

Monitoring and evaluation

Authors

Heidi Hung, Division of Global Health and Humanitarian Medicine, Jockey Club School of Public Health and Primary Care, CUHK, Hong Kong SAR, China.

Gloria K.W. Chan, CCOUC, CUHK, Hong Kong SAR, China.

Emily Y.Y. Chan, CCOUC, CUHK, Hong Kong SAR, China; GX Foundation, Hong Kong SAR, China.

4.15.1 Learning objectives

To understand key factors in the development of studies that focus on monitoring and evaluation (M&E) of Health EDRM interventions, including:

1. Meaning and significance of M&E.
2. Existing M&E frameworks in DRR and health.
3. Methodologies for Health EDRM M&E studies.
4. Challenges in developing Health EDRM M&E studies.

4.15.2 Introduction

The relatively new discipline of Health EDRM emerged from the cross-over between health and DRR. Health EDRM is “the systematic analysis and management of health risks, posed by hazardous events, including emergencies and disasters, through a combination of hazard, exposure and vulnerability reduction to prevent and mitigate risks, preparedness, response, and recovery” (1). M&E studies are an important means for assessing the impact of Health EDRM, drawing on lessons from the disciplines of both health and disaster management.

The concept of M&E is widely applied across disciplines and by different organizations. It includes a wide range of investigations, from M&E of the impact of national climate adaptation policies to the outcomes of reproductive health programmes of a local NGO. In general, M&E aims to assess the performance of an initiative, programme, project or intervention and to provide evidence to improve future ones. Monitoring and evaluation are two distinct processes involving different methodologies and techniques. According to the disease-specific M&E toolkit issued by WHO, monitoring is the routine tracking of an intervention’s inputs (such as financial resources, staff time, cost of medical supplies) and outputs (such as new health services, improved drug supply system, new skills among health workers), which includes regular record-keeping, reporting and surveillance. Evaluation, meanwhile, is the assessment of the contribution

made by the various factors of an intervention given the output or outcome (2). A similar concept of M&E has been applied in DRR, as elaborated in the 2015 Monitoring and Evaluation Framework published by the United Nations Office for Disaster Risk Reduction (UNDRR – formerly known as UNISDR). Definitions adopted by UNDRR emphasize the function of monitoring as providing an early indication on the progress, or lack thereof, of an intervention, and that evaluation should not be an one-time event but should be carried out at several time points in response to evolving needs in relation to the achievement of the intended outcomes (3).

This chapter begins with an overview of some of the available frameworks for M&E, before setting out choices that have to be made when developing a M&E study, and concluding with a discussion of the major challenges.

4.15.3 M&E frameworks in disaster management and health

The Sendai Framework for Disaster Risk Reduction 2015 – 2030 (Sendai Framework) was adopted on 18 March 2015 at the Third World Conference for Disaster Risk Reduction, in Sendai, Japan (4). The Sendai Framework guides the global direction in DRR until 2030 and its emphasis on monitoring and accountability illustrates the critical role of M&E in relation to disasters. The Framework has seven targets, with 38 global indicators that were recommended by an Open-ended Intergovernmental Expert Working Group for measuring the implementation progress of these targets. National governments are also required to define custom targets and indicators to measure their progress, addressing the four priorities of the Sendai Framework, based on national priorities. UNDRR built the Sendai Framework Monitor Initiative (see Chapter 2.1), under which Member States have to report on the indicators and global assessments (4).

While the Sendai Framework has emphasized the significance of M&E in DRR efforts, M&E is considered a relatively weak area in DRR research. There are a number of possible reasons for this, such as the lack of common terminology and methodology, insufficient training of relevant personnel and researchers and a large range of intervention forms (5). In fact, in disasters, different agencies often have their own M&E frameworks involving different approaches, indicators and outcomes. In 2016, Scott and colleagues (6) proposed a common framework for DRR programmes to strengthen M&E quality in this field and suggested three outcomes:

- i) whether the ability of participants to employ DRR-related knowledge, innovation, education, communication or technology has been enhanced;
- ii) whether the DRR institutional framework has been strengthened (for example, development of DRR policies and strategies, range of stakeholders involved in the process); and
- iii) whether the motivation to achieve effective DRR has been improved.

While there are benefits of a common framework, it has to be pointed out that the appropriateness of an M&E system depends on a range of factors, including the level of development in the country involved, the scale and nature of the disaster in question, the capacity of the agency, and the funding sources (7). Notably, existing M&E frameworks in Health EDRM

are mainly designed for assessing programme effectiveness and not intended for research purposes. However, they still provide valuable and practical information on the purpose of M&E and the major components to be covered; the following sections provide a brief overview of UNDRR's M&E Framework, WHO's M&E Toolkit for the health component in Health EDM, and The Sphere Standards .

UNDRR's M&E Framework

UNDRR's M&E Framework is very comprehensive, with detailed description of terminologies, types of indicators, criteria for selecting indicators, implementation plan, data collection methods and reporting mechanisms (3). The guiding factors contained in the Framework for selecting indicators are:

- i) linkage between indicator and expected outcomes;
- ii) temporal stability of the definition of the indicator;
- iii) availability of data and cost-effective data collection instruments;
- iv) comprehensibility of the indicator; and
- v) quantitative nature of the indicator.

UNDRR classifies evaluations into formative evaluations (such as needs assessments or process evaluations) and summative evaluations (such as impact evaluations, cost-benefits analyses). Formative evaluations aim to improve implementation quality of the project under evaluation, while summative evaluations examine the outcome and impact of the project. Recommended data collection methods for evaluations generally include questionnaires, surveys, checklists, interviews, documentation review and observations.

WHO's M&E Toolkit

The M&E Toolkit developed by WHO adopts a similar M&E approach, although it is disease-specific (2). In particular, the WHO M&E Toolkit emphasizes the importance of comparable indicators across time and countries, data collection supported by a surveillance system, with a data dissemination plan.

The Logical Framework Approach

The logical framework ('logframe') approach, while not explicitly elaborated in the UNDRR M&E framework, is an M&E management tool commonly adopted in development projects. Under the logframe approach, project strategy, objectives and outputs are clearly defined, with objectively verifiable indicators developed under each category, and they are all presented in a single matrix. Such an approach has been adopted in projects by several agencies in the United Nations family, including the United Nations Development Programme (UNDP) (5) and the United Nation Children's Fund (UNICEF).

The Sphere Standards

The Sphere Standards are the most widely recognized principles and minimum standards for improving the quality of humanitarian operations and the accountability of the humanitarian sector. They comprise the Core Humanitarian Standard as well as standards in four technical areas, including health, water supply, sanitation and hygiene promotion; food security and nutrition; and shelter and settlement (8). M&E is emphasized,

with performance indicators listed for each of the nine commitments under the Core Humanitarian Standard. Moreover, “key indicators” is one of the main components of all the technical standards, covering various aspects of a humanitarian operation. For example, for an operation to control communicable diseases, relevant indicators include initiation of outbreak investigation within 24 hours of notification and case fatality rate maintained at 1% or lower in the case of cholera (9). These indicators act not only as guidelines for designing health services during a disaster, but also for monitoring and evaluating the operations. In order to strengthen the role of M&E in humanitarian operations, the Sphere Monitoring and Evaluation guide has been published to elaborate on how the Sphere Standards could be used for M&E. M&E under Sphere emphasizes in particular monitoring of the context of the humanitarian operation, the activities and processes, and the impact of the operation on the affected population; and that there should be real-time evaluation, mid-term evaluation and final evaluation (10).

4.15.4 Designing M&E studies for Health EDRM: the choices to be made

Although the existing frameworks provide some assistance to researchers in designing their own M&E studies, it is critical that the actual design must reflect the specific context of the study and fit the needs of the research. For any M&E study to serve its purpose, choices must be made in relation to a range of issues (such as approach, components, methods and process, and so on). This section elaborates on this aspect, in particular in regards to nature of the evaluation, evaluation framework, levels, data sources and study design.

Nature of the evaluation

There are many different types of evaluation, depending on the objective of the study. These include formative and summative evaluations, which are further divided into various subtypes (Table 4.15.1):

Table 4.15.1 Subtypes of formative and summative evaluations (3)

Formative evaluation	Summative evaluation
Needs assessment	Outcome evaluations
Evaluative assessment	Impact evaluation
Structured conceptualization	Cost-effectiveness and cost-benefit analysis
Implementation evaluation	Secondary analysis
Process evaluation	Meta-analysis

As discussed in Chapter 3.5 for research generally, M&E researchers need to be clear about their main objective and research question and select the most suitable type of evaluation accordingly. M&E studies in Health EDRM require appropriate and practical research methods for monitoring and evaluating the interventions implemented to strengthen DRR capacity. In general, more attention has been paid to impact evaluation than process monitoring.

4.15

Evaluation framework

There are many evaluation frameworks available to Health EDRM researchers to help them plan systematic data collection, analysis and interpretation. After deciding on the nature of the evaluation, researchers could consider the study focus and draw up the most suitable evaluation framework. For example, an M&E study might focus on the outcome of interventions (11) or the cost and benefit of a DRR measure (12). Moreover, researchers might wish to conduct a theory-driven study (see Chapter 4.10) to answer pre-defined questions (13) or to test hypotheses about what they expect to find. Case Study 4.15.1 provides an example of an M&E framework for disaster management at national level.

Case Study 4.15.1

South Africa Disaster Management M&E Framework (14)

South Africa is exposed to various weather hazards (such as drought and cyclones) and the country's extensive coastline means that there are potential marine and coastal threats (such as floods). Furthermore, people living in poor and ecologically fragile areas are faced with additional risks.

In light of severe disasters experienced in the 1990s, discussions and consultations at local, national and international levels led to the adoption of the Disaster Management Act in South Africa in 2002. The National Disaster Management Centre (NDMC) was established under the Act. To ensure the performance of disaster management-related policies and programmes, the South African Government issued the Disaster Management Monitoring and Evaluation Framework in 2014 to "provide a comprehensive and integrated strategic monitoring and evaluation direction to the entire Disaster and Fire services management to determine, on an on-going basis, how best to maximize the value of prevention, reduction, response and intervention". This Framework outlines:

- Key processes, mechanisms, tools, templates, strategies and methods for M&E;
- M&E architecture, system design and performance monitoring and evaluation plans;
- How evaluation findings will be used to enhance evidence-based decision making and accountability, and give feedback to policy development or implementation review mechanisms.

South Africa's Disaster Management Monitoring and Evaluation Framework requires that all evaluations should comprise three components: internal rapid assessment, long-term impact and multi-dimensional evaluation projects, and joint venture evaluation projects with strategic partners.

Levels

Given the potential complexity of a programme, it is essential for researchers to decide which "levels" to focus on. The "level" might be at the activity or output level as opposed to the outcome or strategic level. It might also be at the organizational level, which may be international, national, inter-agency, community or the individual level (for example,

patients or other beneficiaries). The choice of level affects the data sources and the study design. Multilevel M&E studies are possible.

Data sources

Research methodologies for M&E studies in Health EDRM cover various components, including formulation of hypothesis and research questions, drawing up study design, recruiting subjects, designing research tools and data collection methods (what, when, how and by whom), choosing indicators, and data analysis and dissemination. Health EDRM researchers must be prepared to secure access to a wide range of data sources which could be fed into different components of the study. For example, different types of indicators require different categories of data sources, and whenever possible, there should be at least two sources of data for each indicator (primary and secondary sources) to allow triangulation of information (3). In general, data sources can be grouped into three types:

- Documents: existing evaluations, progress reports and policy documents, media coverage;
- Qualitative data: in-depth interviews, focus groups, participants observation.
- Quantitative data: surveys, routine surveillance data, national or local registries, clinical samples.

M&E studies for Health EDRM must be considered in light of the unique setting of emergencies and disasters, which distinguishes them from conventional research. Randomized trials (Chapter 4.1) are the gold standard, in particular for evaluating clinical effectiveness, but cohort studies, case control studies, ecological studies and case series (15) might also be done (Table 4.15.2).

For disaster-related M&E research, quasi-experimental designs (QEDs) in which random assignment is not used, merit special attention. These have been increasingly used and encouraged in non-clinical and routine practice settings to test attribution of the intervention to the outcome change, including in public health (16), and disaster and humanitarian settings (17). Health research in disaster and humanitarian settings has used a range of research designs (17–18). Random allocation, which is the essence of the randomized trial, may not be feasible in some disaster settings, but in using a QED, a balance must be achieved between the practicality of doing the study and its internal and external validity in order to reduce the risk of bias, especially selection bias (16). One of the most important techniques to improve the validity of QEDs is to identify a comparison group that resembles the characteristics of the intervention group as closely as possible (19). Some commonly used QEDs are discussed in Chapter 4.5; Case Study 4.15.3 is an example of quasi-experimental M&E study of a programme run by an international NGO among vulnerable population.

Study design

In general, M&E studies in Health EDRM can take the form of quantitative study (Chapter 4.1), qualitative study (Chapter 4.12) and a concurrent or sequential mixed methods study (Chapter 4.13). Case Study 4.15.2 is an example of the use of a mixed methods M&E study in disaster management.

4.15

Table 4.15.2 Experimental and quasi-experimental study designs in M&E for DRR Study design

	Application to M&E studies in DRR context
Randomized trial	<p>Gold standard for evaluating clinical effectiveness and demonstrating causality</p> <p>May have restricted application in routine practice, due to its tightly controlled research environment</p> <p>New or experimental intervention is provided to experimental group but not the control group</p> <p>May be ethical issues involved in depriving the control group of the intervention, especially in a disaster setting</p>
Cohort study	<p>Comparison of the same group of people before and after the follow-up period</p> <p>Some evidence in demonstrating causality</p> <p>Limited application in some disaster setting due to difficulties in identifying the same group of people after a period of time</p>
Case control study	<p>Comparison between two groups of people which are categorized by their outcome</p> <p>Efficiency for rare diseases or outcomes is low</p> <p>Prone to selection bias because the selection of cases and controls is dependent on the criteria defined for the outcome</p>
Non-randomized comparative trial	<p>A form of QED without random allocation</p> <p>Lower level of generalizability of results than randomized trials</p> <p>Wider application in disaster setting</p>
Uncontrolled pre/post and interrupted time series study	<p>A form of QED with no control group</p> <p>Minimizes ethical issues in disaster setting by providing interventions to all groups</p> <p>Cannot demonstrate causality</p>
Cross-sectional study	<p>Practicable in many routine practice settings and for a population-based overview</p> <p>Cannot demonstrate causality</p> <p>Provides analysis at a specific time point only</p>

Case Study 4.15.2**Mixed methods quasi-experimental study of outcomes of a large-scale multilevel economic and food security intervention on HIV vulnerability in rural Malawi (25)**

An impact evaluation study was conducted to assess the impact of an economic and food security intervention on health outcomes and HIV vulnerability in rural Malawi, implemented by CARE International Malawi from 2008 to 2010. This used a quasi-experimental non-equivalent control group design to compare 598 intervention participants with 301 participants in unrelated programmes in similar geographical areas. They were interviewed at baseline and again 18 and 36 months later.

The intervention was found to increase HIV testing and HIV case finding, decrease food insecurity, increase nutritional diversity and improve economic resilience. Most effects were sustained over the 36-month period.

The findings of this impact evaluation study allowed the NGO (CARE International) to identify areas for improvement in their programmes to create greater impact in reducing health risks in a vulnerable setting. The study also demonstrated the possibility of collaboration between an international NGO, local community and academia to conduct a controlled evaluation on locally tailored programmes in routine practice setting, providing scientific evidence on the impact of health-related development programmes.

Pre/post designs with non-equivalent control groups involves collecting data before an intervention and again after the intervention, and then a comparison of these two datasets, with the control group not being randomly assigned (21). Interrupted time series involves multiple observation points over a period of time before and after the intervention with the same group of people (22). A stepped-wedge design generally involves a staggered introduction of the intervention for different groups, and could involve serial cross-sectional data collection, either by site or by a cohort of individuals over a period of time and might include randomization to determine when the intervention is introduced (Chapter 4.3) (23). Some advantages of these designs might include lower cost and greater flexibility, and that they avert the need for a control group of people who will not be provided with the intervention (16). However, measures must be taken to enhance the validity of these studies in the absence of random assignment in order to ensure that biases have been sufficiently minimized to demonstrate attribution and techniques such as propensity score matching and the regression discontinuity design might be used. These are discussed in Chapter 4.5 and the Handbook on Impact Evaluation published by the World Bank (24).

4.15

Case Study 4.15.3 Heat wave plan for England (20)

Heatwaves are considered a public health threat in the United Kingdom, especially after the 2003 European heatwave which caused 2000 deaths in the UK. As a result, the Heatwave plan for England has been published and reviewed annually since 2004, with the latest version published in 2018. The Policy Innovation and Evaluation Research Unit of the United Kingdom's National Institute for Health Research was tasked by the Department of Health and Public Health England to conduct an independent evaluation of the plan (2015 version). A mixed methods study was adopted, involving time series analysis, case studies, and quantitative surveys.

Study design: mixed methods evaluation:

1. Time-series analysis of health data to examine the association between temperature and mortality/morbidity before and after the introduction of the Heatwave plan;
2. Detailed case studies focusing on implementation of the plan;
3. National survey to understand knowledge, attitudes and behaviour of the general population during heatwaves.

Research questions:

1. "Has the introduction of the plan in 2004 had any effect on mortality?"
2. "How well is the plan being implemented locally, including at the 'frontline' of health and care services?"
3. "Is the general population aware of the risks of heat and overheating buildings, do they change their behaviour as a result of hearing heat alerts or advice, and do they take any actions to prevent potential effects of hot weather?"

Although quasi-experimental studies have been conducted in disaster management, the relevant requirements might still be unrealistic for M&E studies in some contexts, such as during the impact stage of a large-scale disaster in a vulnerable State or after a disaster when there is an absence of baseline data. Under such circumstances, researchers might need to use other non-experimental designs, such as participatory monitoring and evaluation, where stakeholders at various levels are engaged to deliberate the relevant process, results or policy of an intervention (26), or ethnographic methodologies (27). Case Study 4.15.4 is an example of participatory action research in a disaster management context.

Case Study 4.15.4**Participatory action research: The World Trade Center evacuation study (26)**

Participatory action research (PAR) was employed to identify the individual, organizational and structural factors that affected evacuation from the World Trade Center (WTC) Towers 1 and 2 on 11 September 2001.

1767 people who worked in one of the towers at the time of the terrorist attack completed the study questionnaire and 11 participated in the PAR teams, which also included professional investigators, experts and specialists. The study investigated the time taken to initiate and complete the evacuation and the incidence of injury.

Quantitative data were collected through questionnaires completed by evacuees. Qualitative data were collected through structured deliberations by the PAR teams, with participation by researchers and consultants.

The PAR teams identified the key risk factors associated with the three study outcomes and prepared 83 recommendations that addressed the risk factors. More than half of the recommendations were aimed at the organizational level, 26% at the structural level, and 23% at the individual level.

This study attempted to illustrate the effectiveness of the PAR methodology for identifying risk-reduction interventions, emergency preparedness and response strategies in disaster research.

4.15.5 M&E studies in Health EDRM: practical examples of challenges

Health EDRM research often takes place in unconventional settings, which calls for innovative and practical methodologies that are nonetheless sufficiently robust. Some of the critiques on impact evaluation regarding humanitarian assistance are also applicable to DRR. For example, Puri and colleagues (28) considered that impact evaluation in such unorthodox settings faces methodological, practical, and ethical challenges. In regard to methodology, there are concerns about the potential to compromise the validity of the findings, in particular as to whether the outcome can be causally attributed to the intervention if randomization is not used to minimize biases between the intervention and control groups (29). In regard to ethical challenges, the foremost concern is the need to have a control group, which could mean that some individuals may be deprived of the experimental interventions that might actually be life-saving through its effects on reducing disaster risks.

Apart from the many complexities that a disaster setting presents to Health EDRM researchers (for example, the wide range of possible interventions, different natures and scales of disasters, and potentially large number of collaborators and funders involved), one distinct practical difficulty for M&E studies is the conceptual challenge of demonstrating the impact of an intervention that had prevented something from happening or reduced the health risks. Outcomes related to knowledge, attitudes and behaviours, and proxy indicators are therefore commonly adopted (6). Furthermore, the

4.15

availability and quality of data is a long-standing issue for all DDR researchers, including those doing M&E studies. The data readiness review conducted under the Sendai Framework revealed that data quality remains an issue and data accessibility is also highly limited in many low- and middle-income countries (30). This makes it difficult to select reliable indicators (31) and the relevance of indicators is critical to the success of any M&E study. The seven targets and 38 indicators under the Sendai Framework (4) have become the globally recognized foundation for researchers to develop indicators suitable to their needs. With health explicitly recognized and mainstreamed in the Sendai Framework, this provides a useful reference point for Health EDRM researchers developing indicators for their M&E studies.

An increasing number of M&E studies are moving towards multi-country or multi-agency settings, given the encouragement for stronger multidisciplinary and cross-country collaboration in DRR. M&E studies are particularly important for DRR initiatives involving multiple actors, but these studies have special challenges. For instance, different actors may have different priorities (Chapter 2.7) and study direction will need to be relevant to all partners. Moreover, all parties need to adopt the same approach if they are to generate comparable data. Differences in the data availability and data quality between actors are often another concern. One way to ensure relevance is to set up standardized key M&E questions and indicators, while allowing different partners to develop their own supporting M&E questions and sub-indicators (32).

4.15.6 Conclusions

M&E provides evidence to help inform understanding of the effectiveness of DRR interventions. Robust and practical M&E studies are essential if Health EDRM initiatives are to be effective and sustainable. This chapter has described the important role of M&E research, existing M&E frameworks related to health and DRR programmes, and described some of the research designs that can be used for such studies and related challenges. While experimental and controlled studies remain the mainstream research methodologies most widely recognized in academia, researchers may need to consider how they can develop studies that are feasible in emergency and disaster settings without compromising strength in demonstrating causality. Researchers need to have the courage and expertise to develop and continuously enhance research methodologies that fit the needs of routine practice if the findings of their M&E studies are to meaningfully guide the allocation of limited resources in Health EDRM.

4.15.7 Key messages

- o **M&E studies can be used to demonstrate the effectiveness of Health EDRM interventions and be instrumental in providing evidence and justifications for sustainable resource allocation.**
- o **The M&E framework chosen by a researcher will determine the study focus during data collection, analysis and interpretation of its findings.**
- o **Randomized trials might not be practical for some Health EDRM M&E studies and quasi-experimental designs are increasingly used.**
- o **In conducting M&E studies with quasi-experimental design, measures must be taken to minimize bias and ensure the internal and external validity of the study, and findings must be interpreted in light of the specific context of the study.**
- o **The poor availability of high-quality data and the selection of indicators are two major challenges for M&E studies in Health EDRM.**

4.15.8 Further reading

Health in Humanitarian Crisis. *Lancet*. June 8, 2017: Vol. 390: No.10109.

Scott Z, Wooster K, Few R, Thomson A, Tarazona M. Monitoring and evaluating disaster risk management capacity. *Disaster Prevention and Management*. 2016: 25(3): 412–22.

Shek DT, Wu J. Quasi-experimental Designs. In: Frey BB, editor. *The SAGE encyclopedia of educational research, measurement, and evaluation*. Thousand Oaks, CA: SAGE Publications. 2018: pp.1353-6.

Monitoring and Evaluation Framework. United Nations Office for Disaster Risk Reduction. 2015. <https://www.unisdr.org/we/inform/publications/49324> (accessed 13 January 2020).

4.15.9 References

1. WHO and Public Health England. Health Emergency and Disaster Risk Management: Overview. 2017. <http://www.who.int/hac/techguidance/preparedness/who-factsheet-overview-december2017.pdf> (accessed 13 January 2020).

2. Monitoring and Evaluation Toolkit: HIV/AIDS, Tuberculosis and Malaria. WHO. 2004. <https://www.who.int/malaria/publications/atoz/a85537/en> (accessed 11 February 2020).

3. Monitoring and Evaluation Framework. United Nations Office for Disaster Risk Reduction. 2015. <https://www.who.int/malaria/publication/monitoring-and-evaluation-framework> (accessed 11 February 2020).

4. The Sendai Framework for Disaster Risk Reduction 2015-2030. United Nations Office for Disaster Risk Reduction. 2015. <https://www.undrr.org/implementing-senai-framework/what-sf>.

5. Villanueva P. Learning to ADAPT: monitoring and evaluation approaches in climate change adaptation and disaster risk reduction – challenges, gaps and ways forward. Brighton, UK: Institute of Development Studies. 2010.

6. Scott Z, Wooster K, Few R, Thomson A, Tarazona M (2016). Monitoring and evaluating disaster risk management capacity. *Disaster Prevention and Management* 25(3): 412–22.

7. Katich K. Monitoring and Evaluation in Disaster Risk Management. Washington, DC: World Bank. 2010.

8. Sphere Association. The Sphere Handbook: Humanitarian Charter and Minimum Standards in Humanitarian Response, fourth edition, Geneva, Switzerland, 2018. www.spherestandards.org/handbook

9. Sphere Association. Minimum Standards in Health Services in The Sphere Handbook: Humanitarian Charter and Minimum Standards in Humanitarian Response, fourth edition, Geneva, Switzerland, 2018. www.spherestandards.org/handbook

10. Mountfield B., Sphere for Monitoring and Evaluation. Published by the Sphere Project in Geneva. 2015. <https://www.spherestandards.org/resources/sphere-for-monitoring-and-evaluation/>

11. Pawson, R. The science of evaluation: A realist manifesto. : SAGE Publications Ltd, London 2013. [Doi.org/10.4135/9781473913820](https://doi.org/10.4135/9781473913820). 2013.

12. Shreve CM & Kelman I. Does mitigation save? Reviewing cost-benefit analyses of disaster risk reduction. *Int J Disaster Risk Reduct* 10:213–235 . doi.org/10.1016/j.ijdr.2014.08.004. 2014.

13. Dückers MLA, Thormar SB, Juen B, Ajdukovic D, Newlove-Eriksson L, Olff M (2018) Measuring and modelling the quality of 40 post-disaster mental health and psychosocial support programmes. *PLoS One* 13:e0193285 . [doi: 10.1371/journal.pone.0193285](https://doi.org/10.1371/journal.pone.0193285)

14. Disaster Management M&E Framework – March 2014, Department of Cooperative Governance and Traditional Affairs, The Republic of South Africa. 2014. <http://www.ndmc.gov.za/Frameworks/Disaster%20Management%20Monitoring%20and%20Evaluation%20Framework.pdf>.

15. Guest G, Namey EE. *Public Health Research Methods*. London, UK: SAGE Publications. 2015.

16. Handley MA, Lyles CR, McCulloch C, Cattamanchi A. Selecting and Improving Quasi-Experimental Designs in Effectiveness and Implementation Research. *Annual Review of Public Health*. 2018; 39: 5-25.

17. Blanchet K, Ramesh A, Frison S, Warren E, Hossain M, Smith J, et al. Evidence on public health interventions in humanitarian crises. *Lancet*. 2017; 390: 2287–96.

18. Chan EYY, Lam HCY, Chung PPW, Huang Z, Yung TKC, Ling KWK, et al. Risk Perception and Knowledge in Fire Risk Reduction in a Dong Minority Rural Village in China: A Health EDRM Education Intervention Study. *International Journal of Disaster Risk Science*. 2018; 9: 306-18.

19. White H, Sabarwal S. *Quasi-experimental design and methods. Methodological Briefs: Impact Evaluation No. 8*. Florence. 2014.

20. Policy Innovation and Evaluation Research Unit of National Institute for Health Research Heat Wave Plan Evaluation. <https://piru.lshtm.ac.uk/projects/current-projects/heatwave-plan-evaluation.html> (accessed 13 January 2020).

21. Cousins K, Connor JL, Kypri K. Effects of the Campus Watch intervention on alcohol consumption and related harm in a university population. *Drug and Alcohol Dependence* 2014; 143: 120–6.

22. Chan EYY, Lam HCY, Lo ESK, Tsang SNS, Yung TKC, Wong CKP. Food-Related Health Emergency-Disaster Risk Reduction in Rural Ethnic Minority Communities: A Pilot Study of Knowledge, Awareness and Practice of Food Labelling and Salt-intake Reduction in a Kunge Community in China. *International Journal of Environmental Research and Public Health* 2019; 16(9): E1478.

23. Killam WP, Tambatamba BC, Chintu N, Rouse D, Stringer E, Bweupe M, et al. Antiretroviral therapy in antenatal care to increase treatment initiation in HIV-infected pregnant women: a stepped-wedge evaluation. *AIDS* 2010; 24: 85-91.

24. Khandker S, Koolwal G, Samad H. *Handbook on impact evaluation: quantitative methods and practices*. Washington, DC: World Bank. 2010.

25. Weinhardt LS, Galvao LW, Yan AF, Stevens P, Mwenyekonde TN, Ngui E, et al. Mixed-Method Quasi-Experimental Study of Outcomes of a Large-Scale Multilevel Economic and Food Security Intervention on HIV Vulnerability in Rural Malawi. *AIDS and Behavior*. 2017; 21(3): 712-23.

26. Gershon RRM, Rubin MS, Qureshi KA, Canton AN, Matzner FJ. Participatory Action Research Methodology in Disaster Research: Results From the World Trade Center Evacuation Study. *Disaster Med Public Health Prep* 2:142–149 . doi/org: 10.1097/DMP.0b013e318184b48f. 2008.

27. Turkon D, Himmelgreen D, Romero-Daza N, Noble C. Anthropological perspectives on the challenges to monitoring and evaluating HIV and AIDS programming in Lesotho. *African J AIDS Res* 8:473–480 . doi: 10.2989/AJAR.2009.8.4.11.1048. 2009.

28. Puri J, Aladysheva A, Iversen V, Ghorpade Y, Brück T. Can rigorous impact evaluations improve humanitarian assistance? *Journal of Development Effectiveness*. 2017: 9: 519-42.

29. Delgado-Rodriguez M. Bias. *J Epidemiol Community Heal* 58:635–641 . doi: 10.1136/jech.2003.008466. 2004.

30. Osuteye E, Johnson C, Brown D. The data gap: An analysis of data availability on disaster losses in sub-Saharan African cities. *International Journal of Disaster Risk Reduction*. 2017: 26: 24-33.

31. Chan EYY, Murray V. What are the health research needs for the Sendai Framework? *Lancet*. 2017. 390: e35-6.

32. Pasanen T, Shaxson L. How to design a monitoring and evaluation framework for a policy research project. London, UK: A Methods Lab publication. 2016
