

A Framework to Assess the Quality of Non-traditional Articles in the Field of Disaster Response and Management

Mary Hall, MPH; Chris Cartwright, MPH; Andrew C. K. Lee, MD

ABSTRACT

Objective: While carrying out a scoping review of earthquake response, we found that there is no universal standardized approach for assessing the quality of disaster evidence, much of which is variable or not peer reviewed. With the lack of a framework to ascertain the value and validity of this literature, there is a danger that valuable insights may be lost. We propose a theoretical framework that may, with further validation, address this gap.

Methods: Existing frameworks – quality of reporting of meta-analyses (QUORUM), meta-analysis of observational studies in epidemiology (MOOSE), the Cochrane assessment of bias, Critical Appraisal Skills Programme (CASP) checklists, strengthening the reporting of observation studies in epidemiology (STROBE), and consensus guidelines on reports of field interventions in disasters and emergencies (CONFIDE)–were analyzed to identify key domains of quality. Supporting statements, based on these existing frameworks were developed for each domain to form an overall theoretical framework of quality. This was piloted on a data set of publications from a separate scoping review.

Results: Four domains of quality were identified: robustness, generalizability, added value, and ethics with 11 scored, supporting statements. Although 73 out of 111 papers (66%) scored below 70%, a sizeable portion (34%) scored higher.

Conclusion: Our theoretical framework presents, for debate and further validation, a method of assessing the quality of non-traditional studies and thus supporting the best available evidence approach to disaster response. (*Disaster Med Public Health Preparedness*. 2019;13:147-151)

Key Words: best available evidence, disaster management, quality assessment

Effective disaster response depends upon good quality, reliable evidence.¹ The 2016 launch of the United Nations International Strategy for Disaster Reduction (UNISDR) Science and Technology Partnership aimed to advance the role of science and technology for the implementation of the Sendai Framework and highlighted the need for a strong evidence-based approach. However, disasters are random by nature and not easy to predict. Similarly, disaster studies are difficult to organize in a timely manner. This results in a lack of robust, empiric studies and a preponderance of observational or “non-traditional” articles (eg, field reports, letters to the editor, narratives, commentaries, evaluations, needs assessments, or case reports).²

These non-traditional articles, as well as gray literature (literature that is unpublished or not published commercially), are often deemed to be of low quality and their findings dismissed, as a result. The “best available evidence” approach advocates the collation of information from “all available sources without restriction by hierarchy or grade.”³ It recognizes that

these articles, although of undefined quality, may often contain valuable and useful information relevant to the field. There is considerable diversity in the literature base, ranging from “disaster tourism” commentaries and opinion pieces⁴ to more detailed field reports. If all articles are summarily dismissed regardless of content, there is a real risk that valuable insights could be missed as a result.

While carrying out a scoping review of earthquake response,⁵ we found that there was no universal standardized approach for assessing the quality of disaster evidence. A scoping review uses a systematic review methodology but allows for the review of a broader, less restrictive range of evidence and is useful for disaster-related reviews where literature may be of a “non-traditional” type. In the absence of a quality assessment measure, we were unable to distinguish between those articles in our scoping review that may have more “weight” in contributing to the evidence base and those with little added value, relevance, or reliability. We attempted to address this gap by identifying the key

domains of quality in existing quality frameworks and using these to develop a framework for non-traditional studies.

METHODS

There currently exists a range of quality assessment tools for traditional studies such as the quality of reporting of meta-analyses (QUORUM),⁶ meta-analysis of observational studies in epidemiology (MOOSE),⁷ the Cochrane assessment of bias,⁸ Critical Appraisal Skills Programme (CASP) checklists,⁹ strengthening the reporting of observational studies in epidemiology (STROBE),¹⁰ and the disaster reporting framework: consensus guidelines on reports of field interventions in disasters and emergencies (CONFIDE).⁴

We identified the common domains of quality that these existing tools encompass. We then extrapolated those common domains that might be applicable to non-traditional study types. We tested the selected domains using a data set of 152 publications from a separate scoping review⁵ to assess alignment with identified themes and categories emerging from the scoping review.

All published material that was not of a traditional study type (ie, trial, cohort, case control, longitudinal, systematic review, or meta-analyses) was classed as non-traditional. Studies included field reports, first/third person narratives, letters to the editor, needs assessments, and commentaries. The characteristics of these non-traditional articles were then mapped to the main domains of the existing quality frameworks. Using an inductive approach, we identified 3 initial domains of quality and 11 quality indicators based on the originally identified common domains of quality, the 152 publications, and by the authors' discussions and consensus. These formed the basis of our theoretical framework.

In our proposed framework, each quality indicator was given a defined measure with a numerical value assigned. Each indicator was accorded equal weighting. Articles were graded for each indicator from A to D or N (*not applicable* or *not relevant*) and a numerical value applied depending on the grading (A=3, B=2, C=1, D=0, N=-3). A scoring system was devised so that each assessed article would be given a percentage score equivalent to the total proportion of points allocated.

We piloted the proposed framework on an initial 20 non-traditional articles identified by the scoping review. Subsequently, an additional category was added to the overall framework, making 4 in total. The final framework was applied independently by 2 researchers to all non-traditional articles (n=111) identified in the data set of publications from the scoping review.⁵

RESULTS

Common themes identified in existing frameworks included study characteristics, study population, internal and external

validity, study design, and study reporting mechanisms. Resulting domains of quality identified for our proposed framework were robustness, generalizability, added value, and ethical consideration (Table 1). The 11 indicators included triangulation to literature, use of emotive language, level of lessons learned, author perspective and bias, time period, sample population, disease description, implications, applicability, and ethics.

Of the 152 articles identified in the scoping review, 41 (27%) were of a "traditional" design, including cross sectional (n=26), cohort (n=1), and mixed methods (n=6). The majority (73%) of articles were classed as *other* or *non-traditional* (n=111), including field reports (n=69), letters to the editor or opinion pieces (n=22), reviews of support provided (n=9), and audits (n=2). Our draft framework was applied to these 111 non-traditional articles.

Whereas 65.8% of articles achieved less than 70% of the total possible score (Figure 1), 38 out of the 111 (34.2%) articles reviewed scored higher, with 2 scoring 90% or higher. One was a retrospective case review of injuries seen in a rural hospital immediately after the disaster, and the second was a letter to the editor detailing disaster preparedness in rural hospitals. Alternative quality frameworks would usually rate both of "low quality," yet both have the potential to contribute to knowledge and learning around disaster management.

Articles tended to score highly (grade "A") on measures of "study time period recorded" and "use of language" (ie, mostly factual language, less than 10% emotive), both of which reflect the articles' robustness. Of those articles that provided a time frame, 64.9% (72/111) described events in the 6 months following the disaster, 4.5% (5/111) in the 6-12 months post-disaster, and 18.9% (21/111) in the year following the disaster; 16.2% (18/111) of articles described multiple time periods. While authors were clear regarding of whom they were writing on behalf (eg, an international organization), few discussed whether this would have any implications or bias on their reporting. This accounts for the predominance of "B" ratings for author bias; 73.9% (82/111) of articles were written "in-country" by expatriate staff working as part of the response, 5.4% (6/111) in-country by native staff, and 19.8% (22/111) externally by non-native staff. Referencing findings with the literature or evidence was not evident in 91% of articles, and, while 42.5% discussed system-wide lessons learned, 54.7% either did not discuss any potential lessons learned or gave only limited attention to possible learning for a future disaster response.

DISCUSSION

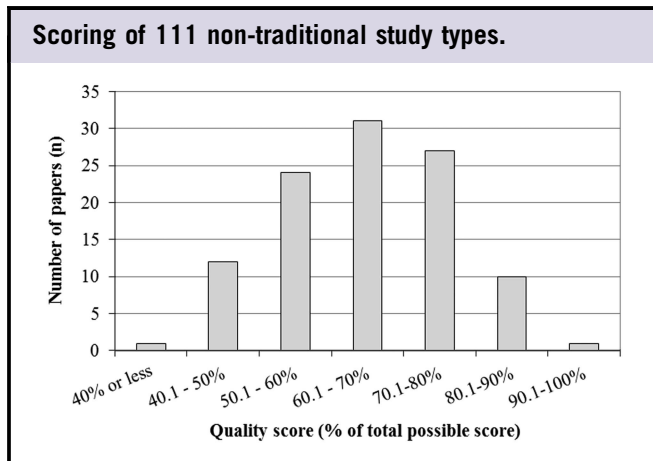
In the hierarchy of evidence, articles such as case reports, expert views, field reports, or gray literature are classed as bottom of the triangle, of low quality, and, by implication, of little value in their contribution to future practice.¹¹ Various

TABLE 1

Quality Assessment Framework of Non-Traditional Study Type

Robustness: Study Information and Context	Score	Time Period of Study/ Data Collection	Definition
	A	Clearly defined:	Exact time frame of data collection given
	B	Some definition:	Approximate time frame given
	C	No definition:	No time frame given
	Score	Sample Population	Definition
	A	Clearly defined:	3 out of 4 parameters from age, gender, number, and study population taken from
	B	Some definition:	1 or 2 parameters defined from above list
	C	No definition:	No parameters defined
	N	N/A:	Not applicable/none population study
	Score	Disease Description/ Patient Condition	Definition
	A	Clearly defined	
	B	Some definition	
	C	No definition	
	N	N/A	
Score	Author Bias	Definition	
A	Clearly defined	Clear of whom the author is writing on behalf, with some reflection of potential bias	
B	Some definition	Clear of whom the author is writing on behalf, with minimal reflection of potential bias	
C	None	Either not clear of whom the author is writing on behalf, or clear but with no acknowledgment of potential bias	
Score	Type of Language	Definition	
A	Minimal	Factual; less than 10% emotive language	
B	Low bias	Less than 30% of language includes below	
C	Moderate bias	30%-50% of language includes below	
D	High bias	Greater than 50% emotive language used: use of "I"; personal beliefs, reflections or experiences; inner experiences	
Generalizability	Score	Author Perspective	Definition
	A		The study was written and reported in-country by a native of the country.
	B		The study was written and reported in-country by an expatriate working as part of the response.
	C		The study was written and reported externally (to the country) by an expatriate working as part of or observing the response.
	Score	Applicability	Definition
	A	Very applicable	Results/findings are applicable, relevant, and likely to be similar in other settings.
B	Moderately applicable	Results/findings are applicable, relevant, and likely to be similar in other settings, but some aspects are specific to the study setting only.	
C	Low applicability	Results/findings are only applicable and relevant to the study setting and not likely to be similar in other settings.	
N	N/A	Opinion piece	
Added Value	Score	Lessons Learned	Definition
	A	High	System/organizational level lessons learned, described, and discussed
	B	Moderate	Individual lessons learned
	C	Low	None or limited lessons learned
	Score	Triangulation to the Literature	Definition
	A	High	Findings linked to previous studies, guidance, or literature as part of discussion
	B	Moderate	Some linkage to previous studies, guidance, or literature
	C	Low	No linkage to previous studies, guidance, or literature
Score	Implications	Definition	
A	Good	Contributes something new and/or different in terms of understanding/insight or perspective; suggests ideas for further research; suggests implications for policy and/or practice	
B	Fair	1 or 2 of the above	
C	None	None of the above	
Ethics	Score	Ethics	Definition
	A	Good	Where relevant ethical consideration has been discussed and approval gained (eg, in use of photos, patient identifiable data, informed consent)
	B	Fair	Some mention of ethical considerations but no clear ethical approval sought or gained
	C	None	No mention of ethical consideration; unclear if permission sought or gained
	N	N/A	Not relevant/none population study
Overall Score	The following values are applicable to scores: A=3, B=2, C=1, D=0, N=-3 The denominator for each paper is 33+N. The numerator for each paper is A+B+C+D. The total score for each paper = numerator/denominator * 100.		

FIGURE 1



frameworks have been developed to appraise published articles, but these favour the traditional study types such as trials, systematic reviews, cohort, and case-control studies. Other article types are ignored, leading to considerable loss of information, particularly for fields (such as disasters) where more robust study types are difficult to conduct and are consequently rare. Attempts have been made to try and encourage and capture lessons learned from disasters such as the CONFIDE statement for disaster reporting. However, it is limited because it does not assess the quality of the reports.⁴

Ideally, practice should be evidence-based, that is, based on the best evidence. But, in reality, it is more likely that practice is rooted in the “best available evidence,” implying a need to incorporate the wider body of published articles and studies into the evidence base. In the disaster field, there are a number of facilities that collate such evidence, including Evidence Aid¹² and the Disaster Information Management Resource Center.¹³ With such quantity of literature, there remains a requirement to balance the need for collating insights and minimizing information loss, with the need to critically appraise the quality of what is published. This is both the science and art of evidence-based practice. We put forward a framework to support this process.

The use of a single overall score provides an opportunity to flag those articles that may have added value out of the overall body of evidence. Further categorization could be applied, such as banding by scores (eg, high $\geq 75\%$, moderate 50%-75%, low $< 50\%$), or the scores could be further separated out into the 4 quality domains to provide a more detailed breakdown. The CASP appraisal checklist purposefully does not use a scoring system, and this may be applicable to our framework. Further piloting and validation would support the identification of the most useful approach.

The proposed framework has a number of strengths and limitations. In developing this framework, we could have introduced our own biases regarding disaster literature. For

example, we awarded an “A” for papers that had been written in-country and by the respective country’s own nationals for author perspective and a “C” for papers that had been written external to the country where the disaster was and by non-native authors. This, in part, was based on our previous review experience where we found many opinion pieces written by non-native authors using a journalistic style of writing, often with high levels of emotive language and in recognition that a native author may offer a distinct perspective and insights that are not always apparent to an external author.

We used existing quality frameworks to help develop this framework to introduce an element of robustness. However, we acknowledge that such tools were meant for more conventional study types and do not neatly fit the types of evidence that disaster reports usually fall under. We are aware that we only looked at a selection of existing quality assessment frameworks. That said, our aim was not to collate all possible frameworks but to identify the main quality domains common to most of them. By applying our framework to a fairly large data set of articles, we believe this allowed us to comprehensively test the quality measures used. Nevertheless, we acknowledge that our framework is an initial starting point only, and further studies will be required to validate the framework further.

CONCLUSION

Evidence-based interventions should be a cornerstone of disaster management and response. Where robust evidence is sparse, the principle of “best available evidence” becomes more important. Hidden within the plethora of field reports, surveys, opportunistic studies, and other non-traditional articles may be important lessons for practice that need to be mined. This proposed framework is a tool for this purpose and invites further debate on how the disaster management community can tap into this vein of past learning and experience.

About the Authors

School of Health and Related Research, University of Sheffield, UK (Ms Hall, Mr Cartwright, Dr Lee).

Correspondence and reprint requests to Ms Mary Hall, School of Health and Related Research, University of Sheffield, 30 Regent St, Sheffield, S1 4DA, UK (e-mail: mary.hall5@nhs.net).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interests.

REFERENCES

1. Knox Clarke P, Darcy J. *Insufficient Evidence? The Quality of Use of Evidence in Humanitarian Action. ALNAP Study*. London: ALNAP/ODI; 2014.
2. Challen K, Lee ACK, Booth A, et al. Where is the evidence for emergency planning: a scoping review. *BMC Public Health*. 2012;12:542.

3. Bradt DA. Evidence-based decision making (part 1): origins and evolution in the health sciences. *Prehosp Disaster Med.* 2009;24(4):298-304.
4. Bradt DA, Aitken P. Disaster medicine reporting: the need for new guidelines and the CONFIDE statement. *Emerg Med Australas.* 2010;22:483-487.
5. Cartwright C, Hall M, Lee ACK. The changing health priorities of earthquake response and implications for preparedness: a scoping review. *Public Health.* 2017;150:60-70.
6. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUORUM statement. *Br J Surg.* 2000;87:1448-1454.
7. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) Group. *JAMA.* 2000;283:2008-2012.
8. Cochrane assessment of bias. Published. http://handbook.cochrane.org/chapter_8/table_8_7_a_possible_approach_for_summary_assessments_of_the.htm Published March 2011. Accessed February 2, 2017.
9. CASP checklists. Published 2018. <http://www.casp-uk.net/casp-tools-checklists>. Accessed February 27, 2017.
10. von Elm E, Altman DG, Egger M, et al. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ.* 2007;335:806-808.
11. De Brún C. Finding the evidence: a key step in the information production process. *The Information Standard;* 2013. <https://www.england.nhs.uk/wp-content/uploads/2017/02/tis-guide-finding-the-evidence-07nov.pdf>. Accessed October 11, 2017.
12. Evidence Aid. Published 2018. <https://www.evidenceaid.org>. Accessed May 10, 2018.
13. Disaster Information Management Resource Center. <https://www.disaster.nlm.nih.gov>. Accessed May 10, 2018.