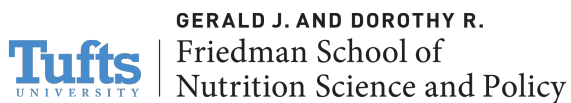

Synthesizing practices of evidence appraisal in the humanitarian field

Roxanne Krystalli and Lauren Emerson



The Humanitarian Evidence Programme

Table of Contents

Executive Summary.....	3
Section I: Premises and Considerations of Evidence Appraisal.....	5
Section II: Methodology of Appraisal Scheme Synthesis.....	7
Section III: Summary of Reviewed Appraisal Approaches.....	8
Appendix: Guidance for Humanitarian Evidence Programme Reviewers	13

Suggested Citation: Roxanne Krystalli and Lauren Emerson, "Synthesizing practices of evidence appraisal in the humanitarian field." Oxfam – Feinstein International Center. November 2015.

Executive Summary: Synthesizing practices of evidence appraisal in the humanitarian field

This paper synthesizes evidence appraisal practices to inform evidence syntheses in the Humanitarian Evidence Programme.¹ It is addressed to all audiences interested in humanitarian evidence, including but not limited to researchers, humanitarian practitioners, and policymakers. The paper helps answer the question, 'How confident are we in the quality of evidence supporting a finding?'

Evidence appraisal is integral to the review process: when clear and context-appropriate, it contributes to the transparency and rigor of the review.² Rather than prescribing a particular evidence appraisal approach, the Humanitarian Evidence Programme summarizes evidence appraisal practices relevant to the humanitarian field and offers some suggestions in critically applying them to the realities of humanitarian data analysis, synthesis, and interpretation.

As discussed in relevant literature and in the programme's own [Guidance Note](#),³ there are challenges in data collection that set evidence syntheses in the humanitarian field apart from reviews in other disciplines. In brief, these limitations include:

- Defining key terms, including interventions and outcomes, may be complex in ways that affect the scope of the review question, the eligibility criteria and their interpretation, and the search strings;
- The design and implementation of studies varies from that of a controlled, laboratory setting in ways that affect data collection, biases, errors, and results;
- The vast 'grey literature' (e.g., programme documents, needs assessments, and internal reports) is difficult to search;
- Data may be limited or of poor quality, and methodologies are often not clearly discussed.

This is not to suggest that research in the humanitarian context cannot achieve or strive for high standards of rigor and validity. Rather, based on the above, it becomes apparent that, while humanitarian evidence syntheses can borrow insight from existing appraisal schemes, these approaches need to be tailored according to the following assumptions:

- Evidence appraisal is an integral part of systematic evidence synthesis;
- Evidence appraisal approaches depend on the evidence synthesis question and context;
- All studies included in a review, regardless of their design or publication status, should be appraised. This may require the combination of appraisal tools to reflect a diversity of methodologies and approaches;

¹ The authors would like to thank Dr. Elizabeth Stites, Dr. Ellie Ott, Dr. Dan Maxwell, and Dr. Patrick Webb for their thoughtful comments and guidance through the process of compiling this paper. The authors are also grateful to the members of the Humanitarian Evidence Programme Advisory Board for their insight, as well as to key informants who anonymously shared their insight on evidence appraisal.

² The term 'review' is used in this document to refer to all evidence synthesis outputs of the program. The term 'systematic review' is used to refer only to that particular approach to evidence synthesis. For more on the programme's approach and outputs, please consult Roxanne Krystalli, Eleanor Ott, Elizabeth Stites, "Evidence Synthesis in the Humanitarian Field: Opportunities, Challenges, and Guidance," *Oxfam GB – Feinstein International Center*, 2015.

³ Roxanne Krystalli and Eleanor Ott, "Guidance Note: Evidence Synthesis in the Humanitarian Evidence Programme," Oxfam GB and Feinstein International Center, http://policy-practice.oxfam.org.uk/~media/Files/policy_and_practice/conflict_disasters/HEP_call%20for%20proposals/HumEvi_Guidance_Note.ashx.

- Those conducting evidence syntheses should attempt, where possible, to address missing methodological information and data;
- A study appraised as 'low confidence' can still be instructive for an evidence synthesis;
- Evidence appraisal decisions should be documented and justified, similarly to other decisions in evidence syntheses.

These parameters are discussed in greater length in the full report. The full report also summarizes a series of existing evidence appraisal schemes that may be relevant to the humanitarian field. An Excel catalogue of the reviewed approaches is available on the Oxfam and FIC programme websites.⁴

Structure of this paper: Section I of this paper explores the Humanitarian Evidence Programme's proposed approach to evidence appraisal, while Section II discusses the methodology of reviewing key evidence appraisal approaches that are relevant to the humanitarian field. An accompanying catalogue of these approaches should be read concurrently.⁵ Section III provides a brief, narrative summary of the evidence appraisal schemes reviewed in the catalogue. Finally, the Appendix provides further guidance that specifically applies to teams carrying out reviews as part of the Humanitarian Evidence Programme.

The Humanitarian Evidence Programme at a glance: The Humanitarian Evidence Programme is a partnership between Oxfam and the Feinstein International Center at the Friedman School of Nutrition Science and Policy at Tufts University (FIC). Between June 2014 and December 2016, the Humanitarian Evidence Programme is commissioning a set of evidence syntheses that synthesize humanitarian research around different questions of interest to the sector. Some of these evidence syntheses follow a strict, more classical systematic review approach, while others are practice reviews or other evidence synthesis outputs, depending on the nature of the question and the amount, quality, and type of data available to answer it.⁶ The topics of these reviews arose from consultations with researchers, humanitarian practitioners, and policymakers to identify priority areas for evidence synthesis. Findings are communicated to researchers, policymakers and practitioners in the humanitarian field, with the ultimate goal of improving humanitarian policy and practice. The programme is funded by UK aid from the UK government; however, the views expressed in this document do not necessarily reflect the UK government's official policies.

This paper is part of a series of documents on evidence synthesis in the humanitarian field. Documents of particular interest in this series include a [mapping](#) of existing humanitarian evidence syntheses and corresponding discussion of the [methodology](#)⁷ and a [guidance note](#) discussing opportunities, challenges, and approaches to evidence synthesis in the humanitarian field.⁸

⁴ See <http://www.oxfam.org.uk/hep> and <http://fic.tufts.edu/research-item/the-humanitarian-evidence-program/>.

⁵ Lauren Emerson, "Catalogue of Evidence Appraisal Schemes," (Feinstein International Center – Oxfam GB: 2015).

⁶ The term 'review' is used in this document to refer to all evidence synthesis outputs of the program. The term 'systematic review' is used to refer only to that particular approach to evidence synthesis. For more on the program's approach and outputs, please consult Roxanne Krystalli and Eleanor Ott, "Evidence Synthesis in the Humanitarian Field: Opportunities, Challenges, and Guidance," *Oxfam GB – Feinstein International Center*, 2015.

⁷ Kristin Bushby, "Map of Humanitarian Evidence Syntheses, 2009-2015," (Feinstein International Center – Oxfam GB: 2015), http://fic.tufts.edu/assets/Mapping-Humanitarian-Evidence_Insights-and-Challenges.pdf.

⁸ Kristin Bushby, Roxanne Krystalli, "Mapping Evidence Syntheses in the Humanitarian Sector: Insights and Challenges," (Feinstein International Center – Oxfam GB: 2015), http://fic.tufts.edu/assets/Mapping-Humanitarian-Evidence_Insights-and-Challenges.pdf.

SECTION I. PREMISES AND CONSIDERATIONS OF EVIDENCE APPRAISAL

Evidence appraisal offers opportunities for intra-field learning, as shown from the list of existing evidence appraisal schemes spanning a range of fields and disciplines. Rather than prescribing a particular evidence appraisal approach, the Humanitarian Evidence Programme summarizes the most relevant evidence appraisal practices and offers some suggestions in critically applying them to the realities of humanitarian data analysis, synthesis, and interpretation.

As discussed at length in relevant literature,⁹ there are many challenges in data collection that set evidence syntheses in the humanitarian field apart from reviews in other disciplines. While a full list of these challenges can be found in the program's [Guidance Note for Humanitarian Evidence Synthesis](#),¹⁰ the limitations most relevant to evidence appraisal in the humanitarian field are outlined below. Not all limitations apply to all types of reviews, as the challenges and opportunities for evidence synthesis depend on the nature and type of data available in response to each evidence synthesis question.

- **Defining key terms, including interventions and outcomes, may be complex** in ways that affect the scope of the question, the eligibility criteria and their interpretation, and the search strings. Definitions and indicators may vary across studies, thus complicating the process of appraisal for a body of evidence.
- **The vast 'grey literature' (e.g., programme documents, needs assessments, and internal reports) is difficult to search** in a standardized, comprehensive way and may not be publically available. Many of the existing evidence appraisal schemes reviewed either do not explicitly address the utility of the grey literature for evidence syntheses in this field or do not include methods typically used in grey literature.
- **Data may be limited or of poor quality and methodologies are often not discussed** explicitly, clearly, or at length in studies. This renders it challenging to assess the process and rigor employed in each study.
- **Sampling is often statistically biased, and comparison or control groups are weak or non-existent.** Many appraisal processes are built upon the premise that randomized control trials (RCTs) are the 'gold standard' of evidence. However, in humanitarian settings, the ethics and feasibility of a randomized intervention with a control group are contested.¹¹ As a result, there are limited or no RCTs on many interventions and the applicability of appraisal schemes that exclusively or primarily pertain to experimental designs (or that place only such designs at the top of hierarchies of evidence) is limited.

The particularities of evidence in this field, however, do not suggest that evidence appraisal is impossible or that producers and users of research should not strive for high standards of rigor

⁹ See: Mallett, R., Hagen-Zanker, J., Slater, R., and Duvendack, M. "The benefits and challenges of using systematic reviews in international development research." (Journal of Development Effectiveness: 2012), 445-455, <http://dx.doi.org/10.1080/19439342.2012.711342>.

¹⁰ "Guidance Note: Evidence Synthesis in the Humanitarian Evidence Programme," Oxfam GB, http://policy-practice.oxfam.org.uk/~media/Files/policy_and_practice/conflict_disasters/HEP_call%20for%20proposals/HumEvi_Guidance_Note.ashx.

¹¹ See Proudlock, K, Ramalingam, B, and Sandison, P, "Improving humanitarian impact assessment: bridging theory and practice," in ALNAP 8th Review of Humanitarian Action Ch.2 (2009), pp.26-29 AND Clarke, M, Allen, C, Archer, F, Wong, D, Eriksson, A and Puri, J, 2014. *What Evidence is available and what is require, in humanitarian assistance? 3ie Scoping Paper 1.* (New Delhi: International Initiative for Impact Evaluation (3ie): 2014) AND Knox Clarke, P. and Darcy J., "Insufficient Evidence? The quality and use of evidence in humanitarian action," ALNAP Study (2014), London: ALNAP-ODI.

and validity.¹² Rather, based on the above, it becomes apparent that while much can be learned from existing evidence appraisal schemes, these approaches need to be tailored to the type and state of evidence in the humanitarian field. Some premises and assumptions for evidence appraisal in the humanitarian field are introduced below.

- 1. Evidence appraisal is an integral part of systematic evidence synthesis.** Even if the evidence in the humanitarian field is of different strength or types than, for example, laboratory-produced data, it still merits synthesis and appraisal. A strength of evidence synthesis approaches is that they bring together existing evidence in ways that highlight convergence and divergence among individual studies. This synthesis allows gaps in evidence to emerge, informing opportunities for future research. It also showcases areas of agreement and disagreement within the research in ways that may be difficult to discern when selectively considering individual studies. However, not all evidence yields same confidence in the results. Evidence appraisal allows researchers, policymakers, and practitioners to consider the weight of different types of evidence and findings and to decide with what confidence a certain type of evidence may be applicable to the context at hand.
- 2. Evidence appraisal approaches depend on the research question.** The selection of an evidence appraisal approach depends on the nature and scope of the research question, and the quantity, quality, and type of data available to answer it. No type of data is exempt from appraisal; conversely, no single appraisal system is necessarily suitable for all types of evidence found in humanitarian synthesis outputs. It is likely that review teams will have to employ a combination of tools to appraise the full spectrum of data included in their evidence synthesis. To that end, the accompanying catalogue¹³ summarizes quantitative, qualitative, and mixed methods appraisal approaches. Similar to the process of determining which studies are potentially includable in an evidence synthesis, there is a degree of subjectivity in how to select, interpret, and apply appraisal criteria. It may be possible to minimize this subjectivity in the appraisal process, while also acknowledging that any appraisal approach fundamentally requires reviewers to make judgments.
- 3. Reviewers should attempt to address missing information and data.** Methodology sections in study write-ups may provide less information than is required to fully appraise data. To the extent possible, reviewers should contact study authors to obtain a full picture of the methodological approach and study. For certain reviews, when little information about the methodology is included in the study and/or the authors cannot be reached, reviewers may need to conclude that there is insufficient information to appraise the evidence in question.
- 4. A study appraised as 'low confidence' can still be instructive for an evidence synthesis.** Valuable lessons can still be drawn from evidence with a high chance of bias or evidence that is deemed not generalizable. Studies can be used to illustrate patterns that appear elsewhere in the synthesis, or to provide a model for further exploration and replication in other contexts. Moreover, the lack of high quality evidence, based on these appraisal determinations, can be instructive in designing future studies and questions for research and evidence synthesis.

¹² Webb, P. "How strong is our evidence for effective management of wasting? A review of systematic and other reviews," *Food and Nutrition Bulletin Supplement* 1 (2015), 65s-71s (7).

¹³ Emerson, "Catalogue of Evidence Appraisal Schemes."

- 5. Evidence appraisal decisions should be documented and justified, as with other decisions in evidence syntheses.** As part of the evidence synthesis process, reviewers will need to appraise both the individual studies that are included and the body of includable evidence as a whole. Reviewers should provide a narrative explaining the criteria and process for these determinations to ensure that decisions are systematic and consistent, and to avoid, to the extent possible, 'cherry-picking' of evidence synthesis or appraisal approaches.

II. METHODOLOGY OF APPRAISAL SCHEME SYNTHESIS

The process for synthesizing and cataloguing evidence appraisal schemes began with a literature review of such schemes across disciplines. The purpose of this review was to sketch the landscape of evidence appraisal, understand the terminology and hierarchies of evidence codified in different approaches, and map the debates within and across fields on how evidence should be assessed.

For each appraisal scheme, the programme team recorded the approach's name; the author or institution affiliated with developing that approach (where relevant); a link to further information about the approach; the study designs appraised by the approach; quality criteria for appraisal; the hierarchy of evidence or rating system used by the approach (if any); and fields in which the approach has been used or applied. The corresponding catalogue of these approaches is arranged as such and ordered alphabetically according to the author of each appraisal approach.

The catalogue of evidence appraisal schemes relevant to the humanitarian field does not include every approach encountered during the literature review stage. This section provides additional information on the inclusion and exclusion criteria. Exclusion of a particular evidence synthesis approach from this catalogue does not suggest the approach is not useful or instructive; rather, that approach may not be as applicable to evidence synthesis in the humanitarian field. Additional information on these determinations is provided below.

- **Appraisal schemes that explicitly limited their applicability to a single (non-humanitarian) field or discipline were excluded, unless they contained criteria that were evidently transferable to other fields.** While evidence appraisal approaches that are specific only to a single field or discipline (such as health) were considered and have informed this document, they are not included in the matrix of approaches because they were deemed less immediately applicable and transferable to evidence synthesis in the humanitarian field. However, certain appraisal schemes -- such as the EPC Approach, GRADE, and the JBI Levels of Evidence and FAME Scale (all explained below) -- were designed for the health care field, but they do not contain health-specific quality criteria and were, therefore, included in the document.
- **Appraisal schemes that were specific to a single study design, as opposed to a variation of data, designs and evidence, were excluded.** The Humanitarian Evidence Programme embraces diverse types of study designs in its evidence syntheses and no systematic review is limited to a particular type of data or a specific study design from the onset. As a result, appraisal schemes that indicate applicability to only a single type of study (such as checklists for appraising randomized control trials or case control studies) were not included in the catalogue because they did not reflect the wide variety

of designs and sources from which humanitarian evidence syntheses can potentially draw data and insights. Such appraisal schemes may still be relevant to reviewers seeking to appraise a particular type of evidence, but they are less relevant for appraising a body of humanitarian evidence, considering the full spectrum of study designs and data included within it.

- **Appraisal schemes for systematic reviews were excluded.** Given that the purpose of this document is to highlight considerations in appraising evidence when conducting an evidence synthesis, appraisals of the quality of systematic reviews were not included in this catalogue. Appraising the quality of a particular systematic review is important for decision-makers, and it may be the subject of future inquiry for the Humanitarian Evidence Program, but appraising a systematic review as a standalone output is a fundamentally different project than appraising the individual studies and body of evidence within the review.
- **There is a distinction between appraisal of evidence-based research and the effectiveness of interventions.** A series of tools and criteria exist to guide the evaluation of particular interventions in humanitarian settings, largely referring to monitoring and evaluation of the implementation of these interventions on the ground during the cycle of a humanitarian crisis. Data collected through evaluations is critical to the body of evidence found in the humanitarian field and is eligible for inclusion in evidence syntheses. However, criteria used to guide evaluations, such as the Organization for Economic Cooperation and Development – Development Assistance Committee (OECD-DAC) criteria, are not sufficient to appraise the quality of the resulting studies; rather, these criteria focus on evaluating the quality of the *program*. Evidence appraisal approaches below, even those used for the purposes of meta-evaluation or the appraisal of evaluation quality, determine the quality of the evidence or data collected, rather than the quality of the programme being evaluated.

III. SUMMARY OF REVIEWED APPRAISAL APPROACHES

This section should be read in conjunction with the catalogue of appraisal systems. The following is a brief narrative summary of each approach, arranged in alphabetical order according to author.

The EPC Approach (Agency for Healthcare Research and Quality): The Agency for Healthcare Research and Quality's Evidence-based Practice Center (EPC) Approach "is based in large measure on the approach developed by the GRADE working group for assessing evidence," with differences in "some terminology, purposes of grading evidence, and characteristics of domains."¹⁴ Such differences, according to the authors of the approach, account for the fact that "EPCs often need to assess evidence from both trials and observational studies in evaluating a single outcome. They frequently encounter substantial heterogeneity in populations,

¹⁴ Berkman, ND, Lohr, KN, Ansari, M, McDonagh, M, Balk, E, Whitlock, E, Reston J, Bass, E, Butler, M, Gartlehner, G, Hartling, L, Kane, R, McPheeters, M, Morgan, L, Morton, SC, Viswanathan, M, Sista, P, Chang, S, "Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update" in *Methods Guide for Comparative Effectiveness Reviews* (Rockville, MD: Agency for Healthcare Research and Quality: 2013), 4, <http://effectivehealthcare.ahrq.gov/ehc/products/457/1752/methods-guidance-grading-evidence-131118.pdf>.

interventions, or outcomes that may preclude conducting meta-analyses.”¹⁵ The approach was designed for the field of healthcare, but could presumably be applied to reviews in other fields.

Evidence Quality Guidelines (Active Learning Network for Accountability and Performance in Humanitarian Action - ALNAP): ALNAP’s Evidence Quality Guidelines, published in an article by Clarke and Darcy entitled “Insufficient Evidence? The quality and use of evidence in humanitarian action,” consist of six main criteria that can be used to assess evidence quality in the humanitarian context. These criteria are: Accuracy, Representativeness, Relevance, Generalizability, Attribution, and Clarity around Context and Methods. The criteria were developed in an effort to standardize the language used to assess the strength of evidence in the humanitarian field, though the authors do not claim to present an authoritative approach for humanitarian evidence appraisal.

Quality Proforma (ALNAP): The ALNAP Quality Proforma were developed “as a way of assessing humanitarian evaluation reports drawing on current thinking and good practice in the evaluation of humanitarian action.”¹⁶ Aiming to improve the quality of evaluations in the humanitarian field, it provides an assessment tool for conducting meta-evaluations and a checklist for evaluation managers and evaluators. “Although originally designed with programme evaluations in mind, the Proforma can also be used to review evaluations of such activities as humanitarian management processes, funding partnerships and sectoral approaches.”¹⁷

Bond Evidence Principles (BOND): The Bond Evidence Principles and corresponding checklist are intended to assess and improve “the quality of evidence in evaluation reports, research reports and case studies” in the international development and humanitarian fields. “They have been designed specifically for NGOs and can be used when commissioning, designing and reviewing evidence-based work. The principles help ensure that decisions about projects and programmes are made on the highest quality basis.” The principles are: Voice and Inclusion, Appropriateness, Triangulation, Contribution, and Transparency.

C2 Checklist for Assessment of Methodological Quality in Economic Evaluation Studies (Campbell Collaboration): The Campbell Collaboration (C2) Checklist is a guide for appraising economic evaluations, though the authors indicate that the checklist alone is not sufficient and must be used in conjunction with expert consultation. In fact, C2 researchers presenting on “quality assessment” recommended against using a “quality scale” for evidence appraisal in systematic reviews, and instead suggest reviewers create *a priori* inclusion criteria and conduct statistical meta-analysis to assess variations in determined study quality criteria and effect size estimates.¹⁸ As this is only relevant for quantitative studies, the assessment of other study designs is yet to be determined.

CEBMA Critical Appraisal Questionnaires (Center for Evidence-Based Management): CEBMA’s [checklists](#) are a series of appraisal questions specific to study designs such as Meta-Analysis or Systematic Reviews, Controlled Studies, Cohort or Panel Studies, Surveys, Qualitative Studies, and Case Studies. When appraising evidence, CEBMA first recommends asking initial appraisal

¹⁵ Ibid, 5.

¹⁶ “Assessing the Quality of Humanitarian Evaluations, The ALNAP Quality Proforma 2005,” (London: ALNAP), 1, <http://www.alnap.org/resource/5320>.

¹⁷ Ibid.

¹⁸ Jeff Valentine, “Measuring and Assessing Study Quality,” Presentation for The Campbell Collaboration, http://www.campbellcollaboration.org/artman2/uploads/1/Quality_Assessment_Valentine_2013.ppt.pdf.

questions such as: "Is the evidence from a known, reputable source?"; "Has the evidence been evaluated in any way? If so, how and by whom?"; "How up-to-date is the evidence?".¹⁹ The organization then recommends asking more specific questions about the study design and results, potentially using one of their design-specific checklists if applicable.

The GRADE Approach (The Cochrane Collaboration): GRADE "is a well-developed formal process to rate the quality of scientific evidence in systematic reviews and to develop recommendations in guidelines that are as evidence-based as possible."²⁰ According to authors of an introductory paper, "GRADE was designed for reviews and guidelines that examine alternative clinical management strategies or interventions...the system can also be applied to rehabilitation, public health, and health systems questions."²¹ Beyond a rating system, GRADE "...offers a transparent and structured process for developing and presenting evidence summaries for systematic reviews and guidelines and for carrying out the steps involved in developing recommendations."²²

Confidence in the Evidence from Reviews of Qualitative Research (CERQual) (GRADE – CERQual Project Group, a subset of the GRADE Working Group): This is an approach for "assessing the confidence of evidence from reviews of qualitative research."²³ The four pillars of the approach are assessing methodological limitation in the design or conduct of the primary study, establishing the relevance of the study to the context of the review question, determining the adequacy of the data, and appraising the coherence of findings.²⁴ The approach is still under development to operationalize each of the components.

Qualitative Research Appraisal (The Cochrane Collaboration): The Cochrane Collaboration, acknowledging that qualitative research can be valuable to systematic reviews, created the Cochrane Qualitative Methods Group and published a broad guidance note on the inclusion and appraisal of qualitative studies as complementary to quantitative evidence in reviews.

CASP Critical Appraisal Checklists (Critical Appraisal Skills Programme): CASP's "eight critical appraisal tools are designed to be used when reading research...[such as] Systematic Reviews, Randomized Control Trials, Cohort Studies, Case Control Studies, Economic Evaluations, Diagnostic Studies, Qualitative Studies and Clinical Predication Rule."²⁵ The checklists are study design specific and offer guidance in determining the credibility of a study, but do not include a systematic rating process.

DFID Principles of High Quality Research Studies (UK Department for International Development (DFID): DFID's Principles provide general guidance on research quality criteria to be used when assessing the quality of single studies when seeking to determine the broader strength of a body of evidence. The guidance note also includes a systematic way for reviewers to describe and categorize studies according to type (primary, secondary, theoretical or conceptual), design (experimental, quasi-experimental, observational, etc.), and method. To summarize, the principles are: Conceptual Framing, Transparency, Appropriateness, Cultural

¹⁹ "What is Critical Appraisal?," (CEBMA), <http://www.cebma.org/frequently-asked-questions/what-is-critical-appraisal/>.

²⁰ Marcel Dijkers, "Introducing GRADE: a systematic approach to rating evidence in systematic reviews and to guideline development," (The Center on Knowledge Translation for Disability and Rehabilitation Research: 2013) http://ktdrr.org/products/update/v1n5/dijkers_grade_ktupdatev1n5.pdf.

²¹ Ibid., 1.

²² Ibid., 2.

²³ "About CERQual." <http://cerqual.org>

²⁴ "What is the CERQual Approach?" <http://cerqual.org>

²⁵ "CASP Checklists," (Oxford: CASP UK), <http://www.casp-uk.net/#!casp-tools-checklists/c18f8>.

Sensitivity, Validity, Reliability, and Cogency. According to the Note, "A really rigorous review of the evidence on a given topic should give due consideration to all seven of these aspects of study quality. It is possible to construct checklists or scorecards to grade evidence based on these criteria, and it is expected that DFID Evidence Papers will do so."²⁶

Prompts for Appraising Qualitative Research (Dixon-Woods et al): The Prompts presented in a paper by Dixon-Woods, Shaw, Agarwal et al. entitled "The problem of appraising qualitative research" were developed by a project team (including the authors) in the ESRC Research Methods Programme. The prompts were designed to address "...the need to avoid commitments to particular methodological approaches, and have distinguished between aspects of reporting and aspects of study design and execution."²⁷ Referring to the rationale for developing the prompts for qualitative research, the authors write, "Because these prompts are generic, it would also be possible for them to be complemented by prompts that are specific to different methods of data collection and qualitative methodologies. So far there have been few attempts to develop such methodology-specific approaches."²⁸

Review Specific Assessments (EPPI-Centre): EPPI-Centre's Review Specific Assessments were developed for the purposes of determining the trustworthiness and strength of evidence in relation to the systematic review question. Rather than assessing the trustworthiness of a study independently from consideration of the review question, the EPPI-Centre approach suggests taking a composite result of the following three criteria: methodological quality, methodological relevance for addressing the review question, and topic relevance in relation to the review question.²⁹ Results of each study along these criteria are then synthesized to determine the overall weight of evidence.

The Maryland Scientific Methods Scale (SMS) (Farrington et al.): The Maryland SMS "ranks policy evaluation from 1 (least robust) to 5 (most robust) according to the robustness of the method used and the quality of its implementation. Robustness, as judged by the Maryland SMS, is the extent to which the method deals with the selection biases inherent to policy evaluations."³⁰

JBI Levels of Evidence and the FAME Scale (The Joanna Briggs Institute): The JBI Levels of Evidence and grades of recommendation through the FAME scale provide a systematic framework for evaluating the strength of a wide variety of evidence when conducting systematic reviews in the healthcare field. Levels are determined for 'Effectiveness' (Study Design), Diagnosis, Prognosis, Economic Evaluations, and Meaningfulness.³¹ Resulting recommendations from systematic reviews are then assigned a grade of 'strong' or 'weak,' also applying the FAME

²⁶ "Assessing the Strength of Evidence," DFID How to Note (London: DFID: 2014), 13, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/291982/HTN-strength-evidence-march2014.pdf.

²⁷ Dixon-Woods, M, Shaw, R L, Agarwal, S, Smith, J A, "The problem of appraising qualitative research," *Qual Saf Health Care*, Vol. 13 (QSHC: 2004), 224, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1743851/pdf/v013p00223.pdf>.

²⁸ Ibid.

²⁹ "EPPI-Centre Methods for Conducting Systematic Reviews," (EPPI-Centre: 2007), 14, <http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=hQBu8y4uVwI%3d&tabid=88&mid=6162>.

³⁰ Margarida Madaleno and Sevrin Waights, "Guide to scoring methods using the Maryland Scientific Methods Scale," (What Works Centre for Local Economic Growth), 2, <http://www.whatworksgrowth.org/public/files/Scoring-Guide.pdf>.

³¹ "Levels of Evidence," *The JBI Approach*, (The Joanna Briggs Institute: 2014), <http://joannabriggs.org/jbi-approach.html#tabbed-nav=Levels-of-Evidence>.

scale (Feasibility, Appropriateness, Meaningfulness, Effectiveness) to assess the strength of recommendation.³²

JCSEE Program Evaluation Standards (Joint Committee on Standards for Educational Evaluation): JCSEE's Standard Names and Statements are intended to guide the quality, utility, and accountability of programme evaluations. They can be used to assess the quality of programme evaluations along the following criteria: Utility, Feasibility, Propriety, Accuracy, and Evaluation Accountability. They were initially intended to inform programme evaluation in the education field, but could be applied to evaluations in other fields as well.

The Meta-Evaluation Checklist: The Meta-Evaluation Checklist (MEC) developed by Michael Scriven seeks to answer the question of "What are the criteria of merit for an evaluation in any field, including program evaluation?"³³ and generate general standards that can be applied across fields. The criteria that he proposes are: Validity, Clarity, Credibility, Propriety, Cost-utility, and Generalizability.

Framework for Assessing Qualitative Evaluations (Spencer et al.): In a report titled "Quality in Qualitative Evaluation: A framework for assessing research evidence" by Spencer et al, the authors propose a comprehensive framework for appraising the quality of qualitative evaluations "concerned with the development and implementation of social policy, programmes and practice."³⁴ The authors write, "This framework draws heavily on previously developed quality criteria, both from the general methodological literature and from pre-existing guidelines. However, it also takes heed of the persistent concern that formalized criteria should avoid being rigidly procedural or over-prescriptive. It has therefore been devised to aid informed judgment, not mechanistic rule-following."³⁵ The framework includes numerous appraisal questions, each with quality indicators for consideration, and room for notes on the study being appraised. The authors of the framework leave any grading or scoring mechanism to the reviewer's discretion.

The Humanitarian Evidence Programme hopes the accompanying catalogue and narrative summary of appraisal approaches are useful to those seeking to review evidence appraisal approaches of relevant to the humanitarian field. Please contact [Roxanne Krystalli](#), Programme Manager of the Humanitarian Evidence Programme at the Feinstein International Center, with questions, comments, or suggested amendments to these documents. Recommendations specific to Humanitarian Evidence Programme reviewers follow in the Appendix.

³² "Grades of Recommendation," *The JBI Approach*, (The Joanna Briggs Institute: 2014), <http://joannabriggs.org/jbi-approach.html#tabbed-nav=Grades-of-Recommendation>.

³³ Michael Scriven, "Evaluating Evaluations: A Meta-evaluation Checklist," (London: ALNAP: 2011), 1, <http://www.alnap.org/resource/8154>.

³⁴ Spencer, L, Ritchie, J, Lewis, J, Dillon, L, "Quality in Qualitative Evaluation: A framework for assessing research evidence," (London: Government Chief Social Researcher's Office: 2003), 2, http://www.civilservice.gov.uk/wp-content/uploads/2011/09/a_quality_framework_tcm6-7314.pdf.

³⁵ Ibid.

APPENDIX I. INSTRUCTIONS FOR HUMANITARIAN EVIDENCE PROGRAMME REVIEWERS

This section specifically addresses reviewers carrying out humanitarian evidence syntheses on behalf of the Humanitarian Evidence Programme and outlines questions reviewers should consider when selecting an evidence appraisal approach for their reviews.

Please draw on your past experience and the experience of your colleagues in determining which evidence appraisal scheme is relevant for your review question. As a starting point, you may find it useful to browse the catalogue of select evidence appraisal approaches reviewed by the programme team. Though all approaches contained in the catalogue are deemed to have some applicability to the humanitarian field, the approaches you may be most likely to use are the EPPI-Centre's Review Specific Assessments, ALNAP's criteria, DFID's guidance for evidence appraisal, and CERQual. For reviews that lend themselves to more experimental data, the GRADE approach can be a suitable fit. As you review the approaches, please consider the following questions:

- ✓ What is the predominant study design included in your review? Which evidence appraisal best corresponds with it?
- ✓ How will you account for other study designs in the review that may not be best appraised through your predominant appraisal approach?
- ✓ Which appraisal systems have been used in reviews on similar topics or themes? Do those systems appropriately account for all types of evidence, emerging from different study designs or publication forums, in your own review?
- ✓ Which approach or combination of approaches may assist in the appraisal of qualitative data and/or the grey literature?

After you finished appraising studies, please consider:

- ✓ If your appraisal suggests there may be low confidence in the results, how may the studies in your evidence synthesis still be able to inform humanitarian practitioners, researchers, and decision-makers? How would you caution these audiences on any findings?
- ✓ Have you contacted study authors, to the extent possible, to gather information on the methodology and findings, when such information is not readily available or complete in the published version of the study?
 - When such information remains unavailable, have you adequately documented why a full appraisal of that evidence was not possible based on the information given and what additional information would have been useful in completing such an appraisal?
- ✓ Based on your appraisal of the respective studies, what comments may you be able to make on the state of evidence in your review on the whole? Please pay particular attention to gaps or divergences in the research findings that future research may be able to address.