for other factors that might influence poverty at follow up, for example health or education levels.
Yes=1, No=0	Was 'analysis of change' conducted? In other words, did the study design and analysis capture change in the outcome within individual subjects? Following from the example above, health and education levels at baseline might influence poverty outcomes at follow-up. However, the most likely predictor of poverty at follow-up would be poverty at baseline. Therefore, proper 'analysis of change' would control for outcome levels at baseline (in this case, poverty). This type of analysis allows us to capture *change in outcome levels within each study participant* rather than within the population as a whole.
Total score:	
Note: There are no cut-off points for this checklist. Such cut-off points would be arbitrary because the purpose of this checklist is to help practitioners and program developers think critically about the quality of the evidence and what the findings of HES studies mean within their context. We recommend that this tool be used to evaluate studies relative to one another and the evidence base as a whole. *Program developers and implementers should use this checklist to think about the quality of available evidence. The checklist should also help decipher what the results of HES studies mean in practical terms. For example, if the study resulted in positive effects but convenience sampling was used, perhaps the intervention has potential but, at present, findings are not generalizable to the population of interest. Similarly, the geographic scope of a systematic review might be helpful in determining to what extent the results can be applied to the country or context of interest.*	

[1] It is understandable that most HES studies with OVC would recruit via service providers as the population of interest needs to meet certain vulnerability criteria. However, it is important that papers and reports state explicitly how the participants were actually sampled. Ideally, either all OVC from an organization were sampled, or they were randomly selected.

ABOUT THE AUTHORS

Marija Pantelic *is a DPhil candidate in the Department of Social Policy and Intervention, University of Oxford. Her research interests lie in methods to minimize bias from observational studies and finding evidence-based interventions to improve the psychosocial wellbeing of orphaned and vulnerable children.*

Janina Steinert *is a DPhil candidate in the Department of Social Policy and Intervention, University of Oxford. Her research is focused on HES through microsavings and financial literacy and is analyzing economic resilience promotion in a low-income, HIV/AIDS-affected population in rural South Africa.*

About the HES Research Dialogues:

In 2014, FHI 360's ASPIRES Project and the SEEP Network recognized that, while HES was a growing area of practice and research, gaps in HES research and evidence remained. To respond to this evidence gap, SEEP facilitated an HES Research Dialogues initiative, bringing together HES researchers and practitioners to define a collaborative learning agenda. Through a series of collaborative activities, the initiative aimed to identify key research questions within HES, as well as draw on existing experience related to appropriate research methods and tools.

This document is complemented by a series of research methods and evidence briefs developed out of the HES Research Dialogues initiative. Access them on FHI 360's ASPIRES Project web page on Microlinks at: http://bit.ly/1rwRue3

References

Balshem H., Helfand M., Schunemann H.J., et al. (2011). GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology* 64(4): 401-406. http://dx.doi.org/10.1016/j.jclinepi.2010.07.015

Campbell Collaboration (2016). What is a systematic review? http://www.campbellcollaboration.org/what_is_a_systematic_review/index.php (accessed 19 May 2016).

Critical Appraisal Skills Programme (2016). CASP Checklists. http://www.casp-uk.net/#!checklists/cb36 (accessed 22 May 2016).

Chalmers, I. (2003). Trying to do more Good than Harm in Policy and Practice: The Role of Rigorous, Transparent, Up-to-Date Evaluations. *The Annals of the American Academy of Political and Social Science*, 589(1), 22–40. http://www.jstor.org/stable/3658559 (accessed 24 July 2016)

Cluver, L., Boyes, M., Orkin, M., Pantelic, M., Molwena, T., & Sherr, L. (2013). Child-focused state cash transfers and adolescent risk of HIV infection in South Africa: a propensity-score-matched case-control study. *The Lancet Global Health*, 1(6) , e362 - e370. http://linkinghub.elsevier.com/retrieve/pii/S2214109X1370115 (accessed 24 July 2016).

Des Jarlais D.C., Lyles C., Crepaz N., & the TREND Group (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health*, 94, 361-366. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1448256/ (accessed 24 July 2016).

Gibbs, L., & Gambrill, E. (2002). Evidence-Based Practice: Counterarguments to Objections. *Research on Social Work Practice*, 12(3), 452–476. http://rsw.sagepub.com/content/12/3/452.full.pdf+html (accessed 24 July 2016).

Greenhalgh, T. (1997). How to read a paper : getting your bearings (deciding what the paper is about). *BMJ*, 315(7102), 243–246. http://doi.org/10.1136/bmj.315.7102.243

Guyatt G.H., Oxman A.D., Sultan S., et al. (2011). GRADE guidelines: 9. Rating up the quality of evidence. *Journal of Clinical Epidemiology*, 64(12), 1311-1316. http://doi.org/10.1016/j.jclinepi.2011.06.004.

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K.F., Weeks, L., & Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal*, 343, d5928. http://doi.org/10.1136/bmj.d5928

Higgins J.P.T., Green S. (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

Murray, J., Farrington, D. P., & Eisner, M. P. (2009). Drawing conclusions about causes from systematic reviews of risk factors: The Cambridge Quality Checklists. *Journal of Experimental Criminology*, 5(1), 1–23. http://doi.org/10.1007/s11292-008-9066-0.

National Institute for Health and Care Excellence (2016). Glossary. https://www.nice.org.uk/Glossary (accessed 17 May 2016).

USAID (3015). Strategy and Planning. https://www.usaid.gov/results-and-data/planning (accessed 17 May 2015).

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ*, 312(7023), 71–72. http://doi.org/10.1136/bmj.312.7023.71.

Schulz, K.F., Altman, D.G., Moher, D., for the CONSORT Group. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Annals of Internal Medicine*,152, Epub 24 March. http://dx.doi.org/10.1136/bmj.c332