Study design

4.1 Basic principles in designing studies to assess the effects of interventions

Authors

Mike Clarke, Centre for Public Health, Queen's University, Belfast, United Kingdom; Evidence Aid, London, United Kingdom.

Dimuthu Rathnayake, Ministry of Health, Colombo, Sri Lanka.

4.1.1 Learning objectives

To understand key factors to consider when developing a study to assess the effects of an intervention, action or strategy for health emergency and disaster risk management (Health EDRM), including:

1. The importance of reliable and robust estimates of the effects of interventions.
2. Minimizing the risk of bias.
3. The role of randomized trials.
4. Aspects of conducting prospective, comparative studies.

4.1.2 Introduction

This chapter will show how research can provide reliable and robust evidence about the likely effects of different interventions in order to help people choose between alternatives when there is more than one intervention suitable for an individual, or a variety of actions or strategies that are appropriate for a population. To be reliable, this evidence needs to come from studies in which the interventions were compared in ways that minimize the effects of biases *(1)*, such as biases that might arise from using information about a participant's likely outcomes to select who will or will not receive one of the interventions being compared. To be robust, the studies also need to be large enough to minimize the effects of chance.

This chapter outlines how such studies might be carried out in the Health EDRM context and highlights important features for the design, conduct and interpretation of such studies. The various types of research design that might be used to study different areas of importance to Health EDRM are discussed in Chapter 3.5. In this chapter, particular emphasis is placed on a type of comparative effectiveness study called a randomized trial, because this design seeks to minimize bias and generate reliable and robust estimates of the relative effects of interventions. It does this by creating comparison groups that differ only in regard to the interventions being compared. In randomized trials, some of the individuals who join the study are randomly allocated to receive the intervention being tested,

which might be a new way of protecting people from contaminated water, a treatment for fractures or a treatment to minimize anxiety, for example; this is often called the experimental group. The other participants in the trial would be allocated to an alternative intervention or a control group. Cluster randomized trials are a modified version of this design, when randomization is done at the level of clusters (such as families, villages or hospital wards), rather than at the individual level. These are discussed in Chapter 4.3.

Randomized trials seek to answer research questions about cause and effect in a controlled manner. Their aim is to produce an estimate of the impact or effect of the intervention by comparing the outcomes in the experimental group to those in the control group. The purpose of this is to generate evidence, which can then be used to make assumptions about how the intervention might affect people who are similar to those in the trial.

However, although we focus here on randomized trials, many of their key features discussed below are also applicable to other prospective studies in which individuals are recruited and followed up.

4.1.3 Why do we use randomized trials?

Randomized trials are prospective studies in which eligible participants are randomly allocated to one of the two or more groups that are to be compared, with each group receiving a different intervention. This allows a comparison to be made of how each intervention affects the outcomes that are measured — such as the speed of a person's recovery, their quality of life, or how well they understand information about a disaster-related threat to their health or livelihood. However, for some research questions – on topics such as estimating the proportion of people who have different levels of mental or physical trauma after an earthquake, for example – other study designs would be used; these are discussed elsewhere in this book, such as in Chapter 3.2 for assessing risk factors.

4.1.4 Planning the trial: eligibility criteria

Chapter 3.5 discussed the importance of having a clear research question for a study, including the need to match the research question to the comparison to be made in a randomized trial, using the example of fish oil for PTSD. Case Study 4.1.1 describes a randomized trial undertaken with rescue workers after the Great East Japan Earthquake in 2011. It illustrates both the comparison that was made and the decision about the population to study. The decisions about who to study are set out in the inclusion and exclusion criteria for a trial, which may be broad or narrow, and determine who is and is not eligible for the study *(2).*

Case Study 4.1.1
The APOP randomized trial of fish oil for attenuating post-traumatic stress disorder (PTSD) symptoms among rescue workers after the Great East Japan earthquake

The Great East Japan Earthquake and tsunami of 11 March 2011 caused tremendous damage to the north-eastern coast of Japan, leaving 20 000 people dead or missing. Many rescue workers were exposed to traumatic experiences. Researchers decided to investigate whether PTSD symptoms might be attenuated by the use of fish oil. The same researchers had previously shown that PTSD symptoms at 12 weeks after injury were significantly alleviated if patients with physical injury took fish oil. The new study was done among disaster medical assistance team (DMAT) members who were deployed during the acute disaster phase of the earthquake. The randomized trial was approved on 1 April 2011 and started the following day.

After providing informed consent, participants were randomly assigned to one of two groups - one group that received the fish oil supplementation plus psychoeducation, or the other group, which received psychoeducation alone *(3).* The fish oil was given as seven capsules per day, each containing 320mg of fish oil. 172 rescue workers joined the trial between 2 and 12 April 2011 and were followed up over the next few months.

The primary outcome was measured using the Impact of Event Scale-Revised (IES-R), and this showed no significant difference at 12 weeks between the decline in scores for participants in the fish oil group compared to those in the control group *(4).*

In an explanatory trial – also known as an efficacy trial – the inclusion criteria might be kept narrow to ensure that the people recruited to the study are all similar to one another. Such a trial would determine whether, in such ideal circumstances, there is a difference between the interventions being compared. Examples of such studies include: randomized trials to compare the speed of onset of pain control when two formulations of an analgesic drug are used in people with specific types of minor injury; a comparison of surgical techniques for managing fractures of the lower leg; or a test of a psychological therapy in school-aged children following a tsunami. In studies of this kind, the participants would be carefully chosen so that they have the characteristics that are felt to be most receptive to the intervention being tested. One rationale for such trials is that, if the experimental intervention is no better than the routine intervention in these “ideal” circumstances, it is unlikely to be better in a much broader population.

However, in health emergencies and when seeking to manage disaster risk, randomized trials are more likely to take the form of an effectiveness or pragmatic trial. This is because a wide range of participants is likely to be recruited, and there would likely be less strict control over the specific elements of the interventions being tested, in order to make the trial as close as possible to routine practice. In effectiveness studies, the eligibility criteria are broad enough to ensure that many of the types of people who are likely to be considered for the intervention in the future are included.

Such trials might use the "uncertainty principle" to set the eligibility criteria *(5)*, meaning that people would be eligible for a trial if there is sufficient uncertainty about what effects the interventions would have for them. This is also a fair way to allocate interventions when a choice has to be made about who is given or not given the intervention, as is often the case in Health EDRM. When deciding on the eligibility criteria for a trial, and its feasibility, careful consideration is also required of what number of participants will be needed to answer the research question: researchers deal with this when calculating the necessary sample size, which is discussed in Chapter 4.2.

4.1.5 Participant selection and informed consent

The success of any prospective study relies on the cooperation of the people who are participating in it. In medicine, one major difference between treating patients inside or outside a research study is the formal process of informed consent that is likely to be required for the study (see Chapter 6.4); this can be challenging in disaster situations where the intervention has to be administered quickly, there is little time to provide detailed information or no opportunity for a full discussion with potential participants. However, there are several examples of ethically acceptable trials conducted in such difficult circumstances. For instance, the CRASH trial recruited patients with serious head injuries and showed that a widely used treatment, steroids, was not beneficial *(6).*

The uncertainty principle can also be employed in deciding whether or not a trial is ethical (see Chapters 3.4 and 6.4 for a discussion of the ethics of research). For example, it can be used when considering whether it is ethical to not do a randomized trial. If there is uncertainty about the relative effects of two interventions, and both are available and suitable for the target population, the most ethical approach may be for them to enter a randomized trial. This ensures that participants have a fair chance of receiving the more beneficial intervention (since it will be unknown when they join the trial which this will be) and the data collected should help to resolve uncertainty in the future, as was the case with the aforementioned CRASH trial for people with head injuries *(6).*

4.1.6 Randomizing participants

The key feature distinguishing randomized trials from other prospective studies is the use of a random process to determine which of the interventions is received by each participant. This process ensures that any differences between the outcomes for those in the randomized groups will be due either to the effects of the interventions being compared, or to the effects of chance.

Randomization can be achieved in a variety of ways, and some methods are described here. The key elements are the use of a random sequence to allocate participants to one of the groups, and ensuring that no-one knows which group a person will be allocated to before they join the trial. If an individual's allocated group is known in advance, this may lead to a different decision being made about whether they join the trial, or to some other form of manipulation, such as delaying their joining until a different allocation is available.

Generating a random sequence

In simple randomization, each participant has the same probability of being allocated to each intervention being tested. This can be achieved using simple physical techniques such as flipping a coin, rolling a dice or drawing lots. It might also be done by shuffling envelopes into which information about the allocation has been placed. Mathematical techniques, using random numbers, can also be used. Simple randomization is completely unpredictable, provided that the allocation for an individual participant is concealed up until the point that they enter the trial. However, the disadvantage of simple randomization is that, particularly in a small trial, it can lead to large, chance imbalances between the groups. For example, if a coin is flipped 100 times, it is likely that at some point in the sequence there will be a consecutive run of 6, 7 or 8 heads or tails. If this occurred in a trial, it could lead to an imbalance in the number of people in the groups, making analysis of the trial difficult. It could also lead to imbalances in participant characteristics between the groups, which might also make the analysis of the trial more difficult.

These potential problems can be overcome by using a technique called blocked randomization, which allows stratification of the allocated interventions (or a more complex, computer-based technique called minimization *(7).* Blocked randomization means that after a particular number of participants have been allocated, the numbers in the different intervention groups will be balanced. For example, a block size of four in a trial with two intervention groups guarantees that for each sequence of four people joining the trial, two will be allocated to one group and two to the other group; using that block size for a trial as a whole will therefore ensure that the difference between the number of people in each of the two groups will be no more than two (if, at the start of the final block, two are both allocated to the same group). Similarly, using blocks for different types of people in the trial (for example, young and old, or those living in rural, semi-urban and urban settings) can ensure balance within those groups.

Concealing the random sequence until the participant joins the trial

Allocation concealment is not the same as blinding or masking the intervention, which is discussed below and happens after the person has entered the trial. Allocation concealment takes place earlier, before the person enters the trial. It means that no-one involved in recruiting potential participants can know what they will receive until they have joined the trial. Allocation concealment prevents manipulation that might arise if knowing the allocation leads to a different decision about someone's eligibility or their willingness to join the trial.

One way to implement adequate allocation concealment is to use sealed, opaque, sequentially numbered envelopes, which must be used in the predetermined sequence and cannot be opened to reveal the allocation until the person has entered the trial. Researchers might also use randomization systems in which an online or computer-based system, or a telephone call, is used to first capture data on the participant before their allocation is given.

4.1

4.1.7 Blinding or masking

In some studies, it is important that the people involved in conducting the trial do not know which intervention a participant is receiving. This is usually called 'blinding' or, particularly when the research is related to eyesight, 'masking', and might be achieved by giving patients in the control group a dummy intervention or placebo. However, adding placebos or blinding to trials can be difficult, because doing so increases the resources needed for the trial and can make the interpretation of the results more difficult because after the trial, in routine practice, those receiving or administering an intervention would know what is being taken or given *(8)*.

There are a number of different people involved in a trial who might be kept blind to the intervention and there are a variety of reasons for doing so. Typically, the participant might be kept blind in order to reduce the risk that they will either report outcomes differently because they know which intervention they are receiving or, through a placebo effect, will actually respond differently simply because of their knowledge of the intervention rather than as a result of the intervention itself. Problems can also arise if participants knowing which intervention group they are in makes them change their behaviour in ways that would not happen outside of the trial.

To illustrate the potential impact of blinding: in a randomized trial of an iron-fortified biscuit for children with iron deficiencies, those who know they are in the control group might try to change their eating habits, while those in the intervention group might change in a different way, perhaps assuming that the biscuits will provide the nutrition that they need. Blinding might be achieved by giving those in the control group a biscuit that is identical in every way except for the ingredient being tested, to act as a placebo.

It might also be important to keep people other than the participant blind to the allocated intervention. This can include those treating and caring for patients in a study and the people measuring outcomes. Keeping the practitioners blind ensures that they are less likely to do other things differently for a patient – just as the participant might modify their behaviour if they know which intervention they have been allocated, practitioners might add extra treatments if they know a patient is in the control group or monitor them more carefully if they are receiving the experimental intervention.

Likewise, if the people assessing the participants' outcomes or collecting data know that someone is receiving the experimental intervention, they might look more closely for side effects. If someone is in the control group, unblinded assessors might be more pessimistic when recording their outcomes. For example, in a trial testing different types of dressing for wounds after surgery, it could be important that the outcome assessor responsible for classifying the level of infection in a wound did not know which dressing was used when they made their assessment. Problems can also arise if the statistician analysing the trial's results is influenced in how they do this by knowing which group is the experimental group. In such circumstances, it would be important to keep them blind to which group is which.

4.1.8 Avoiding publication bias: registering and reporting a study

Even if a researcher is careful to minimize bias when designing and conducting their prospective study, biases can be introduced when they make decisions about reporting its findings. These can lead to problems when the results are used by others. Publication bias arises when the results of a study have an influence over whether it is published. Selective reporting bias can mean that, even though the study is published, some of its findings remain unpublished, while others are given more prominence. Chapter 6.6 describes some of the elements to consider when reporting a research study, and the importance of publishing research in ways that will help people and organizations such as United Nations agencies, NGOs and others involved in Health EDRM to use the findings in their future decision making.

During recent decades, efforts to combat the problems of publication and selective reporting bias have led to the development of prospective registers of research studies *(9).* Registering the study before the first participant is recruited makes the existence of the study public knowledge in a way that ensures that this could not possibly be influenced by its results. It also requires the researcher to say, in advance, what they are studying. Some journals will not publish the results of trials that have not been prospectively registered. Furthermore, in the context of a sudden-onset disaster, carefully pre-planning the trial, registering and perhaps even publishing its full design in advance, allows a trial to be sitting "on the shelf" ready to be activated. Case Study 4.1.2 presents one such example, where a detailed plan has been prepared for a blinded, randomized trial of regional anaesthesia in earthquake survivors with lower limb trauma.

Case Study 4.1.2
Plan for a randomized trial of anaesthesia and pain management for patients with lower limb trauma after an earthquake

After an earthquake, the largest burden of injuries is due to trauma of the legs and feet, and pain management for these patients is a substantial challenge. The Regional Anaesthesia for Painful Injuries after Disasters (RAPID) trial has been designed to evaluate whether regional anaesthesia, either with or without ultrasound guidance, can reduce pain from earthquake-related lower limb injuries in a disaster setting *(10)*. The plan for the trial was prospectively registered in February 2016.

After informed consent has been obtained, study participants will be randomized in a 1:1:1 allocation to standard care (parenteral morphine at 0.1 mg/kg), standard care plus a landmark-guided fascia iliaca compartment block, or standard care plus an ultrasound-guided femoral nerve block. In order to blind participants and healthcare providers who are not part of the research to a patient's allocated group, sham ultrasound activities will be used in the first two groups and a normal saline injection will be given to the first group (the control group). The primary outcome measure will be a standard pain intensity score over the first 24 hours, with secondary outcome measures including analgesic requirements, adverse events, and participant satisfaction.

If the trial shows that regional anaesthesia is effective in a disaster setting, its future use for survivors of earthquake trauma could reduce both their acute suffering and the long-term complications of the injury.

4.1.9 Other types of prospective, comparative study

When it is not feasible to use randomization to allocate individuals to different interventions, there are other methods that can be used. For example, for a research question relating to a comparison of different methods of coordinating the multidimensional response to a disaster, randomly assigning individuals or groups of people to coordinate their actions in very different ways would be likely to lead to chaos. Instead, the new method of coordination could be implemented and then its impact assessed using a "counterfactual" to estimate what might have happened without the intervention in order to decide whether it improved, worsened or made no difference to outcomes. This might also be the case for other interventions; methods for conducting such studies are discussed in Chapter 4.15. To illustrate the planning of such a study, Case Study 4.1.3 describes how the findings from research into a surge of dengue cases at a hospital in Sri Lanka might be used in the evaluation of future changes to hospital strategy and health systems research.

Case Study 4.1.3
Planning an evaluation of strategies that would be implemented in a future health emergency

Dengue is the most important infectious disease-related public health concern in Sri Lanka. A massive outbreak occurred at the time of the south-western monsoon rains in 2017: approximately 185 000 dengue cases were reported and more than 400 people died *(11)*. The National Institute of Infectious Diseases, as the leading hospital for managing infectious diseases in Sri Lanka, played a major role during the outbreak and researchers there studied the size and effects of the dengue epidemic *(12)*. Their study identified particular challenges and, along with a systematic review *(13)*, has led to proposals for implementation in the future. These include the need for public health systems to use robust systems approaches with sufficiently detailed managerial approaches. It would not be possible to assess the effects of these systems-level strategies in a randomized trial because it would not be feasible to allocate them to some individuals or hospitals, and not to others. However, it would still be useful to know how effective they are. In order to assess this, a prospective study would be put in place to gather outcome measures that could then be compared with the earlier data. This would seek to answer a research question about whether the new systems were an improvement on the old systems, and provide evidence to inform the decision to continue with them or refine them further for future dengue epidemics. However, caution would be needed when deciding whether the comparison of the future epidemic with that in 2017 was a valid comparison of "like with like" in relation to everything except the new strategies. The prospective study would collect information on the dengue cases, the use of hospital resources and outcomes for patients. It would include attendance at the outpatient department, admissions to hospital and bed occupancy before and during the next outbreak, and demand on services such as the haematology laboratories. These data would then be compared with the findings from 2017, with care being taken to ensure that any differences were not merely due to differences in the way in which the data were gathered.

Two other types of prospective study that might be used when randomized trials are not feasible are described below.

Controlled before-after study

In a controlled before-after study, the decision about whether a person will be in the intervention or the control group is not made by the researcher. The outcomes of the people in both groups are measured before and after the intervention is introduced for one of the groups. For example, if some people who lost their homes after a windstorm are given a new type of shelter, their respiratory health would be monitored before and after the delivery of the new shelters, as well as that of a control group of people provided with the usual shelter. One disadvantage of these studies is that they have a high risk of bias because there may be differences between the intervention and control groups. If these differences not only determined whether a person went into the intervention or the control group but also had an effect on their outcomes, it is possible that the

study's findings might simply arise from these underlying differences rather than from the effects of the intervention.

Interrupted time series

In an interrupted time series design, outcomes are collected at multiple time points, before and after the intervention is introduced. A single setting or group of participants is used, and there is no control group. The effect of the intervention would then be estimated by comparing the trend in the outcomes after its implementation with the trend beforehand. For example, if the level of gender-based violence was holding steady or slowly declining in a displaced person's camp, but declined rapidly after a new strategy was put in place, this would suggest that the new strategy is beneficial. However, a disadvantage of this design is that if any other features of the camp had changed close to the time that the intervention was introduced, it would not be known whether those changes may have caused (in full or in part) any detected improvement (or conversely, if the new intervention did not appear to have an impact, may have cancelled out what would have been a benefit).

4.1.10 Conclusions

For many centuries, decisions about interventions and policies intended to improve the health of populations were based mostly on personal experience, anecdotal case histories and comparisons of people who had received one intervention with an entirely separate group who had not received it or had received something different. Although these sources of knowledge are still in use today, they are subject to biases which mean that the information they provide may be unreliable.

In recent decades, routine health care and policy making has relied increasingly on randomized trials and systematic reviews (see Chapter 2.6) of these as a source of reliable and robust estimates of the relative effects of different interventions. Provided the trial is sufficiently large, random allocation ensures that any differences in outcomes between groups must be due to the effects of the interventions. This allows future decision makers to have greater confidence in the answer provided by the trial when they are choosing interventions or setting policy.

4.1.11 Key messages

- **People choosing between different interventions, actions and strategies need reliable and robust evidence on their relative effects.**
- **Such evidence needs to come from research that has minimized the effects of bias and chance.**
- **Randomized trials provide a means for testing interventions in such a way that any difference between the outcomes of the participants in the groups being compared are due to the effects of the intervention, or chance.**
- **Pre-planning a trial, or other prospective study, allows it to be ready to be activated when needed, for example in a sudden-onset disaster.**

4.1.12 Further reading

Clarke M, Allen C, Archer F, Wong D, Eriksson A, Puri J. What evidence is available and what is required, in humanitarian assistance? Scoping Paper 1. New Delhi: International Initiative for Impact Evaluation (3ie). December 2014. https://www.3ieimpact.org/sites/default/files/2019-01/3ie_scoping_paper_1-humanitarian-top.pdf (accessed 9 February 2020).

White H. An introduction to the use of randomized control trials to evaluate development interventions. Journal of Development Effectiveness. 2013; 5(1): 30-49.

4.1.13 References

1. Clarke M, Atkinson P, Badenoch D, Chalmers I, Glasziou P, Podolsky S et al. The James Lind Library's Introduction to Fair Tests of Treatments. Oxford, UK: James Lind Library. 2018 http://www.jameslindlibrary.org/download/18790 (accessed 9 February 2020).
2. Yusuf S, Held P, Teo KK, Toretsky ER. Selection of patients for randomized controlled trials: implications of wide or narrow eligibility criteria. Statistics in Medicine. 1990; 9: 73-86.
3. Matsuoka Y, Nishi D, Nakaya N, Sone T, Hamazaki K, Hamazaki T et al. Attenuating posttraumatic distress with omega-3 polyunsaturated fatty acids among disaster medical assistance team members after the Great East Japan Earthquake: the APOP randomized controlled trial. BMC Psychiatry. 2011; 11: 132.
4. Nishi D, Koido Y, Nakaya N, Sone T, Noguchi H, Hamazaki K, et al. Fish oil for attenuating posttraumatic stress symptoms among rescue workers after the Great East Japan earthquake: a randomized controlled trial. Psychotherapy and Psychosomatics. 2012; 81: 315-7.
5. Peto R, Baigent C. Trials: the next 50 years. Large scale randomized evidence of moderate benefits. BMJ. 1998; 317: 1170-1.
6. Edwards P, Arango M, Balica L, Cottingham R, El-Sayed H, Farrell B, et al. Final results of MRC CRASH, a randomized placebo-controlled trial of intravenous corticosteroid in adults with head injury - outcomes at 6 months. Lancet. 2005; 365: 1957-9.
7. Treasure T, MacRae KD. Minimization: the platinum standard for trials. BMJ. 1998; 317: 362-3.
8. Anand R, Norrie J, Bradley JM, McAuley DF, Clarke M. Fool's gold? Why blinded trials are not always best. BMJ. 2020; 368: l6228.
9. Ranawaka UK, de Abrew A, Wimalachandra M, Samaranayake N, Goonaratna C. Ten years of clinical trial registration in a resource-limited setting: Experience of the Sri Lanka clinical trials registry. Journal of Evidence Based Medicine. 2018; 11(1): 46-50.
10. Levine AC, Teicher C, Aluisio AR, Wiskel T, Valles P, Trelles M, et al. Regional Anesthesia for Painful Injuries after Disasters (RAPID): study protocol for a randomized controlled trial. Trials. 2016; 17(1): 542.
11. Tissera HA, Jayamanne BDW, Raut R, Janaki SMD, Tozan Y, Samaraweera PC, et al. Severe Dengue Epidemic, Sri Lanka, 2017. Emerging Infectious Diseases 2020; 26(4): 682-91.
12. Rathnayake D, Wijewickrama A, Aluthge K. Response of the National Institute of Infectious Diseases, Sri Lanka to an unexpected dengue epidemic in 2017. Ceylon Medical Journal. 2018; 63(3): 108-112.
13. Rathnayake D, Clarke M, Jayasooriya L. Hospital surge capacity: The importance of better hospital pre-planning to cope with patient surge during dengue epidemics – A systematic review. International Journal of Healthcare Management. 2019; 1-8 (first published online 21 November 2019).

4.2 Measuring the problem: Basic statistics

Authors

Christopher Garimoi Orach, Makerere University, School of Public Health, Kampala, Uganda.

Ngoy Nsenga, WHO Regional Office for Africa, Nairobi, Kenya.

Olushayo Olu, WHO Country Office, Juba, South Sudan.

Megan Harris, Public Health Wales, Swansea, Wales.

4.2.1 Learning objectives

To understand the following in the context of Health EDRM:

1. Basic statistical concepts.
2. Epidemiologic study designs.
3. Commonly used sampling methods.
4. Estimation of sample size.

4.2.2 Introduction

Statistics are used to describe the health status of population groups, quantify disease burden and estimate the effects of interventions. This is especially important in Health EDRM, where health authorities making decisions about the use of limited resources need to be able to identify the best possible programmes for prevention and care so that they can prioritize key interventions. One of the prerequisites of data analysis is to collect data that will allow the research questions to be answered and hypotheses to be tested (Chapter 3.5). The kind of statistical analyses chosen will depend on the type of data that were collected through research, routine data collection or surveillance data.

Case study 4.2.1 provides an example of how the data collection for statistics was conducted in humanitarian settings.

4.2

Case study 4.2.1
Measuring the public health problem in a human-made disaster in Sub-Saharan Africa

An armed conflict in Sub-Saharan Africa resulted in a major humanitarian crisis. The conflict internally displaced more than one million people into camps which were largely managed by the United Nations *(1).* Camps for internally displaced persons may have poor living conditions, overcrowding and inadequate access to social services that predispose the displaced populations to outbreaks of infectious diseases such as measles, cholera, malaria, and hepatitis E *(2)*. The Early Warning Alert and Response System (EWARS) was established to address the need for good quality and real-time data for timely detection and response to epidemics in support of the Early Warning Alert and Response Network (EWARN) *(3)*, a system that supports surveillance and response in humanitarian settings where routine systems are unavailable or underperforming *(4).*

The system collects real-time data on infectious diseases, injuries, trauma and nutrition from health facilities managed by frontline health partners in the camps and conflict-affected areas. Data are entered at the facility level and automatically uploaded into a central database. Automated analysis is conducted, a weekly bulletin is generated and disseminated to all health partners on a regular basis. The system resulted in drastic improvements in the timeliness (69%) and completeness (73%) of reporting from the camps and conflict-affected locations and timely detection of several outbreaks including the cholera epidemic of 2016 and measles outbreaks of 2018 to 2019 *(5)*. The system also provides detailed case-based and laboratory data which are used for better characterization and response to outbreaks and for research purposes. Furthermore, the system contributes to improvements in the national Integrated Diseases Surveillance and Response System and has been expanded to generate monthly information on health service functionality and nutrition status. Poor mobile network coverage in the conflict-affected areas of the country remains a key challenge as data are transmitted electronically.

The EWARS has proven to be a good tool in the generation of data for public health decision making during humanitarian crises while also serving as foundation for strengthening disease surveillance during the transition from humanitarian to development programming. The system is also a major repository of secondary research data.

4.2.3 Types of quantitative data

The two main types of quantitative data are categorical and continuous. Categorical data can be either dichotomous (taking only one of two possible values) or polytomous (having more than two distinct categories). Dichotomous data are considered binary – for example, vital status might be either alive or dead, a community might have either been exposed or not exposed to a toxic spill and someone might have either received or not received an intervention. Polytomous data have more than two categories and have a number of different attributes. It may be ordinal, being rank-ordered, typically based on a numerical scale that is comprised of a small set of discrete classes or integers, but may not always have a specific set interval between integers (for example, socio-economic status or income level). Alternatively, the categories might not be in any order (for example, types of injury or cause of death).

Continuous data are measured on a continuum and, theoretically at least, can have any numeric value over a continuous range, with the level of granularity dependent on the precision of the measurement instrument. Interval data are a form of continuous data in which equal intervals represent equal differences in the property being measured, for example temperature. Ratio data are another form of continuous data, which have the same properties as interval data, plus a true definition of an absolute zero point – for example weight or height *(6)*.

4.2.4 Types of statistical analysis

Statistical methods can be divided into two main branches: descriptive and inferential. Descriptive statistics are commonly used to categorize, display and summarize data; inferential statistics are used to make predictions based on a sample obtained from a population or some large body of information. These inferences can be used to test specific research hypotheses *(7)*. This chapter covers the basic statistical principles that should be considered when choosing a study design and conducting the study. It includes examples and definitions of issues such as summary statistics and the calculation of the sample size needed for a study. Other chapters in this book deal with the development of the research question for a study (Chapter 3.5), study design (Chapter 4.1) and data collection (Chapter 4.4); more advanced statistical techniques are covered in Chapter 4.5.

4.2.5 Descriptive statistics

Descriptive statistics are typically used simply to calculate, describe and summarize the collected data in a logical, meaningful, and efficient way. Descriptive statistics do not allow any conclusions to be drawn regarding the validity of research hypotheses. They might include measures of central tendency (such as the mean, the median and the mode) to show the most representative value of the data set. They are usually accompanied by a measure of dispersion (such as the standard deviation or inter-quartile range) to indicate the degree of variation of values within a data set or the level of dispersion of observations around the measure of central tendency. Some of these are described below.

4.2

Measures of central tendency

Mean: the mean (sometimes referred to as the arithmetic mean) is the most common measure of central tendency. It is calculated by the dividing the sum total of all observations by the number of records. One advantage of the mean is that, because its calculation includes the summing of all the observations, its value takes into account all the data. However, this characteristic of the mean also makes it especially sensitive to extreme values among the observations, which can skew this central tendency towards extreme outliers. Thus, the mean can be a misleading measure if the data set contains such outliers.

Median: this the observation that divides the distribution into two equal parts. In other words, when all observations are ranked from the lowest to the highest, the median is the observation that is located at the half way point. Therefore, the median can only be determined for observations that are ranked by value or size and is less influenced by extreme values. The median can be used to compare groups on certain characteristics (for example, to compare the age between two groups of children or to compare number of days of exposure to extreme weather for people in different regions).

Mode: this is the observation or value that appears most frequently in a set of data. The mode is identified by noting the observation that occurs the most or value that has the highest number of records. The mode has the advantage of being easy to identify by simply counting the frequency of the records presenting that value. However, its main disadvantage is its potential lack of stability as a measure of central tendency because it can change if the data set is categorized or even defined in different ways. The mode can be used to determine, for instance, which socioeconomic group has the highest number of individuals.

Measures of dispersion

Standard deviation: this is the square root of the deviance, which is calculated by squaring and summing the difference between each observation and the arithmetic mean. The sum is then divided by the total number of observations. In the same population, the standard deviation is more stable from one sample to another. When comparing two groups or samples, a group or sample with a relatively smaller standard deviation indicates that the members of this group are more homogenous (or similar to each other) than the group with a large standard deviation. If the observations in a data set have a normal distribution, 70% of observations will lie within one standard deviation of the mean and 95% within two standard deviations *(8)*.

Standard error: This measures the amount of variance in the sample mean and is calculated by dividing the standard deviation by the square root of the number of observations in the sample. The standard error is used to indicate how well the true population mean is likely to be estimated by the sample mean.

Range: This represents the difference between the highest and the lowest values of the distribution and can be used to give complementary information to other statistics, such as the mean. When two distributions seem to have similar means, the range can provide an additional layer of information to distinguish the characteristics of the two distributions.

However, one important disadvantage of the range is that it will be influenced by extreme values. This means that a change in a single record that was the highest or lowest value could have a substantial impact on the range. The range can also be expressed in quartiles or in percentiles to show the highest and lowest values in different parts of the distribution (such as the range of ages for children and for adults in a sample).

Interquartile range: Just as for calculating the median as the half-way point in a series of observations, the interquartile range requires the observations to be ranked from the lowest to the highest. The interquartile range median is then the difference between the lower (25th percentile) and the higher (75th percentile) quarters of the observations.

Confidence interval: This is derived from the standard error of the mean. The confidence interval (usually 95%) shows the range within which the true population value is likely to fall, based on the sample statistical values and probability data distributions.

4.2.6 Inferential statistics

In the context of research into the effects of interventions (as discussed in Chapters 4.1 and 4.3), inferential statistics allow researchers to make a valid estimate of the association between an intervention and its effect in a specific population, based upon their representative sample data. Inferential statistics allow researchers to make generalizations or inferences from the results obtained from the sample to the populations from which the samples were drawn. Approaches to inferential statistics include the estimation of parameters, and the testing of research hypotheses. Inferential statistics vary depending on the type of statistical tests applied in the analysis. For instance, they might use correlation coefficients to assess the correlation and association between risk factors and outcome, or use an odds ratio to measure the probability of an event occurring.

4.2.7 Rapid needs assessments

Rapid needs assessments (as also discussed in Chapter 2.1) will usually require basic statistical analyses to be conducted. For instance, in disaster settings, rapid needs assessments often use survey sampling techniques in the field to rapidly determine the health status and basic needs of an affected community. Emergency response requires immediate information on health status and community needs. Such information must be gathered and analysed quickly. In many cases, an assessment may need to be initiated and completed within 72 hours. Speed is critical because circumstances can change dramatically with time, and outdated information may therefore be of little use to response personnel *(9)*. However, these surveys need to be conducted in a statistically robust and valid manner to support decisions about the response. Various areas of consideration (such as disease states or conditions) might need to be measured using various statistical parameters – such as prevalence, incidence and attack rate (see below).

A rapid health needs assessment is often carried out at a single point in time, using a cross-sectional study design. Key stakeholders should be involved in the survey process, and it is important to identify specific

targeted groups as the study population, depending on the objective of such needs assessment. For example, when undertaking a nutrition assessment, the study population may include all children under the age of 5 years and their parents. The sample size for the study (see below) might not be estimated statistically but may simply be based on the population who are being studied.

Rapid needs assessments will collect data on the population and may include the number of displaced or affected people and their demographic characteristics (for example, the number of women, men, children, pregnant women and persons with disability). It might also be important to collect data on the proportion of people with shelter, in order to establish the shortfall in shelter requirements for the displaced population (such as refugees or internally displaced persons). Data should also be collected on the available resources (see also Chapter 3.1), including health systems. This might include the number and type of health facilities, number and category of health workers and types of health services available. Depending on the situation, data may also be collected from other sectors such as water and sanitation, education, food security, protection and so on. It might also be gathered to establish a picture of other baseline features, such as numbers of medical staff still working per 1000 people in the population, vaccination rate for key vaccines or rate of severe acute malnutrition. During emergencies, the values of these indicators are usually compared to reference values and norms, such as the Sphere standard to evaluate the status of population humanitarian condition *(10)*. There is more information on health indicators in Chapter 2.2.

4.2.8 Epidemiologic Measures

This section provides a brief review of some key terms used in epidemiology to describe data about diseases.

Population

In the epidemiology of disasters (Chapter 2.1), the definition of the "population" can vary depending on the situation. In general, the term is used to refer to people living in a defined area, such as a refugee camp, settlement, village or neighbourhood. However, in some situations, it may refer to groups of people being affected by an emergency, who do not necessarily live in a well-defined area. For instance, in an infectious disease outbreak, population may refer to groups of people with a specific characteristic, such as a profession, lifestyle or activity that predisposes them to the disease (for example, farmers, butchers, or those in school settings). It might also be necessary to count subgroups of the population, such as the number of women or the number of children under 5 years of age.

In some cases, the total population figure will be the denominator for calculating health indicators (Chapter 2.2). For example, it might be used to estimate the proportion of people out of the total population who were made homeless after an earthquake, the proportion of pregnant women who are likely to give birth in the days after a disaster, or the proportion of children in an internally displaced person (IDP) camp who have not been vaccinated against measles.

Usually, the census or a registration system might be relied on as the most accurate method of estimating the population. However, in an emergency, it might be necessary to use other methods (Chapter 2.4), such as mapping the IDP camp and dividing it into smaller sections, with the population size of each section estimated using sample surveys.

Depending on the type of data being collected and the context, gathering information from individuals can sometimes be perceived as intrusive. It is, therefore, important to identify and implement methods to count people and cases that maintain the dignity of the individuals involved, using appropriate ethical oversight (Chapter 3.4). This is especially important if public health priorities (speed, accurate information) and human rights priorities (privacy, consent) might come into conflict during data collection.

Data Analysis

Basic data analysis can be used to provide information to guide the development and implementation of operational plans for Health EDRM. The information is often summarized into a minimum set related to person, place and time. Minimum data analysis can generate basic answers to questions such as: who is affected or most at risk? Where are those affected or at most risk? What is the trend of the impact of the events on the target population? Subsidiary, basic analysis can provide insight into major risk factors making the target population vulnerable or rendering them resilient to the effect of the hazard. In addition to the descriptive statistics outlined above, epidemiology uses measures of morbidity and mortality and these rely on the quantification of various aspects of health, outlined below.

Prevalence

This is useful for understanding the overall burden of a disease on a population, since it describes how common a particular condition is at a given point in time (point prevalence) or the existing and new cases that happen over a set period of time, such as 12 months (period prevalence). Prevalence is a calculation of the existing cases and is determined by the rate of new cases occurring, the rate of recovery and the rate of deaths. Prevalence is often used for conditions that are longer lasting or for which an on-set date may be more difficult to recall (for example, the number of people suffering anxiety related to a disaster).

Incidence

This is the number of new cases of the condition occurring in a given population during a defined period of time. There are different ways to calculate incidence, based on the condition, issue or disease. The most common is the cumulative incidence, which is the number of new cases in a specific time period divided by the number of people who were initially disease (or condition) free at the start. For example, if there were 120 new measles cases in one week among 18 000 people in an IDP camp, this would give an incidence rate of 6.7 per 1000 per week. The incidence rate is useful when discussing or comparing acute, communicable diseases of short duration.

Attack rate

This is the cumulative incidence rate of a disease in a specified population over a given period of time. It is usually used during epidemics and is calculated as the percentage of the population with a condition out of the

whole population (for instance, those with the condition and healthy, susceptible people) (Table 4.2.1). The attack rate can help when calculating the resources needed to respond to an outbreak. It also provides an idea as to the magnitude of an outbreak in a community or a geographic entity. If immunity to the disease (as a result of vaccination or prior infection for instance) is measured, this may allow some of the population to be removed from the denominator.

Table 4.2.1 Example of incidence and attack rate for measles among 18 000 refugees

Week	New cases per week	Weekly Incidence Rate	Attack Rate
1	120	6.7 per 1000	0.67%
2	150	8.3 per 1000	1.50%
3	80	4.4 per 1000	1.94%

Case fatality rate

This is the number of deaths from a specific disease during the observational period, divided by the number of cases of that disease during that period, multiplied by 100 (to calculate a percentage). The case fatality rate is used mainly in infectious diseases, such as cholera, dysentery, malaria and measles. It provides a useful guide to assess the virulence of the disease, its severity and the effectiveness and quality of care.

Mid-interval population

This can be estimated by adding together the number of people in the population at the start of the period of observation and the number at the end, and dividing this by two. Alternatively, it can be calculated as the average size of the population during the period. Population data are usually collected from official government census reports or other administrative documents, such as the birth and deaths registry. It may already be available from national statistical offices and published online.

Benchmarks

These are standards or reference values for indicators that serve as signposts to let the researcher, or other interested people such as policy makers, know what has been achieved or how severe a situation is. They can include key mortality indicators such as the infant mortality rate, cause-specific mortality rate and case fatality rate discussed below.

4.2.9 Demographic indices

Demographic indices include statistics such as fertility rates, birth rates, growth rates and mortality rates.

Crude birth rate

This is calculated as a proportion by dividing the number of live births by the number of people in the mid-interval population, and multiplying the value by 1000 (or other amount depending on the population size) to create a rate.

Crude growth rate
This is the crude birth rate minus the crude mortality rate. It provides information on the growth or decline of a population, in the absence of migration.

Crude mortality rate
This is calculated as a proportion by dividing the number of deaths at all ages by the number of people in the mid-interval population, and multiplying the value by 1000 (for annual or monthly rates) or 10 000 for daily crude mortality rate. This crude rate does not adjust for the age distribution of the population, and should not be used to compare across different populations.

Infant mortality rate
This is calculated by dividing the number of deaths in children under one year of age by the number of live births during the same period and multiplying this by 1000 (or other amounts depending on the population size). Although this is conventionally referred to as a rate, it is really a ratio. This is because in a rate, those counted in the numerator must also be part of the denominator (for example, the number of deaths due to measles divided by cases of measles). However, in the infant mortality rate, some of those children who die during the specified interval (the numerator) might not have been born during the same interval (the denominator).

Cause-specific mortality rate
This is the number of deaths from a specific cause during the observational period divided by the number of people in the mid-interval population (or other denominator of the population), multiplied by 100 to provide a percentage.

Age-specific mortality rate
Because different populations have different characteristics and age structures it is not meaningful to compare the crude mortality rate for different settings or countries. For example, a high proportion of elderly people in a population will give it a high crude mortality rate and, as a result, the crude mortality rate of the Plurinational State of Bolivia and that of the USA may be very different because of the underlying age-distribution rather than the likelihood of an individual dying. To overcome this, age-specific mortality rates are calculated. There are two different methods of standardizing population statistics – direct standardization and indirect standardization. More information on these methods can found in Gerstmann *(11)*.

4.2.10 Epidemiological Studies

Epidemiological studies can be descriptive, analytical or both. Descriptive studies are used to describe exposure and disease in a population (see Chapter 3.2), and can be used to generate hypotheses, but they are not designed to test hypotheses. Analytical studies are designed to test hypotheses, and are designed to evaluate the association between an exposure or intervention and a disease or other health outcome (see Chapters 4.1 and 4.3).

Epidemiological studies can be cross-sectional, prospective or retrospective. A cross-sectional study is taken at a specific point in time. A

prospective study is one where the study starts before the exposure and outcomes are measured moving forward in time. A retrospective study is one where the study starts after the exposure has begun and, in some cases, the outcomes have occurred and been measured. It works backwards in time. Epidemiological studies can also be experimental or observational and some of the terminology important for epidemiological studies is described below.

Exposure
This is the risk factor (agent, experience or procedure for example) that is suspected to have caused the disease or condition. In statistical terms, exposure is often called the independent variable.

Outcome
This is the disease, condition or other endpoint being measured. In statistical terms, the outcome is often called the dependent variable.

4.2.11 Descriptive studies

Descriptive studies describe an event, condition or disease state in terms of time, place and person. They include:

- Case series or record review.
- Descriptive incidence study (active surveillance) (for example, collecting information on all cholera cases, by age, sex, location of hut, nearest water source and duration of stay in an IDP camp).
- Descriptive prevalence study (cross-sectional survey) (for example, a study of prevalence of acute malnutrition among children under 5 years of age).
- Ecological study (for example, times series analysis of the impact of air pollution on respiratory morbidity and mortality).

4.2.12 Analytical studies

Analytical studies examine the relationship between a possible cause (or exposure or intervention) and its effect (disease or condition). These are generally developed to test a hypothesis, which could have been developed from descriptive studies previously undertaken. Two common examples of analytical studies are cohort studies and case-control studies:

Cohort study
In a cohort study, a population is followed over time (either prospectively or retrospectively). There are usually two study groups: those exposed to a certain exposure – which may be either a risk factor (such as diet deficient in vitamin C) or a protective factor (such as measles immunization) – and those not exposed. The cohort study follows both groups over a period of time and estimates incidence of the outcome in each group. The measure of association in this study design is the relative risk, which is the ratio of the incidence of disease in the exposed group to the incidence of disease in the non-exposed group. Cohort studies can be carried out in many time frames, from days to decades.

Case-control study
In a case-control study, the two groups being compared are people who meet the criteria (or case definition) of the disease or other outcome and people from the same or similar population who do not, as a control group. This retrospective design is used to determine who was exposed to certain factors (contaminated water, for example) and who was not exposed and whether exposure in those who have the outcome is different to those without. The measure of association in this study design is often the odds ratio, which is the ratio of the odds of disease in the exposed group to the odds of disease in the non-exposed group. The odds of disease is the proportion of people with the disease divided by those without it.

4.2.13 Sampling Methods

When choosing the people to include in a study, a variety of sampling methods are available:

Non-probability or judgemental sampling
For example – convenience, snowballing or quota sampling.

Probability sampling
Probability sampling includes simple random sampling, systematic sampling and cluster sampling; Table 4.2.2 shows the advantages and disadvantages of each of these specific methods.

Simple random sampling:
This would lead to a fully random sample by using a method such as a random number table to draw the sample from a whole population to which all the members belong.

Systematic sampling
This involves choosing the first member of the sample of the whole population using a random number and choosing the rest of the sample by proceeding at a fixed interval.

Cluster sampling
This involves the random selection of a cluster (such as a village, school or hospital) and then random sampling of the individuals from within the selected clusters.

Table 4.2.2 Advantages and disadvantages of different types of probability sampling

Type of probability sampling	Advantages	Disadvantages
Simple random sampling	Minimal bias. Every member has an equal chance of being included (which can balance confounding factors).	Must enumerate all members of the population, which is expensive and sometimes not feasible. Can miss geographical clusters (such as people from a minority ethnic group living in one part of an IDP camp).
Systematic sampling	Guarantees a broad geographical representation. Do not have to have prior knowledge of the total number of people who could be selected for the study.	May be expensive and time consuming to ensure full randomization.
Cluster sampling	Easier to conduct, less travel time and cost. Do not need a complete list of the sampling units.	Bias toward more dense areas, such as town centres.

If a sample is used for a study, rather than the whole population, this leads to an estimate of what the results might be for the population as a whole. If a series of samples is taken, these are likely to give different values, but providing the samples have been selected correctly there should be little variation between them. However, in order to provide an estimate of this variation, confidence intervals are often used to show the extent of the variation. The confidence intervals provide the upper and lower limits of this range. For example, if the mean for a sample was 12% and the standard deviation was 2%, the 95% confidence interval would be shown as 10 to 14%.

4.2.14 Sample size calculation

If it were possible for a research study to include the whole population of interest, sampling would not be necessary, but covering a whole population would usually require too much money, time or personnel. Therefore, researchers need to rely on a population subset: the sample. This allows them to seek reasonably valid answers to their research questions, but they first need to estimate the size of the sample needed to achieve this. Determining the appropriate sample size for a study is a fundamental aspect of all research; this is because having an adequately-sized sample ensures that the information the study yields will be reliable, regardless of whether the data ultimately suggest an important difference between the impact of a disaster on different types of people, or the effects of intervention and control in a randomized trial.

Two types of false conclusion may occur when inferences about the whole population are derived from a study of a sample of the population. These are called Type 1 and Type 2 errors, whose probabilities are denoted by the symbols σ and β. A Type 1 error occurs when one concludes that a difference exists between the groups being compared when, in reality, it

does not. This is akin to a false positive result. A Type 2 error occurs when one concludes that a difference does not exist when, in reality, a difference does exist, and it is equal to or larger than the effect size defined by the alternative to the null hypothesis *(12)*.

The calculation of a sample size for a research study depends on the type of study being planned, the data to be collected, the outcomes being measured and the hypothesis being tested *(13)*. More information is available in the texts listed in the further reading section (4.2.17) but, in general, sample size estimation depends on the level of confidence and precision. The following formula can be used to calculate the sample size for a binary outcome:

$$n = \frac{Z^2\, pq}{d^2}$$

n corresponds to the sample size in each of the groups; *Z* is the level of confidence chosen (95% confidence, *Z* = 1.96; 90% confidence: *Z* = 1.68); *g* is the design effect and a usual value for this situation is 2; *p* is expected proportion of the population with the characteristic of interest (such as acute malnutrition), *q* is 1-*p*; and d is the precision (in proportion of one; if 5%, *d* = 0.05).

This formula shows that in order to increase the level of confidence or precision, the sample size must be increased. Therefore, when a study is trying to detect a small effect with high precision (such that the entire width of confidence interval would be consistent with a beneficial effect of an intervention, for example), the study will need to be much larger than when the study is testing a hypotheses that there is a large effect.

4.2.15 Conclusions

This chapter presents an introduction to basic statistical concepts, epidemiologic study designs, commonly used sampling methods and estimation of sample size. It provides basic statistical knowledge to support effective Health EDRM.

4.2

4.2.16 Key messages

- **Statistical analyses of quantitative data from research studies and the results these generate are vital to a variety of types of research in Health EDRM. They help by estimating disease burden (to help with the distribution of humanitarian assistance, for instance), the health consequences of disasters for populations (to help with planning for future needs, for example) and the effects of interventions, actions and strategies (to prioritize the elements to include in humanitarian assistance, for example). They often require the contribution of partners with diverse disciplines.**
- **Practitioners need to understand a variety of methods of data collection and analysis, and apply those most relevant to their research question if they are to answer it reliably. This might include surveys, cohort studies, case control studies or experimental studies such as randomized trials for quantitative research and the use of qualitative methods where appropriate.**
- **Research in emergency settings is constrained by ethical concerns (Chapter 3.4) and limited resources, increasing both the challenges of conducting rigorous epidemiological research and the importance of reliable statistical analysis of the data that are available.**

4.2.17 Further reading

Gerstman B. Basic Biostatistics: Statistics for Public Health Practice (2nd edition). Burlington, MA: Jones & Bartlett Learning. 2014.

Horney JA . Disaster Epidemiology: Methods and Applications. London, UK: Elsevier. 2017.

Ricci EM, Pretto EA. Disaster Evaluation Research: A field guide. Oxford, UK: Oxford University Press. 2019.

4.2.18 References

1. Global Emergency Overview Snapshot, 6 - 12 May 2015 – World. ReliefWeb. 2015. https://reliefweb.int/report/world/global-emergency-overview-snapshot-6-12-may-2015 (accessed 12 June 2019).
2. Outbreak surveillance and response in humanitarian emergencies. WHO. 2012. http://www.who.int/diseasecontrol_emergencies/publications/who_hse_epr_dce_2012.1/en/ (accessed 13 April 2018).
3. South Sudan health crisis worsens. WHO. 2016. https://www.who.int/news-room/feature-stories/detail/south-sudan-health-crisis-worsens-as-more-partners-pull-out-and-number-of-displaced-rises (accessed 12 June 2019).
4. South Sudan weekly disease surveillance bulletin 2019. WHO Regional Office for Africa. 2019. https://www.afro.who.int/publications/south-sudan-weekly-disease-surveillance-bulletin-2019 (accessed 12 June 2019).
5. WHO and MoH. South Sudan (EWARN) Early warning and disease surveillance bulletin. 2015. http://www.who.int/hac/crises/ssd/epi/en/index3.html (accessed 22 June 2017).
6. Vetter TR. Fundamentals of Research Data and Variables: The Devil Is in the Details. Anesthesia and Analgesia. 2017: 125: 1375-80.
7. Overholser BR, Sowinski KM. Biostatistics primer: part I. Nutrition in Clinical Practice. 2007: 22: 629-35.
8. Kirkwood BR, Sterne JAC. Essential Medical Statistics (2nd edition). Malden MS: Blackwell Science. 2003.
9. Malilay J, Heumann M, Perrotta D, Wolkin AF, Schnall AH, Podgornik MN, et al. The role of applied epidemiology methods in the disaster management cycle. American Journal of Public Health 2014: 104: 2092-102.
10. Sphere. The Sphere Handbook: Humanitarian Charter and Minimum Standards in Humanitarian Response (4th edition). Geneva, Switzerland. 2018. https://handbook.spherestandards.org/en/sphere/#ch001 (accessed 16 January 2020).
11. Gerstmann B. Basic Biostatistics: Statistics for Public Health Practice (2nd edition). Burlington, MA: Jones & Bartlett Learning. 2014.
12. Hazra A, Gogtay N. Biostatistics Series Module 5: Determining Sample Size. Indian Journal of Dermatology. 2016: 61: 496-504.
13. Devane D, Begley CM, Clarke M. How many do I need? Basic principles of sample size estimation. Journal of Advanced Nursing. 2004: 47(3): 297-302.

4.3 Cluster Randomized Trials

Authors

Matthew Coldiron, Epicentre, Paris, France.

Rebecca F. Grais, Epicentre, Paris, France.

4.3.1 Learning objectives

To understand the role that cluster randomized trials can play in health emergency and disaster risk management (Health EDRM), including:

1. The advantages and disadvantages of the cluster randomized trial methodology.
2. Situations in which cluster randomized trials could be used.
3. Potential difficulties in the implementation of cluster randomized trials and solutions for overcoming them.

4.3.2 Introduction

Chapter 4.1 discussed the role of individually randomized trials in resolving uncertainties about the effects of interventions, actions and strategies, and focused on studies in which the allocation to groups is determined at the level of each individual participant. However, in cases where this is not possible or appropriate, studies may be designed to randomize groups of participants ("clusters") rather than individuals, in what are called cluster randomized trials – sometimes also known as group-randomized trials or place-randomized trials – and these are the focus of this chapter.

In a cluster randomized trial, the intervention is directed at a group of people, which makes this design well-adapted for performing research in Health EDRM situations. Common examples of clusters include villages, schools, doctors' offices, and different wards or services of a hospital. A variety of designs have been used *(1)*. For example, cluster randomized trials have been used to evaluate the effectiveness of:

- Mass vaccination *(2)*
- Mass antibiotic prophylaxis during epidemics *(3)*
- Water and sanitation packages designed to prevent diarrhoeal disease *(4–5)*
- Population-based interventions aimed at decreasing the incidence of acute malnutrition *(6)*.

4.3.3 Design of cluster randomized trials

Most people are more familiar with individually randomized trials (Chapter 4.1) than with cluster randomized trials. However, many of the same considerations apply to their design. These include:

- ensuring that there is not already evidence that would support the hypothesis being tested (ensuring "equipoise", or genuine uncertainty about the potential effects of an intervention);
- conducting a scoping review (Chapter 3.6) or systematic review (Chapter 2.6) if needed;
- defining relevant outcomes;
- estimating the expected effect size of the intervention;
- developing an appropriate strategy for randomization and, if appropriate and necessary, for blinding participants and others involved in the trial to a person's allocated group.

There are however some important differences between cluster randomized trials and individually randomized trials. For example, the risk of an imbalance in potential confounding factors may be higher in a cluster randomized trial, because the number of clusters included is usually smaller than the number of individuals included in an individually randomized trial. Identifying and mitigating selection bias can also be more difficult in cluster randomized trials, where the study intervention is allocated at cluster level, but some individuals within the clusters may choose not to participate. It is also usually impractical (and often impossible) to keep study participants and researchers blinded to intervention allocation in a cluster randomized trial.

There are several additional considerations specific to the cluster randomized trial design. The first concerns the timing of the interventions in the different groups. Clusters are most commonly randomized in parallel, with group allocation happening at the same time. However, in some cases it is not desirable or feasible to carry out parallel randomization. If an intervention takes a long time to put into place (for example, a sanitary system or a new monitoring system in a hospital ward), researchers will sometimes perform what is called a stepped-wedge cluster randomized trial *(7)*. In this type of trial, the different clusters receive the intervention sequentially, and the outcomes of interest are compared across the different clusters, taking into account when the intervention was implemented, with all clusters having received the intervention by the end of the trial.

Secondly, crossover between individuals in different clusters needs to be minimized. The potential for individuals not in a given cluster to receive the intervention, or to have second-hand or spillover benefit from it, must be considered when designing a cluster randomized trial. If clusters are physically distant and there is little contact between them, significant crossover (or contamination) effects are unlikely. Separation of clusters can be integrated into trial design from the beginning, as was done in a trial of emergency room care for acute stroke in which hospitals were purposefully selected to minimize movement of physicians between emergency departments *(8)*. However, if clusters are contiguous

neighbourhoods of a city, or if there are important cultural links between two distinct villages, it is reasonable to expect that some crossover may occur. Researchers should strive to reduce this risk as much as possible.

Thirdly, the effects of clustering need to be accounted for during statistical analysis. In an individually randomized trial, participants receive their intervention (medication, vaccine and so on) and are evaluated individually. In a cluster randomized trial, the intervention is performed at the cluster level, but the outcome of interest is often measured at an individual level. For instance, in a cluster randomized trial evaluating village-level sanitation interventions, where the outcome of interest is diarrhoea, inherent characteristics of the villages, such as socioeconomic level and proximity to a floodplain, might play an important role in the risk of developing diarrhoea. Quantifying the similarities between individuals in a cluster in the intra-cluster correlation coefficient is an essential factor when calculating the sample size and the results of a cluster randomized trial *(9–10)*. Finally, it is important to recognize that inferences made from results of cluster randomized trials are often applied at an individual level, despite the cluster-level randomization. This has important consequences for data analysis, and for communication of trial results. Case Study 4.3.1 describes a novel cluster randomized trial of Ebola vaccines.

Case Study 4.3.1
A novel cluster randomized design for evaluating Ebola vaccines *(2)*

A relatively novel cluster randomized design was used to evaluate experimental vaccines early during the 2014 West Africa Ebola outbreak. The trial was a cluster randomized trial modelled on the ring vaccination approach used in the 1970s to eradicate smallpox. Ring vaccination involves vaccinating individuals who are socially or geographically connected to a confirmed case of an infectious disease, thereby creating a "ring" around infected individuals to prevent spread. In the ring trial, contacts of Ebola cases were enrolled and randomized into two groups, one of which was vaccinated immediately with an experimental vaccine, while the other was assigned to receive the vaccine 21 days after enrolment. The delay of 21 days was based on Ebola's maximum incubation period of 21 days after infection and on the fact that it takes some time for vaccine-induced protection to develop. The design was chosen because the time delay provided a non-placebo comparator group. Incidence of Ebola was compared between the rings (clusters) vaccinated immediately and those vaccinated with a 21-day delay. This design was controversial among scientists and ethicists, but was seen as an acceptable compromise between scientific rigour and providing hoped-for benefits of an unproven vaccine.

4.3.4 Advantages of cluster randomized trials

The most obvious advantage of cluster randomized trials over individually randomized trials is that they allow the evaluation of study interventions that cannot be directed toward selected individuals. This may be because of feasibility (for example, radio advertisements about smoking cessation,

or nursing protocols in a hospital ward), or biological mechanisms (such as interventions that aim to induce herd protection in a population). In certain situations, they may also be easier to implement than an individual-level intervention. For example, providing an intervention about hand hygiene to mothers in a rural village would reasonably be expected to have indirect spillover effects to other members of her household *(11)*.

4.3.5 Disadvantages of cluster randomized trials

The disadvantages of cluster randomized trials compared with individually randomized trials include the greater complexity of their design, as discussed above, as well as the need to include larger numbers of individual participants to obtain the same statistical power *(11)*. Specifically, the intra-cluster correlation coefficient is the main driver of the differences in sample size and clustering must also be considered during analysis of trial data. An example would be an educational intervention in which schools are randomized to one of several new teaching methods. When comparing differences in outcome achieved under the new methods, researchers must account for the fact that two students sampled from the same school are more likely to be similar in terms of outcomes than two students sampled from different schools. Multilevel or other similar statistical models are typically used to correct for non-independence of this kind.

On a more practical level, the hierarchical nature of cluster randomized trials can lead to a duplication of upstream preparation and sensitization efforts – first at cluster-level, and then among individuals in the clusters. This may have cost and time implications for researchers.

Cluster randomized trials are generally not designed to show individual-level effectiveness as a primary objective because the interventions happen at population level. For this reason, it is unusual to use a cluster randomized design with non-licensed products. Nonetheless, in some cases, it is possible to estimate individual effectiveness of an intervention by comparing outcomes among persons who are known to have received the intervention with those who are known not to have received it.

4.3.6 When to use a cluster randomized trial design

Cluster randomized trials are best suited for testing interventions intended for a group of people. Any population-based, mass distribution or administrative activity, such as those used in Health EDRM, lends itself well to cluster-based randomization. Health promotion activities and other interventions aiming to change behaviour are often tested in cluster randomized trials. This is also the case for interventions with a high risk of contamination. In this context, the term “contamination” refers to when individuals randomized to different comparison groups are in frequent contact with one another and thus may be influenced (contaminated), in either or both directions. Contamination is likely to occur in comparisons of interventions within the same community, but randomizing at community-level is an effective solution to this problem.

4.3

Cluster designs can also have practical advantages over individual randomization. They are easier to understand conceptually for policymakers who may be less familiar with the statistical and scientific properties of different trial designs, because they mirror more closely how interventions are implemented at scale. This is one of the reasons they are also a design that should be considered in an emergency, disaster or public health crisis. The design provides easy-to-understand information for groups of people and policy-makers, and can reach more participants due to the larger sample size. It is also important to consider that cluster randomization can capture both direct and indirect effects of an intervention. This is important when assessing effectiveness in a population and means that cluster randomized trials are well-suited to infectious diseases, when there might be direct benefits to those who receive the intervention as well as indirect benefits to those around them, who may benefit from a reduction in exposure *(12)*.

Case Study 4.3.2 describes how a cluster randomized trial was used to test village-wide antibiotic prophylaxis for meningococcal meningitis.

Case Study 4.3.2
Testing a strategy of village-wide antibiotic prophylaxis during a meningococcal meningitis outbreak *(3)*

Mass vaccination campaigns have been part of the standard response to meningococcal meningitis outbreaks in the African meningitis belt for decades, but vaccine supply is not always guaranteed. Antibiotic prophylaxis for contacts of cases is recommended in high-income countries but is not recommended in the meningitis belt because of a lack of evidence. As meningitis epidemics are seasonal, a cluster randomized trial protocol was prepared to test whether a village-wide prophylaxis strategy would work in this setting. When an epidemic hit the Madarounfa District of the Republic of the Niger, the trial started. After the first case was notified in each village, that village was randomized to receive either no prophylaxis, prophylaxis with single-dose ciprofloxacin for household contacts of meningitis, or a village-wide distribution of single-dose ciprofloxacin. The primary outcome was overall meningitis attack rate in the villages at the end of the epidemic. Household prophylaxis did not reduce attack rates, but village-wide prophylaxis reduced attack rates by 60%.

This trial is an example of research performed in an emergency setting. Not all emergencies can be predicted in advance, but in this setting, it was reasonable to be prepared for a meningitis epidemic. The advance preparation, including ethical review, meant that the trial could start very quickly after the beginning of the epidemic. A cluster randomized design was appropriate because the village-wide distributions were implemented across an entire population. Clustering within the individual villages was weaker than expected, which allowed for greater statistical power to discern differences in the meningitis attack rate. Because the villages included in the trial had a reasonable degree of separation, there was little evidence of spillover, which added to the reliability of the main results. If the villages had been closer to each other or there had been more social contact between them, it is likely that more persons from villages randomized to no prophylaxis or household-prophylaxis would have received prophylaxis, which could have influenced the results.

4.3.7 Informed consent in cluster randomized trials

Ethical issues relating to informed consent for participation in research are discussed more fully in Chapters 3.4 and 6.4. In an individually randomized trial (Chapter 4.1), a researcher approaches a potential study participant, explains the nature of the study, potential harms and benefits of participation, and underscores the potential participant's freedom to choose whether to participate in the study without negative consequence. If the participant provides informed consent, they are randomized and receive the study intervention and follow study procedures.

However, this procedure can be difficult – or even impossible – to replicate in cluster randomized trials, which generally take place at a larger scale, and in which many participants will not directly receive the study intervention which is to be given at the cluster level. Researchers and ethicists have therefore established a set of guidelines for the ethical conduct of cluster randomized trials, including issues related to obtaining informed consent from participants: the Ottawa Statement on the Ethical Design and Conduct of Cluster Randomized Trials *(13)*.

The guidelines require that trial protocols be reviewed by ethics committees, and address some of the inherent challenges with trials where the level of intervention (cluster level) may differ from the level of outcome ascertainment (individual level). All individuals living in participating clusters are considered to be research participants, which may prove problematic given the size of some cluster randomized trials. Crucially, the guidelines lay out specific criteria for justifying the use of "gatekeepers" who may provide permission for a cluster to participate in a trial (such as a village chief or a nurse manager of a hospital ward). The permission of a gatekeeper should not be confused with proxy consent for individuals to participate, but does allow for most cluster randomized trial interventions to proceed without the individual-level informed consent that is required in individually randomized trials.

Nonetheless, even if a gatekeeper provides permission to participate, researchers have an obligation to communicate openly with individuals in the randomized clusters about the objectives of the research, their individual risks and benefits, and their autonomy to decide whether to participate in study activities, including simply being counted as a study participant. The Ottawa Statement is very clear that any derogation of individual consent must be reviewed and approved by ethical review committees (Case Study 4.3.3).

If unlicensed or investigational medicines or vaccines are used in a cluster randomized trial, it is likely that individual written informed consent would be required from all participants, just as in an individually randomized trial. Given the comparatively larger size of most cluster randomized trials, researchers should consider this during trial design and when they are planning the number of staff that they will need.

4.3

Case Study 4.3.3
Permission to participate and informed consent process in a cluster randomized trial

In the antibiotic prophylaxis trial described in Case Study 4.3.2, 49 villages were included in the trial over the course of only 27 days. The total population of these villages was 71 308, including 22 177 who lived in villages that were randomized to receive village-wide distributions of antibiotic prophylaxis.

Even without the emergency situation caused by the ongoing epidemic, it would have been impossible to obtain individual written consent from all persons living in the randomized villages over that brief time period. During study protocol development, the researchers reviewed the Ottawa Statement, and after consultation with the ethical review committees, determined that the criteria for the waiver of individual consent were met. During the trial, village chiefs served as "gatekeepers" and were asked to provide permission for the randomization of their villages.

At the same time, community health workers shared information about the trial in all participating villages. In villages allocated to receive ciprofloxacin distributions, the same community health workers passed through the village before the distribution to give information about the potential harms and benefits of single-dose ciprofloxacin prophylaxis and underscored that there was no obligation to take the prophylaxis. During the village-wide distributions, 77% of the target population received ciprofloxacin. The researchers believed that this was partly due to absences and partly due to individuals choosing not to participate, suggesting that the overall informed consent process of the trial was successful.

4.3.8 Special design and analysis considerations

Cluster randomized trials require careful reflection during their design and analysis. This is primarily because data collected about individuals in clusters are almost always correlated. The outcomes of an individual within a cluster may be likely to be the same as that of other individuals in the same cluster. This needs to be accounted for in the analyses, and subsequent interpretation of the results must consider both intra-cluster correlation and between-cluster variability. Between-cluster variability can be summarized using the coefficient of variation between clusters, and the intra-cluster correlation coefficient. These intuitive statistical properties require the guidance of a researcher experienced in these techniques who can help guide the design of the trial.

4.3.9 Conclusions

Cluster randomized trials have become more common and have been implemented for a variety of Health EDRM issues. Although they are similar to individually randomized trials, cluster randomized trials have important design differences that have implications for data analysis and interpretation of results.

4.3.10 Key messages

- **Cluster randomized trials are interventional studies well-adapted for many emergency situations, and are ideal for evaluating population-level interventions.**
- **Compared to individually randomized trials, cluster randomized trials usually require larger numbers of participants and can be more complex to design and analyse.**
- **Cluster randomized trials can be parallel randomized or sequentially randomized, such as in a stepped-wedge design *(7).***
- **The fundamental ethical principles are similar to those in individually randomized trials, but the Ottawa guidelines consider the particularities of cluster randomized trials** *(13)***.**
- **Design and analysis of cluster randomized trials requires careful reflection and the guidance of experienced researchers.**

4.3.11 Further reading

Campbell MK, Elbourne DR, Altman DG, CONSORT group. CONSORT statement: extension to cluster randomized trials. BMJ 2014: 328: 702–8.

Coldiron ME, Assao B, Page A-L, Hitchings MDT, Alcoba G, Ciglenecki I et al. Single-dose oral ciprofloxacin prophylaxis as a response to a meningococcal meningitis epidemic in the African meningitis belt: a three-arm, open-label, cluster-randomized trial. PLoS Medicine 2018: 15: e1002593.

Hayes R, Moulton L. Cluster randomized trials (2nd edition). Boca Raton, Florida, USA: Chapman and Hall/CRC Press; 2017.

Henao-Restrepo AM, Camacho A, Longini IM, Watson CH, Edmunds WJ, Egger M, et al. Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomized trial (Ebola Ça Suffit!). Lancet 2017: 389: 505–18.

Isaakidis P, Ioannidis JP. Evaluation of cluster randomized controlled trials in sub-Saharan Africa. American Journal of Epidemiology; 2003: 158(9): 921-6.

4.3.12 References

1. Crespi CM. Improved designs for cluster randomized trials. Annual Review of Public Health. 2016: 37(1): 1-16.
2. Henao-Restrepo AM, Camacho A, Longini IM, Watson CH, Edmunds WJ, Egger M, et al. Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomized trial (Ebola Ça Suffit!). Lancet 2017: 389: 505-18.

3. Coldiron ME, Assao B, Page A-L, Hitchings MDT, Alcoba G, Ciglenecki I, et al. Single-dose oral ciprofloxacin prophylaxis as a response to a meningococcal meningitis epidemic in the African meningitis belt: a three-arm, open-label, cluster-randomized trial. PLoS Medicine 2018: 15: e1002593.

4. Lin A, Ercumen A, Benjamin-Chung J, Arnold BF, Das S, Haque R, et al. Effects of water, sanitation, handwashing, and nutritional interventions on child enteric protozoan infections in rural Bangladesh: A cluster-randomized controlled trial. Clinical Infectious Diseases 2018: 67(10): 1515–22.

5. Lindquist ED, George CM, Perin J, Neiswender De Calani KJ, Norman WR, et al. A cluster randomized controlled trial to reduce childhood diarrhea using hollow fiber water filter and/or hygiene-sanitation educational interventions. American Journal of Tropical Medicine and Hygiene. 2014: 91(1): 190–7.

6. Dewey KG, Mridha MK, Matias SL, Arnold CD, Cummins JR, Khan MSA, et al. Lipid-based nutrient supplementation in the first 1000 d improves child growth in Bangladesh: A cluster-randomized effectiveness trial. American Journal of Clinical Nutrition. 2017: 105(4): 944–57.

7. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomized trial: Rationale, design, analysis, and reporting. BMJ. 2015: 350: h391.

8. Weston VC, Meurer WJ, Frederiksen, SM, Fox AK, Scott PA. Prevention of emergency physician migratory contamination in a cluster randomized trial to increase tissue plasminogen activator use in stroke (the INSTINCT trial). American Journal of Emergency Medicine. 2014: 32(12): 1460-63.

9. Donner A and Klar N. Statistical considerations in the design and analysis of community intervention trials. Journal of Clinical Epidemiology; 1996: 49(4): 435–9.

10. Rutterford C, Copas A, Eldridge S. Methods for sample size determination in cluster randomized trials. International Journal of Epidemiology; 2015: 44(3): 1051–67.

11. Campbell MK, Piaggio G, Elbourne DR, Altman DG. CONSORT 2010 statement: Extension to cluster randomized trials. BMJ; 2012: 345: e5661.

12. Hayes R, Moulton L. Cluster randomized trials (2nd edition). Boca Raton, Florida, USA: Chapman and Hall/CRC Press; 2017.

13. Weijer C, Grimshaw JM, Eccles MP, McRae AD, White A, Brehaut JC, et al. The Ottawa Statement on the Ethical Design and Conduct of Cluster Randomized Trials. PLoS Medicine; 2012: 9(11): e1001346.

4.4 Collection and management of good quality data

Authors

Fernando Gouvea-Reis, Public Health England, London, United Kingdom.

Marcelo Farah Dell'Aringa, CRIMEDIM–Research Center in Emergency and Disaster Medicine, Università del Piemonte Orientale, Novara, Italy.

Virginia Murray, Public Health England, London, United Kingdom.

4.4.1 Learning objectives

To understand key aspects of data collection for research in health emergency and disaster risk management (Health EDRM), including:

1. Different sources and methods for data collection, along with their advantages and limitations.
2. Challenges involved in collecting data in disaster settings, and how these might be overcome.
3. The importance of data quality, data storage and data sharing.

4.4.2 Introduction

The timely collection of good quality data on key aspects relevant to disaster risk management, including emergency response is critical to Health EDRM research, as research outcomes are dependent on data quality and outputs. High quality research and data are invaluable to enable:

- Planners and responders to implement Health EDRM for effective and efficient action in the areas where their work is most needed.
- Policymakers to influence evidence informed best policy and practice in Health EDRM.

Good quality research requires data that are relevant to the research question and objectives, which may include demography, morbidity, mortality, infrastructure, different health factors, environmental characteristics, and so on. Such data are needed to manage disaster risk so that future disasters can be avoided or their impact minimized. It also supports the planning, management, and evaluation of post-disaster interventions. Poor quality data will lead to poor quality research and, potentially, to misinformed policies. Therefore, it is key to ensure the collection of high-quality data during any study.

This chapter discusses important aspects that should be considered before, during and after the process of data collection in order to ensure that good quality data are used and available in disaster research. It explores the planning and preparation processes, different methods for data collection, as well as the challenges that a researcher may face when studying disasters and tools that might help them to address these challenges. Finally, it will discuss how to ensure good quality data are stored and made accessible to others so that it can bring additional benefits.

4.4.3 Preparation

Successful data collection demands careful preparation. It is important to formulate a clear and specific research question or hypothesis to be tested, and then to plan what specific data and what collection strategy will provide adequate and sufficient information to answer that question or allow the hypothesis to be accepted or rejected. Although it can be tempting to adapt the data collection or methods of an ongoing study to collect additional data to test another hypothesis, without proper reflection and planning, this can result in the presence of confounding factors in the collected data, leading to biased results. Alternatively, it can also compromise the statistical power of the results. Having a clear research question and aim at the start of the planning process can help to avoid such issues (Chapter 3.5).

It is also important to have a clear, written protocol before data collection begins, and this may be needed when seeking ethics approval (Chapter 6.4). This includes the research question, aim and objectives, definitions of exposure, outcome, and other terms, the expected sample size, the methods to be used, how participants will be recruited and how the data will be curated and analysed after collection. Furthermore, agreement on clear hazard definitions is key to allow comparability on data collected from different sources. Conducting a literature or scoping review before you write your protocol is an important tool to understand how other researchers studying disasters and disaster risk management have collected data on similar contexts, or how they answered similar questions (Chapters 2.6 and 3.6). This can help in the understanding of what strategies work best, as well as with anticipating the main challenges as encountered by others, so that the researcher is prepared to deal with these should they appear during their study.

4.4.4 Data collection methods

Once a research question and the variables of interest have been defined, the next step is to determine how these parameters will be measured. Depending on factors such as the study design, funding, time and human resources available, the researcher may decide between collecting new data or studying data that have been previously collected by others. These different approaches are also known as primary and secondary data collection methods.

4.4.5 Primary data

Primary data are data collected for the first time and for the purpose of a specific study. The researchers conducting that study decide where, how and when the data will be collected to specifically address their research question. However, this approach can be expensive and time consuming, and may demand technical resources. Methods of primary data collection can be broadly divided into two approaches:

Quantitative methods are used for numerical data. They include analysis of the data using descriptive and comparative statistical techniques (see Chapters 4.2 and 4.5) to answer specific questions about, for example, how commonly something occurs, or differences between groups. In Health EDRM, this approach can be used to estimate morbidity and mortality. It can also be used in the construction of more complex models to estimate, for example, the economic impact of a flooding in an affected area (Chapters 4.6 and 4.7). Data collection methods in quantitative research can involve surveys (Chapter 3.1), the measurement of outcomes in experiments or observational studies (Section 2 and Chapter 4.1), and the use of routinely collected data from different monitoring systems (Chapter 2.4). It usually requires large sample sizes and appropriate sampling of the participants from whom the data will be collected, in order to ensure the desired generalizability of the results.

Qualitative methods, which are discussed in detail in Chapter 4.12, are most often used to study research questions about how and why phenomena occur, and use observed and recorded non-numerical data, such as words and images, to understand meaning. The collection of such data is usually performed through in-depth interviews, focus groups, key-informant interviews, and observations. Because statistical methods are not used for qualitative research, there is no predetermined sample size. A qualitative approach is particularly useful when the objective is to understand underlying reasons, opinions and motivations during exploratory research, or to develop a theory. For example, it can be employed during a study that aims to understand the drivers of behaviour change related to the implementation of safe burial practices during an Ebola outbreak. It is also useful in the development of hypothesis to be tested in later quantitative studies.

4.4.6 Secondary data

Secondary data comprises data already collected or produced by others. Common sources of secondary data are government databases and publications, books, scientific papers, media channels and routine data. Routine data are data collected in a periodic, systematic manner by the government or other organizations (Chapter 2.4) and include:

- **Demographic data**, describing variables such as age, sex, ethnicity, migration patterns, marital status, and so on.
- **Health event data**, describing health variables that affect individuals or populations, including births, deaths, and population interaction with the health sector at different levels.
- **Circumstantial data**, describing factors associated with the social determinants of health, including data on education, employment, housing and environmental data.

- **National reference data**, which covers data that has not been issued purely for health purposes, but when integrated and combined with other variables can be useful in the understanding of different health issues.

Using secondary data means the researchers do not have full control over data quality, making it more difficult to ensure that the dataset they use is complete, unbiased, time accurate, and reliable. Table 4.4.1 highlights important key points on data quality that must be considered when using secondary data.

Table 4.4.1 Important considerations for the use of routine data *(1)*

Accuracy: to what extent is the dataset accurate? What are the potential biases?
Precision: Have appropriate measures of uncertainty been included (such as 95% confidence intervals)?
Completeness: how much of the data is missing?
Timeliness: were the data collected in a period that is relevant to the study?
Coverage: is the whole population of interest covered? If not, how does this impact the study?
Accessibility: who has access to the data, and how is this access controlled?
Confidentiality: have individual-level data been anonymized?
Original purpose of collection: can the data be used for a different purpose to the one for which it was collected? Who collected the data and how?
Analysis: have the data been standardized and presented in a comparable way?

4.4.7 Dealing with challenges in disaster data collection

Researchers can anticipate facing different challenges during data collection. Some examples are:

- limited access to certain areas due to infrastructural collapse (destruction of roads and other transportation systems, for example).
- Persistence of the hazard that originated the disaster, which might pose a risk for the research team (radiation after nuclear incidents, for example).
- emergence of infectious diseases outbreaks due to damaged or poorly functioning water and sanitation infrastructure, which can become a threat to the local community and researchers (cholera epidemics after floods, for example).
- political barriers (local authorities attempt to minimize or change disaster-related statistics, such as mortality estimates, or refuse access to the planned research site, for example).
- language barriers, when the researchers do not speak the local language, leading to the possibility of bias in the use of translators.

Case Study 4.4.1 illustrates how researchers in the field can face some of these barriers. The early consideration of the challenges that are most likely to be encountered can help choosing the most appropriate data collection strategy.

Case Study 4.4.1
Challenges in disaster data collection after the 2004 Indian Ocean Earthquake and Tsunami *(2)*

The 2004 earthquake and tsunami that occurred in the Indian Ocean affected 12 countries and left almost 230 000 people dead and approximately 1.7 million people displaced *(3)*. In the post-disaster environment, different groups conducted research aiming to understand how the event affected factors such as the health status of the local communities and their health needs. These groups faced various challenges in data collection.

For example, a study was conducted to determine the public health impact of the tsunami on the population of three communities in Aceh Jaya District, Republic of Indonesia. However, all health facilities in the three communities were destroyed during the tsunami, and the only health professionals to survive the disaster were two midwives. As a result, much of the data had to be obtained from secondary sources, such as reports from local authorities, and the results of the study were thus susceptible to recall, reporting and misclassification biases *(4)*. Another study found that poor health record keeping in facilities prior to the tsunami limited the comparative effectiveness of the health data collected after the tsunami. This led to issues in determining which health-related issues were the result of the disaster and which reflected pre-existing problems *(5)*.

In another study, the French Army medical service carried out an epidemiological survey to estimate health indicators in children during the weeks following the tsunami in Meulaboh. They reported issues with communication and translation during interviews, where sometimes it was difficult to communicate directly with the children or their parents, leading to errors of interpretation. Furthermore, the researchers also faced barriers related to the transportation of the data collection teams among the disaster settings *(6)*.

There are different approaches that can support researchers in gathering good quality data and overcoming the challenges involved in data collection for disaster research. The use of routine data, for example, is a useful tool in contexts where time and resources are constrained (Chapter 2.4). For example, using secondary, routine data can rapidly provide the necessary information to compare before and after disaster scenarios, demonstrate change in demand for specific healthcare services, and to evaluate its impact on local health systems, as demonstrated by Case Study 4.4.2.

4.4

Case Study 4.4.2
An ecologic study to evaluate the impact of the 2011 Rio de Janeiro landslides in the utilization of public mental health services *(7)*

Many areas of the south and south-eastern regions of Brazil are hit frequently by heavy rains during the summer months. These regions have some of the places with the highest population density in the country and many people living in disaster-prone areas. This leads to important vulnerabilities and thus many communities are under extensive disaster risk of landslides and floods. The 2011 landslides in the mountainous region of Rio de Janeiro State were the largest disaster by immediate death count in recent Brazilian history, with a report counting 845 immediate deaths, mostly by mud burial. Moreover, around 30 000 people were left homeless in 11 different municipalities and there was important damage to agricultural and industrial activities.

An ecologic study was performed using routine data from DATASUS (Departamento de Informática do SUS - Informatics Department of the Brazilian Public Health System in free translation). DATASUS comprises a wide range of open access data, and allows researchers to gather and analyse datasets regarding health outcomes, the incidence of diseases and on the utilization of the health services in different levels.

The study analysed data from the affected region of Rio de Janeiro state two years before and after the event and comparing it with unaffected regions of the state. The analysis of the data suggested a sustained increase in the search for mental health services by the affected population after the landslides, which was not found in the other regions of the state.

The use of routine data can also be helpful in the construction of models to leverage disaster risk reduction strategies. Case Study 4.4.3 presents an example where this approach was used to better prevent and respond to infectious diseases outbreaks.

Case Study 4.4.3
The combination of cholera outbreak data and satellite environmental information to estimate cholera risk *(8)*

Cholera is an infectious disease caused by the ingestion of contaminated water or food with the bacteria *Vibrio cholerae*. Water-related diarrheal diseases like cholera are estimated to kill approximately 1.5 million people every year. They are the second leading cause of death in children under five years old. The impact of cholera is higher in settings with poor availability of clean water, as well as places susceptible to floods and with heavy rainy seasons.

Scientists combined in an algorithm data related to the time and location of previous cholera outbreaks in sub-Saharan Africa with different satellite datasets, including precipitation, air temperature, and land surface temperature. The algorithm was tested in five cholera epidemic regions of Sub-Saharan Africa (Mozambique, Central African Republic, Republic of Cameroon, South Sudan, and Republic of Rwanda), and was able to identify and predict regions most at risk for an outbreak at least four weeks in advance *(8)*.

In the Republic of Yemen, this model has been used to predict where and when the next increase in cases of cholera will happen. When risk areas are identified, local partners can work in managing disaster risk by directing emergency resources to the most critical areas, improving infrastructure where needed, chlorinating water and running educational and vaccination campaigns *(9–10)*.

To build a complete picture related to the hazard or disaster of interest, information from several data sources are likely to be needed. It is also important to note that, in different countries and contexts, the data of interest may be collected and curated by different organizations, which can include the Ministry of Health, National Statistics Offices, or even be fragmented through different levels of regional and local health departments (Chapter 2.4). This can result in extra time and resources needed to collect and standardise data provided by different sources.

However, in settings where local data collection for relevant parameters is poor or absent, the use of secondary data might be constrained. Depending on the availability of time and resources, you might choose to perform the primary data collection yourself using protocols with relevant ethical consent (Chapters 3.4 and 6.4). If this is not suitable to your context, the development of models can also be considered as an alternative strategy to fill the information gaps (Chapter 4.6). This can be an important opportunity to raise awareness among local governments, universities and independent organizations about the importance of initiating and maintaining good routine data collection and how this might help them prevent and respond to disasters.

4.4

4.4.8 Different approaches in data collection

There are a growing number of useful tools to support disaster research, and big data can be leveraged to provide important information in a variety of contexts. Big data includes data such as satellite imagery, images and videos from unmanned aerial vehicles (UAVs), sensor web and Internet of Things (IoT), airborne and terrestrial Light Detection and Ranging (LiDAR), simulation, crowdsourced information, social media, and mobile global positioning system (GPS) and Call Data Records (CDR) *(11).*

For example, the management of disaster risk can be supported through images and videos captured by satellites or UAVs to develop hazards maps and risk assessments. Similarly, the assessment of post-disaster damage through change detection, for instance, provides enhanced situational awareness, supporting and guiding action from rescue teams. It may also be possible to use crowdsourcing to gather these types of data (Chapter 5.2).

4.4.9 Data storage and data sharing

When the data has been collected and cleaned, the next step is to store it securely for current and future analysis, and to consider how it might be shared so that others can also benefit from it.

According to the type of research study, it is possible that data will be collected from multiple sources. Therefore, the design of a curation system should account for such differences and allow standardization. This can be achieved by a computerized database with clear rules for data entry. This involves facilitating the user role by requiring only the needed information to be added. For example, for discretionary variables, the adoption of drop down lists to be selected by the user instead of empty spaces for free text can help reducing entry errors and ensure standardization. Similarly, the implementation of rules such as limiting the valid range for variable fields and flagging errors if information is not adequately entered in a core field exemplify how the adoption of simple, good practices, help the achievement of a complete and accurate dataset *(12)*.

It is also important to consider that the usefulness of a dataset to others can be enhanced by providing data as disaggregated as possible, but while still safeguarding individual privacy. A simple example to understand this principle is when reporting on residents who have been affected by a local flood, a dataset which can be filtered according to sex, age, socioeconomical factors, health status and disability allows a much broader set of analysis to be made, such as developing hypothesis on the correlation of the outcome with possible risk factors. The more disaggregated a dataset can be to the individual level, the more invisible persons can be made visible. It can then be used as reliable evidence to inform policymaking, for example helping to direct resources to those affected who need it the most.

There is currently a widespread call across research for making data open and transparent, improving its usefulness so that others can also benefit from it. The ‘data revolution’ comprises the large increase in the volume and types of data that are currently collected by governments, private companies, NGOs, researchers and citizens. This is leading to an

unprecedented possibility of transforming such data into knowledge to not only manage disaster risk but also to better respond to disasters *(13)*. However, important data are often not released rapidly, or not shared at all, which compromises the potential re-usability of many datasets. The FAIR principles of data sharing were developed to assist in the production of good-quality data, with practical actions that can be adopted to increase findability, accessibility, interoperability and reusability of datasets *(14)*.

Examples of actions that can improve data quality and interoperability include the use of clear standards and definitions, as well as the use of data dictionaries to describe the variables and values present in a given dataset. A challenge faced by Health EDRM researchers is the great variety of hazards and the lack of agreed definitions on them. Different definitions for a given hazard hampers the comparability of results from different studies, for example. As a result, it is important to have clear case and hazards definitions when conducting research in emergencies and disasters, and to present data in a machine-readable format, so that it can be retrieved and processed by computers.

4.4.10 Conclusions

Overall, data collection in the context of disasters is a challenging task that demands careful preparation and planning. Different methods can be used to gather data, and the local context, time and resources available should be considered in selecting the most suitable approach for a specific study. Science-based policy making depends on high quality research, which in turn is dependent on high quality data. Therefore, it is important to ensure that data are collected, stored and shared at high standards. A careful preparation is essential to achieve this, including the construction of a research protocol containing a clear and specific research question, objectives, the strategy to be used during data collection and how the data will be curated and analysed at a later stage.

4.4.11 Key messages

- **A specific research question and a data collection strategy that will provide adequate and sufficient information to answer this with the available resources are important for high quality research.**
- **It is fundamental to acknowledge that despite good preparation, challenges may occur. Anticipating how to deal with them can help researchers to overcome future barriers.**
- **A careful plan on how the collected data will be stored and shared in the long term will ensure that others benefit from the study.**

4.4

4.4.12 Further reading

Fakhruddin B, Murray V, Gouvea-Reis F. Disaster loss data in monitoring the implementation of the Sendai Framework [Policy brief]; 2019. https://council.science/publications/disaster-loss-data-in-monitoring-the-implementation-of-the-sendai-framework (accessed 18 January 2020).

4.4.13 References

1. Goodyear M, Malhotra N. Collection of routine and ad hoc data. 2007. https://www.healthknowledge.org.uk/public-health-textbook/health-information/3a-populations/collection-routine-data (accessed 18 January 2020).
2. Morton M, Levy JL. Challenges in disaster data collection during recent disasters. Prehospital and Disaster Medicine; 2011: 26(3): 196-201.
3. Inderfurth AK, Fabrycky D, Cohen S. The 2004 Indian Ocean Tsunami: One Year Report. The Sigur Center Asia Papers. 2005.
4. Brennan RJ, Rimba K. Rapid health assessment in Aceh Jaya District, Indonesia, following the December 26 tsunami. Emergency Medicine Australasia; 2005: 17(4): 341-50.
5. Centers for Disease Control and Prevention. Assessment of health-related needs after tsunami and earthquake--three districts, Aceh Province, Indonesia, July-August 2005. MMWR: Morbidity and mortality weekly report; 2006: 55(4): 93-97.
6. Meynard JB, Nau A, Halbert E, Todesco A (2008). Health indicators in children from Meulaboh, Indonesia, following the tsunami of December 26, 2004. Military Medicine 173(9): 900-5.
7. Dell'Aringa M, Ranzani O, Bierens J, Murray V. Rio's Mountainous Region ("Região Serrana") 2011 Landslides: Impact on Public Mental Health System. PLoS Currents: Disasters; 2018: 10.
8. Khan R, Aldaach H, McDonald C, Alam M, Huq A, Gao Y, et al. Estimating cholera risk from an exploratory analysis of its association with satellite-derived land surface temperatures. International Journal of Remote Sensing; 2019: 40(13): 4898-909.
9. Department for International Development. World first as UK aid brings together experts to predict where cholera will strike next; 2018. https://www.gov.uk/government/news/world-first-as-uk-aid-brings-together-experts-to-predict-where-cholera-will-strike-next (accessed 18 January 2020).
10. National Aeronautics and Space Administration (NASA). NASA Investment in Cholera Forecasts Helps Save Lives in Yemen. 2018. https://www.nasa.gov/press-release/nasa-investment-in-cholera-forecasts-helps-save-lives-in-yemen (accessed 18 January 2020).
11. Yu M, Yang C, Li Y. Big data in natural disaster management: a review. Geosciences; 2018: 8(5): 165.

12. Busby A. Data Quality. In: Kreis IA, Busby A, Leonardi G, Meara J, Murray V, editors. Essentials of environmental epidemiology for health protection: a handbook for field professionals. Oxford University Press: Oxford; 2012.

13. A World that Counts – Mobilising the Data Revolution for Sustainable Development. New York: United Nations; 2014, Nov. http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf (accessed 18 January 2020).

14. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data; 2016: 3: 160018.

4.5 Advanced statistical techniques

Authors

Marcella Vigneri, Centre of Excellence in Development Impact and Learning, London School of Hygiene and Tropical Medicine, London, United Kingdom.

Howard White, Campbell Collaboration, New Delhi, India; and Centre of Excellence in Development Impact and Learning, London School of Hygiene and Tropical Medicine, London, United Kingdom.

4.5.1 Learning objectives

To understand the following more advanced factors to consider in developing an impact evaluation for health emergency and disaster risk management (Health EDRM):

1. Different approaches to estimating impact in the absence of random assignment.
2. Advantages and disadvantages of these different approaches.
3. Importance of baseline data for both intervention and comparison groups.

4.5.2 Introduction

Random assignment usually provides the most robust method for comparing the effectiveness of interventions (Chapter 4.1). However, it may not be possible in some settings related to Health EDRM. For example, the implementing agency might not be willing to accept randomization, or the impact evaluation may have to be designed after an intervention is already underway or even completed. When randomization is not possible, impact can still be estimated through a range of non-experimental techniques, which may be broadly divided into two categories: quasi-experimental methods (see also Chapters 4.14 and 4.15) and regression-based approaches.

Quasi-experimental (QE) methods identify a comparison group using statistical matching, such as propensity score matching and coarsened exact matching. Matching is also used to increase the power of designs such as difference in differences, which are explained below. Matching ensures that the comparison group is as similar to the intervention group as possible, such that the average characteristics (age, location and education, for example) of the intervention and control groups are similar at baseline (that is, pre-intervention). Impact is then calculated as either the difference in outcomes after the intervention (ex-post single difference) or the difference in the change in outcomes between baseline and endline (difference-in-differences).

Regression-based approaches include instrumental variables, Heckman sample selection models, endogenous switching regressions and fixed effects models. These approaches require the use of data in untreated or less treated units. Endogenous switching models and Heckman selection models are not covered in this chapter, and information on them is available elsewhere *(1)*. Regression based approaches are usually the only option if the intervention is measured as a continuous indicator (for example changes in the amount of exposure to the intervention), rather than as a binary indicator (that is, the intervention is either provided or not provided).

Non-experimental approaches are best based on specifying the underlying structural model, that is the set of behavioural relationships which lead to intervention impact (see Chapter 4.10). Applying non-experimental approaches requires data from both an intervention and a comparison population. Moreover, more reliable impact estimates are usually possible if baseline data are available that provide variables for matching that are unaffected by the intervention, since such data were collected before the intervention took place.

This chapter introduces three common matching techniques: propensity score matching, regression discontinuity and interrupted time series, as well as one regression-based approach: instrumental variable estimation. First, the following section explains how impact can be estimated using differencing.

4.5.3 Double difference estimates

When the intervention has taken place, impact can be estimated by single or double difference. Table 4.5.1 shows the different stages of an intervention (top row) and the data that are required to apply these approaches.

Table 4.5.1 Timing of intervention and surveys for large impact evaluations

Start of intervention	During intervention	At end of intervention	After intervention
B: Baseline	M: Mid-term	E: Endline	P: Post-endline

Description

Ex-post single difference impact estimators are calculated as the difference between the outcome indicator after the intervention (that is, at endline, time E) in the intervention group and the outcome indicator in the comparison group which did not receive the intervention. The double difference impact estimate is the difference in the change in the outcome indicator for the intervention and for the comparison groups between baseline and endline, rather than the difference in their endline values, as is the case for the single difference. Double differencing removes any difference in the indicator between intervention and comparison groups that was present at baseline. This is useful because these baseline differences cannot be a result of the intervention. If the values of the outcome indicators for the intervention and the comparison groups are the same at baseline, then the single and double difference estimates are equivalent.

Double differencing is a means of calculating the estimated impact. It is also used as an impact evaluation method. Double difference estimates require baseline data that should be collected immediately prior to the intervention. The validity of this approach relies on the 'parallel trends assumption', that is, the trend in the outcome in intervention and comparison populations should be the same without the intervention. The parallel trends assumption can be tested *(2)* if trend data from before the intervention are available, but unfortunately this is often not the case. Acquiring more data points (observations) before and after the intervention allows a visual inspection of whether the parallel trend assumption holds. If the assumption can be tested and does not hold, then using double differencing without matching cannot be expected to be free of bias. Matching can help to control for observable determinants of differences in changes over time and make the analysis less dependent on this assumption. Implementation of the method requires data on outcomes from the intervention and comparison groups at baseline and endline. If matching is to be used, then data for matching are also required.

Advantages and disadvantages of double differencing

Double differencing is easy to implement and easy to understand. However, pre-intervention trend data may not be available to test its validity. Hence, it is more rigorous when used with a matching technique.

4.5.4 Propensity score matching

Propensity score matching (PSM) creates a comparison group from observations on a population that did not receive the intervention by matching intervention observations to one or more observations from the sample without the intervention, based on observable characteristics. Matching is based on the propensity score, which is the estimated probability of being in the intervention group given the observable characteristics. The propensity score is estimated using a regression model of participation (taking part in the intervention). Propensity score matching cannot incorporate selection on unobservables, so may give biased estimates if these are important. Additional information is available elsewhere *(3–5)*.

Description

Perfect matching would require matching each individual or unit in the intervention group with a person or unit in the comparison group that is identical on all relevant observable characteristics (for example, age, education, religion, occupation, wealth, attitudes to risk and so on). Clearly, this is not possible nor is it necessary. 'Balance' between intervention and comparison group units (which is necessary for unbiased estimates) requires that the average characteristics of the intervention and comparison groups are the same before the intervention. A good example on the methods used for variable selection in PSM is provided by Brookhart and colleagues *(6)*.

In PSM, matching is not achieved on every single characteristic but on a single number: the propensity score. This is the likelihood of a person taking part in the intervention given their observable characteristics. This probability is obtained from the 'participation equation': a probit or logit regression in which the dependent variable is dichotomous, taking the

value of 1 for those who took part in the intervention and 0 for those who did not. The right-hand side of the equation includes all observed variables (individual, household or firm and community or market) that may affect participation, but that are not affected by the intervention. Baseline values of all variables, including outcomes, cannot be affected by the intervention, so having baseline data helps to obtain a stronger match.

Observations outside the 'region of common support' are discarded before matching. The region of common support is the area of overlapping propensity scores. Therefore, those observations with very low scores (which typically come from the comparison group) or very high scores (typically from the intervention group) are discarded. The observations retained from those who did not receive the intervention are used as the comparison group, which ensures that the comparison is 'like with like'.

Each member of the intervention group is matched to one or more members of the comparison group. This is done through a variety of matching algorithms such as the nearest neighbour matching, caliper matching and kernel matching. An example is the study by Boscarino and colleagues *(7)* which uses PSM to estimate the impact of mental health interventions received by employees at the worksite after the World Trade Center attacks among workers in New York City. The authors used data from telephone interviewees with adults in a household survey conducted one and two years after 9/11 to match intervention cases to non-intervention control cases based on a bias-corrected nearest-neighbour algorithm. Their findings from matching with PSM suggest that about 7% of approximately 425 000 adults reported positive outcomes (such as reduced alcohol dependence, binge drinking, depression, severity of post-traumatic stress disorder and anxiety symptoms) resulting from receiving employer-sponsored, worksite crisis interventions related to the attacks.

In PSM, those members of the comparison group that do not match those in the intervention group are discarded. Once matching is completed, a balancing test is performed to ensure there is no statistically significant difference between the mean characteristics of the matched intervention and comparison groups. Finally, the impact is estimated by calculating the difference between the outcome indicator of interest for the intervention units and the average value for the matched comparison individuals, and then averaging over all these differences. Another interesting application of PSM is the study by Gomez and colleagues *(8)* which exploits data collected as part of a large-scale evaluation of an early childhood education intervention related to earthquakes in Santiago, Chile. The data included 4-year old children who had experienced, and who had did not experienced, the severe earthquake episodes of 2010. These children were then matched through PSM to find that the earthquake affected lower scores on some early language and pre-literacy assessments of children that had experienced the earthquake. A further example is provided as Case Study 4.5.1, which assessed the impact of humanitarian aid on food security in the Republic of Mali.

There are several statistical packages (such as Stata and R) that allow to implement PSM analysis through pre-built commands.

4.5

Advantages and disadvantages of propensity score matching

The two main advantages of PSM are that it easily lends itself to establish the propensity score of being treated through a binary model, and that it can be done ex post, including in the absence of baseline data. If baseline data are not available, matching uses time invariant characteristics (such as sex and religion) and recall information on pre-intervention characteristics that can be reliably recollected. These features suggest the greater flexibility of the PSM model to accommodate many covariates.

Case study 4.5.1
Using PSM to measure the impact of humanitarian aid on the food security of rural populations in Mali *(9)*

PSM was used to measure the impact of humanitarian aid on the food security of rural populations in the Mopti region of Northern Mali.

The evaluation exploited data from a unique pre-crisis baseline in the region to use matched difference-in-difference methods to estimate whether access to different forms of food assistance improved household food expenditures, food and nutrient consumption, and the long-term nutritional status of children. The existence of baseline data enabled the matching of 'intervention' households with comparable 'comparison' households.

The measures used for matching were all pre-intervention (and so unaffected by it) and relate to both the selection into intervention and the outcome of interest (household expenditures, food consumption and a proxy for child nutritional status). The matching variables were both village-level measures (the presence of a secondary school within 5 km and the presence of a market within 5 km) and household-level measures (including whether children were involved in past projects, feelings of safety and age of the household head).

The impact evaluation found that food assistance increased household non-food and food expenditures and micronutrient availability.

A disadvantage of PSM is that it relies upon matching on observables. If selection (participation) into the intervention is affected by unobservables, PSM will yield biased impact estimates for ex-post single difference estimates. When panel data are available, PSM is biased if the unobservables are time varying or affect differences over time. However, time invariant observable factors can be removed by double differencing, so that PSM would again be unbiased.

4.5.5 Regression Discontinuity Design and Interrupted Time Series

Regression discontinuity designs (RDD) are used when there is a threshold rule for allocation to the intervention (such as administration of a drug if patient has a heartrate or temperature above a specific value, or the poverty line, or villages on either side of an administrative boundary). The assumption, which is tested as part of the procedure, is that units in proximity to either side of the boundary are sufficiently similar for those excluded from the intervention for these to be a valid comparison group. The difference in outcomes between those near either side of the boundary, as measured by the discontinuity in the regression line at that point, is attributable to the intervention, and so is the measure of the intervention's impact.

Interrupted time series (ITS) is a specific application of RDD in which the threshold is the point in time at which the intervention came into effect. This can be a particularly relevant method where intervention effectiveness is sudden, rather than gradual, such as the completion of a bridge or major power transmission connection, or the sudden availability of relief services.

Description

RDD can be used when there is a threshold rule that determines eligibility for the intervention, where the threshold is based on a continuous variable assessed for all potentially eligible units of assignment (such as individuals, households or communities). For example, households above or below the poverty line, children born before or after the cutoff date for school enrolment in a specific academic year, or students above a certain test score are awarded a scholarship. If the threshold is imperfectly applied, a variation on the approach, called 'fuzzy RDD', can be used.

The threshold variable must not be one which can be manipulated to become eligible for the intervention, as that might lead to selection bias. As an example, an impact evaluation of the Tropical Cyclone Winston social protection top up transfers was conducted by the World Bank in 2016 *(10)*. The goal of the intervention was to provide additional assistance in the form of top-up transfers to the most vulnerable, as a key component of its disaster response, and the intervention and control groups were constructed based on the Poverty Benefit Scheme (PBS) eligibility (poverty score) threshold. The treatment group was formed from PBS recipient households (20% below threshold) in affected areas in the Republic of Fiji that would also receive the intervention (top-up PBS benefit) after the cyclone. The control group was formed from the PBS-evaluated (before the cyclone) households in affected areas that were not eligible for PBS, as they were above (but within 20%) the threshold. The disaster responsive social protection intervention, in the form of top-up transfers to beneficiaries, was found to be an effective response following the cyclone.

In ITS, the threshold is the point in time at which the intervention or policy was introduced. In the case of a policy, this point in time is common to all households but other interventions (such as electrification or connection to a sewage disposal system) may affect different communities at different points in time. The threshold should be unique to the intervention. Clearly, those on either side of the threshold have some differences. In addition, the threshold criteria may be correlated with the outcome, so that there is

selection bias if simple comparisons are made. For example, scholarships are awarded to improve learning outcomes, but those with better learning outcomes are given the scholarships. Older women are more likely to get breast cancer, and it is older women who are selected for screening for this cancer. However, those near either side of the threshold are also much more similar. Regression discontinuity is based on a comparison of the difference in average outcomes for these two groups.

Another interesting application of this method comes from the study of Mezuk and colleagues *(11)* who used the September 11 2001 attack as the discontinuity (cut-off) point to investigate its impact on the average monthly suicide rate in New York City. Using average monthly suicide rates data between 1990 and 2006, the study found no net change in suicides rates just before and immediately after the attacks, suggesting that factors other than exposure to that particular traumatic event may have been driving the risk of suicide in the population studied.

An iterative approach is used to determine the margin around the eligibility threshold. Initially, one sets a small margin and checks for balance of the resulting intervention and comparison group units. If the match is good, the margin may be widened a little and balance checked again. This can be repeated until the samples start to become dissimilar (that is, there is no longer balance between the two groups). When the sample is established, a regression line is fitted to the sample around the threshold. The sample for the regression is restricted to observations just on either side of the threshold. Specifically, the outcome indicator is regressed on the selection variable (such as test scores and an intercept dummy). The intercept dummy is a dichotomous variable, taking the value 0 for observations below the threshold and 1 at the threshold and above it.

Advantages and disadvantages of RDD

RDD controls unobservables better than other quasi-experimental matching methods. It can also often use administrative data, thus reducing the need for data collection (see Chapters 2.4 and 4.4). The main limitation of RDD is that it is usually valid only for observations relatively close to the discontinuity point. Hence, a challenge for RDD is often to find a sufficiently large sample of observations on either side of the threshold. Further, the impact is being estimated only for the population close to the threshold. The estimate is what is called a local area treatment effect (LATE), rather than an average effect for the whole population in the intervention group. In principle, this limitation restricts the external validity of the approach.

Case Study 4.5.2 provides an example of how RDD was used to measure the impact of a winter cash assistance programme for Syrian refugees in Lebanon.

Case Study 4.5.2
Using RDD to measure the impact of a winter cash assistance programme to Syrian refugees in Lebanon *(12)*

The evaluation assessed the impact of cash on household well-being among Syrian refugees in Lebanon and whether cash might attract refugees to regions with assistance. The RDD design exploited the targeting approach of the cash assistance programme itself. Cash was given at high altitudes to target assistance for those living in the coldest areas during the winter months (households did not know beforehand that there would be an altitude eligibility cutoff). When the eligibility cutoff was set at 500 meters, households residing at 501 meters and above (intervention group) were included, while households residing at 499 meters or below (comparison group) were excluded. Intervention and comparison groups had very similar characteristics before the start of the programme, so differences measured after the programme's implementation represent the causal impact of cash assistance.

The impact evaluation found that the current value of cash assistance was inadequate because beneficiaries' income was so low that they were forced to use the cash assistance to satisfy other basic needs, in particular food. It also found that cash assistance increased access to school, reduced child labour and that the cash assistance programme had no pull factor on refugees settling in communities where cash was distributed.

4.5.6 Instrumental variables approach

The instrumental variable (IV) method is a regression-based estimation of the outcome variable of interest on either a project dummy or a measure of participation in the intervention group *(13)*.

In the conventional ordinary least squares (OLS) approach, the outcome is regressed on a dichotomous intervention dummy variable. The problem with this approach is that selection bias can affect the estimate of the impact coefficient. If selection is entirely based on observables, and the regression has included variables on all those observables, then OLS will indeed yield a valid impact estimate. However, if – as is more frequently the case – there are time varying unobservables, then cross sectional OLS models on differences will yield biased impact estimates. IV estimation is the technique used to remove the bias. It is an OLS regression in which the variable which is the source of the endogeneity problem is replaced by an instrument satisfying the following two conditions:

i. To be correlated with the probability of intervention (programme participation)
ii. To be uncorrelated with the outcome, except through its effect on the intervention.

When more than one instrumental variable is identified, the procedure is implemented as two-stage least squares: first one regresses the endogenous variable (the one measuring intervention participation) on the instruments and calculates its fitted value, then the outcome equation is estimated replacing the endogenous variable with the fitted values from the first stage. The estimated impact is the coefficient on the instrument. It is

important to have determined the instruments before data collection starts, so that the relevant questions are included in the survey instruments.

Advantages and disadvantages of IV

The advantage of IV is that if a valid instrument is found, both observable and unobservable sources of selection bias are controlled for. The main disadvantage of the method is that it may be difficult to find a valid and defendable instrument, because many factors that affect decisions to use an intervention typically also affect outcomes.

Case Study 4.5.3 provides an example of the use of IV to measure the political effects of environmental change.

Case Study 4.5.3
Using instrumental variables to measure the political effects of environmental change to understand the disaster–violence nexus *(14)*

In 2004, Sri Lanka was hit by a massive tsunami that killed more than 35 000 people and destroyed over 78 000 homes in that country alone. By May 2006, the Government of Sri Lanka had spent more than US$200 million on recovery, reconstructing at least 40 000 houses *(14)*. This study examined whether post-disaster reconstruction triggered further intrastate violence to explain civil unrest after the disaster.

The author addressed the endogeneity problem between reconstruction processes and violence (that is, that reconstruction is endogenous to violent events, but noted that there may be also a reverse causation if future violence limits current reconstruction efforts in disaster zones) by using the wave heights in the tsunami as an IV for post-war housing reconstruction.

The results suggest that an increase in housing construction is associated with the number of violent events, while the number of destroyed houses has no discernible impact on violence. Therefore, the paper plausibly concludes that reconstruction is a manipulable strategy that policy makers can use to respond to disasters through different post-disaster measures.

4.5.7 Conclusions

The chapter introduces some of the non-experimental quantitative methods that are available for impact evaluation studies in Health EDRM. These approaches are likely to be appropriate in establishing impact of interventions when random assignment is not be possible. Strengths and limitations of these approaches are illustrated with references to specific studies from disasters and other health emergencies. In general, best practice in planning a research study is to consider which approach is most appropriate and feasible at the design stage in order to prepare data collection tools and think of the best sampling strategy to get a good match. For example, PSM requires that data collection includes suitable matching variables and IV requires that data is available for one or more valid instruments. Oversampling will be necessary if observations will be discarded in establishing the regional of common support.

Moreover, where possible, it is best to use a combination of methods to ensure the most reliable and credible results on the impact of the intervention being assessed. For example, it is much better when possible to exploit baseline data for matching and using the difference-in-difference strategy. Similarly, if an assignment rule exists for the project, it would be ideal to match on this rule and subsequently do a regression discontinuity design.

4.5.8 Key messages

- **Impact estimates are possible in the absence of randomization, but still need data from a comparison group that did not receive the intervention.**
- **The available methods may be subject to selection bias.**
- **It is important to test for baseline balance to check if bias based on observables has been removed.**
- **The reliability of matching and the ability to calculate a double difference estimate are enhanced by the availability of baseline date for the intervention and comparison groups.**

4.5.9 Further reading

Allaire MC. Disaster loss and social media: Can online information increase flood resilience? Water Resources Research; 2016: 52(9): 7408-23.

White H, Sabarwal S. Quasi-experimental Design and Methods, Methodological Briefs: Impact Evaluation 8. Florence, Italy: UNICEF Office of Research. 2014.

4.5.10 References

1. White H, Raitzer D, editors. Impact Evaluation of Development Interventions: A Practical Guide. Asian Development Bank. 2017.
2. Angrist JD, Pischke JS. Mostly Harmless Econometrics. Princeton University Press. 2019.
3. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behavioral Research. 2011: 46(3): 399-424.
4. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology: 1974: 66(5) 688-701.
5. Rubin DB. Estimating causal effects from large data sets using propensity scores. Matched Sampling for Causal Effects. 2006.
6. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. American Journal of Epidemiology. 2006: 163(12): 1149-56.
7. Boscarino JA, Adams RE, Foa EB, Landrigan P. A propensity score analysis of brief worksite crisis interventions after the World Trade Center disaster: Implications for intervention and research. Medical Care. 2006: 44(5): 454-62.
8. Gomez CJ, Yoshikawa H. Earthquake effects: Estimating the relationship between exposure to the 2010 Chilean earthquake and preschool children's early cognitive and executive function skills. Early Childhood Research Quarterly. 2017: 38: 127-36.
9. Tranchant JP, Gelli A, Bliznashka L, Diallo AS, Sacko M, Assima A, et al. The impact of food assistance on food insecure populations during conflict: Evidence from a quasi-experiment in Mali. World Development. 2019: 119: 185-202.
10. Mansur A, Doyle J, Ivaschenko O. Cash Transfers for Disaster Response: Lessons from Tropical Cyclone Winston. SSRN Electronic Journal. 2018. https://ssrn.com/abstract=3143459 (accessed 29 February 2020).
11. Mezuk B, Larkin GL, Prescott MR, Tracy M, Vlahov D, Tardiff K, Galea S. The influence of a major disaster on suicide risk in the population. Journal of Traumatic Stress. 2009: 22(6): 481-8.
12. Masterson D, Lehmann C. Emergency Economies: the Impact of Cash Assistance Program in Lebanon. International Rescue Committee. 2014. https://www.rescue.org/sites/default/files/document/631/emergencyeconomiesevaluationreport-lebanon2014.pdf (accessed 29 February 2020).
13. Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. Journal of the American Statistical Association. 1996: 91(434): 444-55.
14. Kikuta K. Postdisaster Reconstruction as a Cause of Intrastate Violence: An Instrumental Variable Analysis with Application to the 2004 Tsunami in Sri Lanka. Journal of Conflict Resolution. 2019: 63(3): 760-85.

4.6 Health-related risk modelling

Authors

Holly C.Y. Lam and **Zhe Huang**, CCOUC, Faculty of Medicine, CUHK, Hong Kong SAR, China.

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

4.6.1 Learning objectives

To have a basic understanding of some modelling methods that might be applied in research studies relevant to the following issues for health emergency and disaster risk management (Health EDRM):

1. Short-term environmental health associations.
2. Factors associated with the uptake of protection behaviours.
3. Trends of influenza.
4. Health-related vulnerability index.

4.6.2 Introduction

Health EDRM is an important approach for reducing the numerous public health impacts of disasters and emergencies (Chapter 1.2). Other chapters in this book describe research methods that require the collection of new data in prospective studies; this chapter complements these by discussing the use of statistical modelling to establish mathematical associations between variables. The chapter focuses on health-related risk models that are applicable to Health EDRM and discusses models for four particular topics: short-term environmental health associations; factors associated with the uptake of protection behaviours; trends in influenza; and health-related vulnerability index.

4.6

4.6.3 Models for evaluating short-term environmental health associations

Hydrometeorological hazards (that is, hazards related to water and weather-related events) are common triggers of disasters and account for about 95% of the people affected by disasters caused by natural hazards in the past 50 years *(1)*. Climate change is causing these extreme events to become more common and severe, leading to further impacts on human health. Improved weather forecasting and better understanding of the health risks of extreme environmental events is allowing for the implementation of effective health protection plans and improvements in resource allocation. These are supported by modelling methods for evaluating short-term associations between environmental exposures and health outcomes, and this section uses ambient temperature as an example to illustrate this. Extreme temperatures are a silent killer, due to people's lower awareness compared to other hazards *(2)*, and have caused substantial public health problems *(3-7)*.

Similar to other environmental exposures (air-pollutants, storms, for example), ambient temperature usually has a short-term association with health outcomes, ranging from hours *(8)* to weeks *(9)*, depending on the degree of exposure and the health outcome considered. A delayed effect is commonly reported for the temperature-health association, but it is not always linear. For instance, since both extreme high and low temperature may cause adverse effects on human health, there may be a v-shaped association between ambient temperature and the risk of adverse health outcomes. Combined with a seasonal effect and some other confounding effects (such as air-pollutants and population-level demographic factors), the estimation of a temperature-health association is complicated. A time series design is the most common method to reveal these short-term temperature-health associations *(10)*.

Time series data are a series of sequential records in equal time units, such as the number of deaths and the average daily or weekly temperature within a specific time period. Bhaskaran and colleagues discussed and compared time series designs used in environmental epidemiology, identifying three main types of time series study: time stratified model, periodic functions and flexible spline functions *(11)*.

For the time stratified model, exposure and outcome are associated in stratified time units. Time intervals are indicated by indicator variables (such as time period: 1, 2, up to "n") instead of the true date record. This type of model is relatively easy to understand but many parameters are included in the model and it cannot facilitate the calculation of the continuous effect from one time unit to another *(11)*.

Periodic functions (Fourier terms) model exposure and outcome by using periodic functions such as sine and cosine function to represent the periodic characteristics (such as calendar months). This model type creates smooth predictions but the period of the pattern is fixed, and this might not be appropriate for representing trends that are mathematically complicated and do not have a fixed pattern *(11)*.

Flexible spline function is a modelling approach combining different polynomial curves *(11–12)*. This design is most commonly applied in assessing short-term associations between temperature and health

outcomes *(13–14)*. It allows the health outcome to be linked to a number of exposures with different non-linear associations at the same time. This is an important feature, because most temperature-outcome associations and long-term trends are non-linear and non-periodical. Another reason for using a flexible spline to model long-term trends is that it helps control the long-term demographic factors at a population level. For example, smoking is a potential confounder of the risk of admissions to hospitals for breathing problems when studying the association with temperature but, because the proportion of people in a population who smoke does not change significantly from day to day, it does not affect the daily association between temperature and these admissions. Therefore, overall changes in the proportion of smokers can be captured by fitting a spline function for the long-term trend.

To incorporate the non-linear delayed effects of ambient temperature on health outcomes into the spline model, Armstrong *(15)* and Gasparrini *(16)* introduced the Distributed Lagged Non-linear Model (DLNM) and the corresponding R package dlnm, respectively. This modelling approach is a three-dimensional data analysis. It considers the exposure, health outcome and the delay (time) dimension at the same time. In DLNM, spline functions can also be applied on the time dimensions, thereby addressing the need to model the non-linear delayed effect in exposure-outcome associations. The combination of flexible spline approach and DLNM tackles most of the concerns when evaluating short-term environmental health associations but is complicated because it involves one more dimension than other traditional time series designs. Similar to other time series approaches, the problem of effect modification by other factors (such as age and gender) still exists and needs to be handled separately (for example, by conducting subgroup analysis). More details about the method and some examples are available elsewhere *(17)*.

4.6.4 Identifying factors associated with the uptake of protective behaviours during extreme events

Applying appropriate protective behaviours during extreme events can lower exposure to hazards and hence reduce health risk. Sociodemographic factors *(19)* and knowledge, attitude and practice (KAP) *(20-21)* are common examples of determinants of health behaviours. Identifying associated sociodemographic factors and understanding KAP for a protective behaviour provides evidence to support health promotion policies. This section introduces a statistical method for identifying factors associated with the uptake of protective behaviours, using data collected from a cross-sectional KAP survey.

Logistic regression is a regression model characterized by one binary dependent variable (outcome) and multiple independent variables (explanatory variables) *(22)*. It allows users to investigate the association between the outcome variable and an explanatory variable with adjustment for other confounders. It is used widely for identifying factors (such as knowledge and gender) that might be associated with the likelihood of a group of people acting in a certain way (taking or not taking action, for example) and comparing this to a reference group of other people.

4.6

In Health EDRM, there are usually several explanatory variables to consider but including too many explanatory variables in the model compromises its power to reveal the real associations. A general guide is that there should be at least ten cases for each explanatory variable in each outcome group *(22)* and the power increases with increasing numbers of cases. To reduce the number of explanatory variables in a regression model, univariate analysis, such as the chi-square test (for categorical variables) and t-test (for continuous variables), can be used to provide a quick assessment of the potential associating factors. Explanatory variables showing potential association with the outcome in the univariate analysis, together with some core explanatory variables (supported by literature or hypothesis) are then entered into the logistic regression model. Model selection (the process of selecting explanatory variables for a model) can also be done by removing non-significant variables from a full model or adding variables and keeping those that are significant (see Case Study 4.6.1).

Case Study 4.6.1
Data collection by telephone survey

For a community with a high level of landline telephone penetration, data collection through a telephone survey might be an appropriate way to examine knowledge, attitude and practice (KAP) in community behaviour patterns. A population-based telephone survey among the Hong Kong population investigated their weather information acquisition pattern during an intense cold spell *(23)*. The Chi-square test and a logistic regression model were used to identify independent associated factors in a two-stage analysis. Univariate analyses were used to identify potential associated factors with the outcome and factors with a p-value from the chi-square test of less than 0.20 were entered to the second stage of the analysis, the multiple logistic regression analysis, to assess their independent association with the outcome. In the univariate analyses, educational attainment, age and marital status were significantly associated with current use of smartphone apps to acquire weather information. In multiple logistic regressions, only older age and lower education level remained significantly associated with lower smartphone app usage.

4.6.5 Prediction and forecasting of influenza trend

Influenza is a global public health burden, usually associated with cold-like symptoms but leading to serious illnesses in vulnerable groups (for example, young children and the elderly) *(24)*. Influenza causes health and economic burdens, with loss of work or school hours for patients and caretakers, large numbers of emergency room visits, hospitalizations and deaths *(25–27)*. Influenza viruses gradually mutate and when a new contagious strain emerges in a community without immunity, this may lead to an epidemic. To reduce the risk of disease outbreak and disease burden, accurate prediction of strain types and the number of cases is important for primary prevention strategies. Accurate prediction facilitates effective vaccine strain selection and resource planning for the healthcare sector, and various prediction models have been developed to meet different purposes and region-specific environmental conditions. This section

introduces predictive models for vaccine selection and the forecast of influenza activity *(28–30)*.

Vaccine selection is conducted annually, in general, and is a year-long process because of the long production time for the vaccines (approximately 6 to 8 months). The process is managed collaboratively between WHO and professionals around the world, supported by global surveillance data related to influenza virus circulating in humans *(29)*. Employing present and past data, predictive models are used to identify and predict emerging influenza clades (that is, groups of virus strains that are believed to comprise of evolutionary descendants of a common virus ancestor) that may be dominant in the following year. Most of the predictive models focus on the biological determinants of the evolution of influenza, with scale from molecular, within-host, population, regional to global level. Some models infer phenotypic properties of the current population *(29)*.

Antigenicity-stability fitness model *(31)*, Epitope Clade Growth *(32)* and Local Tree Shape *(33)* are probabilistic evolutionary focused models for predicting future viral populations *(29)*. Antigenicity-stability fitness model is a validated model estimating expected growth rate (fitness) of viral clades by input of a few years of genetic and antigenic data and is able to predict frequency of trajectory of clades for about one year ahead *(31)*. Epitope Clade Growth, a model based on genealogical tree, estimates antigenic differences by extrapolating recent growth hemagglutinin clades seeded by epitope mutation *(32)*. Local Tree Shape is another genealogical tree-based model. It estimates recent clade growth from information stored in the local shape of a hemagglutinin genealogical tree *(33)*.

Linking antigenic properties and genetic data, and identification of proposed vaccine strains are two ways of inferring phenotypic properties *(29)*. They estimate the effectiveness of current vaccines for the emerging influenza strains and identify new antigenic variants at an early stage of expansion *(29)*. Strain selection involves complex decisions that require the integration of the results from different models at different scales. Integration and interpretation of data for decisions are key challenges *(29)*.

Forecasts of influenza activity have been conducted worldwide to support preparedness activities *(28, 30)*. These forecasts can be based on single or multiple measures. Typical measures are peak periods (time), peak and outbreak magnitude and case counts by day or week *(30)*.

There are two main modelling approaches: (i) statistical models without consideration of the epidemiology process and (ii) epidemiological models *(28)*. The common statistical models are time series models, generalized linear models, Bayesian network and classification methods *(28)*. The susceptible-infections-removed (SIR) models and agent-based models (AMBs), which include exposure, infection, transmission and behaviours in the calculations, are the common epidemiological approaches for forecasting influenza activity *(28)*. Agent-based models can be operated by simulation algorithm to estimate key epidemiological parameters and then to forecast future activity (see Case Study 4.6.2). While time series models can capture the temporal dependence of health outcomes, epidemiological approaches are able to account for health-related human behaviours and address questions related to the impact of prevention measures on health. Dynamic virological data and syndromic influenza-like

illness are common input data for surveillance data forecast models *(28)*. Real-time forecast models, making use of retrospective forecast information have been developed for temperate regions, with seasonal winter epidemics such as the USA *(34–35)*. However, these real time models performed less well in subtropical regions, such as Hong Kong SAR, with a two peak or year-round pattern *(36)*.

Case Study 4.6.2
Forecast Model - Simulation Optimization (SIMOP)

Nsoesie and colleagues *(37)* introduced a **sim**ulation **op**timization (SIMOP) approach for forecasting influenza epidemic infection curves. This combines the individual-based epidemiology model and the optimization technique for model parameters estimation (Nelder-Mead simplex method). The three model parameters estimated were the disease transmissibility, incubation and infectious period distribution. The individual-based model consisted of a dynamic social contact network (representing Montgomery County in Virginia, Miami, Seattle and surrounding metropolitan regions of the USA) and a disease model with the several assumptions.

There were three main steps for the SIMOP: (i) initialize the individual-based model and the Nelder-Mead simplex method, (ii) run the Nelder-Mead algorithm to find new parameter sets, and (iii) simulate an epidemic using the proposed parameter set and evaluate the objective function. Steps 2 and 3 were repeated for convergency. The input measures were the sequential daily or weekly number of cases during the period of epidemic, which were simulated by the estimated disease transmissibility, incubation and infectious period distribution. The model was used to forecast the epidemic peak timing, counts of infected individuals and cumulative infected individuals.

The model predicted the peak time at seven weeks before the actual peak. Forecasting the peak count of infected and cumulative infected individual was more challenging because of the possibilities of the epidemic curve trajectories, but the forecast was found to be accurate for Montgomery County.

4.6.6 Compositing indicators/index to measure vulnerability

Climate change is set to increase the frequency and intensity of disasters due to natural hazards *(38)*. Risk assessment tools are important for saving lives and reducing losses in disasters. During disasters, the number of deaths, the number of people affected and economic loss are not only determined by the hazard itself, but also by the proportion of population exposed and the vulnerability of the community (Chapter 1.3). Understanding risk in all its dimensions is essential for effective Health EDRM, and as such, the collection of large volumes of data is a major focus of research and public interest, because it presents opportunities to describe reality accurately (Chapter 2.4). However, although large amounts of data provide information from many perspectives,there may be too many variables for a clear understanding. This problem is sometimes known as the “curse of dimensionality”.

If there are a large number of variables in a dataset, a dimension reduction method can be applied. This maps the numerous original variables into fewer independent dimensions, based on their correlation to each other. It is therefore more meaningful to summarize data as a few independent dimensions, while preserving as much of the original information as possible *(39)*.

On some occasions it is easier to interpret one composite index resulting from dimension reduction, rather than indicators from multiple perspectives, despite the simplification of the original data. A composite index can allow multi-country comparisons for complex issues, such as society development, vulnerability to environmental hazards and urban heat islands. A good quality composite index is based on careful variable selection and appropriate use of the dimension reduction method, and can facilitate communication and policy making.

Principal components analysis (PCA) and factor analysis (FA) are two examples of linear dimension reduction methods. They attempt to explain a multivariate dataset by reducing them into a smaller number of dimensions. PCA is one of the oldest multivariate techniques and is useful for displaying multivariate data as a set of dimensions (called 'principal components'). It simplifies the complexity by transforming correlated variables into a set of uncorrelated principal components *(40)*. Each principal component is rated according to the extent to which it represents the original dataset, and most of the information from the original variables is captured by the principal components rated the highest (see Case Study 4.6.3). In summary, PCA provides a concise summary of the original variables, with no probabilistic or statistical assumptions.

Case Study 4.6.3
Principal components analysis (PCA) to develop a Heat Vulnerability Index

PCA was used to combine socioeconomic indicators into a Heat Vulnerability Index in London, United Kingdom *(41)*. Nine variables were identified: households in rented tenure, households in a flat, population density (persons/hectare), households without central heating, population above 65 years old, population with self-reported health status, receiving any kind of social benefit, single pensioner households and ethnic group. These were included in the principal components analysis. Four principal components were then identified, which could be interpreted as high-density housing, poor health and welfare dependency, being elderly and isolated, and poor housing quality. Principal component loadings are weighted according to the variance they explain and summed to form the Heat Vulnerability Index. In this way, the number of independent factors (dimensions) associated with the outcomes could be decreased and interpretation of the findings was simplified.

If statistical assumptions are added into principal components analysis, the principal components analysis becomes a factor analysis (*42)*. The results from principal components analysis and factor analysis would not differ dramatically if the specific variances added are small. Like principal components analysis, factor analysis is a classical technique used to

derive fewer dimensions from a large set of variables. However, unlike principal components analysis, factor analysis can allow for further statistical inference and support assertions about a population (see Case Study 4.6.4). Although the use of factor analysis draws considerable criticism (due to the lack of uniqueness of the factor loadings, for example), it is a useful approximation for the truth and a suitable starting point for further investigation.

Case Study 4.6.4
Factor analysis to develop a Health Vulnerability Index

By using FA to create a linear combination of indicators, a Health Vulnerability Index for disaster risk reduction along the Belt and Road Initiative was developed *(17)*. The index is based on three latent factors: population status, disease prevention and coping capacity. These were derived from nine indicators: proportion of the population below 15 and above 65 years, under-five mortality ratio, maternal mortality ratio, tuberculosis prevalence, age-standardized raised blood pressure, physician ratio, hospital bed ratio, and coverage of the measles-containing-vaccine first-dose (MCV1) and diphtheria tetanus toxoid and pertussis (DTP3) vaccines.

Non-linear dimension reduction methods are an extension of the linear methods and are useful if Euclidean distances (that is, straight-line distance between two points) fail to capture the dissimilarity between the observations. These methods reduce the volume of data by simplifying it into a set of low-dimensional coordinates that preserve distances in the high-dimensional space as much as possible, but involves non-linear transformations of the data.

4.6.7 Conclusions

Risk modelling is well established and can be used in helping resource allocation in Health EDRM. In recent years, it has been applied to a wide range of temperature-related studies, but consistent associations were not often found for other climate-related topics such as rainfall or sea level rise *(17)*. Risk modelling in other contexts (such as complex emergencies) or between varying contexts (such as rural versus urban) is also needed to understand health-related impact of hazards and disasters.

4.6.8 Key messages

- **Time series analysis is widely used for establishing short-term associations between exposures and health outcomes.**
- **Factors associated with protective or preparedness behaviours can be identified by applying the multiple logistic regression method.**
- **Linking Antigenic Properties and Genetic Data, and Identification of Proposed Vaccine Strains are two ways of inference of phenotypic properties for influenza vaccine selection. They estimate the effectiveness of current vaccine strains for the emerging strains and identify new antigenic variants at an early stage of expansion.**
- **In predicting influenza trends, epidemiological approaches, such as the susceptible-infections-removed models and agent-based models, consider human behaviours and address questions related to the impact of prevention measures.**
- **In constructing a health-related risk index, dimension reduction approaches such as principle component analysis (PCA) and factor analysis are widely used to simplify the display of multivariate data.**

4.6.9 Further reading

Jackson JE. A user's guide to principal components. New York, NY: Wiley. 1991.

Wood SN. Generalized additive models: An introduction with R. Chapman and Hall/CRC. 2006.

Gasparrini A. Distributed lag linear and non-linear models in R: the package dlnm. Journal of Statistical Software. 2011: 43(8): 1.

Vynnycky E, White R. An introduction to infectious diseases modelling. Oxford, UK: Oxford University Press. 2010.

McSharry P. Parsimonious risk assessment and the role of transparent diverse models. In Risk modeling for hazards and disasters Elsevier. 2018. pp. 263-9.

4.6

4.6.10 References

1. Centre for Research on the Epidemiology of Disasters (CRED). EM-DAT: The Emergency Events Database. 2020. www.emdat.be (accessed 9 March 2020).
2. National Oceanic and Atmospheric Administration (NOAA). Excessive heat, a 'silent killer'. 2017. www.noaa.gov/stories/excessive-heat-silent-killer (accessed 9 March 2020).
3. Johnson H, Kovats RS, McGregor G, Stedman J, Gibbs M, Walton H, et al. The impact of the 2003 heat wave on mortality and hospital admissions in England. Health Statistics Quarterly. 2005: (25): 6-11.
4. Rey G, Jougla E, Fouillet A, Pavillon G, Bessemoulin P, Frayssinet P et al. The impact of major heat waves on all-cause and cause-specific mortality in France from 1971 to 2003. International Archives of Occupational and Environmental Health. 2007: 80(7): 615-26.
5. Yang J, Liu H, Ou C, Lin G, Ding Y, Zhou Q, et al. Impact of heat wave in 2005 on mortality in Guangzhou, China. Biomedical and Environmental Sciences. 2013: 26(8): 647-54.
6. Hajat S, Haines A. Associations of cold temperatures with GP consultations for respiratory and cardiovascular disease amongst the elderly in London. International Journal of Epidemiology. 2002: 31(4): 825-30.
7. Ryti NR, Guo Y, Jaakkola JJ. Global association of cold spells and adverse health effects: a systematic review and meta-analysis. Environmental Health Perspectives. 2015: 124(1): 12-22.
8. Bhaskaran K, Armstrong B, Hajat S, Haines A, Wilkinson P, Smeeth L. Heat and risk of myocardial infarction: hourly level case-crossover analysis of MINAP database. BMJ. 2012: 345: e8050.
9. Lam HCY, Chan EYY, Goggins WB. Comparison of short-term associations with meteorological variables between COPD and pneumonia hospitalization among the elderly in Hong Kong—a time-series study. International Journal of Biometeorology. 2018: 62(8): 1447-60.
10. Shrestha MS, Khan MR, Wagle N, Babar ZA, Khadgi VR, Sultan S. Chapter 13 - Review of Hydrometeorological Monitoring and Forecasting System for Floods in the Indus Basin in Pakistan. In: Indus River Basin. Elsevier. 2019: pp. 309-33.
11. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. International Journal of Epidemiology. 2013: 42(4): 1187-95.
12. Wood SN. Generalized additive models: An introduction with R. Chapman and Hall/CRC. 2006.
13. Chan EYY. Climate Change and Urban Health: The Case of Hong Kong as a Subtropical City. Routledge. 2019.
14. Gasparrini A, Guo Y, Hashizume M, Lavigne E, Zanobetti A, Schwartz J, et al. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. Lancet. 2015: 386(9991): 369-75.

15. Armstrong B. Models for the relationship between ambient temperature and daily mortality. Epidemiology 2006: 17: 624-31.

16. Gasparrini A . Distributed lag linear and non-linear models in R: the package dlnm. Journal of Statistical Software. 2011: 43(8): 1.

17. Chan EYY, Huang Z, Lam HCY, Wong CKP, Zou Q. Health vulnerability index for disaster risk reduction: application in belt and road initiative (BRI) region. International Journal of Environmental Research and Public Health 2019: 16(3): 380.

18. Chan EYY, Ho JY, Hung HH, Liu S, Lam HC. Health impact of climate change in cities of middle-income countries: the case of China. British Medical Bulletin. 2019: 130(1): 5-24.

19. Ulla Díez SM, Perez-Fortis A . Socio-demographic predictors of health behaviors in Mexican college students. Health Promotion International. 2009: 25(1): 85-93.

20. Health education: theoretical concepts, effective strategies and core competencies. WHO. 2012. http://applications.emro.who.int/dsaf/emrpub_2012_en_1362.pdf (accessed 9 March 2020).

21. Launiala A. How much can a KAP survey tell us about people's knowledge, attitudes and practices? Some observations from medical anthropology research on malaria in pregnancy in Malawi. 2009. www.anthropologymatters.com/index.php/anth_matters/article/view/31/53 (accessed 9 March 2020).

22. Sperandei S. Understanding logistic regression analysis. Biochemia Medica. 2014: 24(1): 12-8.

23. Chan EYY. Public health humanitarian responses to natural disasters. Routledge. 2017.

24. Influenza: are we ready? WHO. 2019. https://www.who.int/influenza/spotlight (accessed 9 March 2020).

25. Descalzo MA, Clara W, Guzmán G, Mena R, Armero J, Lara B, et al. Estimating the burden of influenza-associated hospitalizations and deaths in Central America. Influenza and Other Respiratory Viruses. 2016: 10(4): 340-5.

26. Vestergaard LS, Nielsen J, Krause TG, Espenhain L, Tersago K, Sierra NB, et al. Excess all-cause and influenza-attributable mortality in Europe, December 2016 to February 2017. Eurosurveillance. 2017: 22(14): 30506.

27. Young-Xu Y, van Aalst R, Russo E, Lee JK, Chit A. The annual burden of seasonal influenza in the US Veterans Affairs population. PloS ONE. 2017: 12(1): e0169344.

28. Chretien JP, George D, Shaman J, Chitale RA, McKenzie FE. Influenza forecasting in human populations: a scoping review. PloS ONE. 2014: 9(4): e94130.

29. Morris DH, Gostic KM, Pompei S, Bedford T, Łuksza M, Neher RA, et al. Predictive modeling of influenza shows the promise of applied evolutionary biology. Trends in Microbiology. 2018: 26(2): 102-18.

30. Nsoesie EO, Brownstein JS, Ramakrishnan N, Marathe MV. A systematic review of studies on forecasting the dynamics of influenza outbreaks. Influenza and Other Respiratory Viruses. 2014: 8(3): 309-16.

31. Łuksza M, Lässig M. A predictive fitness model for influenza. Nature. 2014: 507(7490): 57.

32. Steinbrück L, Klingen TR, McHardy AC. Computational prediction of vaccine strains for human influenza A (H3N2) viruses. Journal of Virology. 2014: 88(20): 12123-32.

33. Neher RA, Russell CA, Shraiman BI. Predicting evolution from the shape of genealogical trees. Elife. 2014: 3: e03568.

34. Hu H, Wang H, Wang F, Langley D, Avram A, Liu M. Prediction of influenza-like illness based on the improved artificial tree algorithm and artificial neural network. Scientific Reports. 2018: 8(1): 4895.

35. Reich NG, McGowan CJ, Yamana TK, Tushar A, Ray EL, Osthus D, et al. A Collaborative Multi-Model Ensemble for Real-Time Influenza Season Forecasting in the US. 2019. bioRxiv 566604.

36. Yang W, Cowling BJ, Lau EH, Shaman J. Forecasting influenza epidemics in Hong Kong. PLoS Computational Biology. 2015: 11(7): e1004383.

37. Nsoesie EO, Beckman RJ, Shashaani S, Nagaraj KS, Marathe MV. A simulation optimization approach to epidemic forecasting. PloS ONE. 2013: 8(6): e67164.

38. Intergovernmental Panel on Climate Change (IPCC) (2014). Climate Change 2014: Impacts, Adaptation, and Vulnerability. 2014. http://www.ipcc.ch/report/ar5/wg2 (accessed 9 March 2020).

39. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. New York, NY: Springer. 2013.

40. Everitt B, Hothorn T. An introduction to applied multivariate analysis with R. Springer Science & Business Media. 2011.

41. Wolf T, McGregor G. The development of a heat wave vulnerability index for London, United Kingdom. Weather and Climate Extremes. 2013: 1: 59-68.

41. Cosma S. Advanced Data Analysis from an Elementary Point of View. (in press). https://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/ (accessed 9 March 2020).

4.7 Evaluating economic impacts in health emergency and disaster risk management

Authors

Lorcan Clarke, School of Medicine, Trinity College Dublin, Dublin, Ireland; Department of Health Policy, London School of Economics and Political Science, London, United Kingdom.

Michael F. Drummond, Centre for Health Economics, University of York, York, United Kingdom.

4.7.1 Learning objectives

The learning objectives of this chapter are to:

1. Understand how economic evaluations and economic impact studies can support decision making in health emergency and disaster risk management (Health EDRM).
2. Know the methods available to researchers conducting these studies.
3. Be aware of research limitations, including evidence gaps and methodological challenges.

4.7.2 Introduction

Economic evaluations and economic impact studies are important because they can help decision makers manage competing spending priorities and maximize the value of their financial budgets. Economic impact studies quantify the costs and consequences of past or potential events. Economic evaluations are a structured way to evaluate costs and consequences of a programme or policy compared to an alternative course of action. Conducting these studies and applying their findings can be part of prevention, preparedness, response and recovery activities in Health EDRM.

This chapter provides an introduction to economic evaluations. It outlines the value of evaluating economic impacts, key concepts involved in conducting economic evaluations, and current limitations in the context of Health EDRM. In this chapter, the term “researchers” refers to individuals and groups undertaking economic studies.

4.7

4.7.3 Why conduct economic evaluations and economic impact studies?

Economic studies describe and explain the implications of a specific event or health issue, and potential risk management actions, in terms of financial and non-financial resources. This information can help justify the size of overall spending and support specific resource allocation decisions about which policies and programmes to use to improve health outcomes *(1)*.

4.7.4 Informing decision making

Economic studies that can help inform Health EDRM include economic evaluations and economic impact studies. Economic evaluations explicitly compare the costs (use of resources) and consequences (effects) of a programme or policy with an alternative course of action *(2)*. This alternative may incorporate another programme or policy, or simply reflect the current situation. Economic impact studies evaluate actual or potential economic outcomes related to a specific intervention, event or health-related issue, such as those associated with a heatwave or an infectious disease outbreak. Findings from both economic evaluations and economic impact studies can be inputs for decision-making tools that account for broader economic and non-economic evidence, such as multi-criteria decision analysis (MCDA). In such cases, MCDA combines findings from economic studies with additional decision-making factors, such as budget constraints or implications for equity and fairness *(3)*.

Various stakeholders can use the information created by economic studies to evaluate past events, manage current challenges or plan for future risks. These stakeholders include government agencies, private companies and civil society groups. For example, findings from economic studies can inform the costing tools used to plan and implement measures to prevent, prepare, respond to and recover from health emergencies and disasters *(4)*. Economic studies also help to describe inequality and hardship, which might link to socioeconomic and demographic characteristics such as income status, gender and age. Section 4.7.5 “Understanding the economic impact of health emergencies and disasters” discusses these topics further.

Economic evaluations help support population-level decisions about which health services, medicines and other medical technologies should be funded and made available. Economic studies can help offer a reference point for balancing and aligning different stakeholders’ priorities, such as those of patients and the public, taxpayers and politicians, insurance providers, healthcare providers, and health technology producers *(5)*. The term “health technology” refers to the application of organized knowledge and skills in the form of devices, medicines, vaccines, procedures and systems developed to solve a health problem and improve quality of lives *(6)*.

4.7.5 Understanding the economic impact of health emergencies and disasters

Health emergencies and disasters lead to economic impacts on households, health systems and the economies as a whole (macroeconomic impacts). Economic studies help describe these impacts.

Illness or injury can create healthcare costs and income losses that put stress on families and households. Healthcare costs create direct economic impacts through spending on health services or medicines, which limit funds available for other household expenditures or create the need for raising additional funds, potentially via incurring financial debt. An inability to work, due to illness or caring for others who are sick, can create indirect economic impacts (see 4.7.7) through income losses and associated financial distress.

Proactive policies to guarantee healthcare access and support wellbeing can help reduce household and community impacts and hardship, which may be distributed inequitably between different socioeconomic and demographic groups *(7–8)*. For example, after Super Typhoon Yolanda devastated parts of the Philippines in 2013, the response included rapidly adapting existing healthcare funding systems. The national insurance agency (PhilHealth) guaranteed hospital services to all affected persons seeking access, regardless of whether insurance policies already covered the person's healthcare costs *(9)*. This meant that people who did not have the necessary health insurance could still access healthcare, without concern about further typhoon-related hardship due to additional costs.

Damage and disruption can restrict healthcare services and, at the same time, create increased demand due to direct and indirect health impacts (see 4.7.7). Damage to infrastructure, constrained workforce capacities and disruption to physical supply-chains can limit the availability and accessibility of health care *(10)*. This can mean that illness and injury are not treated, leading to worse health outcomes and higher long-term health-related costs *(11–12)*. Economic studies can support proactive risk management policies, ensuring that healthcare services can adapt to restrictions and meet sudden increases in healthcare requirements *(13)*.

Disasters and emergencies also create macroeconomic impacts, by disrupting the functioning of government institutions, private organizations and the overall economy. Government institutions are stressed by responses to challenging public priorities, while private organizations lose potential revenues from the goods and services they produce, and the supplies of labour and other inputs needed to produce them. This disruption will negatively impact both economic output and people's general welfare *(14)*. Examples of research into the macroeconomic impacts of climate change, natural hazards, and infectious disease outbreaks have found that climate change-related increases in exposure to extreme heat in South-East Asian countries may restrict feasible annual working hours by 15% to 20% by 2030 *(15)*, that disasters due to natural hazards lead to impacts to wellbeing and losses to economic consumption that result in over US$520 billion in economic losses per year *(16)*, and a severe pandemic outbreak of infectious disease could reduce global economic output by US$500 billion if there were 720 000 associated deaths in a single year (forecast conducted in 2017) *(17)*.

4.7

Table 4.7.1 provides examples of the sorts of economic impact studies that can inform decision makers and help address economic impacts on households, health systems, and the economy as a whole. These studies were obtained from two evidence reviews of economic studies and are a sample of the (limited) available economic evidence in health emergency and disaster risk management published prior to 2020 *(18-19)*. Two studies focused on infectious disease outbreaks (Ebola Virus Disease) and four focused on extreme weather events (hurricanes and heatwaves). Some of these studies offer a range of estimates, which reflects their accounting of potential uncertainty in their findings (see 4.7.10 Ten steps to conducting an economic evaluation).

Table 4.7.1 Examples of economic impact studies

Infectious disease outbreaks: Ebola Virus Disease (EVD) in West Africa (2014-2016)

Bartsch and colleagues *(20)* estimated costs associated with individual patient cases of EVD.

- The authors looked at individuals who survived and who died after receiving care for EVD, in Guinea, Liberia and Sierra Leone during the 2014-2016 outbreak. Estimates of costs included supportive care, personal protective equipment, wages for health workers, and productivity losses linked to health-related absence from work.
- They compiled costs associated with 17 908 cases of EVD and 6373 deaths caused by EVD, as of December 2014, to estimate total societal costs of between US$82 million and US$356 million.

Kirigia and colleagues *(21)* estimated economic losses associated with EVD deaths.

- The authors focused on individuals who died in Guinea, Liberia, Mali, Federal Republic of Nigeria and Sierra Leone during the 2014-2016 outbreak. They measured losses based on expected overall losses of economic outputs, excluding those related to the provision of health care.
- They compiled costs associated with 11 234 deaths from 27 543 EVD cases, as of 28 June 2015, and estimated that cumulative future economic losses would be over US$155 million.

Extreme weather events: Hurricanes in the USA

Fonseca and colleagues *(11)* forecast economic impacts associated with hurricane-related disruption to health care.

- The authors focused on individuals with diabetes impact by Hurricane Katrina, which made landfall in the USA in August 2005. Estimates of health outcomes included measures of blood sugar, blood pressure and lipids. They drew on a previous study to combine these measures to estimate life expectancy, quality-adjusted life expectancy, and future costs of diabetes-related complications *(22)*.
- They forecast that disruption to diabetes patients' access to healthcare services and supplies because of the damage to the health system might lead to US$504 million in additional healthcare costs over the lifetimes of affected individuals.

Zahran and colleagues *(23)* assessed mental health resilience and related economic impacts for individuals exposed to hurricanes.

- The authors focused on population impacts, specifically for single mothers, of two hurricanes which made landfall in the USA in 2005: Hurricane Katrina and Hurricane Rita.
- They measured costs by calculating expected declines in productivity and wages following the hurricane events. The authors found that, following the hurricane events, single mothers had over three times more poor mental health days and five times more days absent from work than the general population. These effects were linked to economic losses of US$4200 per person and a total of US$130 million for all single mothers in the affected population.

Extreme weather events: Heatwaves in Australia and the USA

Toloo and colleagues *(24)* forecast healthcare costs associated with more common and more intense heatwaves.

- The authors focused on emergency department use by individuals impacted by heatwaves in Brisbane, Australia. They estimated emergency department use for a younger and older age group and linked use to health issues such as exacerbated cardiovascular issues, diabetes, and renal complaints. They estimated costs by combining data from 2012 and 2013, which described the costs of excess emergency department visits with forecasts for extreme temperature prevalence in 2030 and 2060.
- They forecast that expected heatwaves could increase emergency healthcare costs in Brisbane by between AU$78 000 and AU$260 000 in 2030 and between AU$215 000 and AU$1 985 000 in 2060, without adjusting for inflation.

Lin and colleagues *(25)* forecast healthcare costs associated with hospital admissions linked to a range of different heatwave scenarios.

- The authors focused on respiratory-related hospital admissions in New York, USA. They combined estimates of daily hospitalization costs with excess days of hospitalization per year attributable to extreme heat, using a range of scenarios forecast by the Intergovernmental Panel on Climate Change (IPCC).
- They estimated that heatwave-related annual admissions created additional costs of US$0.64 million per year from 1991-2004, with estimated excess costs of between US$5.5 and US$7.5 million per year from 2045-2065, and between US$26 and US$76 million per year from 2080-2099.

4.7

4.7.6 Key concepts involved in conducting economic evaluations

This section introduces the key concepts and steps involved in conducting an economic evaluation and offers some guidance on how to conduct an economic evaluation in the context of Health EDRM. Although the focus is on economic evaluations, some of the concepts discussed under the headings of 'Population' and 'Economic Outcomes' are relevant to researchers conducting economic impact studies. This information is a complement to, rather than a substitute for, established guidance on conducting and reporting economic evaluations *(26–27)*.

The following sections outline three elements involved in economic evaluations comparing the value for money of alternative programmes or policies: the target population, the economic outcomes, and the comparison methods. Other important elements include the interventions, comparison groups, and the time horizon for evaluating outcomes; as discussed in other chapters in this book. Specific concerns for researchers conducting economic evaluations are highlighted in the "Research limitations" section of this chapter (4.7.11).

4.7.7 Population

An economic evaluation focuses on the outcomes of a specific group of individuals, namely the study's target population. Researchers can define this population by its size and using factors such as the socioeconomic or demographic characteristics (such as income status or age) of the people within it, the interventions they receive and geographic area covered by the population. Researchers should also consider whether they define this population based on whether a health emergency or disaster directly or indirectly affected the people in the population. The meanings of "directly affected" and "indirectly affected" are outlined below.

Directly affected
People who have suffered injury, illness or other health effects; who were evacuated, displaced or relocated or have suffered direct damage to their livelihoods, economic, physical, social, cultural and environmental assets *(28)*. Examples of direct health effects include immediate illness due to an infectious disease or injuries such as wounding, blunt force trauma, and burns *(10)*.

Indirectly affected
People who, over time, have suffered consequences other than or in addition to direct effects. These may be due to disruption or changes in economy, critical infrastructure, basic services, commerce or work, and include social, health and psychological consequences *(28)*. Examples of indirect health effects include post-emergency sanitation issues leading to infectious disease outbreaks and disrupted access to healthcare services leading to untreated health issues *(10)*.

4.7.8 Perspective

Researchers use a variety of measures to estimate costs and consequences. One way to group these measures is to take a "payer perspective", which focuses on healthcare use. Another way is to use a "societal perspective", which accounts for a broader set of economic impacts *(2)*. The choice as to which economic outcomes should be included in a study is influenced by the amount of time and effort required to conduct the study, due to analysis requirements and the intended audience for the results of the study. For example, a payer perspective may meet the needs of a health insurance company focused on managing healthcare costs, whereas a government agency may prefer to take a societal perspective to account for broader impacts on health, wellbeing, and economic welfare. The choice of perspective for a study is often discussed in terms of the range of costs considered, but can also account for consequences considered.

Payer Perspective

Payer perspective focuses on costs and consequences linked to the use of (and payment for) healthcare. Payers can include a variety of actors directly involved in the provision and receipt of healthcare services. The main payers are usually government agencies or health insurers, depending on how healthcare is organized and financed in the country concerned. However, in many settings, patients and family members will incur costs associated with accessing or receiving health care. Medical costs and consequences may involve payments for access to care, medical supply costs, salaries for health workers and expected future healthcare costs related to changes in health outcomes. Non-medical costs and consequences may involve spending on transport, accommodation, and food by individuals receiving care and informal nursing care provided by their families.

Societal perspective

Societal perspective focuses on the costs and consequences, including but not limited to those measured in a payer perspective, which can be linked to health outcomes and healthcare use. Societal costs and consequences include broader societal concerns – such as employment, labour productivity, and consumption of goods and services other than health care.

Economic costs and consequences are measured based on the value of market or non-market resources. Market resources are purchased with money and have a defined price. They include wages for health workers and the cost of drugs. Non-market resources are not purchased with money and do not have a defined price. These include household work, volunteer services, and donated medical supplies. One way that researchers can estimate the economic outcomes associated with non-market resources is by using a proxy measure. A proxy is a variable that is more readily measurable and can act as a substitute estimate of costs and consequences, such as values of similar goods and services.

4.7

Figure 4.7.1 Components of an economic evaluation of a healthcare programme (adapted from *(2)*)

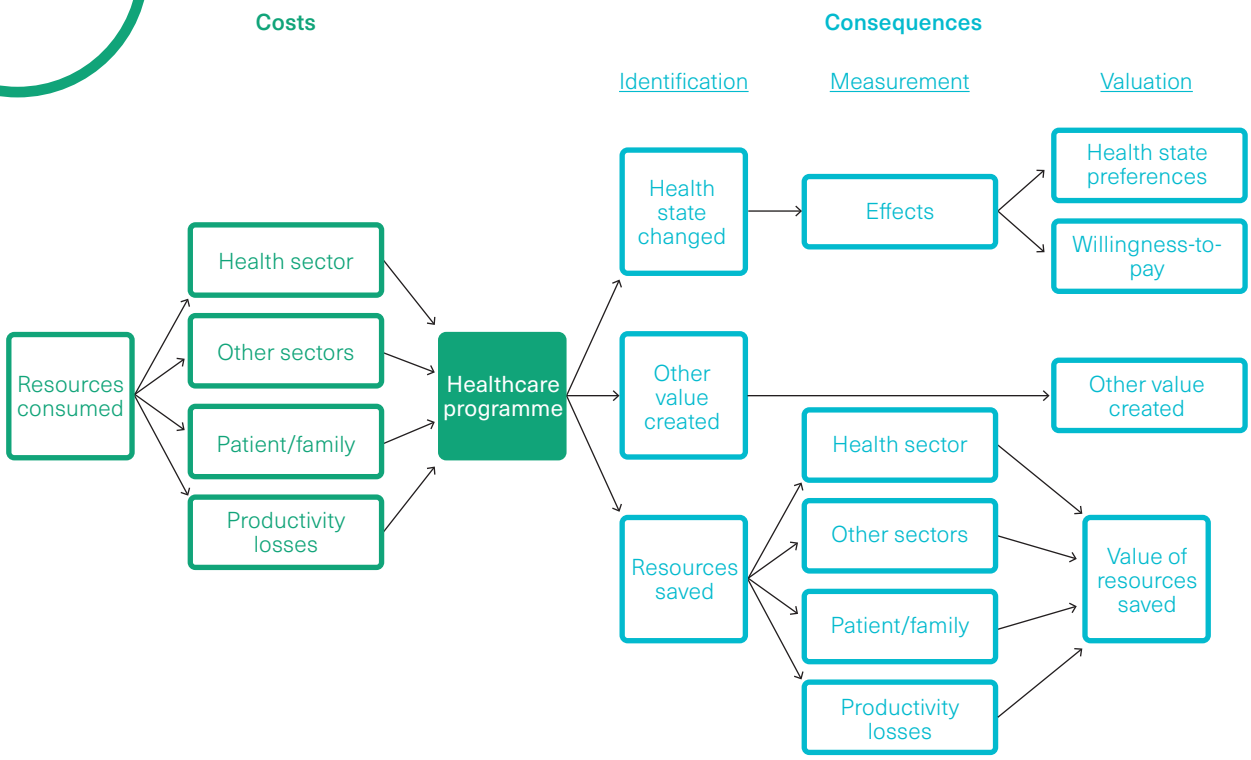


Figure 4.7.1 displays key pathways involved when estimating the costs and consequences involved in an economic evaluation of a healthcare programme. Costs reflect resource use across different sectors to deliver the programme, consequences reflect outcomes related to the programme's impacts on health and wellbeing. For example, costs for a vaccination programme might involve vaccine manufacture, delivery of vaccines to a health facility, and health workers providing vaccination services. Consequences for this programme might include immunization preventing future healthcare costs and losses to labour productivity. Researchers can adapt these pathways and the interaction between different nodes to vary their study perspective and focus on prevention, preparedness, response or recovery activities in Health EDRM.

4.7.9 Comparison Methods

There are several established methods for combining data on costs and consequences to evaluate economic outcomes *(2)*. The following paragraphs outline some of them. Other approaches to evaluating economic outcomes, not discussed in detail here, include extended cost-effectiveness analysis and the use of social welfare functions *(1)*.

Cost-benefit analysis

Cost-benefit analysis combines costs, positive consequences, and negative consequences to calculate a cost-benefit ratio or measure of net-benefit (benefits minus costs). Both costs and benefits are measured in monetary terms. This approach provides a clear estimate of relative economic outcomes, but only if it is possible to estimate the monetary value of costs and consequences.

Cost-consequence analysis
Cost-consequence analysis compares costs and outcomes by placing them in discrete categories. Estimates are not combined to create a single measure or ratio. This approach allows the user of the research to make their own interpretation about the relative importance of different costs and consequences.

Cost-effectiveness analysis
Cost-effectiveness analysis compares costs measured in monetary terms with outcomes measured via natural units. Examples of natural units for health-related outcomes include clinical endpoints (see Chapter 2.2), such as end of viral infection or alleviation of symptoms of depression, or life-years gained (which is the additional number of years of life that a person lives as a result of receiving a treatment). Case Study 4.7.1 summarizes a study that used cost-effectiveness analysis to compare antiviral stockpiling approaches for pandemic influenza preparedness.

Cost-minimization analysis
Cost-minimization analysis compares interventions based on costs measured in monetary terms. This approach does not measure consequences and is only appropriate if the compared interventions have the same effect.

Cost-utility analysis
Cost-utility analysis compares costs measured in monetary terms with consequences measured via a measure of health gain or 'utility'. Examples of utility measures include:

- **Quality-Adjusted Life-Years (QALYs)** are a measure of additional life expectancy combined with the health-related quality of life. QALY measures are determined by surveying people's evaluations of being in different health states, accounting for factors such as pain or mobility, through surveys and instruments such as the EQ-5D *(2)*.
- **Disability-Adjusted Life-Years (DALYs)** are a measure of life expectancy combined with years of healthy life lost due to mortality and/or morbidity associated with a health issue. DALY measures reflect the difference between a given health state and a benchmark that is based on the experience of a healthy life that reaches full life expectancy.

Return on investment analysis
Return on investment analysis calculates the size of the difference between positive consequences and costs. Return on investment involves calculating net consequences (positive consequences minus negative consequences) and then expressing this figure as a proportion of costs. Typically, these studies consider only those costs and consequences that can easily be expressed in monetary terms. Case Study 4.7.2 describes a return on investment calculation for vaccine interventions, focusing on the resource costs and savings of a potential flu outbreak in Chicago, USA.

4.7

Case Study 4.7.1
Comparing the value of stockpiling approaches

Carrasco and colleagues (*29)* conducted an economic evaluation to assess arrangements for stockpiling antiviral medicines in anticipation of an influenza pandemic across ten high- and middle-income countries. They examined different stockpile sizes and impacts on eligible recipients of antivirals for prophylaxis and treatment. They focused on estimates of mortality associated with infectious disease outbreaks and the costs of antiviral stockpiles. Health risks were estimated by forecasting morbidity and mortality associated with pandemic risks over a 30-year time horizon, accounting for factors including seasonality and development of an effective vaccine. Economic outcomes included treatment costs and work absenteeism.

The authors estimated that stockpiles in higher income countries had a greater potential avoidance of expected costs, while stockpiles in lower income countries had more potential avoidance of mortality. Their findings showed that the USA could potentially avert potential future costs by US$22 billion, and that improved stockpiling in Indonesia could reduce expected mortality by more than 9 million deaths.

Case Study 4.7.2
Preparing for public health emergencies

Dorratoltaj and colleagues *(30)* conducted an economic evaluation to understand vaccination priorities and economic outcomes during disease outbreaks. They examined vaccine use versus a base case scenario of no vaccine intervention during moderate, strong, and catastrophic influenza outbreaks. They focused on people living in Chicago, USA and examined impacts across different population sub-groups based on age and levels of health risk. They estimated economic outcomes by linking expected health impacts associated with an influenza-like illness with healthcare costs and productivity costs taken from another study *(31)*.

The authors included cost-benefit and return on investment methods in their analysis. High-risk people under 19 years of age had the highest return on investment in a catastrophic influenza pandemic scenario, with US$249.16 saved for each US$1 invested in vaccinations. The lowest return on investment in a catastrophic influenza pandemic scenario was among non-high risk people aged between 20 and 64 years, with US$5.64 saved for each US$1 invested in vaccinations. Net benefits were highest among high-risk people aged between 20 and 64 years in all pandemic scenarios.

Having identified and implemented a comparison method, researchers can account for uncertainty their economic study results by conducting sensitivity analyses. A sensitivity analysis measures variations in results based on changes to the inputs informing the costs and consequences in an economic evaluation. Changes can involve varying the value of an input (such as implementation cost or population characteristics) or other features of the study, such as the time horizon (number of months or years over which costs and consequences are estimated).

Variance in results displayed by sensitivity analyses, can help decision makers to understand how the variance in their input values affects the results of their economic evaluation and help researchers to reduce uncertainty in their inputs data (such as intervention effectiveness or costs). Researchers may also compare findings from different statistical models to help understand how different approaches to estimating costs and consequences will impact their results.

4.7.10 Ten steps to conducting an economic evaluation

The process of conducting an economic evaluation can be set out as a series of ten steps. These steps, adapted from questions created to help guide assessments of economic evaluations, are outlined below *(2)*.

These steps complement guidance elsewhere in this book on study design, such as in Chapter 3.5 on determining the research question. They can also be considered alongside other published, and well-established, recommendations for conducting economic evaluations *(2, 32-36)*.

Step 1: Define a research question (see also Chapter 3.5) which:

- identifies the population involved;
- outlines the costs and consequences of the compared courses of action over an appropriate time horizon;
- defines the analytic perspective and decision-making context.

Step 2: Describe the interventions and identify any that were not considered, such as specific interventions for population subgroups (see also Chapter 3.3).

Step 3: Establish the effectiveness of the intervention or policy. Note how data were synthesized and any factors that may influence the reliability of primary data. If no primary data are available, researchers could draw upon relevant evidence syntheses, such as a systematic review and meta-analysis, to inform estimates of effectiveness (see also Chapter 2.6).

Step 4: Describe the relevant costs and consequences for each alternative intervention or policy.

Step 5: Measure relevant inputs, for costs and consequences, using appropriate and comparable units. Justify the included measures and their information sources.

Step 6: Estimate values for costs and consequences. Record the source of these values and whether they are market values (such as specified drug costs), or non-market values (such as unpaid work) and if values were adjusted, this is often done to account for differences between costs that healthcare providers actually incur, versus the amount they charge.

Step 7: Adjust estimates of costs and consequences to account for their changing value over time. This is also known as discounting. Discounting involves individuals placing a lower value on a future cost or consequence versus an immediate one, such as a health benefit today versus one obtained five years in the future. Recommended discount rates vary

between countries and organizations. It may also be appropriate to first adjust for inflation, which is the rate of change in average prices over time.

Step 8: Compare the costs and consequences of different interventions by combining estimates using an established analysis method. Examples include the incremental cost-effectiveness ratio used in a cost-effectiveness analysis or net benefit used in a cost-benefit analysis.

Step 9: Describe uncertainty in estimates of costs and consequences by:

- analysing statistical variance within population level estimates (if available);
- accounting for heterogeneity in results between different population subgroups (if applicable);
- assessing the effect of altering the values of inputs to measures of costs and consequences on overall study findings (via sensitivity analysis).

Step 10: Describe results and discuss:

- basing conclusions on an overall index (such as a value in US dollars) or ratio of costs and consequences (such as cost-effectiveness ratio);
- differences between the methods and findings of the study with those in comparable studies;
- the generalizability of results to other settings and populations;
- important factors influencing decision making, such as equity implications;
- wider resource implications, such as budgetary impacts;
- implications of any uncertainty in the study's findings, including the need for future research.

4.7.11 Research limitations

Evidence gaps and methodological challenges have limited the prevalence and use of evaluations of economic impacts in Health EDRM research. Reviews of research on infectious disease outbreak preparedness and the impacts of extreme weather events have identified several gaps in economic evidence *(18, 37–38)*. These gaps include a lack of studies that incorporate economic evaluations (most are economic impact studies), use a societal perspective for economic outcomes, or are set in low- and middle-income countries. Addressing evidence gaps is important, especially for those populations that are expected to suffer most from increasing hazard risks, such as heat stress in South Asia *(39)*.

Researchers often use different methods, or adapt methods to their needs. These actions can limit the ability of others to compare the findings of a study with otherwise similar studies. However, from the researchers perspective, it can be difficult to strike a balance between adhering to standardized approaches (to ensure comparability across different economic studies) and adapting to constraints (because of the availability of data, research aims, and resource limitations).

Methodological challenges include attributing outcomes to interventions, measuring the economic value of outcomes and accounting for how preferences for outcomes vary over time. Addressing these for Health EDRM can draw upon research areas with similar methodological challenges, such as economic studies of public health activities and of natural environment interventions *(40–42)*.

- **Attributing outcomes:** In many circumstances it may not be feasible to use a randomized trial (see Chapters 4.1 and 4.3) to attribute and measure outcomes associated with interventions in Health EDRM. This increases the difficulty involved in conducting a robust economic evaluation. However, if sufficient data can be collected, researchers may be able to create a quasi-experimental study (see Chapter 4.5) by using natural variation in people's exposure to interventions.
- **Measuring economic outcomes**: It is difficult to measure different stakeholders' preferences for health and non-health outcomes and to create a combined measure of economic outcomes. Population preferences for these outcomes may also change over time and need to be accounted for. Future research may expand the scope of existing measures, such as recent efforts to adapt the QALY approach to better account for broader wellbeing *(43)*.
- **Time variance:** It is important to consider how to apply discount rates in economic studies in Health EDRM, given the potential (in) frequency of a given health emergency or disaster. A discount rate accounts for the difference in stakeholder preferences for an outcome today versus one in the future, as well as uncertainty and the time value of money, and discounts the expected value of an intervention appropriately. Recommended time horizons and discount rates are available for specific contexts and uses, but there is persistent debate on the most appropriate values to use *(44–45)*.

4.7.12 Conclusions

Researchers use economic evaluations and economic impact studies to identify and explain the costs and consequences involved in policies and programmes that support Health EDRM. Practitioners and policymakers can then use the evidence generated by these studies to guide their decision making on specific issues and broader strategic planning.

Established methods and concepts are available to researchers to synthesize and improve the current evidence base of economic studies, although there are challenges to expanding research in this area. Nevertheless, there are opportunities for economic studies to fill knowledge gaps and to address the ongoing needs of decision makers. Researchers and stakeholders can use these opportunities to advocate for putting greater effort into assessing and addressing the economic aspects of past, present, and future health emergencies and disasters *(46)*.

4.7

4.7.13 Key messages

- **Evaluating economic impacts in Health EDRM can inform and improve prevention, preparedness, response and recovery activities.**
- **Economic evaluations and economic impact studies are established ways to evaluate the impacts of interventions and events. Researchers can draw upon standardized methods and knowledge built by existing communities of expertise.**
- **Current research gaps mean that researchers have the opportunity to develop specific guidance on how to examine economic outcomes in the context of Health EDRM and to conduct more research that incorporates economic evaluations, uses a societal perspective for economic outcomes, and is set in low- and middle-income countries – all of which can offer useful and usable information to improve Health EDRM practices.**

4.7.14 Further reading

Drummond M, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of healthcare programmes. Oxford, UK: Oxford University Press. 2015.

Sculpher M, Revill P, Ochalek JM, Claxton K. How much health for the money? Using cost-effectiveness analysis to support benefits plan decisions. in Glassman A, Giedion U, Smith PC, editors. What's In, What's Out: Designing Benefits for Universal Health Coverage. Washington DC: Centre for Global Development. 2017.

Madhav N, Oppenheim B, Gallivan M, Mulembakani P, Rubin E, Wolfe N. Pandemics: Risks, Impacts and Mitigation. In: Jamison DT, Gelband H, Horton S, Jha P, Laxminarayan R, Mock CN, Nugent R, editors. Disease Control Priorities (third edition). Washington DC: World Bank. 2017.

Peters DH, Hanssen O, Gutierrez J, Abrahams A, Nyenswah T. Financing Common Goods for Health: Core Government Functions in Health Emergency and Disaster Risk Management. Health Systems & Reform. 2019: 5(4):307–321.

Clarke L, Le Masson V. Shocks, stresses and universal health coverage: pathways to address resilience and health. ODI Working Paper 526. 2017.

4.7.15 References

1. Lauer JA, Morten A, Bertram M. Cost-Effectiveness Analysis. In Norheim OF, Emanuel EJ, Millum J, editors. Global Health Priority-Setting: Beyond Cost-Effectiveness. Oxford, UK: Oxford University Press. 2020.
2. Drummond M, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of healthcare programmes. Oxford, UK: Oxford University Press. 2015.
3. Baltussen R, Niessen L. Priority setting of health interventions: The need for multi-criteria decision analysis. Cost Effectiveness and Resource Allocation. 2006. 4: 14.
4. WHO Expert Group Consultation on Health Systems for Health Security - Meeting Report (6-7 March 2019). WHO. 2019. https://extranet.who.int/sph/docs/file/3559 (accessed 1 March 2020).
5. Timmins N, Rawlins M, Appleby J. A Terrible Beauty: A Short History of NICE - The National Institute for Health and Care Excellence. F1000Research 2016: 6: 915.
6. WHA60.29 - Health technologies. WHO. 2007 https://www.who.int/medical_devices/resolution_wha60_29-en1.pdf (accessed 1 March 2020).
7. Zahran S, Tavani D, Weiler S. Daily Variation in Natural Disaster Casualties: Information Flows, Safety, and Opportunity Costs in Tornado Versus Hurricane Strikes. Risk Analysis; 2013: 33(7): 1265–80.
8. Pagliacci F, Russo M. Socioeconomic effects of an earthquake: does spatial heterogeneity matter? Regional Studies; 2019: 53(4): 490–502.
9. PhilHealth. PhilHealth to pay for hospitalization of 'Yolanda' Survivors (December 4, 2013). 2013. https://www.philhealth.gov.ph/news/2013/hospitalization_yolandasurvivor.html (accessed 1 March 2020).
10. Clarke L, Le Masson V. Shocks, stresses and universal health coverage: pathways to address resilience and health. ODI Working Paper 526. 2017
11. Fonseca VA, Smith H, Kuhadiya N, Leger SM, Yau CL, Reynolds K, et al. Impact of a Natural Disaster on Diabetes: Exacerbation of disparities and long-term consequences. Diabetes Care. 2009: 32(9): 1632–8.
12. De Alwis D, Noy I. The cost of being under the weather: Droughts, floods, and health care costs in Sri Lanka. Asian Development Review. 2017: 36(2): 185-214.
13. Clarke L. Predicting and preparing: Innovating for health-system resilience to extreme weather events. Oasis Loss Modelling Framework. 2018.
14. Strulik H, Trimborn T. Natural Disasters and Macroeconomic Performance. Environmental and Resource Economics. 2019: 72(4): 1069–98.

15. Kjellstrom T. Impact of Climate Conditions on Occupational Health and Related Economic Losses: A New Feature of Global and Urban Health in the Context of Climate Change. Asia Pacific Journal of Public Health. 2015: 28 (2 suppl): 28S-37S.

16. Hallegatte S, Vogt-Schilb A, Bangalore M, Rozenberg J. Unbreakable: Building the Resilience of the Poor in the Face of Natural Disasters. Climate Change and Development. World Bank. 2017.

17. Fan VY, Jamison DT, Summers LH. Pandemic risk: how large are the expected losses? Bulletin of the World Health Organization. 2018: 96(2):129–34.

18. Schmitt LHM, Graham HM, White PCL. Economic Evaluations of the Health Impacts of Weather-Related Extreme Events: A Scoping Review. International Journal of Environmental Research and Public Health: 2016: 13(11): 1105.

19. Huber, C, Finelli, L and Stevens, W. The Economic and Social Burden of the 2014 Ebola Outbreak in West Africa. The Journal of Infectious Diseases. 2018: (suppl 5): S698–S704.

20. Bartsch SM, Gorham K, Lee BY. The cost of an Ebola case. Pathogens and Global Health. 2015: 109(1):4–9.

21. Kirigia JM, Masiye F, Kirigia DG, Akweongo P. Indirect costs associated with deaths from the Ebola virus disease in West Africa. Infectious Diseases of Poverty. 2015: 4(1): 45.

22. Palmer AJ, Roze S, Valentine WJ, Minshallb ME, Hayes C, Oglesby A, et al. Impact of Changes in HbA1c, Lipids and Blood Pressure on Long-term Outcomes in Type 2 Diabetes Patients: An Analysis Using the CORE Diabetes Model. Current Medical Research and Opinion. 2004: 20(suppl 1): S53–8.

23. Zahran S, Peek LA, Snodgrass JG, Weiler S, Hempel L. Economics of Disaster Risk, Social Vulnerability, and Mental Health Resilience. Risk Analysis. 2011: 31(7): 1107–19.

24. Toloo G (Sam), Hu W, FitzGerald G, Aitken P, Tong S. Projecting excess emergency department visits and associated costs in Brisbane, Australia, under population growth and climate change scenarios. Scientific Reports. 2015: 5: 12860.

25. Lin S, Hsu W-H, Van Zutphen AR, Saha S, Luber G, Hwang S-A. Excessive Heat and Respiratory Hospitalizations in New York State: Estimating Current and Future Public Health Burden Related to Climate Change. Environmental Health Perspectives. 2012: 120(11): 1571–7.

26. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. European Journal of Health Economics: HEPAC: Health Economics in Prevention and Care. 2013. 14(3): 367–72.

27. Sculpher M, Revill P, Ochalek JM, Claxton K. How much health for the money? Using cost-effectiveness analysis to support benefits plan decisions. In Glassman A, Giedion U, Smith PC, editors. What's In, What's Out: Designing Benefits for Universal Health Coverage. Washington DC: Centre for Global Development. 2017.

28. United Nations General Assembly. Report of the open-ended intergovernmental expert working group on indicators and terminology relating to disaster risk reduction. 2017. https://www.undrr.org/publication/report-open-ended-intergovernmental-expert-working-group-indicators-and-terminology (accessed 1 March 2020).

29. Carrasco LR, Lee VJ, Chen MI, Matchar DB, Thompson JP, Cook AR. Strategies for antiviral stockpiling for future influenza pandemics: a global epidemic-economic perspective. Journal of the Royal Society, Interface. 2011: 8(62):1307–13.

30. Dorratoltaj N, Marathe A, Lewis BL, Swarup S, Eubank SG, Abbas KM. Epidemiological and economic impact of pandemic influenza in Chicago: Priorities for vaccine interventions. PLOS Computational Biology. 2017: 13(6):e1005521.

31. Carias C, Reed C, Kim IK, Foppa IM, Biggerstaff M, Meltzer MI, et al. Net Costs Due to Seasonal Influenza Vaccination — United States, 2005–2009. PLOS ONE. 2015: 10(7):e0132922.

32. Tan-Torres Edejer T, Baltussen R, Adam T, Hutubessy R, Acharya A, Evans DB, et al, editors. 2003. Making Choices in Health: WHO Guide to Cost-Effectiveness Analysis. Geneva, Switzerland: WHO. 2003.

33. Weatherly H, Drummond M, Claxton K, Cookson R, Ferguson B, Godfrey C, et al. Methods for assessing the cost-effectiveness of public health interventions: Key challenges and recommendations. Health Policy; 2009: 93(2): 85-92.

34. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS): Explanation and Elaboration: A Report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value in Health. 2013: 16(2): 231–50.

35. Wilkinson T, Sculpher MJ, Claxton K, Revill P, Briggs A, Cairns JA, et al. The International Decision Support Initiative Reference Case for Economic Evaluation: An Aid to Thought. Value in Health 2016: 19(8): 921–8.

36. Robinson LA, Hammitt JK, Jamison DT, Walker DG. Conducting Benefit-Cost Analysis in Low- and Middle-Income Countries: Introduction to the Special Issue. Journal of Benefit-Cost Analysis. 2019: 10(S1): 1–14.

37. Ott JJ, Klein Breteler J, Tam JS, Hutubessy RCW, Jit M, de Boer MR. Influenza vaccines in low and middle income countries. Human Vaccines & Immunotherapeutics. 2013: 9(7): 1500–11.

38. Pasquini-Descomps H, Brender N, Maradan D. Value for Money in H1N1 Influenza: A Systematic Review of the Cost-Effectiveness of Pandemic Interventions. Value in Health. 2017: 20(6): 819–27.

39. Im E-S, Pal JS, Eltahir EAB. Deadly heat waves projected in the densely populated agricultural regions of South Asia. Science Advances. 2017: 3(8): e1603322.

40. The Economics of Immunization Policies. WHO. 2014. https://www.who.int/immunization/programmes_systems/financing/analyses/1_economics_of_immunization_policies.pdf (accessed 1 March 2020).

41. Bojke L, Schmitt L, Lomas J, Richardson G and Weatherly H. Economic Evaluation of Environmental Interventions: Reflections on Methodological Challenges and Developments. International Journal of Environmental Research and Public Health. 2018: 15(11): E2459.

42. Watson SI, Sahota H, Taylor CA, Chen Y-F, Lilford RJ. Cost-effectiveness of health care service delivery interventions in low and middle income countries: a systematic review. Global Health Research and Policy. 2018: 3(1): 17.

43. Garrison LP, Neumann PJ, Willke RJ, Basu A, Danzon PM, Doshi JA, et al. A Health Economics Approach to US Value Assessment Frameworks - Summary and Recommendations of the ISPOR Special Task Force Report [7]. Value in Health. 2018: 21(2): 161–5.

44. Dietz S. A long-run target for climate policy: the Stern Review and its critics. Grantham Research Institute on climate Change and the Environment/Department of Geography and Environment, London School of Economics and Political Science, London. 2008.

45. Haacker M, Hallett TB, Atun R. On discount rates for economic evaluations in global health. Health Policy and Planning. 2019: 35(1):107-114.

46. Clarke L. An introduction to economic studies, health emergencies and COVID-19, Journal of Evidence Based Medicine. 2020: 13: 161-167.

4.8 Geographic Information Systems

Authors

Qian Ye and **Shihui Guo**, State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing, China.

4.8.1 Learning objectives

To understand the following about geographic information systems (GIS):

1. The basics of GIS.
2. The role of geospatial analysis in disaster health.
3. The use and challenges of GIS in Health EDRM.

4.8.2 Introduction

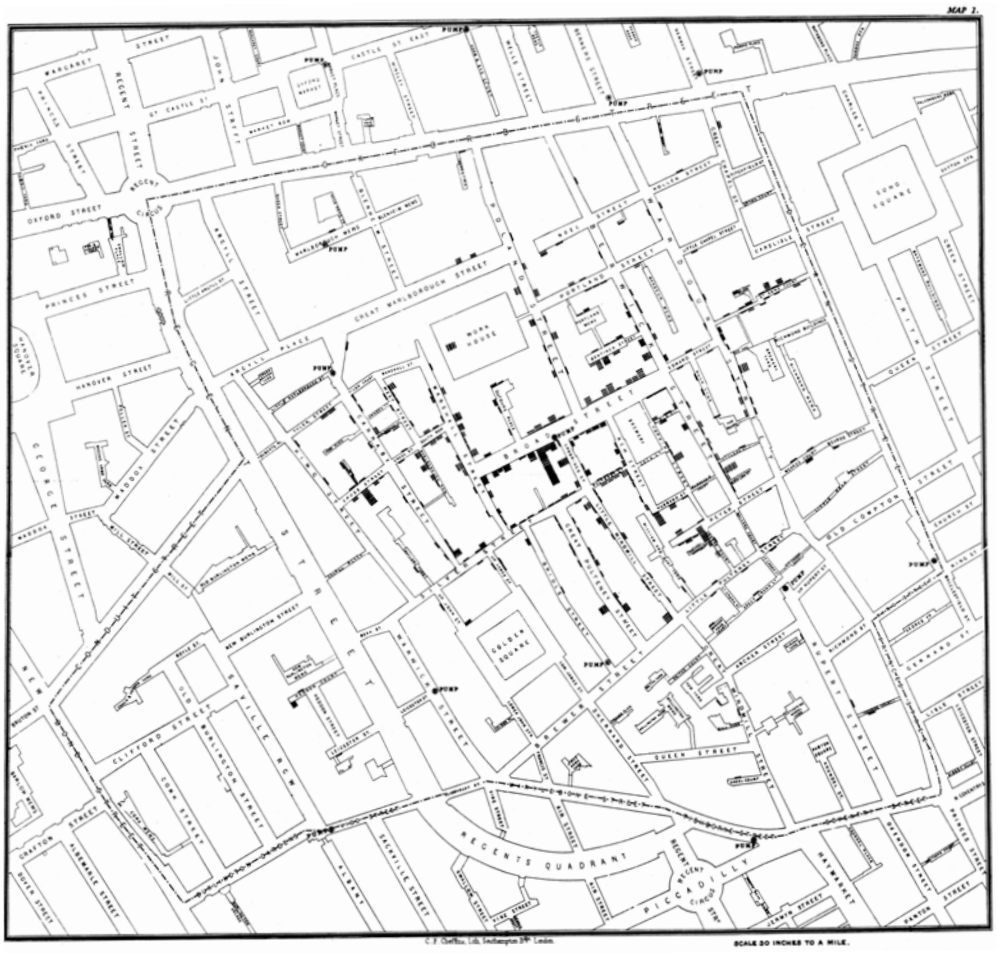
"Location, location, location" is the real estate agents' mantra, emphasizing the overwhelming importance of location on home values. This also provides a framework for the first three questions that should be asked when any disaster occurs, no matter whether it is an earthquake, typhoon, flood or something else. The first question, usually asked by everyone, is *Where has it happened?* The second question, asked mostly by those affected, is *Where are the shelters?* The third question, which is often asked by government emergency management officials, is *Where are the resources?*

The idea that place and location can influence health and safety is old and familiar in many countries and across different cultures. For example, since ancient China, Feng Shui wisdom has offered the understanding that there are a wide variety of energies in different environments, and a variety of Feng Shui methods have been developed for finding places to protect humans and their dwellings from low and attacking energies. In western history, as far back as the time of Hippocrates in the 3rd century BC, physicians have observed that certain diseases seem to occur in some places and not others. More importantly, the spatial nature of epidemiological data has long been understood and used as scientific evidence to support the development of policies to protect and improve human and animal health. In 1854, a cholera outbreak in the Soho district of London, United Kingdom, killed nearly 600 people in just ten days. John Snow, a London physician, identified a contaminated water pump by mapping the locations of water pumps and the homes of people who died of cholera (Figure 4.8.1). After instructing the authorities to remove the handle to the pump, the number of new cholera cases dropped dramatically.

4.8

In the modern digital era, people encounter features of geographic locations (such as parks, bus stops, schools, hospitals, police stations and so on) every day. This is also important for Health EDRM, where emergency preparedness and health risk reduction are essentially spatial problems. With the help of new information technology including remote sensing, computers and the internet, all location-based information can be now visualized. Moreover, based on analysis and interpretation of this information, people can better understand relationships, patterns and trends of various components in social-ecological system. This chapter introduces the general concept of GIS, outlines areas of current application in disaster health and discusses future developments.

Figure 4.8.1 John Snow's cholera map



A contaminated water pump in Broad Street proved to be the source for the spread of cholera.

Map drawn by Dr John Snow in approximately 1854; shown in Stamp, LD. 1964. A Geography of Life and Death. This redrafting leaves out some interesting bits of evidence that appeared on the original map, and in Tufte's version. For instance, there was a building across the street from the pump that had no deaths at all.

4.8.3 What is GIS?

There are many working definitions for GIS. In this chapter, GIS is defined as “a computer system that incorporates hardware, software, and infrastructure for capturing, manipulating, integrating, interrogating, modelling, analysing, and visualizing all forms of geographically referenced information.”

GIS have developed rapidly in recent years, providing powerful tools for policy support in a wide range of areas on almost all geographic and administrative levels. For different users, the effectiveness and success of GIS-based applications depends on the hardware, software, technicalities of its implementation and data quality. The design and upgrading of GIS have a close and two-way relationship with the host organization.

4.8.4 GIS hardware

In general, a complete GIS system comprises individual computers, computer configuration and networks, input devices, storage systems, output devices (such as 3D printers) and virtual reality display systems. It should be emphasized that computers for GIS usage can be mobile telephones and personal computers at the low end, or supercomputers and X-Terminals at the high end. Hardware requirements vary considerably depending on the tasks undertaken.

4.8.5 Software

The choice of software depends on the needs of the organization, the functionality desired and the money available, as well as the period during which the system is planned. There are many high quality and widely used proprietary software applications on the market, which should be compared for their costs and benefits before a particular system is chosen. To date, there are five generations of software have been developed: desktop GIS, Web GIS, GIService, Cloud GIS and Smart GIS.

The development of desktop GIS extends the GIS applications to geographic data management, analysis and visualization. Web GIS is the most used by the general public, and includes products such as Google Earth and Google Map. It allows global access to geospatial data with low barriers for using GIS software in many disciplines, thus delivering real-time data and enabling collaborative data collection and mapping across platform. GIService combines functions of GIS and Web Service. Cloud GIS helps users make better use of the power of cloud computing to provide powerful capability in storage, computation and network. Smart GIS will not only make GIS available everywhere, all the time, and for everything but will also make everything as service.

A summary of different open-source and ESRI GIS programs, showing their capabilities and functionalities is available online *(1)*.

4.8

4.8.6 GIS Database

The database is the heart of any GIS application project. The development of a GIS-based database is the first step of the project, which involves a process of data acquisition, data digitization, data modelling, and data quality assurance and quality control (QA/QC).

Data acquisition is the GIS-related information acquired in the research project area, usually including data on the ecosystem, climatology, geology, hydrology, land form, soil, and social-economy, as well as other specific information. The data are usually comprised of satellite images, hard copy maps, ground observations and data obtained from the literature.

Data digitization is the process of transforming acquired data from a variety of data formats (such as images or drawings) to a relatively standard data format such as vector and raster:

- Vector consists of features such as point, line and polygon, and is usually stored as a shape file.
- Raster consists of grid cells and pixels which can be stored as images and TIN. After this process, new data will have the same coordinates system, projection, and datum, which can be readily used by GIS software for data analysis.

Data modelling is the process of using the available data to derive additional types of data. For example, the Digital Elevation Model (DEM) and river shape files are used to derive slope, aspect and watershed.

Data QA/QC is the process of validating the GIS data transformed from different sources. Transformed data is validated by comparing the geographic coordinates of pre-determined locations to the field survey results.

Case Study 4.8.1
Map of health vulnerability and disaster risk *(2)*

To measure the health vulnerability of each country, three factors were captured from nine health indicators for the 147 countries along the Belt and Road region *(2)*: population status, disease prevention and coping capacity. Population status is related to proportion of the population aged under 15 or over 65 years of age, the mortality ratio for children under 5 years and the maternal mortality ratio. The most vulnerable countries were Sierra Leone, the Republic of Chad and the Central African Republic. Ukraine was shown to be the least vulnerable among all of the studied countries. For the second factor, disease prevention, which is related to coverage of the measles-containing-vaccine first-dose (MCV1) and diphtheria tetanus toxoid and pertussis (DTP3) vaccines, the Republic of Equatorial Guinea and Ukraine are prominent, because they had low MCV1 and DTP3 immunization coverage. For the third factor, coping capacity, which is related to physician ratio and hospital bed ratio, Thailand, the Solomon Islands and Indonesia were at the top of the scale. After combining the three factors into a health vulnerability index, Greece, the Republic of Korea and the Republic of Belarus were the three least vulnerable countries, whereas countries in Africa, including the Federal Republic of Somalia, the Central African Republic and Chad were the most vulnerable.

As disaster risk is a function of exposure, hazard and vulnerability, the top five areas with the highest disaster risk identified in this study were in locations near the Philippines, the Islamic Republic of Afghanistan, Bangladesh, Somalia and Indonesia. Northwest China, North Africa, eastern Europe and Australia were found to have relatively lower risks.

The most common usage of the GIS-based database is to quantify research objects' spatial distributions as shown in Case Study 4.8.1. The distribution of any phenomenon or indicators on the earth's surface (geographically) is called spatial distribution. As shown in this case study, mapping various selected factors, allows the health vulnerability of the country to be shown visually to answer the question "what is where?".

As stated in Tobler's First Law of Geography, "Everything is related to everything else. But near things are more related than distant things". Understanding the spatial correlations of various factors in a research region is another important application of the GIS-based database. Exploration of spatial data involves the use of statistical methods to determine whether observed patterns are random. Visualization is the most commonly used spatial analysis method, resulting in maps that describe spatial patterns as shown in Case Study 4.8.2. Models might also be used to study cause-effect relationships, to explain or predict spatial patterns.

Case study 4.8.2
Chikungunya in Latin America

Transmission of Chikungunya virus became rapidly established during 2014 in Latin America in places where dengue and its main vector, Aedes aegypti, were present. This 2014 outbreak was the start of a new endemic disease, meaning that in the countries which faced this new arboviral disease, some areas saw stabilization of its transmission with decreased incidence, while others observed a significant increase during 2015. This was the case of the Coffee-Triangle region in Colombia.

In this setting, travellers to endemic areas in Latin American countries should be aware of the risk of infective biting exposure. In order to provide advice to travellers, epidemiological maps for Chikungunya virus were developed using GIS for the Coffee-Triangle region, which is a tourist area with three departments (Caldas, Quindı´o and Risaralda) and 53 municipalities.

Use of GIS-based epidemiological maps allows the integration of preventive and control strategies, as well public health policies for control of this vector-borne disease. For example, preparedness on Chikungunya virus for healthcare workers and students in the region have increased through intense continuing education activities, including community participation on vector control for the purpose of controlling and mitigating the effects of Aedes transmission on Chikungunya virus. Because travellers might also spread the virus, GIS maps also provide relevant information to assess the risk of travellers going to specific destinations with high transmission rates. This allows prevention advice to be made available for both government officials and the general public.

4.8

4.8.7 GIS Application in disaster health

Any disaster event creates a significant short-term spike in the demand for emergency services, which will require extraordinary measures. As reported by UNDRR *(3)*, the number of natural, accidental, and intentional disasters is growing globally and is an increasing concern for governments, healthcare organizations and the public. Many research studies, in a variety of countries and regions, have shown how the appropriate use of GIS can enhance the effectiveness of the disaster risk management system, thereby safeguarding the population and the community infrastructure. Much of the responsibility for emergency medical response to emergencies and disasters rests on the healthcare sector, but other sectors need to be involved as well and one of the distinctions of disaster health is its multidisciplinary nature.

In high-income countries, many hospitals and other health care facilities are equipped with new information technologies (IT) such as wireless local area networks (LANs) with disaster medical response capabilities including personal digital assistants, tablets and handheld personal computers. Unfortunately, many disaster events overwhelm or destroy the medical infrastructure by damaging hospitals, limiting emergency supplies and closing medical clinics. Taking advantage of recent advances in IT, hospitals and disaster relief agencies could work together using GIS to develop better plan for disasters.

Studies show that when disasters strike, a comprehensive disaster medical response plan with state-of-the-art IT is essential. This needs to ensure that adequate personnel, supplies, equipment and protocols are established to meet potential threats and are at the correct scale to meet the level of the disaster.

During pre-disaster stages, Health EDRM needs accurate public health data on air, water, sanitation, utilities and community healthcare facilities. Moreover, geo-referenced baseline demographic data and health area boundaries are also important. During a disaster, healthcare organizations need to have an acceptable surge capacity, so that they are able to expand beyond normal service levels to meet an increased demand for medical care. One example of building surge capacity is the development of a national real-time, hospital-bed tracking system named the National Hospital Available Beds for Emergencies and Disasters (HAvBED) system in the USA. The system includes a GIS, established communications protocols, a database and standardized hospital bed definitions.

It is also critical to track patients and essential medical supplies in both pre-disaster and post-disaster phases, as well as during a disaster. A related planning tool is the Emergency Preparedness Resource Inventory, which is a web-based tool that can assess the regional supply of critical resources, prepare for incident management, identify deficiencies in services, and support resource acquisition decisions. The Emergency Preparedness Resource Inventory also includes an inventory checklist to record where emergency equipment and medicines are located, the amount on hand, and how to obtain additional supplies.

GIS can also provide real-time tracking of people both patients and healthcare personnel. Tracking patients during a disaster helps with improved care, family notification and the allocation of emergency services.

This type of technology can help first responders to locate patients during emergency response and transfer patients when hospitals are full, then assist in coordinating patient care as individuals are moved during the emergency care process; as illustrated in Case Study 4.8.3.

Case Study 4.8.3
Health risk distribution of people with high temperature disasters *(4)*

Global climate change is increasing the frequency of extreme weather events, which have substantial impact on human health and social economy *(5)*. As an important type of extreme weather event, extreme summer temperatures have been widespread throughout the world and will continue to increase in frequency, extent and duration *(6)*. High-temperature disasters caused by high summer temperatures directly affect human health. In 1995, the heatwave in Chicago in the USA and the heatwave in Europe in 2003 caused a large number of deaths *(7–8)*. Excessive summer temperatures will increase the incidence of cardiovascular, respiratory, digestive tract and other diseases. High temperature weather will also lead to environmental pollution caused by the accumulation of harmful gases and smoke, further threatening human health *(9)*. However, targeted space control measures such as high-temperature warnings and resource allocation can minimize expected risks. People with different characteristics have different resistance to high temperatures. Therefore, the identification of vulnerable populations and health risk assessment of high-temperature disaster populations are important for targeted disaster prevention and mitigation and resource allocation *(10)*.

Based on the disaster risk assessment framework proposed in the fifth research report of IPCC *(6)* , a conceptual model of “high temperature stress-social vulnerability-population exposure” for population health risk assessment based on high temperature disasters; combined with meteorological data, remote sensing data, and socio-economic statistical data, the GIS and RS platforms have been used to complete the high-temperature disaster risk assessment at the country level. The results show that the hot spots of high temperature disaster vulnerability are mainly distributed in underdeveloped areas, with high temperature stress or poor social economy.

4.8.8 Challenges

Many studies show that healthcare organizations that invest appropriately in IT, including GIS, can improve the quality and efficiency of their healthcare services. In particular, when these investments are incorporated into disaster plans, it leads to benefits for emergency medical response and to other aspects of Health EDRM. However, there are concerns about the safety of the data which GIS collects, stores, analyses and displays; as noted in Case Study 4.8.4.

Case Study 4.8.4
GIS for population-wide health monitoring in the Federal Republic of Germany

In recent years, GIS have become an integral part of public health research. They offer a broad range of analysis tools, which enable innovative solutions for health-related research questions. An analysis of nationwide studies in Germany that applied GIS underlines the potential of GIS for health monitoring in Germany. GIS provide up-to-date mapping and visualization options to be used for national health monitoring at the Robert Koch Institute (RKI). Objective information on the residential environment as an influencing factor on population health and health behaviour can also be gathered and linked to RKI survey data at different geographic scales. Besides using physical information, such as climate, vegetation or land use, as well as information on the built environment, the instrument can link socioeconomic and sociodemographic data to information on health care and environmental stress with the survey data. This allows integration of the data into concepts for analyses. In this way, GIS expands the potential of the RKI to present nationwide, representative and meaningful health-monitoring results. However, in doing so, data protection regulations must always be followed. This balance of the safety of the data with the development of a national spatial data infrastructure and the identification of important data sources that can improve access to high quality data sets relevant for the health monitoring, is an important element in the development of this GIS.

Another challenge associated with implementing GIS in a robust medical disaster response plan is the cost associated with many of the necessary tools. In order to better serve their patients, continuous financial support for accurate, update and sufficient information is needed by healthcare organizations. This is particularly significant in rural US and in low-income countries. The level of regional, national and international efforts to manage disasters also urgently requires a coordinated GIS-based approach that connects local, state, and national emergency programs.

The third challenge to disaster medical response is the effective use of multiple data sources to develop a coordinated management approach *(11)*. The use of wireless LANs, GIS technology, patient-tracking systems and online medical resource databases will improve disaster medical response including early disaster event detection, outbreak management, connecting laboratory systems, response administration, communications and public health alerts; but will need good coordination. These technologies will improve patient care and safety, as well as provide for better command and control, leading to more efficient resource utilization. However, GIS will only make a powerful contribution if they include reliable and representative underlying baseline and situational data. The quality of these data needs to be carefully considered while interpreting the results. To help users better understand the complex situation, the choice of the GIS visualization method (for example, colour or grouping of the variables in a map) can also affect the overall interpretation of the situation.

4.8.9 Conclusions

GIS technology is expanding its application into Health EDRM, covering and going beyond disaster health risk detection, modelling, assessment, response planning and public health policy development. The development and maintenance of disaster health management systems based on GIS, however, not only depend on technology but also involve many components in a complex social-ecological system. Multi- and trans-disciplinary trained professionals equipped with relevant information technologies are crucial to meet the current and future challenges of using GIS in disaster health science.

4.8.10 Key messages

- **A main strength of GIS lies in its powerful ability to combine, analyse and display spatial and attribute data.**
- **This will help to satisfy the need for large-scale data analysis and processing in disaster response planning and improve Health EDRM.**

4.8.11 Further reading

Mansour S. Spatial analysis of public health facilities in Riyadh Governorate, Saudi Arabia: a GIS-based study to assess geographic variations of service provision and accessibility. Geo-spatial Information Science; 2016: 19(1): 26-38.

Yafei Z, Mao L. GIS-based urban fire risk assessment and its application in disaster mitigation planning. Journal of Catastrophology; 2010: 25(S1): 258-63.

Nagata T, Kimura Y, Ishii M. Use of a Geographic Information System (GIS) in the Medical Response to the Fukushima Nuclear Disaster in Japan. Prehospital and Disaster Medicine; 2012: 27(2): 213-15.

Kawasaki A, Berman M L, Guan W. The growing role of web-based geospatial technology in disaster response and support. Disasters; 2013: 37(2): 201-21.

4.8.12 References

1. Comparison of Geographic Information Systems (GIS) software. WHO. 2018. www.who.int/health-cluster/resources/publications/OpenSourceGISComparison.pdf (accessed 12 January 2020).
2. Chan EYY, Huang Z, Lam HCY, Wong CKP, Zou Q. Health vulnerability index for disaster risk reduction: application in Belt and Road Initiative (BRI) region. International Journal of Environmental Research and Public Health; 2019: 16(3): e380.
3. Global Assessment Report on Disaster Risk Reduction, United Nations Office for Disaster Risk Reduction. Geneva, Switzerland. 2019.
4. Pan X, Yanglin W, Yanxu L, Jian P. Incorporating social vulnerability to assess population health risk due to heat stress in China. Acta Geographica Sinica; 2015: 70(7): 1041-51.
5. Quansheng G, Jingyun Z, Hao Z, Zhang X, Fang X, Wang H, Yan J. State-of-the-arts in the study of climate changes over China for the past 2000 years. Acta Geographica Sinica. 2014: 69(9): 1248-58.
6. Intergovernmental Panel on Climate Change (IPCC). Climate Change 2014: Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK and New York, USA: Cambridge University Press. 2014. https://archive.ipcc.ch/report/ar5/wg2/ (accessed 4 March 2020)
7. Semenza et al. Semenza JC, Rubin CH, Falter KH, Selanikio JD, Flanders WD, Howe HL, Wilhelm JL. Heat-related deaths during the July 1995 heat wave in Chicago. New England Journal of Medicine; 1996: 335(2): 84-90.
8. Stott et al, 2004 Stott PA, Stone DA. Human contribution to the European heat wave of 2003. Nature 432; 2004: (7017): 610-4.
9. Patz JA, Campbell-Lendrum D. Impact of regional climate change on human health. Nature; 2005: 438(7066): 310-7.
10. Gulrez A, Shubhayu S, Partha G, et al. Heat Wave Vulnerability Mapping for India. International Journal of Environmental Research and Public Health; 2017: 14(4): 357.
11. Chan TC, Killeen J, Griswold W, Lenert L. Information technology and emergency medical care during disasters. Academic Emergency Medicine; 2004: 11(11): 1229-36.

4.9 Real-time Syndromic Surveillance

Authors

Alex J. Elliot, Real-time Syndromic Surveillance Team, Public Health England, Birmingham, United Kingdom; National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections, University of Liverpool, Liverpool, United Kingdom.

Helen E. Hughes, Real-time Syndromic Surveillance Team, Public Health England, Birmingham, United Kingdom; National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections, University of Liverpool, Liverpool, United Kingdom.

Sally E. Harcourt, Real-time Syndromic Surveillance Team, Public Health England, Birmingham, United Kingdom.

Roger A. Morbey, Real-time Syndromic Surveillance Team, Public Health England, Birmingham, United Kingdom; National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Emergency Preparedness and Response, King's College London, London, United Kingdom.

Sue Smith, Real-time Syndromic Surveillance Team, Public Health England, Birmingham, United Kingdom.

Gillian E. Smith, Real-time Syndromic Surveillance Team, Public Health England, Birmingham, United Kingdom; National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Emergency Preparedness and Response, King's College London, London, United Kingdom.

4.9.1 Learning objectives

To understand the key factors underpinning real-time syndromic surveillance systems and the use of syndromic surveillance data in research, including:

1. The definition of syndromic surveillance;
2. Data sources for syndromic surveillance;
3. Governance issues;
4. Data analysis and statistics;
5. The application of syndromic surveillance in research.

4.9

4.9.2 Introduction

Syndromic surveillance is the near real-time collection, analysis, interpretation and dissemination of health-related data in order to enable the early identification of the impact (or absence of impact) of potential health threats that may require public health action *(1)*. Although a relatively new field in comparison to more established methods of surveillance (such as using laboratory reports), syndromic surveillance is growing in stature internationally as it becomes recognized as an innovative approach to public health surveillance. The advantages that syndromic surveillance brings to the identification and investigation of public health threats, including those relevant to health emergency and disaster risk management (Health EDRM), include early warning, situational analysis, reassurance and flexibility.

Early warning
Many syndromic surveillance systems operate in near-real-time (daily, for example), allowing the timely identification of, and response to incidents.

Situational awareness
During an incident, syndromic surveillance systems enable further description of healthcare seeking behaviour in near real-time (daily, for example) providing key intelligence to incident managers and response teams (such as identifying particularly affected age groups, geographical clusters).

Reassurance
During mass gatherings and other similar events, syndromic surveillance can often provide reassurance that there have been no widespread acute public health problems, particularly where surveillance is long term and a 'normal' or historical baseline level has been established prior to the event.

Flexibility
By using broad and adaptable syndromes, syndromic surveillance systems can be flexible in responding to a variety of public health demands ranging from infectious disease outbreaks to environmental incidents and mass gatherings, in addition to providing measures of impact of public health interventions – vaccination impact, for example. Syndromic surveillance also has the potential to detect newly emerging threats not covered by existing surveillance systems.

In general, syndromic surveillance makes opportunistic use of anonymized data collected either as part of standard patient care from healthcare service providers, or proxies of population health (for example, information on accessing of health advice from other sources; see also Chapter 2.1). This information is collected by the healthcare provider or advisor, usually during the contact with the patient and before any final confirmation of a diagnosis or cause of illness. The data used for syndromic surveillance therefore contain valuable detail of symptoms, chief complaints, clinical diagnoses, or other proxies for healthcare seeking behaviour. Furthermore, as this information is collected contemporaneously these data can be made available and used for syndromic surveillance purposes very quickly – often the following day, if not sooner *(2)*.

Syndromic surveillance collates the information received and groups it into syndromes of public health relevance (Table 4.9.1). Each syndrome is constructed from the symptoms, chief complaints or clinical diagnoses, as they have been recorded in the patient record. The format of the data is often data provider specific, based on how information is organized and stored in the local patient record, which may use a standardized coding system, a locally used list of clinical terms or even free text. For example, general practitioners (GPs) managing a patient with acutely presenting asthma use clinical codes (such as ICD-10, SNOMED-CT or Read codes *(3–5)*) to record the clinical management of the patient. Asthma monitored in a syndromic surveillance system would be based on the identification of those patient contacts including clinical asthma codes.

Table 4.9.1 Examples of syndromic surveillance syndromes that are flexible in responding to a range of public health threats

Syndrome monitored	Related public health threats
Asthma	Respiratory pathogens, air pollution, chemical incidents, wild or industrial fires, severe thunderstorms
Fever	Influenza, respiratory pathogens, heatwave (infants)
Difficulty breathing	Air pollution, respiratory pathogens, chemical incidents, wild or industrial fires
Diarrhoea and vomiting	Gastrointestinal pathogens, flooding
Conjunctivitis	Respiratory pathogens, chemical incidents, wild or industrial fires, allergic rhinitis
Cough	Influenza, respiratory syncytial virus (children aged <5 years), respiratory pathogens, chemical incidents, wild or industrial fires

Syndromic surveillance does not generally monitor laboratory confirmed reports. Although a lack of laboratory confirmation (and therefore the absence of a direct link to a causal pathogen) presents a potential limitation in the *specificity* of reporting (particularly around infectious diseases), it can also be an advantage as the flexibility of the systems enables greater *sensitivity* due to the broadness of data collected and the volume of information available. The flexibility of syndromic surveillance systems enables them to respond to a variety of public health incidents, ranging from infectious diseases *(6–7)* to environmental events *(8)*, mass gatherings *(9–10)*, terrorism *(11)*, recovery from disasters caused by natural hazards *(12–13)* or investigations of vaccination impact *(14)*. A single syndrome may be relevant to several different public health issues (Table 4.9.1). For example, a newly emerging respiratory pathogen may not be detected by existing laboratory tests, but increases in numbers of presentations, or severity of illness, in symptomatic patients presenting to healthcare services would be captured by syndromic data.

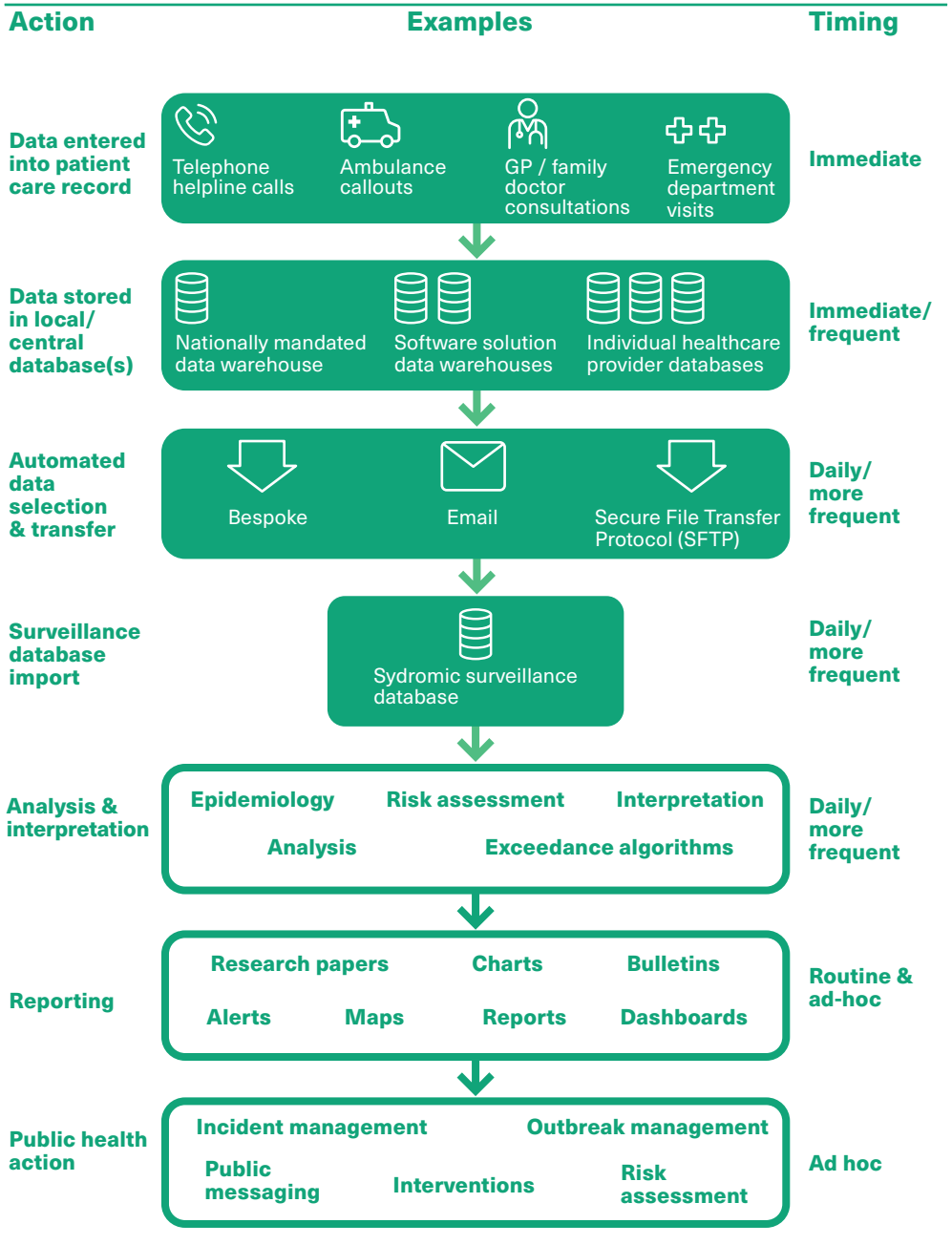
Syndromic surveillance systems also have the advantage of providing wider population surveillance, covering whole regions or countries, at different levels of patient care (from those requesting advice only, to those requiring urgent emergency treatment), providing a picture of the levels of

severity of disease within the community. Laboratory-based surveillance, however, is often biased, based upon only those sampled for testing, which is often limited to patients with ongoing illness, who are more severely ill or hospitalized, or are considered to be at-risk of complications or death. Laboratory surveillance therefore monitors only a fraction of the total burden of disease.

While there are fundamental differences between syndromic and laboratory-based surveillance, it is important that both are synergistic, complementing each other to ensure the delivery of a functioning public health surveillance programme. Without laboratory surveillance, it is difficult to determine the underlying pathogens driving seasonal trends in syndromic data; without syndromic surveillance, it is difficult to establish representative community-based estimates of burden.

The collection of information for syndromic surveillance is normally automated, with electronic transmission of anonymized data from healthcare service providers to public health organizations. Figure 4.9.1 illustrates how health data might flow in a multi-partite syndromic surveillance system. The automation of data collection removes the requirement to ask data providers to undertake additional time-consuming tasks or to remember to flag individual records for inclusion in a syndromic surveillance system. Automation is critical to the success of such systems, especially those based upon healthcare services. Data are collected as part of the usual patient care or advice process. No extra steps or changes to working practices are required by the data providers for syndromic surveillance to be possible.

Figure 4.9.1 Example data flow for a multi-partite syndromic surveillance service



Source: Public Health England Real-time Syndromic Surveillance Team.

4.9

4.9.3 Data sources for syndromic surveillance

Data for syndromic surveillance are commonly sought from a range of healthcare services including primary care providers or GPs, emergency departments (EDs), telehealth services and ambulance services.

Primary care/general practitioners/physicians/family doctors
Primary care surveillance is often considered a gold standard for assessing community morbidity. Syndromes are usually constructed using clinical diagnoses as recorded by the treating physician at the time of the consultation.

Emergency departments
EDs are frequently used for syndromic surveillance, particularly in countries where access to primary care data may not be readily available. ED surveillance provides a metric for more severe presentation of disease or conditions. Syndromes may be constructed from chief or presenting complaints, or clinical diagnoses, depending on the timescale at which the information is available.

Telehealth services
Telehealth surveillance can provide access to populations not captured through ED or primary care surveillance, such as those who are less ill and require advice, rather than urgent care. Traditionally considered to provide early warning over other systems, the syndromes used are based on patient reported symptoms and may have the lowest specificity.

Ambulance services
Monitoring ambulance dispatch calls can provide an additional measure of acute, potentially more severe presentation of diseases or conditions in public health surveillance.

Outside the healthcare setting, many additional data sources have been used for syndromic surveillance. School absenteeism, employee absenteeism and over the counter pharmacy sales are examples where data represent proxies for disease. These sources have been usefully adopted for monitoring the health of the population *(15)*.

In recent years, with the advent and increasing use of digital platforms to access healthcare and advice, more public health resource has focused on assessing the potential benefits of using 'digital data' such as web searches (such as Google *(16)*), social media activity (such as Twitter *(17)*) and online health services (an online 'symptom-checker', for example *(18)*). The methods used for accessing and collecting data continue to develop, evolving from platforms such as messaging services (for example, HL7 *(19)*) to techniques suited for trawling big data (for example, data mining or natural language processing *(20)*.

4.9.4 Governance

Although it is often overlooked in the published syndromic surveillance literature, the adherence to good governance and data security practices around the collection, storage, processing and use of healthcare data for syndromic surveillance is important. Establishing a syndromic surveillance system (either at national or subnational level) requires multiple phases undertaken by a multi-disciplinary group. This has previously been described by experienced exponents of syndromic surveillance *(1)*. However, one of the key areas that will determine the sustainability of a system is establishing appropriate governance arrangements with data providers to assure the correct use and secure storage of data, as well as the competence of trained specialist staff accessing, analysing and interpreting data. Without such assurances, data are unlikely to be made available for syndromic surveillance.

The governance arrangements underpinning syndromic surveillance systems are equally essential for the long-term success of systems. Without appropriate governance, these surveillance systems are not fit for purpose and are likely to fail. Alongside governance, appropriate management and oversight of syndromic surveillance systems is important for their success, with collaboration between data providers and public health intelligence teams to steer the development and management of the systems. Management through steering or strategic groups, including senior members from all organizations involved in delivering the system is crucial to long term success, fruitful outputs and assurance of the public health benefits of the surveillance system. Collaboration may involve a wide range of organizations including data providers, technology firms providing data collection or transfer systems, public health bodies, clinical groups, academics and professional bodies. Furthermore, these steering groups might be used as a conduit to ensure that research undertaken using the syndromic surveillance data is appropriate (that is, with a public health focus), undertaken with appropriate rigour and, most importantly, that it does not undermine any organization involved in the collaborative surveillance system.

4.9.5 Analysis of syndromic surveillance data

There are many methods used to routinely analyse syndromic surveillance data. The underlying principle of syndromic surveillance is the analysis of trends, rather than identifying individual cases. Traditional descriptive epidemiological methods can be used to examine patterns in disease over time, by person and place, and formal statistical tests can be used to detect anomalies (Figure 4.9.2).

Time

Syndromic surveillance data are analysed over time to identify short term increases in syndromes (suggesting outbreaks of disease, for example), environmental impacts (air pollution, for example) and long-term changes in trend (suggesting changes in disease burden).

Person

Data can be broken down by patient demographics (such as age or gender) to identify changes in burden, which may be indicative of public health threats.

4.9

Place

Where possible, links to the location of the patient (either area of residence or place of healthcare consultation) can be used to identify clusters or map the spread of activity.

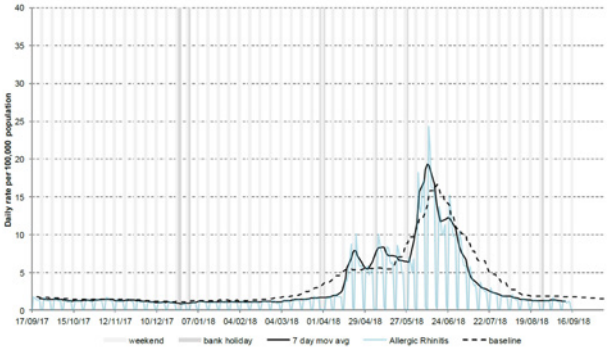
Anomaly detection

Statistical algorithms are used to automatically identify unusual activity. Statistical tests can also be used for anomaly detection or aid interpretation of syndromic data. A wide range of different statistical methods have been used for anomaly detection, including control charts, regression and time series analysis *(21–22)*. Statistical methods can also be applied to the development of historical baselines, which can supplement the interpretation of syndromic data by comparing the observed values to historically expected levels *(23)*.

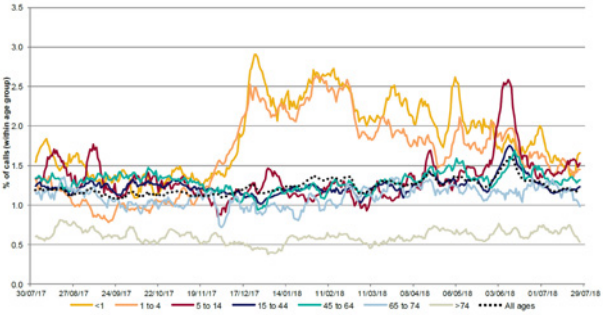
A further important consideration is the translation of complex information (as produced by epidemiological or statistical analyses) into public health action, a core component of the definition of surveillance *(24).* This element of syndromic surveillance is not well described in the literature but there are examples available of risk assessment processes designed to assess statistical exceedances by examining relevant epidemiological information and assigning an appropriate response – for example, whether no further action is required, or whether the information needs to be sent to a relevant public health expert for further action *(25)*.

Figure 4.9.2 Analysis of syndromic surveillance data using A) time, B) person, C) place and D) anomaly detection

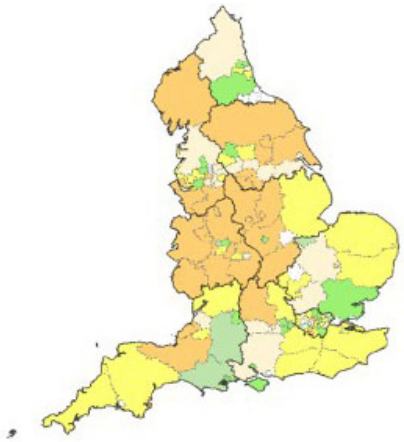
A. Time: daily GP consultation rate for allergic rhinitis



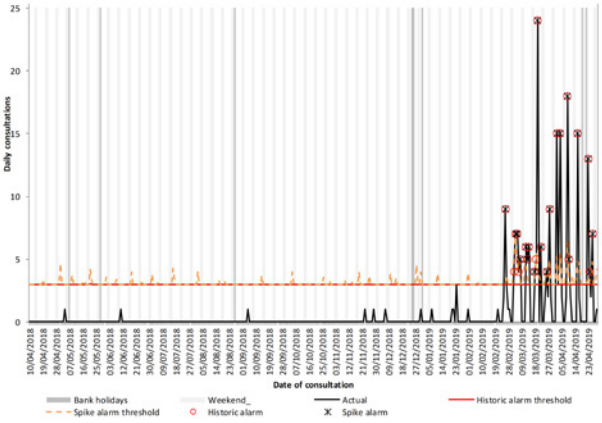
B. Person: telehealth calls for eye problems by age group



C. Place: map of GP consultations for influenza-like illness (England)



D. Anomaly detection: daily statistical exceedances for mumps



Source: PHE Real-time Syndromic Surveillance Team
Map contains Ordnance Survey data. ©Crown copyright and database right 2018. Contains National Statistics data.

4.9

4.9.6 Using syndromic surveillance in research

The collection of health data to deliver real-time syndromic surveillance can provide a rich resource for Health EDRM researchers to address important public health questions. Alongside the use of other sources of public health data, a wide range of research methods outlined elsewhere in this book can be used alongside syndromic surveillance data. However, syndromic surveillance data are not collected specifically for research purposes and therefore when considering the use of syndromic surveillance data in research, it is important to understand several key limitations of these data, which might limit their application in certain research projects (Table 4.9.2).

Table 4.9.2 Limitations of syndromic surveillance data that need to be assessed when considering its use in research projects

Limitation	Detail
Anonymized records	Syndromic data tend to be anonymised and therefore patient-level data cannot be linked to other records or databases and cannot be used to trace patients or undertake further studies (for example, selecting controls for case-control type analysis)
Population level	Syndromic data tend to be aggregated to population level and often cannot be used for secondary analyses on an individual level
System coverage	Some syndromic systems do not have full or representative coverage geographically (country or region), or person level (such as different age groups: paediatric or adult EDs) or other limitations on access to healthcare
Coding	Clinical coding used to define syndromes can be limited or very generic or, if free text is provided this might require additional analytical skills
Symptom based	Syndromic data are not based on confirmed laboratory reports and, therefore, are not directly attributable to specific pathogens
Data quality	Syndromic data are not 'cleaned' before being used for surveillance. Consequently, compared to other health data sources used by researchers, there is a greater risk of data errors (for example, duplications, miss-entry of age data, incorrect coding or incomplete data fields)
Incomplete data	Syndromic data only uses data available in real-time, taking a 'snapshot' of daily activity. Therefore some data will be excluded due to transfer issues or time taken to confirm diagnoses. For example, most GP pneumonia diagnoses occur after laboratory confirmation and are not available in a next-day extract.

Case Studies 4.9.1, 4.9.2 and 4.9.3 describe examples of published research projects where syndromic surveillance data have been used to respond to a public health problem.

Case Study 4.9.1
Assessing potential health impacts of mass gatherings and sporting events *(26)*

Mass gatherings can impact on the health of the public, including both infectious and non-communicable diseases or conditions. Specifically, the increased risk from infectious diseases includes importation, exposure of visitors to endemic diseases in the host country and increased disease transmission across large populations gathered in one location. Surveillance during mass gatherings is needed to identify and quantify any impact (or reassure that there is an absence of impact) on public health in a timely manner. Subsequently, research on specific areas following an event can inform priorities for healthcare providers and public health organizations at future events.

Large sporting events (for example, the Olympics or world or continental football championships) have the potential to influence the behaviour of the population, and increase (or decrease) demand on health services around the timings of individual events. Of particular note, the impact of sporting events on ED attendances has been documented *(26)*. The 2016 European Football Championship (Euro 2016) was hosted in France, involving 24 nations with 51 matches during a four-week period. To assess the potential impact of Euro 2016 on healthcare seeking behaviour in different nations, syndromic surveillance ED data from four participating countries (England, France, Northern Ireland and Wales) were analysed retrospectively to identify any relevant impacts of matches played. This study focussed on hourly ED attendances across each country. In the four hours before matches were played by the national team, attendances were statistically significantly lower than would be expected in all countries, and reduced further during matches. Following the completion of matches, there was no consistent significant increase in attendances. However, these observed impacts were highly variable between individual matches. For example, in the four hours after the final match, involving France, the number of ED attendances in France increased significantly. Overall, these results indicated relatively small impacts of major sporting events upon ED attendances.

4.9

Case Study 4.9.2
Assessing the impact of air pollution on health using syndromic surveillance *(27, 28)*

Globally, air pollution is the biggest environmental risk to health, carrying responsibility for about one in every nine deaths annually. It is estimated that 91% of the world's population lives in places where air quality exceeds WHO guideline limits *(29).* Syndromic surveillance systems present an opportunity to assess the acute impact of air pollution on the health of the population. The utility of syndromic surveillance for this purpose has been demonstrated by the identification and monitoring of healthcare seeking behaviour during periods of poor air quality (air pollution). In this scenario, research involving syndromic surveillance data would require a methodological approach to determine whether existing data collected prospectively over a defined time period can be assessed against air quality data. Different research methods may include using numbers or rates for each syndrome or statistical exceedance data to identify periods of unusual syndromic activity. These events can then be compared to air quality data highlighting periods of poor air quality to identify concurrent activity.

More complex research approaches may incorporate the inclusion of further variables and confounders, which might influence the outcome of the relationship between healthcare seeking behaviour and air quality. For example, meteorological variables (such as temperature), environmental variables (such as pollen and spore counts) or pathogen activity (such as influenza laboratory reports) can all be included in models which explore the relationship between air quality and syndromic data. The results of this research can be used to assure prospective surveillance during air pollution incidents by providing baselines for future interventions and adding to the knowledge base. Furthermore, this research provides information on the specificity and sensitivity of syndromic surveillance systems and uses syndromic surveillance data to explore which pollutants drive changes in healthcare seeking behaviours *(28)*.

Case Study 4.9.3
Determining the likely impact of a new vaccine programme using syndromic surveillance *(14)*

Syndromic surveillance can contribute to research investigating the impact of public health interventions, for example, the impact of the introduction of new vaccines on the health of the population. Whilst national vaccination programmes will employ large scale evaluations to assess the impact of the new vaccine on confirmed outcomes, syndromic surveillance can contribute a rapid assessment of the impact. An anticipated outcome of the introduction of a new vaccine might be reduced disease incidence and thus fewer healthcare visits, something which is measured by syndromic surveillance as standard.

Interrupted time series and 'before-after' study methods (Chapter 4.1) can be used to assess the impact of a new vaccine on the demand for healthcare services. These research methods involve measuring the outcome of interest before and after the programme, service or intervention has been implemented. Syndromic data collected before the introduction of the intervention are compared to equivalent data collected after the event. Statistical comparisons of syndromic surveillance data, for example, in pre- and post-vaccine periods, can inform the interpretation of the likely impact of the intervention or vaccine.

In the United Kingdom, rotavirus vaccine was introduced in 2013 and integrated into the routine immunization schedule for young infants. Syndromic surveillance was used to provide an early indication of the potential impact of the introduction of the rotavirus vaccine. Syndromes were chosen based on the anticipated outcome affected by the introduction of the vaccine: GP and ED gastroenteritis, diarrhoea and vomiting syndromes were retrospectively assessed across different age groups, but particularly focussed on young children. Incidence rate ratios (IRRs) were used to compare (statistically) the period of activity pre-vaccine introduction with activity post-vaccine. IRRs showed an approximate 30% decrease in gastroenteritis incidence in infants and children aged 1 to 4 years.

Syndromic surveillance thus revealed a marked decline in gastroenteritis, coinciding with the introduction of the new rotavirus vaccine programme in England (*14)*. This model for contributing to the assessment of the impact of vaccine has been applied to other areas including the live attenuated influenza vaccine *(30)* and meningococcal B vaccine *(31)*, and will be applied to future vaccines as and when they are licensed and introduced (such as respiratory syncytial virus, norovirus).

4.9

4.9.7 Conclusions

Syndromic surveillance can complement existing public health surveillance programmes, introducing new intelligence for identifying and managing incidents. The flexibility of these systems supports a range of public health issues, including infectious disease activity to Health EDRM. Healthcare service data have traditionally underpinned syndromic surveillance systems, however, novel sources including social media and internet-based data are being explored for their potential added benefit.

4.9.8 Key messages

- **Syndromic surveillance systems can augment existing public health surveillance programmes, providing early warning and introducing real-time intelligence and reassurance at a national, regional and local level.**
- **Compared to traditional surveillance systems, syndromic surveillance can provide a more flexible approach to surveillance, enabling multi-purpose surveillance including emerging threats.**
- **Adherence to good governance and data security practices around the collection, storage, processing and use of syndromic surveillance data is essential for the long-term success of systems.**
- **Syndromic surveillance data are a valuable resource for public health research, including in Health EDRM, but specific limitations of syndromic surveillance for research need to be considered.**
- **Syndromic surveillance systems gain value in research data sources when operated consistently over time enabling comparison to historical data.**

4.9.9 Further reading

Conway M, Dowling JN, Chapman WW. Using chief complaints for syndromic surveillance: a review of chief complaint based classifiers in North America. Journal of Biomedical Informatics; 2013: 46: 734-43.

Josseran L, Fouillet A, Caillere N, Brun-Ney D, Ilef D, Brucker G, et al. Assessment of a syndromic surveillance system based on morbidity data: results from the Oscour network during a heat wave. PLoS One;2010: 5: e11984.

Smith GE, Elliot AJ, Ibbotson S, Morbey R, Edeghere O, Hawker J, et al. Novel public health risk assessment process developed to support syndromic surveillance for the 2012 Olympic and Paralympic Games. Journal of Public Health;2017: 39: e111-e7.

Triple-S: Syndromic Surveillance Systems in Europe. Guidelines for designing and implementing a syndromic surveillance system. 2013. https://webgate.ec.europa.eu/chafea_pdb/assets/files/pdb/20091112/20091112_d08_giss_en_ps.pdf (accessed 15 January 2020).

Yoon PW, Ising AI, Gunn JE, editors. Syndromic surveillance: the value of real-time data for public health action. Public Health Reports; 2017: 132: 1S-126S.

4.9.10 References

1. Triple-S: Syndromic Surveillance Systems in Europe (2013). Guidelines for designing and implementing a syndromic surveillance system https://webgate.ec.europa.eu/chafea_pdb/assets/files/pdb/20091112/20091112_d08_giss_en_ps.pdf (accessed 15 January 2020).
2. Henning KJ. What is syndromic surveillance? MMWR Supplements; 2004: 53: 5-11.
3. de Lusignan S. Codes, classifications, terminologies and nomenclatures: definition, development and application in practice. Informatics in Primary Care; 2005: 13: 65-70.
4. International Health Terminology Standards Development Organisation. SNOMED-CT. 2012 http://www.ihtsdo.org/snomed-ct (accessed 15 January 2020).
5. International Classification of Disease – version 10 (ICD-10). WHO. 2016. https://www.who.int/classifications/icd/icdonlineversions/en (accessed 15 January 2020).
6. Olson DR, Heffernan RT, Paladini M, Konty K, Weiss D, Mostashari F. Monitoring the impact of influenza by age: emergency department fever and respiratory complaint surveillance in New York City. PLoS Medicine; 2007: 4: e247.
7. Hughes SL, Morbey RA, Elliot AJ, McEwen SA, Greer AL, Young I, et al. Monitoring telehealth vomiting calls as a potential public health early warning system for seasonal norovirus activity in Ontario, Canada. Epidemiology and Infection; 2019: 147: e112.

4.9

8. Josseran L, Fouillet A, Caillere N, Brun-Ney D, Ilef D, Brucker G, et al. Assessment of a syndromic surveillance system based on morbidity data: results from the Oscour network during a heat wave. PLoS ONE; 2010: 5: e11984.
9. Fleischauer AT, Gaines J. Enhancing surveillance for mass gatherings: the role of syndromic surveillance. Public Health Reports; 2017: 132: 95S-8S.
10. McCloskey B, Endericks T, Catchpole M, Zambon M, McLauchlin J, Shetty N, et al. London 2012 Olympic and Paralympic Games: public health surveillance and epidemiology. Lancet, 2014: 383: 2083-9.
11. Vandentorren S, Paty AC, Baffert E, Chansard P, Caserio-Schonemann C. Syndromic surveillance during the Paris terrorist attacks. Lancet; 2016: 387: 846-7.
12. Griffith MM, Yahata Y, Irie F, Kamiya H, Watanabe A, Kobayashi Y, et al. Evaluation of an ad hoc paper-based syndromic surveillance system in Ibaraki evacuation centres following the 2011 Great East Japan Earthquake and Tsunami. Western Pacific Surveillance and Response Journal; 2018: 9: 21-7.
13. Tante S, Villa E, Pacho A, Galvan MA, Corpuz A. Which surveillance systems were operational after Typhoon Haiyan? Western Pacific Surveillance and Response Journal; 2015: 6: 66-70.
14. Bawa Z, Elliot AJ, Morbey RA, Ladhani S, Cunliffe NA, O'Brien SJ, et al. Assessing the likely impact of a rotavirus vaccination program in England: the contribution of syndromic surveillance. Clinical Infectious Diseases; 2015: 61: 77-85.
15. Heffernan R, Mostashari F, Das D, Besculides M, Rodriguez C, Greenko J, et al. New York City syndromic surveillance systems. MMWR Supplements; 2004: 53: 23-7.
16. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS ONE; 2014: 9: e109583.
17. Gesualdo F, Stilo G, Agricola E, Gonfiantini MV, Pandolfi E, Velardi P, et al. Influenza-like illness surveillance on Twitter through automated learning of naive language. PLoS ONE; 2013: 8: e82489.
18. Elliot AJ, Kara EO, Loveridge P, Bawa Z, Morbey RA, Moth M, et al. Internet-based remote health self-checker symptom data as an adjuvant to a national syndromic surveillance system. Epidemiology and Infection; 2015: 143: 3416-22.
19. Health Level 7 International. HL7 International. 2019. http://www.hl7.org/about (accessed 15 January 2020).
20. Ye Y, Tsui FR, Wagner M, Espino JU, Li Q. Influenza detection from emergency department reports using natural language processing and Bayesian network classifiers. Journal of the American Medical Informatics Association; 2014: 21: 815-23.

21. Noufaily A, Morbey RA, Colon-Gonzalez FJ, Elliot AJ, Smith GE, Lake IR, et al. Comparison of statistical algorithms for daily syndromic surveillance aberration detection. Bioinformatics; 2019: 35: 3110-8.

22. Unkel S, Farrington CP, Garthwaite H, Robertson C, Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. Journal of the Royal Statistical Society Series A (Statistics in Society) 2012: 175: 49-82.

23. Harcourt SE, Morbey RA, Smith GE, Loveridge P, Green HK, Pebody R, et al. Developing influenza and respiratory syncytial virus activity thresholds for syndromic surveillance in England. Epidemiology and Infection; 2019: 147: e163, 1–7.

24. Porta M. A Dictionary of Epidemiology: Oxford University Press, USA; 2008.

25. Smith GE, Elliot AJ, Ibbotson S, Morbey R, Edeghere O, Hawker J, et al. Novel public health risk assessment process developed to support syndromic surveillance for the 2012 Olympic and Paralympic Games. Journal of Public Health; 2017: 39: e111-e7.

26. Hughes HE, Colon-Gonzalez FJ, Fouillet A, Elliot AJ, Caserio-Schonemann C, Hughes TC, et al. The influence of a major sporting event upon emergency department attendances; A retrospective cross-national European study. PLoS ONE; 2018: 13: e0198665.

27. Harcourt S, Izon-Cooper L, Colón-González FD, Morbey R, Smith G, Bradley N, et al. Using real-time syndromic surveillance to monitor the health effects of air pollution. Online Journal of Public Health Informatics 2018: 10(1): e85.

28. Smith GE, Bawa Z, Macklin Y, Morbey R, Dobney A, Vardoulakis S, et al. Using real-time syndromic surveillance systems to help explore the acute impact of the air pollution incident of March/April 2014 in England. Environmental Research; 2015: 136: 500-4.

29. Ambient air pollution: a global assessment of exposure and burden of disease. WHO. 2016 https://www.who.int/phe/publications/air-pollution-global-assessment/en (accessed 15 January 2020).

30. Pebody RG, Sinnathamby MA, Warburton F, Andrews N, Boddington NL, Zhao H, et al. Uptake and impact of vaccinating primary school-age children against influenza: experiences of a live attenuated influenza vaccine programme, England, 2015/16. Eurosurveillance; 2018: 23: pii=1700496.

31. Harcourt S, Morbey RA, Bates C, Carter H, Ladhani SN, de Lusignan S, et al. Estimating primary care attendance rates for fever in infants after meningococcal B vaccination in England using national syndromic surveillance data. Vaccine; 2018: 36: 565-71.

4.10 Using logic models in research and evaluation of Health EDRM interventions

Authors

Dylan Kneale, **Mukdarut Bangpan**, and **James Thomas**, Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre), University College London, London, United Kingdom.

Hugh Sharma Waddington, International Initiative for Impact Evaluation (3ie), London, United Kingdom; London School of Hygiene and Tropical Medicine, London, United Kingdom.

4.10.1 Learning objectives

To understand the following about the use of logic models in Health EDRM:

1. The importance of logic models for research and evaluation in Health EDRM;
2. Methods for constructing and using a logic model to guide research and evaluation projects.

4.10.2 Introduction

This chapter outlines how logic models can be used to conceptualize how interventions are intended to work, and their relationship with the broader context in which they take place – focusing on Health EDRM settings. Logic models are tools used to outline assumptions about the chains of processes, activities or events expected to occur during the implementation of an intervention, and the way in which these lead to changes in outcomes. They provide an initial set of assumptions about how different components of an intervention are expected to change outcomes, and can be used to develop further sub-research questions to investigate the validity of these assumptions. Logic models can also be used to communicate findings from research and evaluation activities, and can serve as useful tools in planning an intervention, including for the identification of relevant outcomes and monitoring of its delivery. However, this chapter will focus primarily on the use of logic models for research and evaluation purposes.

4.10.3 Why use a logic model in research and evaluation?

Programme theory refers to a number of collaborative approaches that allow stakeholders to work together to identify what should be done about a particular health challenge, how this should be done, and the intended outcomes and impact. A logic model is a framework for programme theory that graphically depicts a series of assumptions or steps about how an intervention is expected to achieve impact.

A logic model provides an accessible way for developing a shared understanding across different stakeholders of what an intervention is intended to achieve and a theory of how this will happen. Although there are several ways in which logic models can be used during the design of research and evaluation studies, they provide a means to explore two issues of relevance to policy makers and healthcare practitioners.

Firstly, logic models help users to theorize how the observed impacts of an intervention reflect factors around the implementation of the intervention and/or to its design *(1)*. For example, an intervention in a flood-prone area that is intended to help people to prepare for a disaster might include raising awareness of what should be included in a household disaster preparedness kit (for example, a torch and a supply of bottled water) *(2–3)*. The intervention as a whole might consist of a series of educational components delivered in community settings and a mass media campaign to improve knowledge of what should be included in the kit. If an evaluation study then found that the intervention did not lead to an improvement in knowledge, a logic model may help the researchers to assess whether this was due to problems with the design of the intervention or with its implementation. Using a logic model in an evaluation study provides a framework for understanding how an intervention works, and for producing evidence that can help to differentiate between an intervention that was not implemented properly and one that was not theorized properly (that is, even though it was properly implemented, it did not have a beneficial effect) *(4)*.

Secondly, using a logic model as the framework for research and evaluation in Health EDRM provides nuanced evidence that can be used to better understand how, where, and among whom the intervention is more likely to succeed *(5)*. For example, if the aforementioned disaster preparedness intervention was found to be successful in a particular setting, a well-specified logic model could be used to design an evaluation to establish if both components (the educational intervention and the mass media campaign) were necessary for success if the intervention were to be implemented elsewhere. Similarly, the logic model might be used to consider whether there were characteristics of the setting or population that facilitated or hindered the success of the intervention.

Chapter 3.3 discusses the design of interventions; using logic models supports researchers and evaluators to consider the factors that make interventions succeed or fail, and how these differ according to the characteristics of the setting or the population. Logic models are therefore frameworks that guide researchers, practitioners and policy makers and inform their decisions through developing theories of what an intervention is trying to achieve and how it will meet this aim.

4.10

4.10.4 When are logic models used?

Logic models can be used at different stages of an intervention, and by different stakeholders for different purposes *(4,6)*. They can be used from the outset, in the planning and design of an intervention, as a framework to underpin research into what the intervention is attempting to achieve and whether this is likely to be successful. Once the intervention is in place, logic models can be used to support implementation and to monitor and evaluate progress and performance. Although logic models are usually presented in graphical form, they can be presented in other formats; when presented in a tabular format, this may align with a logframe, which can serve many of the same purposes as a logic model, but has been described as more challenging to use for complex interventions *(4)*.

For research and evaluation, logic models can be used to guide the overall conduct and design of the evaluation, including as a framework for identifying the questions that should be addressed, the outcomes that should be measured and the data that should be collected. Until recently, as noted in other chapters, the field of disaster medicine has been impeded in its development by a lack of evaluation studies in the peer-reviewed literature *(6)*. Fortunately, there is now greater emphasis on systematically evaluating disasters and emergencies and their impacts across a range of domains, and understanding how different 'vulnerability, capacity, exposure of persons and assets, hazard characteristics and the environment interact to amplify or reduce losses' *(7–8)*. The use of logic models in evaluation studies provides a framework for prioritizing and structuring data collection and analysis and ensuring that an evaluation examines the main components of an intervention and the relationships between them.

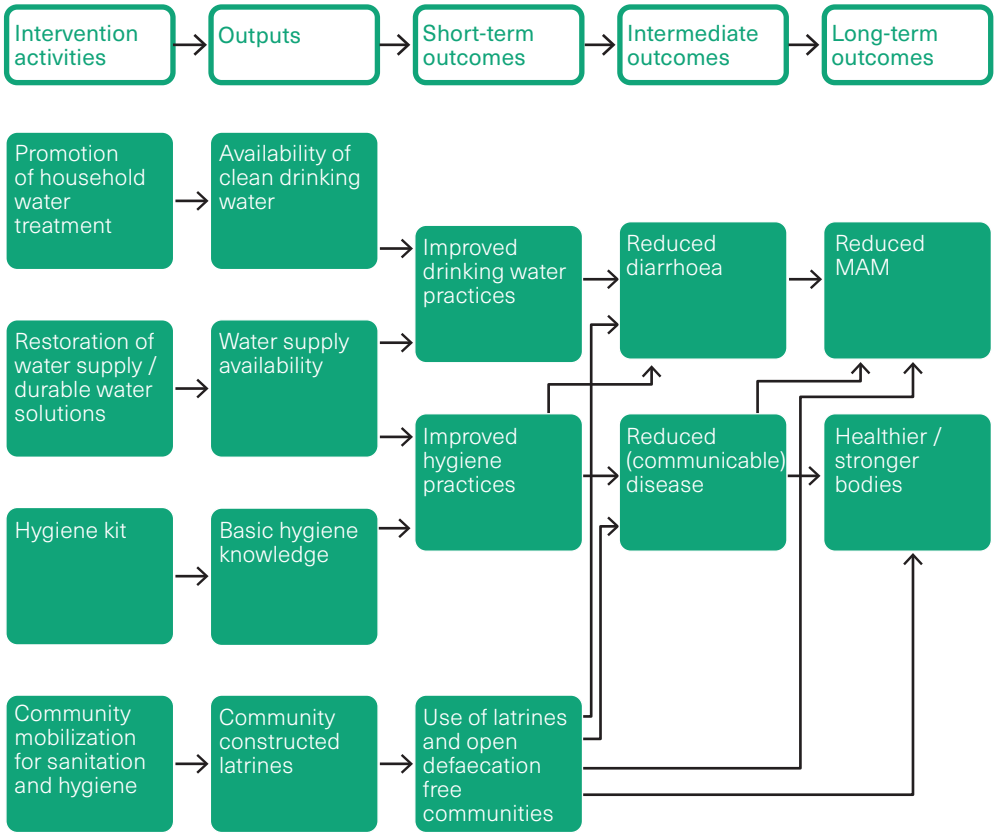
Logic models are also regarded as engagement tools to bring together diverse stakeholders, for example, in allowing them to develop a shared understanding of the priorities and modes of operation of the intervention *(4)*, helping to produce context-specific research knowledge *(9)*, and increasing the likelihood that the results of an evaluation will be accepted and used *(10)*. Logic models are also widely used for communication about evaluation studies *(11)*. Finally, logic models are used in evidence-informed policy and practice when synthesizing evidence from across different studies or settings about the feasibility or impact of a particular intervention approach *(12–13)*, and in making decisions about whether to implement, adapt or innovate a given intervention *(4)*.

Logic models may also be used at different levels, from theorizing how a single intervention might 'work' through to theorizing the impact of a suite of interventions forming a large programme. The latter will likely require the development of complicated multi-strand and multi-level logic models that might seek to depict the actions of several different nongovernmental organizations, institutions and other stakeholders. However, across these different purposes, the processes of interpreting and constructing a logic model follow similar principles.

4.10.5 Interpreting a logic model

A logic model is a graphical representation of intervention processes and how they change outcomes, depicted as chains of cause-and-effect relationships *(14)*. Figure 4.10.1 is an adaptation of a logic model supporting the evaluation of an intervention to increase community resilience to disaster in Pakistan, and is adapted from the work of Avdeenko and Frölich *(15)*. The logic model depicts a programme theory of how multi-component interventions involving Water, Sanitation and Hygiene (WASH) can increase resilience to disasters and improve health. Focusing in on a single pathway at the top of the model, representing a pathway between the restoration of water supplies and a reduction in levels of Moderate-Acute Malnutrition (MAM), we read the model from left to right as a series of 'if...then...' statements *(16)*. These statements are based on the premise that if 'x' occurs, 'y' will occur, and are used to link different sections of the chain. Reading from left to right, *if* water treatments are promoted, *then* there will be greater availability of clean drinking water. In turn, *if* there is greater availability of clean drinking water, *then* drinking water practices will improve; and *if* there are improvements in drinking water practices, *then* levels of diarrhoea will reduce. Finally, *if* there is a reduction in diarrhoea, *then* levels of MAM will reduce.

Figure 4.10.1 Logic model for the impacts of WASH activities in improving health as part of interventions to increase community resilience to natural disasters in Pakistan *(15)*



Our reading of the logic model and focus on a single strand is a simplified interpretation of how the intervention may reduce levels of MAM. For example, it is recognized within forms of guidance around WASH interventions (17) that behaviour change is not automatic with the provision of clean water supplies, and should be explicitly programmed alongside environmental, social inclusion and treatment and care interventions. Furthermore, the logic model actually shows five different potential pathways that might lead to such a reduction, all or only some of which may be needed in order for a reduction to be observed *(15)*. Because the model indicates that a reduction in MAM may be achieved through different pathways or combinations of components (known as equifinality), the intervention can be considered to be complex in nature, requiring a particular suite of analytical tools for its evaluation *(14)*.

4.10.6 Features of a logic model

Logic models depict often highly complex interventions in a manageable and interpretable way. In order for logic models to provide a framework to support research and evaluation studies, they must contain elements that summarize the assumptions of how the intervention works. These elements include:

- The outcomes or the change that the intervention is trying to bring about
- Indicators of implementation that show what was meant to be delivered
- Mechanisms that show how what was being delivered as part of the intervention leads to a change in the outcome
- Characteristics of the context in which the intervention takes place that are likely to influence its implementation or its effectiveness *(18–19)*.

To ensure that a logic model captures these elements, they should represent – at a minimum – intervention activities or inputs, outputs, the intervention outcomes (which may be ordered chronologically), and the relationships between these. These elements are defined in Table 4.10.1, along with other elements that frequently occur in logic models, some of which may be particularly important for Health EDRM interventions.

Table 4.10.1 Definitions of frequently occurring elements of logic models used in intervention research *(6, 13, 20–21)*

Elements of logic models used in intervention research	Definition
Distal or long-term outcomes	Long-term outcomes are those theorized to occur following the initiation of an intervention and reflect broad concepts which are often analogous with the ultimate aims of the intervention.
Intermediate outcomes	Intermediate outcomes are theorized as being necessary pre-conditions of achieving distal or long-term outcomes and occur during follow-up after an intervention has ended. They may reflect behaviours that are among the ultimate aims of the intervention.
Short-term or proximal outcomes	Short-term outcomes are theorized to occur at the end of an intervention or soon after it has ended, as a direct result of the intervention. They are theorized to be necessary pre-conditions for triggering intermediate outcomes.
Outputs	Outputs are descriptive indicators of what the specific activities generate, and quantified and qualified indicators of the implementation of intervention activities. Unlike outcomes, outputs are under the direct control of those delivering the intervention.
Intervention inputs: Activities or intervention components and processes	Activities or components of the intervention that reflect what is being delivered. These are necessary to trigger the expected intervention processes and outputs. They may be represented as sequences of events in themselves, where one intervention component must take place before another component can begin.
Intervention inputs: Resources	Resources that are secured in order to deliver an intervention. They may be financial or may reflect the input or support of different stakeholders and might be identified through asset mapping processes (Chapter 3.1).
Contextual factors or external factors	These include population characteristics and the characteristics of the context or setting where the intervention takes place, which may moderate the way in which the intervention is expected to 'work'. For disaster and emergency interventions, these may reflect pre-existing conditions or new factors that have emerged as a result of an event (for example, the emergence of violence or spread of a communicable disease such as cholera).
Connecting arrows	These form chains, linking intervention inputs with outputs and outcomes. Connecting arrows signal the direction in which the sequence of events take place and can be used to represent more complex relationships
Additions for disaster or emergency interventions *(6)*: Goals	These are broad statements about the long-term expectations of what should happen as a result of the intervention (see Salabarría-Peña, Apt and Walsh *(22)* for a further example).
Additions for disaster or emergency interventions *(6)*: Objectives	Statements describing the changes to be achieved, and the way in which change will occur (linked to the broader goal, with multiple objectives supporting the goal).
Additions for disaster or emergency interventions *(6)*: Impacts	Impacts reflect the way in which the intervention is theorized to meet its overall goal. As Birnbaum and colleagues *(6)* explain, impacts are the 'so-what' of the intervention. They represent the 'high-level' systemic change achieved at a community or population level (in practice, there may also be overlap with long-term outcomes).

4.10

4.10.7 Constructing a logic model de novo

This section will briefly discuss the steps involved in developing a logic model de novo. There are several comprehensive resources to support this process *(4, 6, 13, 16, 20, 23)*, some of which include templates to guide researchers and policymakers *(24),* including one specifically developed for disaster-related interventions *(6)*.

A first step in developing a logic model is to search for existing logic models for the intervention of interest *(13)*. However, despite nearly 60 years of the use of logic models by evaluators *(4)*, existing examples can be difficult to locate. Furthermore, any existing logic model will need to be adapted to reflect different contexts or priorities. Nevertheless, reviewing existing models is a useful preliminary exercise in starting to theorize the outcomes of interest and how they should be sequenced, and in identifying some key intervention processes linking inputs with outputs and outcomes *(25)*.

4.10.8 Steps in creating a logic model

1. Involve stakeholders.
Before developing a logic model, a key step is to secure the involvement of a range of stakeholders, in order to strengthen the salience of the model and its value in subsequent research activities *(8, 10)*. Different stakeholders (such as evaluators, policy-makers, community leaders) tend to hold different views and understandings, which are useful to incorporate when dealing with the uncertainty and complexity in humanitarian crises. Among other benefits, the involvement of a diverse range of stakeholders can:

- Create a useful challenge to the assumptions made in deciding how an intervention changes outcomes.
- Provide an opportunity to develop a consensus as to which outcomes measure the effectiveness of the intervention, and which outputs signal whether the intervention was successfully implemented.
- Ensure that diverse perspectives are represented.
- Help to identify how contextual factors extraneous to the intervention may facilitate or hinder the delivery of the intervention.
- Enhance the usefulness of the evidence produced for different practitioners and policy-makers.

2. Identify the purpose or goal of the intervention.
The overarching research question (Chapter 3.5), purpose or goal of the intervention should be identified and the major assumptions should be outlined. This may include key changes that have taken place in disaster or emergency settings, and theorizing about how these external factors will influence the goal of the intervention.

3. Begin depicting the chain of events, starting with the distal outcomes.
It is customary for the development of a logic model to begin by identifying and representing (usually in boxes) the distal or long-term outcomes that are expected to result from implementation of the intervention.

4. Specify intermediate and proximal outcomes.
The next step involves working backwards to identify or hypothesize the necessary preconditions (intermediate and proximal or short-term outcomes) that are needed to reach these distal outcomes. For example, in Figure 4.10.1, it was hypothesized that reducing the levels of MAM (long-term outcome) required reduction in levels of diarrhoea (intermediate outcome).

5. Continue to develop outcome chains.
The steps needed to reach longer-term outcomes may involve a number of pre-conditions (changes in outcomes) that are needed. Several "links", represented as boxes or other shapes, could be added to the outcome chain.

6. Add intervention outputs.
After identifying outcomes, outputs are identified and represented within the model. Outputs are descriptive indicators of what the specific intervention activities generate, and represent necessary pre-conditions to reach outcomes, but are not necessarily goals in themselves (see Table 4.10.1).

7. Develop intervention inputs/activities.
Continuing to work backwards from the outcomes and outputs, chains of intervention inputs are specified. Areas of ambiguity about precisely how intervention activities are sequenced (that is, a 'black box' of intervention inputs) may be represented in the logic model as a single box, with the research or evaluation study building understanding of how the intervention is implemented.

8. Complete initial model.
An initial logic model will consist of input chains, comprising an intervention's components and resources and how these are sequentially implemented, outputs, and outcome chains.

9. Consider the nature of mechanisms.
Mechanisms, or pathways of action, describe the nature of the action occurring between intervention inputs and outputs and outcomes. Not all relationships depicted within a logic model are simple linear (cause-effect) relationships, and more complex relationships may need to be included to better represent the likely mechanisms and to help guide data collection or analysis. An example is presented below and further examples are available elsewhere *(26–28)*.

10. Consider the role of context, settings and stakeholders.
Additional external or contextual factors, including the characteristics of populations, communities and other stakeholders involved in interventions, should be theorized and represented. These characteristics may be necessary for the intervention to 'work' (that is to say, without them the intervention cannot be implemented) or may moderate its effectiveness and amplify or dampen its success. In some cases, it may be easier to develop separate causal chains, or even separate models if an intervention is theorized to work very differently across diverse settings, populations or stakeholders.

4.10

11. Continue to iterate
It is expected that several iterations of the logic model will be developed before a preferred model is produced, with iterations representing an improvement in clarity, the conceptual soundness of the logic model, and more logical sequencing and organization of its elements. The logic model should be assessed for its consistency with existing research, broader theory, knowledge about the setting and logical plausibility *(4)*. Further iteration and development of the models may take place while the research or evaluation study is being conducted, on the basis of new knowledge generated (see also 'Update on the basis of new learning' below).

12. Consider unintended consequences.
Outcomes of interventions may deviate from their intended outcomes, and it is important to theorize about these unanticipated and adverse impacts. This process is described as modelling "dark logic" within interventions *(29)*.

13. Update on the basis of new learning.
The research process is expected to generate new knowledge and evidence that may lead to changes in the logic model, or lead to an entirely new way of understanding how the intervention works. For examples of how logic models were updated based on new evidence see Harris et al *(25)* and Waddington and White *(30)*.

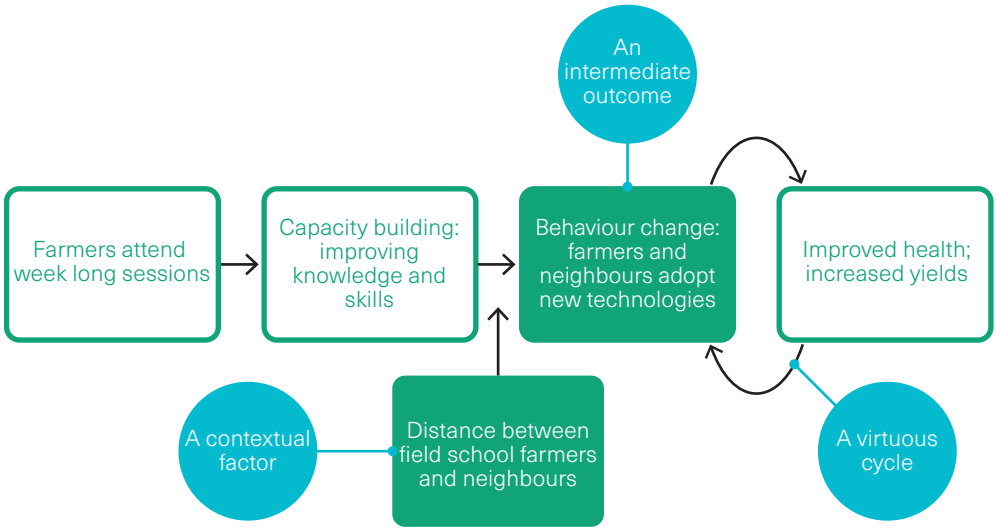
4.10.9 Representing more complex relationships in a logic model

To show how more complex relationships can be included in a logic model, we draw on the example of farmer field schools (FFS). FFS bring together groups of farmers in a community to empower them through learning about best practices in agriculture and, increasingly, about prevention, preparedness and response to disasters. The approach uses participatory models of education, including field-based experiments on neighbouring plots of land through a growing season, in order to examine the impact of best-practice techniques. FFS are believed to be useful in mitigating exposure to disasters and climate change *(31)*. The interventions have been considered as a means of reducing the risks of pesticide-related health emergencies *(30)* and as post-recovery measures for disaster-affected farmers *(32)*.

A systematic review of the effectiveness of FSS on outcomes including health was supported using a logic model *(30)*. A simplified and adapted version of that logic model is shown in Figure 4.10.2. The pathway outlines the steps between attending a FFS and improved health and yields, with three features of interest highlighted that can be represented in logic models. The first is the explicit mention of the intermediate outcome, which represents a factor that lies on the causal pathway between the intervention and distal outcomes. This demonstrates the functionality of a logic model being developed through theorizing a chain of pre-conditions needed to link the intervention with the outcome. The second feature of interest is the inclusion of hypothesized contextual factors (geographical and social distance between farmers) that are expected to moderate the extent to which new skills and behaviours developed among FFS participants will lead to improved knowledge and behavioural change

among neighbouring farmers as well. Here, this contextual factor may amplify or dampen the relationship between exposure to the intervention and the outcomes. Such factors may interrupt or support the chain of events, but are not integral links in the causal chain. The third feature of interest is a 'virtuous circle', which is depicted in Figure 4.10.2 as a process whereby the adoption of new technologies among field school farmers and neighbour farmers leads to better health outcomes, by reducing farmers' exposure to pesticide in the environment, and better yields, via community-wide adoption of improved practices. This suggests that the impacts of the intervention could strengthen over time, and as the benefits of new technologies become apparent, this stimulates further adoption of new technologies. Virtuous cycles are activated when initial changes in the outcome create the opportunities for further positive self-reinforcing changes. Negative changes can be represented as 'vicious cycles'. Virtuous and vicious cycles are two of several more complex relationships that can be depicted in a logic model *(4, 26–28)*.

Figure 4.10.2 Logic model adapted from a systematic review of the effectiveness of farmer field schools *(30)*



4.10.10 Logic model variants

As tools in research and evaluation studies, logic models offer flexibility and a spectrum of forms and uses are available in the literature. Some different variants of logic models are outlined below, drawing in part on work by Rehfuess et al *(21)*. These variants arise from differences in the priorities of the logic model at different stages of the research or evaluation study, or its scope.

Variant 1: Static, staged and iterative logic models

A static logic model is one that is specified before the research or evaluation study, and remains in place without iteration throughout the study (although there may be an assessment at the end of the study as to how well the theory explained the evidence). A staged logic model is one where the theory is adapted or changed on the basis of interim findings or new knowledge, at planned stages of the research or evaluation study. Iterative logic models are those that are adapted at any point in the

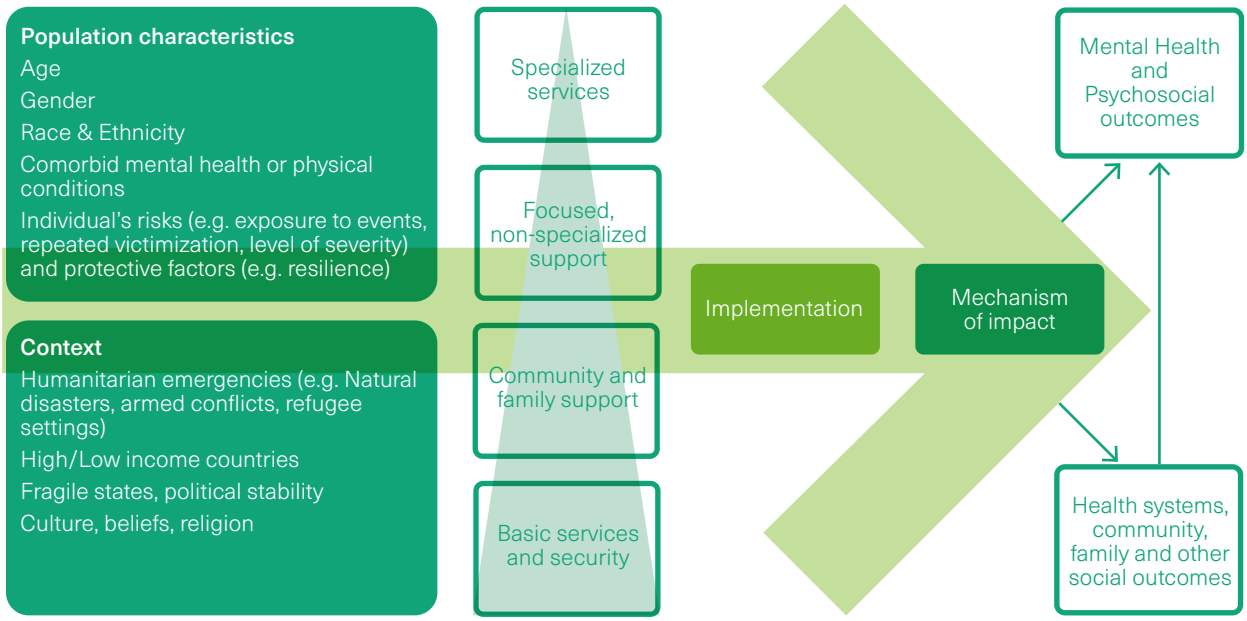
research or evaluation study to reflect findings or new knowledge. This latter approach is more organic and responsive to new insights that may emerge, new questions that may arise, and any change in the priorities of the intervention *(25)*. A logic model should be assessed, and updated as appropriate, based on the findings of the research or evaluation study, with both the original and updated versions made available.

Variant 2: System-based and process-based logic models

A second distinction is between system-based logic models that aim to theorize aspects of complexity around the relationship between an intervention and the broader context and how these interact, and process-based logic models that focus more theorizing aspects of complexity between the processes occurring as part of an intervention and its multiple outcomes. Clearly there is some overlap between these types of model, and a single research or evaluation study may draw on both *(21)*. Process-based logic models tend to represent input and output chains in greater detail, reflecting a priority around understanding temporal sequences of intervention processes. Meanwhile, a system-based logic model depicts the system as 'the interaction between the participants, the intervention, and the context [in which it] takes place' *(21, p.15)*.

A system-based logic model may be particularly useful in accounting for the myriad ways in which different interconnecting components of health systems are impacted by health emergencies and disasters, and theorizing how interventions can restore these systems and 'build back better' to improve health. An example of a system-based logic model is reproduced from the paper by Bangpan, Chiumento, Dickson, and Felix *(33)*, which highlights in a simplified way the types of population characteristics, contextual and implementation factors and the combinations of these factors which may influence the effectiveness of mental health and psychosocial support interventions on people affected by humanitarian emergencies. Interventions in Health EDRM are often complex and sensitive to the context in which they are undertaken. This means that an intervention that is effective in one type of setting may be ineffective (or even harmful) in another population or setting, without modification *(34)*. A system-based logic model provides a starting point for theorizing whether there are aspects of the context (setting and existing health infrastructure) or population that are likely to facilitate or hinder the implementation and effectiveness of an intervention (Figure 4.10.3).

Figure 4.10.3 Components to consider in a system-based logic model *(33)*



Terminology used when theorizing how interventions work

Although we outline the use of logic models, there are a number of different, overlapping, terms for tools that have been used to conceptualize how interventions work. Table 4.10.2 provides definitions for some of the alternative terminology in use, although in practice there are several overlaps between these concepts.

4.10

Table 4.10.2 Definitions of frequently occurring terms around the use of programme theory

Type of (Programme) Theory	Definition
Programme theory	A hypothesis explaining how an intervention is expected to lead to a change in the outcome. Graphical representations of programme theory can be developed into logic models or theories of change.
Logic model	A graphical representation of intervention processes, and outcomes linked by arrows indicating the direction of effect, which are developed into chains of cause-and-effect relationships.
Theory of change	Theories of change are used to represent complex interventions. Although there is overlap, unlike logic models, theories of change are more explanatory as they require all of the underlying assumptions of how and why different components, activities and outputs lead to a change in outcomes to be specified at the outset, as well as an indication of the context and the stakeholders affected. There can be multiple causal chains for different stakeholders. While there are differences between logic models and theories of change *(35)*, these differences are fuzzy and in practice the terminology is often used interchangeably.
Logical framework	The term logical framework or logframe is used to describe an array of different approaches. In some cases, the term is conflated with logic models. However, there are examples of logframes that are used more as project management tools that track how outputs, outcomes and impact are achieved according to different activities *(36)*. While useful as a project management tool, logframes are likely to be less useful as a tool for theorizing how interventions work, and particularly as a tool for theorizing aspects of complexity in interventions *(37)*.
Middle-range theory	Middle-range theories connect high-level sociological theories with empirical knowledge. In the context of interventions, middle-range theories include general principles about the ways in which interventions will 'work' across a range of situations (drawn from high-level theory), but also include some granular detail around intervention causal chains that can inform specific decisions about an intervention. Nevertheless, they are usually more generalized than programme theory, although there are several commonalities between middle-range theory and programme theory more generally *(38)*. There are few specific examples in the development literature of middle-range theories *(39)*.
Conceptual framework	Conceptual frameworks outline the main elements of the intervention and how it is meant to work, and may include a description of the context in which an intervention is expected to take place. A conceptual framework is not necessarily a graphical outline and the nature of the relationships between different components may not be explicitly articulated.

4.10.11 Using logic models in evaluation and research

Logic models can be useful, practical aids for conducting a variety of research and evaluation studies, in several ways:

- As an engagement tool with stakeholders in the design of research and evaluation studies, ensuring that a diverse set of views are represented from the outset.
- Helping to design specific research and evaluation questions that can be used to guide studies and, similarly, in helping to identify the types of research approaches and methods that are suitable for answering questions that emerge from the logic model.
- Helping to decide what information needs to be collected about intervention inputs and activities, the characteristics of the contexts, and outputs and outcomes.
- Helping to design plans of how the research or evaluation data will be processed and for interpreting the findings.
- Communicating the results of the study through updating or redrawing of logic models on the basis of new knowledge.

Using a logic model provides a framework for understanding how an intervention channels an effect between the inputs and outcomes (40–41). Logic models are useful in unpacking the intervention 'black box' to aid understanding of the processes by which interventions can generate impact *(42).* This approach to producing evidence can help to move "beyond 'business as usual', generic programme designs through [developing] a greater awareness of the context", with the logic model being a useful tool "to test the assumptions, demonstrate impact and learn from [interventions]" (*43,* p11).

4.10.12 Conclusions

There is increasing concern around improving the availability and use of evidence for Health EDRM *(44–45).* At the core of good quality evidence is the use of theory to increase the robustness of the findings, the applicability and validity of any recommendations and enhance the generalizability (external validity) of the findings to other settings.

Using a logic model to theorize how an intervention works and how it interacts with context, and designing a research or evaluation study to test this theory, can be a useful basis for making decisions about which interventions to implement in which areas and for which types of emergency, as well as identifying whether interventions may need adaptation. Furthermore, for interventions that do not appear to be effective, evidence that is driven by theory is more likely to help distinguish between failures in intervention design and failures in intervention implementation (potentially due to context). Logic models represent a practical and applied approach for developing a theory of how interventions work which can be updated to incorporate new learning obtained through research and evaluation.

4.10

4.10.13 Key messages

- **Logic models provide a useful basis for thinking conceptually about how an intervention should 'work' to change outcomes. They are a graphical representation of the stages linking intervention inputs and outputs, outcomes and impacts.**
- **Logic models can be used to reflect assumptions about contexts and to illustrate more complex relationships.**
- **There are a number of steps to follow when developing a logic model, but perhaps one of the most important elements of good practice is that logic models should be developed with the input of stakeholders to challenge some of the (potentially erroneous) assumptions made by the research team.**

4.10.14 Further reading

Resources that include logic model templates

Birnbaum ML, Daily EK, O'Rourke AP, Kushner J. Research and evaluations of the health aspects of disasters, part VI: interventional research and the Disaster Logic Model. Prehospital and Disaster Medicine. 2016: 31(2): 181-94.

Rohwer A, Booth A, Pfadenhauer L, Brereton L, Gerhardus A, Mozygemba K, et al. Guidance on the use of logic models in health technology assessments of complex interventions. 2016 https://www.integrate-hta.eu/wp-content/uploads/2016/02/Guidance-on-the-use-of-logic-models-in-health-technology-assessments-of-complex-interventions.pdf (accessed 6 February 2020).

Resources on how to develop a logic model afresh

Kneale D, Thomas J, Harris K. Developing and Optimising the Use of Logic Models in Systematic Reviews: Exploring Practice and Good Practice in the Use of Programme Theory in Reviews. PLoS ONE. 2015: 10(11): e0142187.

Resource on using logic models in research on complex interventions

Kneale D, Thomas J, Bangpan M, Shemilt I, Waddington H, Gough D. Causal chain analysis in systematic reviews of international development interventions. CEDIL Inaugral Papers. Centre of Excellence for Development Impact and Learning, London. 2018. https://cedilprogramme.org/wp-content/uploads/2017/12/Inception-Paper-No-4.pdf (accessed 6 February 2020).

Applied examples

Bangpan M, Chiumento A, Dickson K, Felix L. The impact of mental health and psychosocial support interventions on people affected by humanitarian emergencies: a systematic review. In Humanitarian Evidence Programme. Oxfam GB, Oxford. 2017: https://policy-practice.oxfam.org.uk/publications/the-impact-of-mental-health-and-psychosocial-support-interventions-on-people-af-620214 (accessed 6 February 2020).

Waddington H, White H. Farmer field schools: from agricultural extension to adult education. 3ie Systematic Review Summary 1. International Initiative for Impact Evaluation, London. 2014. https://www.3ieimpact.org/sites/default/files/2019-05/srs1_ffs_revise_060814_final_web_2.pdf (accessed 17 July 2020).

4.10.15 References

1. Bamberger M, Rao V, Woolcock M. Using mixed methods in monitoring and evaluation: experiences from international development. 2010. https://openknowledge.worldbank.org/handle/10986/3732 (accessed 6 February 2020).
2. British Red Cross. Prepare an emergency kit: Be ready to cope with a crisis. 2020 https://www.redcross.org.uk/get-help/prepare-for-emergencies/prepare-an-emergency-kit (accessed 6 February 2020).
3. Chan EYY, Guo C, Lee P, Liu S, Mark CKM. Health emergency and disaster risk management (Health EDRM) in remote ethnic minority areas of rural China: The case of a flood-prone village in Sichuan. International Journal of Disaster Risk Science. 2017: 8(2): 156-63.
4. Funnell SC, Rogers PJ. Purposeful program theory: effective use of theories of change and logic models (volume 31). San Francisco, CA: John Wiley & Sons. 2011.
5. White H. Theory-based impact evaluation: principles and practice. Journal of development effectiveness. 2009: 1(3): 271-84.
6. Birnbaum ML, Daily EK, O'Rourke AP, Kushner J. Research and evaluations of the health aspects of disasters, part VI: interventional research and the Disaster Logic Model. Prehospital and Disaster Medicine. 2016: 31(2): 181-94.
7. Aitsi-Selmi A, Egawa S, Sasaki H, Wannous C, Murray V. The Sendai framework for disaster risk reduction: Renewing the global commitment to people's resilience, health, and well-being. International Journal of Disaster Risk Science. 2015: 6(2): 164-76.
8. Sendai Framework for Disaster Risk Reduction 2015 - 2030. UNISDR. 2015. https://www.unisdr.org/we/inform/publications/43291 (accessed 6 February 2020).
9. Oliver S, Roche C, Stewart R, Bangpan M, Dickson K, Pells K et al. Stakeholder Engagement for Development Impact Evaluation and Evidence Synthesis. 2018. https://cedilprogramme.org/wp-content/uploads/2018/10/Stakeholder-Engagement-for-Development.pdf (accessed 6 February 2020).

4.10

10. Dwyer JJ, Makin S. Using a program logic model that focuses on performance measurement to develop a program. Canadian Journal of Public Health. 1997: 88(6): 421-5.
11. Jones ND, Azzam T, Wanzer DL, Skousen D, Knight C, Sabarre N. Enhancing the Effectiveness of Logic Models. American Journal of Evaluation (published online first). 2019: doi: 10.1177/1098214018824417
12. Anderson LM, Petticrew M, Rehfuess E, Armstrong R, Ueffing E, Baker P, et al. Using logic models to capture complexity in systematic reviews. Research Synthesis Methods. 2011: 2(1): 33-42.
13. Kneale D, Thomas J, Harris K. Developing and Optimising the Use of Logic Models in Systematic Reviews: Exploring Practice and Good Practice in the Use of Programme Theory in Reviews. PloS ONE. 2015: 10(11): e0142187.
14. Kneale D, Thomas J, Bangpan M, Waddington H, Gough D. Conceptualising causal pathways in systematic reviews of international development interventions through adopting a causal chain analysis approach. Journal of Development Effectiveness. 2018: 10(4): 422-37.
15. Avdeenko A, Frölich M. Impacts of increasing community resilience in the face of natural disasters through humanitarian aid in Pakistan. 2019. https://www.3ieimpact.org/sites/default/files/2019-06/IE100-TW6.1028-humanitarian-ACTED-Pakistan.pdf (accessed 6 February 2020).
16. WK Kellogg Foundation. Logic model development guide. 2004 https://www.bttop.org/sites/default/files/public/W.K.%20Kellogg%20LogicModel.pdf (accessed 6 February 2020).
17. WASH and health working together: a 'how-to' guide for neglected tropical disease programmes (9241515007). WHO. 2018. https://apps.who.int/iris/bitstream/handle/10665/279913/9789241515009-eng.pdf (accessed 6 February 2020).
18. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ. 2015: 350: h1258.
19. Morgan-Trimmer S, Smith J, Warmoth K, Abraham C. Introduction to logic models. 2018. https://www.gov.uk/government/publications/evaluation-in-health-and-well-being-overview/introduction-to-logic-models (accessed 6 February 2020)
20. Knowlton LW, Phillips CC. The logic model guidebook: Better strategies for great results. Thousand Oaks, CA: Sage. 2012.
21. Rehfuess EA, Booth A, Brereton L, Burns J, Gerhardus A, Mozygemba K, et al. Towards a taxonomy of logic models in systematic reviews and health technology assessments: a priori, staged and iterative approaches. Research Synthesis Methods. 2017: 9(1): 13-24.
22. Salabarría-Peña Y, Apt B, Walsh C. Practical use of program evaluation among sexually transmitted disease (STD) programs. 2007. https://www.cdc.gov/std/program/pupestd.htm (accessed 6 February 2020).

23. Pfadenhauer L, Rohwer A, Burns J, Booth A, Lysdahl KB, Hofmann B, et al. Guidance for the Assessment of Context and Implementation in Health Technology Assessments (HTA) and Systematic Reviews of Complex Interventions: The Context and Implementation of Complex Interventions (CICI) Framework. 2016. https://www.integrate-hta.eu/wp-content/uploads/2016/02/Guidance-for-the-Assessment-of-Context-and-Implementation-in-HTA-and-Systematic-Reviews-of-Complex-Interventions-The-Co.pdf (accessed 6 February 2020).

24. Rohwer A, Booth A, Pfadenhauer L, Brereton L, Gerhardus A, Mozygemba K, et al. Guidance on the use of logic models in health technology assessments of complex interventions. 2016 https://www.integrate-hta.eu/wp-content/uploads/2016/02/Guidance-on-the-use-of-logic-models-in-health-technology-assessments-of-complex-interventions.pdf (accessed 6 February 2020).

25. Harris K, Kneale D, Lasserson TJ, McDonald VM, Grigg J, Thomas J. School-based self-management interventions for asthma in children and adolescents: a mixed methods systematic review. Cochrane Database of Systematic Reviews. 2019: (1): 1:CD011651.

26. Davies R. Representing Theories Of Change: A Technical Challenge With Evaluation Consequences. 2018. https://mande.co.uk/wp-content/uploads/2018/09/2018-08-31-Inception-Paper-No-15-MandE-NEWS-PDF-copy-2.pdf (accessed 6 February 2020).

27. Kneale D, Thomas J, Bangpan M, Waddington H, Gough D. Causal chain analysis in systematic reviews of international development interventions. 2018. https://cedilprogramme.org/wp-content/uploads/2017/12/Inception-Paper-No-4.pdf (accessed 6 February 2020)

28. Rogers PJ. Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. Evaluation. 2008: 14(1): 29-48.

29. Bonell C, Jamal F, Melendez-Torres GJ, Cummins S. "Dark logic": theorising the harmful consequences of public health interventions. Journal of Epidemiology and Community Health. 2014: 69(1): 95-8.

30. Waddington H,White H. Farmer field schools: from agricultural extension to adult education. 3ie Systematic Review Summary 1, International Initiative for Impact Evaluation, London. 2014: https://www.3ieimpact.org/sites/default/files/2019-05/srs1_ffs_revise_060814_final_web_2.pdf (accessed 16 July 2020).

31. Meijboom M, Tiwari S, Dubbeling M. Enhancing climate resilience of Gorakhpur by buffering floods through climate-resilient peri-urban agriculture. 2016. https://ruaf.org/document/enhancing-climate-resilience-of-gorakhpur-by-buffering-floods-through-climate-resilient-peri-urban-agriculture/ (accessed 6 February 2020).

32. Mariyono J, Luther GC, Bhattarai M, Ferizal M, Jaya R, Fitriana N. Farmer field schools on chili peppers in Aceh, Indonesia: Activities and impacts. Agroecology and Sustainable Food Systems 2013: 37(9): 1063-77.

33. Bangpan M, Chiumento A, Dickson K, Felix L. The impact of mental health and psychosocial support interventions on people affected by humanitarian emergencies: a systematic review. Humanitarian Evidence Programme. Oxford: Oxfam GB. 2017. https://fic.tufts.edu/assets/Mental-Health-Systematic-Review.pdf (accessed 6 February 2020).

34. Greene MC, Jordans MJ, Kohrt BA, Ventevogel P, Kirmayer LJ, Hassan G et al. Addressing culture and context in humanitarian response: preparing desk reviews to inform mental health and psychosocial support. Conflict and Health. 2017: 11(1): 21.

35. Clark H, Anderson AA. Theories of Change and Logic Models: Telling Them Apart. 2004. https://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/TOCs_and_Logic_Models_forAEA.pdf (accessed 6 February 2020).

36. DFID. Guidance on using the revised Logical Framework. 2011. https://www.gov.uk/government/publications/dfid-how-to-note-guidance-on-using-the-revised-logical-framework (accessed 6 February 2020).

37. Floate H, Durham J, Marks GC. Moving on from logical frameworks to find the 'missing middle' in international development programmes. Journal of Development Effectiveness. 2019: 11(1): 89-103.

38. Pawson R. Middle Range Theory and Programme Theory Evaluation. In: Vaessen J, editor. Mind the Gap: Perspectives on Policy Evaluation and the Social Sciences. 2008.

39. CEDIL Call for Proposals Programme of Work 2: Generalising evidence through middle range theory. CEDIL. 2019. https://cedilprogramme.org/wp-content/uploads/2019/03/CEDIL-call-spec-for-PoW-2.pdf (accessed 6 February 2020).

40. Illari P, Russo F. Causality: Philosophical theory meets scientific practice. OUP Oxford. 2014.

41. Reiss J. Causation in the social sciences: Evidence, inference, and purpose. Philosophy of the Social Sciences 2009: 39(1): 20-40.

42. White H. Theory based systematic reviews. Journal of Development Effectiveness. 2018: 10(1): 17-38.

43. Vogel I. Review of the use of 'Theory of Change' in international development. 2012 http://www.theoryofchange.org/pdf/DFID_ToC_Review_VogelV7.pdf (accessed 6 February 2020).

44. Gerdin M, Clarke M, Allen C, Kayabu B, Summerskill W, Devane D, et al. Optimal evidence in difficult settings: improving health interventions and decision making in disasters. PLoS Medicine. 2014: 11(4): e1001632.

45. Kayabu B, Clarke M. The use of systematic reviews and other research evidence in disasters and related areas: preliminary report of a needs assessment survey. PLoS Currents: Disasters. 2013: 22 January. 4.10.15

4.11 Researching communication and communicating research in Health EDRM

Authors

Alistair Humphrey, Canterbury District Health Board, Christchurch, New Zealand.

Lisa Robinson, British Broadcasting Corporation Media Action, London, United Kingdom.

Joseph Bonney, Emergency Medicine Directorate, Komfo Anokye Teaching Hospital, Kumasi, Ghana.

Sue Turner, Canterbury District Health Board, Christchurch, New Zealand.

4.11.1 Learning objectives

To understand the key factors to consider in evaluating and researching emergency risk communication programmes, including:

1. Specific objectives of communication before during and after disasters.
2. Particular challenges and opportunities in Health EDRM communication research.
3. Techniques used in measuring behavioural change inspired by communication programmes.
4. Key principles of quality communication – all of which require further research.

4.11.2 Introduction

At the third session of the United Nations International Strategy for Disaster Reduction (UNISDR) Global Platform in 2011, UN Secretary General Ban Ki Moon noted that success is measured by what does not occur — the school that did not collapse; the building that did not fall; the village that was not destroyed *(1)*. However, the data that are routinely available in Health EDRM research – usually from governments – tend to measure failure: death, destruction and economic loss. This presents a particularly difficult challenge for researchers of communication in disaster risk.

Although it may be reasonably straightforward for an engineer to attribute the survival of buildings to earthquake resilience strengthening, it is much more difficult to attribute human survival in an earthquake to understanding of (and giving effect to) the 'Drop, Cover and Hold' message, for example *(2)*. Public health practitioners are familiar with this conundrum.

They know that measuring interventional practice is easy, but that measuring the success of a preventive programme is always difficult, particularly where human behaviour is involved — and even more so in the case of hazards that occur infrequently. Success is measured by the absence of poor outcomes, but only when a hazard was manifest and risk was minimized. Measuring the absence of an outcome is challenging, particularly when the risk minimization is in the form of a behavioural change made as a consequence of a communication programme. For example, it may be impossible to determine how many cases of enteric disease were prevented by people following advice to wash their hands properly, how many cases of electrocution were prevented by people heeding the message to avoid downed powerlines during a storm, or how many lives were saved by people heading to higher ground on receipt of a tsunami warning.

Communication is one of five key elements of a resilient community, with the others being risk awareness, adaptability, learning and social capital *(3)*. Not only is communication within a community part of resilience in itself, but high-quality communication programmes can also be used to develop the other key aspects of resilience. Communication programmes that identify hazards, quantify risks and convey how to manage them, ideally resulting in population wide-behavioural change, are an essential component of Health EDRM.

Programmes that support communication among experts and general populations can deepen people's understanding of hazards, quantify risks, give guidance on how to manage them, prompt discussions about what can be done at different levels of society and motivate action. They can identify secondary complications of a disaster and ameliorate the psychosocial sequelae of a disaster for months or years afterwards.

WHO has produced a manual titled 'Communicating risk in public health emergencies', which is a guide designed to assist countries in building capacity for risk communication and how risk communication should be carried out before, during and after an emergency *(4)*. However, despite an expert guidelines group and rigorous guideline development methods, including scrutiny of the evidence base for best practice risk communication, the quality of evidence underpinning even the strongest recommendations, using the GRADE (Grading of Recommendations, Assessment, Development and Evaluations) system, was assessed as no better than moderate *(5–6)*. This shows that there is some uncertainty for practitioners of risk communication, but provides opportunities for researchers of risk communication to fill these important knowledge gaps.

Entertainment produces emotional changes, such as laughter, fear and excitement. Art can be entertaining, but goes one step further – with a key requirement of good art being that it makes you think. The art of communication lies in going further still. It requires not only getting people to think, but also inspiring them to change their behaviour. Like art or entertainment, there is a subjective component in the design and the appreciation of a communication campaign. However, communication can also be measured objectively. For example, an objective measure of the success of a communication programme may be whether the target audience have changed their behaviour and whether this behavioural change mitigated the adverse outcomes of an emergency.

There are earlier, intermediary steps to behavioural change. These include whether the communication imparts a greater understanding of the risks of disasters which a population may face and whether the understanding of these risks leads to an improvement in the knowledge required for mitigating them. It is also important to know what beliefs, perceptions, or social norms have shifted, enabling people to translate this knowledge into a change in behaviour, such as improved disaster preparedness kits, actions to build social capital or prompt appropriate responses to early warning systems.

4.11.3 Challenges in doing communication research in disasters

Although these outcome measures may appear to be relatively straightforward to measure, communication research in disasters is difficult for three reasons. First, disasters do not readily lend themselves to interventional studies. Even if a specific intervention can be applied to one group of people while keeping a similar group as a control before, during or after a disaster (which is often logistically impossible), it may be difficult to randomize some to receive a communication programme and some not to receive it *(7)* (see Chapters 4.1 and 4.3). Opportunities for randomization may present themselves through social media (messaging some people but not others, for example) but such randomization in the wake of a disaster would bring ethical challenges. Because the ethical and logistical difficulties of randomization may be insurmountable following a natural disaster, many evaluations of communication programmes are consequently reliant on observational studies, vulnerable to selection biases that can be at best only mitigated, but not entirely remedied, by careful interpretation.

Second, it is impossible to adjust for all the extraneous factors which may impinge upon a particular behavioural change targeted by a communication programme. For example, language skills may be an easily identifiable confounder of a communication programme, affecting both accessibility to a programme and understanding of a programme. Even within a group which uses the same language, subgroups may have a more proficient grasp of both passive (understanding) and active (persuasiveness) use of the language, which may confound results of a communication programme. Thus, the internal validity of a study to assess a communication programme may be compromised.

Third, when the wider social context of a community is considered, including economic and social factors such as employment or education, demographic make-up, ethics, laws and religions, it becomes very difficult to ensure the external validity of a specific communication programme. At best, principles can be learnt, but communication programmes themselves have to be tailored for and developed with the communities they are meant for. There is no such thing as ‘off the shelf’ communication.

4.11

4.11.4 Techniques to use in emergency risk communication (ERC) research

Notwithstanding these challenges, there are techniques that should be employed in ERC research that can provide some insights into how successfully a communication programme has promoted positive behavioural change with respect to Health EDRM. Only with a thorough, planned evaluation – covering formative process, impact and outcomes – of every ERC project, can techniques be refined and benefits demonstrated.

Effective ERC promotes emergency risk literacy, which is analogous to health literacy, as described by Nutbeam*(8)*. Emergency risk literacy represents the cognitive and social skills that determine the motivation and ability to gain access to, understand and use information in ways that promote and maintain good health through the management and mitigation of emergency risk. However, promoting emergency risk literacy in individuals alone (a behavioural change approach) is unlikely to produce the most beneficial results.

The behavioural change approach of health promotion is based on the belief that providing people with information will change their beliefs, attitudes and behaviours *(9)*. Although a popular model, the provision of information on its own is rarely enough to change behaviour because it ignores the factors in the social environment that affect health, including social, economic, cultural and political factors *(10)*. Similarly, without taking into consideration the broader determinants of health in ERC, risk management is likely to be limited. The development of individual responsibility alone is rarely sufficient to effect sustainable behavioural change.

An extension of the behavioural change model is the self-empowerment approach, in which people are encouraged to engage in critical thinking and critical action at an individual level. This model aims to develop 'risk management skills', including decision-making and problem-solving skills, so that the individual is willing and able to maintain control of their life during an emergency. While this model can be successful for some individuals, it is unlikely to be successful across a whole population because it does not address social norms *(11)*.

4.11.5 Taking into account the determinants of health

In order for ERC to be sustainably successful at a population level, the determinants of health must be taken into consideration. The social determinants of health are the conditions in which people are born, grow, work, live, and age, and the wider set of forces and systems shaping the conditions of daily life *(12)*. Determinants of health include education, housing, employment and the environment. They have a far greater effect on health outcomes that the provision of health services alone. Addressing the determinants of health also has a far more profound effect on the ability of a community to manage emergency risk than simply providing the information alone. At the most fundamental level, the three ultimate determinants of disaster risk are poverty, inequity and planetary health (including climate change). These three determinants are also the key

modulators of emergency risk management, and so must be addressed by ERC. These three fundamental issues underpin the great UN initiatives of 2015 – the Sustainable Development Goals, the Paris Climate Change Agreement and the Sendai Framework (Chapter 1.2) *(13)*.

Addressing the determinants of health and disaster risk requires a collective action model – a socio-ecological approach that takes into account the interrelationship between the individual and the environment. Although individual empowerment is necessary, it is not sufficient to generate change at a population level. The collective action model generates population-level change by encompassing ideas of community empowerment and requiring people to individually, but also collectively, acquire the knowledge, understanding, skills, and commitment to improve the societal structures that have such a powerful influence on a community's ability to manage disaster risk *(14)*. It engages people in critical thinking in order to improve their understanding of the factors affecting individual and community well-being. It also engages groups of people in critical action that can contribute to positive change at a collective level.

4.11.6 Components of communicating risk effectively for emergencies

Whichever model is used, there are three essential components to communicating risk effectively for emergencies *(4)*: building trust, integrating communication into prevention, preparedness, response and recovery and specific techniques (including developing a compelling message with the target community, identifying the appropriate balance of media for communicating the message and evaluating the programme).

1. Building trust

Techniques used in both the development and the evaluation and research of communications strategy may be similar, and involve a mix of qualitative and quantitative methods (Chapter 4.13). Audience reach data is often already available from print and broadcast media, which may indicate which media are most trusted for, and used to garner, information. Generally, familiarity engenders trust in individuals, so elders are often more trusted than younger people, but this may need to be confirmed at a local level through surveys, focus groups and interviews.

2. An integrated approach

Communication needs to be integrated into every level of risk management. Bringing media and communication experts into the planning process is more likely to produce messaging which is acted on than simply providing information to the media. Moreover, commercial media have skills in measuring content and effectiveness of messaging beyond that usually found in health organizations. Experts in communication can provide valuable advice when considering the choice and balance of the multiple means by which the message is conveyed. This will also require careful consultation with 'target' communities and their agencies. There is also useful information to be garnered from wider consultation with other agencies (such as government, nongovernmental and private business), which can inform the communication process. For example, some agencies will have information on which people in a community are key

influencers and certain agencies may have specific skills in messaging. The planning process should cover all aspects of emergency risk management, starting with identifying and mitigating risk. During an emergency response, communication will usually focus on immediate survival issues ('drop, cover and hold' in an earthquake; 'seek higher ground in a tsunami', for example). Once the immediate threat is over, there will be a much longer period of recovery involving primary care, maternal and child health and subsequently an even longer period of psychosocial recovery that will involve employment, housing, education and the agencies responsible for the wider determinants of health.

3. Specific techniques

The seven Cs of a good communication were originally described more than sixty years ago in the context of 'public relations' *(15),* but these principles have been adapted for many areas of communication, including humanitarian relief *(16)*:

- Correct – evidence based
- Concise – pithy
- Clear – it says what you mean
- Courteous – cultural values are important
- Complete – as comprehensive as possible
- Considered – with the target community and the agencies which serve them
- Concrete – be specific, not vague.

Despite the apparent objectivity of this schema, developing a good message is more art than science, which is why the involvement of a good communications team in message development is important. Good advertising slogans are often attributed with improving the sales of a product – sales of Nike running shoes went up tenfold in ten years after 'Just Do It' was introduced, for example *(17)* – but slogans to change behaviour during a disaster are more difficult to develop and more difficult to evaluate. 'If it's brown, flush it down; if it's yellow let it mellow' was a slogan used to minimize toilet usage and protect the fragile sewerage system after the Christchurch earthquakes *(18)*. It ticked most of the seven Cs, but there has been no formal evaluation of the message's success.

Deciding which media to use can be difficult. Increasingly, social media is used to convey messages *(19)*, but conventional television, radio and print media still have a place. For example, Katy Perry, Barack Obama and Justin Bieber each have more than 100 million followers on their Twitter accounts, but more than 3.5 billion people watched the FIFA world cup final on television in 2018. In some cases, a 'soapbox' presentation to an audience may be the best way to deliver a message, especially if power is out and buildings are destroyed. Once again, consultation with the target audience is important, using a collective action model of health promotion. Local knowledge can help decide which media mix will gain the greatest attention.

4.11.7 Research and Evaluation

The evaluation and research of any communication campaign requires a mixture of quantitative and qualitative methods (Chapter 4.13). Data gathering tools include focus groups, surveys, interviews, case studies, social media and/or website monitoring ('hits'). A series of measurements – formative (baseline), process, impact and outcome measures – will need to be budgeted for, in order that changes in awareness, knowledge and ultimately behaviour can be tracked over time. Questions about specific communication programmes can be added to routine surveys or market research, as well as specific surveys tailored to the programme be carried out. Well-funded, well-designed and well-implemented surveys should follow a communication strategy over time (before, during and after), and be able to compare different specific subgroups targeted by the strategy.

Such surveys are able to concentrate on positive outcomes of communication programmes, where routinely collected data tends to focus on negative outcomes of disasters. Questions should follow the pattern:

- Are you aware of the programme?
- Did the programme convey knowledge to you?
- Did you change your behaviour as a result of this knowledge?

Behavioural changes can sometimes be corroborated by objective measures. WHO has identified gaps in communication research and evaluation which, although they highlight deficiencies in current knowledge, also identify where there are research opportunities in the future. These gaps/opportunities include a lack of longitudinal studies and of studies of behavioural change (outcomes).

The gaps and research opportunities are particularly marked in low-income countries and among low income or vulnerable groups.

4.11

4.11.8 Case studies

The following four case studies highlight examples of communication research relevant to Health EDRM.

Case Study 4.11.1
The 'All Right?' Campaign, Canterbury, New Zealand 2012

The 'All Right?' campaign is a population-based, multi-media health promotion aimed at improving psychosocial well-being following the 2010-2011 Canterbury earthquakes. It was formatively evaluated and has been continuously evaluated through a series of iterations over several years *(20).* Methods of quantitative and qualitative evaluation include semi-structured interviews for process evaluation, survey questions developed with a market research company aimed at 400 randomly selected Christchurch residents, and specific tailored questions addended to the Canterbury Well-being survey – which is a survey of more than 2000 people carried out initially every two years, then annually to monitor Cantabrians' well-being in the wake of the earthquake sequence *(21).* In May 2018, half of Cantabrians (population 400 000) were aware of the 'All Right' campaign and of those who were aware of it, nearly 90% thought the messages were useful. More than 70% felt that the messages were useful for them personally and 42% claimed to have done at least one of the simple activities advocated by the campaign including, but not limited to, the Five Ways to Well-Being – Communicate, Learn, Be Active, Take Notice and Give (22).

The 'All Right' campaign in conjunction with the Canterbury Well Being survey, is an example of a thoroughly planned and researched communication programme. Inevitably, well-being is often measured subjectively and may require corroboration with more objective measures.

Case Study 4.11.2
'Staying Alive', Health Professional led Urban Radio, Ghana 2015

A formative evaluation identified a gap in information, education and communication about policies and practices in healthcare delivery, healthcare financing, training, ethics, research and environmental issues in Ghana (23). In June 2015, medical practitioners collaborated with a private, local, English-speaking radio station to produce and host a weekly health show whose content was aimed at discussing health from the viewpoint of practitioners, clients, policy makers, administrators and financiers in a simplified language for the general public, including healthcare trainees.

Since July 2015, the show, called 'Staying Alive', has aired weekly with audience analysis demonstrating its appeal to a wide range of active listeners. 'Staying Alive' remains one of the only shows in Ghana with a holistic approach to health hosted by health professionals. The evaluation of the impact of the show was crudely measured by the number of messages received and the number of telephone calls during the call-in segment. The integration of Facebook live expanded the reach of the show and Facebook analytics were useful in determining how many people watch the live show. Listener surveys by a commercial media measurement company (GeoPoll) was able determine a high number of people listening to the 'Staying Alive' compared with other English-language programmes but could not measure the impact of the message and its eventual impact on health.

This shows that where resources are stretched, pre-planning and appropriately detailed research and evaluation are difficult; but that international collaboration may help to address the gaps.

4.11

Case Study 4.11.3
'Amrai Pari' reality TV programme, Bangladesh 2014

BBC Media Action is the BBC's international development charity. It supports media and communication efforts that strengthen governance, improve people's health, increase their resilience and improve emergency response. In Bangladesh, BBC Media Action broadcast a national TV reality show to build resilience alongside roadshows and work with the Bangladeshi Red Crescent to integrate new communication tools into their already established system of long-term, two-way conversations with communities about risk identification and resilience. The Amrai Pari ('Together We Can Do It') reality television programme helps build people's resilience by empowering communities to work together to be prepared for extreme weather conditions. It started as a television programme, but also includes events teaching practical life-saving skills, educational performing arts shows involving music and drama, and festivals with up to 2000 attendees. The programme featured communities adopting low cost, replicable solutions to everyday problems caused by extreme weather and changing weather patterns.

The project reached 22.5 million Bangladeshis, with impact research showing 78% of viewers reporting better understanding of how to prepare for extreme weather – and, more importantly, 47% of viewers reporting they took action after watching the programme *(24)*.

This highlights how BBC Media Action produces communications programmes that are thoroughly formally evaluated and researched and based on tried and tested communication models. Like the 'All Right' campaign in Case Study 4.11.1, the evaluation relied in large part on subjective evaluation.

Case Study 4.11.4
The Pandemic Roadshow, New Zealand 2007

In conjunction with a local virologist, a dietician, a public health physician and an emergency planner, a children's science museum in New Zealand developed six portable exhibits designed to demonstrate the risks of influenza and how they could be mitigated and prepared for. The exhibits were based on the mnemonic CHIRP representing 'Cough etiquette', 'Hand Hygiene', 'Isolation', 'Reducing germs' and 'Preparation'. Cough etiquette showed how far people need to keep apart to prevent the spread of respiratory viruses. Hand hygiene used glow gel to demonstrate how easily germs can spread if hands are not washed properly. Isolation used a domino display to demonstrate how one infected person can lead to many more people succumbing, and how this is prevented by appropriate social distancing. Reducing germs showed how the influenza virus can be transmitted on surfaces which are not cleaned properly. Preparation challenged participants to find appropriate items for an emergency preparedness kit. The sixth display demonstrated a suitable healthy preserved food store for a family of four (and one pet) for a week. This exhibition was circulated among local government leaders, public libraries and schools for two years in conjunction with more conventional preparedness messaging delivered by video or print media.

In the region of New Zealand where the Pandemic Survival Roadshow was used, a random telephone survey demonstrated that the proportion of local population who were aware of the threat of pandemic influenza was almost twice that of the national average. In addition, people who had viewed the Pandemic Survival Roadshow were statistically significantly more likely to have an emergency preparedness kit. Such preparation served the population well during the 2009 H1N1 influenza pandemic and the all hazards approach had spin off benefits when the same population was affected by earthquakes in 2010 and 2011 *(25)*.

This case study shows how awareness following the Pandemic Survival Roadshow was objectively evaluated and compared to other areas across the country. The effectiveness of the programme, particularly with respect to the all hazards approach, was able to be tested (unfortunately) by the Canterbury earthquake sequence, which followed closely after the H1N1 2009 influenza pandemic.

4.11

4.11.9 Conclusions

Health EDRM research is inherently challenging, and nowhere more so than in the area of emergency risk communication. However, by integrating communication programmes into all aspects of the disaster cycle, developing the programmes using evidence-based techniques, using the appropriate balance of media for delivering the programmes and following recognized schema for evaluating such programmes, a valuable contribution can be made not only to disaster risk reduction in the communities served, but also to generating transferable knowledge to inform future emergency risk communication programmes in a diverse range of situations and societies.

4.11.10 Key messages

- **Emergency risk communication (ERC) is an essential part of emergency preparedness.**
- **The essential components of effective communication during emergencies are trust, integration and the seven "C"s of effective communication – correct, concise, clear, courteous, complete, considered, concrete.**
- **Research and evaluation of ERC can be difficult in the pressured environment of an emergency or disaster, but can be achieved with careful advance planning.**
- **In order to learn from and improve ERC, formal evaluation techniques should be applied to ERC, which requires forethought and funding.**

4.11.11 Further reading

Bailey N, Hoque M, Michie K, Ur Rabbi F. How effective is communication in the Rohingya refugee response? An evaluation of the common service for community engagement and accountability. Bridging theory and practice: Research Report British Broadcasting Corporation Media Action, London; 2018 (http://downloads.bbc.co.uk/mediaaction/pdf/research/rohingya-research-report.pdf, accessed 18 January 2020).

Chan EYY. Building bottom-up health and disaster risk reduction programmes. Oxford, UK: Oxford University Press; 2018.

Chan EYY. Public health humanitarian responses to natural disasters. New York, USA: Routledge Humanitarian Studies; 2017.

Sellwood C, Wapling A. Health emergency preparedness and response. Oxfordshire: CAB International; 2016.

4.11.12 References

1. Moon BK. Invest Today for a Safer Tomorrow – Increase Investment in Local Action. Plenary: Global Platform for Disaster Risk Reduction Third Session. UNISDR; 2011, 10 May.
2. Porter K. How Many Injuries can be Avoided Through Earthquake Early Warning and Drop, Cover, and Hold On? Structural Engineering and Structural Mechanics Report Series 16-04, University of Colorado Boulder. 2016. http://www.colorado.edu/ceae/node/1096/attachment (accessed 18 January 2020).
3. Castleden M, McKee M, Murray V, Leonardi G. Resilience thinking in health protection. Journal of Public Health; 2011: 33(3): 369-77.
4. Communicating Risk in Public Health Emergencies: A WHO Guideline for Emergency Risk Communication (ERC) policy and practice; WHO. 2018 https://www.who.int/risk-communication/guidance/download/en (accessed 18 January 2020).
5. Balshema H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J et al. GRADE guidelines: 3. Rating the quality of evidence. Journal of Clinical Epidemiology; 2011: 64: 401e406.
6. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ ; 2008: 336(7650): 924-6. doi: 10.1136/bmj.39489.470347.AD.
7. Bradley DT, McFarland M, Clarke M. The Effectiveness of Disaster Risk Communication: A Systematic Review of Intervention Studies. PLOS Currents Disasters; 2014: 22 August Edition 1.
8. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century Health Promotion International; 2000: 15(3): 259–67.
9. Ngigi, Samuel & Busolo, Doreen. Behaviour Change Communication in Health Promotion: Appropriate Practices and Promising Approaches. International Journal of Innovative Research and Development; 2018: vol. 7. doi: 10.24940/ijird/2018/v7/i9/SEP18027.
10. Laverack G. The Challenge of Behaviour Change and Health Promotion. Challenges; 2017: 82(2): 25.
11. Tengland PA. Behavior Change or Empowerment: On the Ethics of Health-Promotion Goals. Health care analysis: HCA: journal of health philosophy and policy; 2013: 24(1): 24-46.
12. WHO Commission on Social Determinants of Health and World Health Organization. Closing the gap in a generation: Health equity through action on the social determinants of health: Commission on Social Determinants of Health final report. Geneva, Switzerland: WHO, Commission on Social Determinants of Health. 2008.
13. Maini R, Clarke L, Blanchard K, Murray V. The Sendai Framework for Disaster Risk Reduction and Its Indicators—Where Does Health Fit in? International Journal of Disaster Risk Science 2017: 8: 150–5.

14. Racher FE, Annis RC. Community Health Action Model: health promotion by the community. Research and Theory for Nursing Practice; 2008: 22(3): 182-91.

15. Cutlip SM, Center AH. Effective Public Relations; Pathways to Public Favor. Prentice Hall Inc. 1952.

16. Bokhari, A. Training – Communication at USAID. 2015 https://www.slideshare.net/AzharBokhari1/trainingcommunication-at-usaid (accessed 18 January 2020).

17. Conlon J. Analyzing Nike's controversial "Just Do It" campaign. Branding Strategy Insider; 2019 https://www.brandingstrategyinsider.com/2018/09/analyzing-nikes-controversial-just-do-it-campaign.html (accessed 18 January 2020).

18. Ardagh M, Deely J. Rising from the Rubble: A health system's extraordinary response to the Canterbury earthquakes. Canterbury University Press; 2018.

19. Alexander DE. Social Media in Disaster Risk Reduction and Crisis Management. Science and Engineering Ethics; 2014: 20: 717–33.

20. Community & Public Health. Evaluation of the All Right? Campaign for tangata whaiora /mental health service users: An evaluation report prepared for All Right? and Mental Health Advocacy and Peer Support (MHAPS) Canterbury District Health Board; 2018.

21. Morgan J, Begg A, Beaven S, Schluter P, Jamieson K, Johal S et al. Monitoring wellbeing during recovery from the 2010–2011 Canterbury earthquakes: The CERA wellbeing survey. International Journal of Disaster Risk Reduction; 2015: 14: 96-103.

22. Aked J, Marks N, Cordon C, Thompson S. Connect, Be active, Take Notice, Keep Learning and Give...Five Ways to Wellbeing: A report presented to the Foresight Project on communicating the evidence base for improving people's well-being. London, UK: The New Economics Foundation; 2008.

23. Bonney J, Osei-Tutu L, Selormey R, Hammond B, Bonsu P. Public Information, Education, and Communication (IEC) of Health: Active Participation of Health Practitioners in Urban Radio in a Low Resource Setting. Prehospital and Disaster Medicine; 2019: 34(S1): S75.

24. Zuberi S. What has been the impact on Bangladesh of Amrai Pari? British Broadcasting Corporation Media Action, London; 2018. https://www.youtube.com/watch?v=4aNdQz_LQgI (accessed 18 January 2020).

25. Humphrey A . From Pandemics to Earthquakes: Health and Emergencies in Canterbury, New Zealand. In: Sellwood C, Wapling A, editors. Health Emergency Preparedness and Response; CAB International; 2016.

4.12 Qualitative Research

Authors

Christina J. Pickering, Faculty of Health Sciences, University of Ottawa, Ottawa, Canada.

Suzanne Phibbs, School of Health Sciences, Massey University, Manawatu, New Zealand.

Christine Kenney, Joint Centre for Disaster Research IRDR ICoRE, Massey University, Wellington, New Zealand.

Tracey O'Sullivan, Faculty of Health Sciences, University of Ottawa, Ottawa, Canada.

4.12.1 Learning objectives

To understand key factors to consider when developing a qualitative study for health emergency and disaster risk management (Health EDRM) research, including:

1. The epistemological foundations of qualitative research commonly used in disaster research.
2. Common qualitative research methodologies used extensively in disaster research.
3. Different methods used in qualitative data collection.
4. The power of participatory, performatory and arts-based research methods in disaster risk reduction (DRR).
5. Common issues and challenges for qualitative research in a disaster context.

4.12.2 Introduction

This chapter presents an overview of qualitative research methodologies that are commonly used in the study of disasters and relevant to Health EDRM. It highlights different types of qualitative methods and the challenges associated with each type, and explains how qualitative designs can be used to round out the evidence base and fill knowledge gaps. The chapter focuses on the epistemological foundations of the qualitative research methodologies commonly used in disaster research; information on other factors influencing qualitative research is available elsewhere (for example, see Chapter 3.4 and Philips *(1)* on ethical issues in disaster research, Emmel *(2)* on sampling, Saldaña *(3)* on data coding and Curtis and Curtis *(4)* on analysis).

Although disaster research has typically focused on quantitative methods – particularly modelling and survey designs *(5)* – qualitative methods have a

long history of use within disaster research *(1)* and are able to provide different types of evidence. Despite this, qualitative approaches are increasingly marginalized in discussions of evidence-informed practice or DRR policy development, in comparison to the greater attention given to indicators, tools, measurements, computer simulations and technological solutions in discussions of evidence-informed practice or disaster risk reduction (DRR) policy development *(6–7)*. However, the unpredictability of disaster contexts, combined with the need to capture time-sensitive information, means that qualitative research is often more practicable than quantitative study designs *(1, 8)*.

4.12.3 What is qualitative research?

The approaches to qualitative research introduced in this chapter are primarily concerned with 'the exploration of lived experience and participant-defined meanings' *(9)*. This version of qualitative research looks at the world from a naturalistic and interpretive perspective, situating the researcher in the world they are exploring *(1, 10)*. Qualitative research includes primary and secondary data collection and analysis. Primary data is collected face-to-face by the researcher through asking people about their interpretations, understandings, and lived experiences of a particular topic or event. Secondary qualitative data collection involves an exploration of pre-existing sources of information such as websites, publications or media reports *(11)*. Depending on the type of research question, the data generated through qualitative research designs may include participant narratives and field notes from observations, as well as photos, videos or documents. As described by Denzin and Lincoln, qualitative research practices "turn the world into a series of representations, conversations, photographs, recordings, and memos to the self" *(10*, p.3). Qualitative research can help inform and guide evidence-based practice in public health *(12)* and DRR *(13)*.

Rather than focusing on numbers *(14)*, qualitative researchers focus on the qualities of the topic being explored. When a research question seeks to answer 'what?' or 'how?' *(1, 9)*, qualitative research is typically the best strategy *(15)*. Qualitative research contributes by exploring people's meanings, perspectives and experiences, studying how things and systems work, understanding context and unanticipated consequences, as well as discovering important patterns and themes across cases *(16)*. According to Creswell *(15)*, the strengths of qualitative research include:

- Reporting results in the voices of participants
- Placing research in its natural setting to include important contextual factors
- Smaller sample sizes allow greater depth of findings
- Emerging, exploratory and open-ended design allows flexibility in design for different populations
- Good design for marginalized populations
- A starting point when little is known about a topic
- Allowing multiple perspectives on a phenomenon

- Allowing study of sensitive topics
- Allowing for a complex understanding of a phenomenon using inductive and deductive reasoning.

4.12.4 Differences between qualitative and quantitative research

The dominant discourse around research has traditionally been focused on objective measurement, large representative samples and validity; these concepts are embedded in quantitative research designs *(17–18)*. Quantitative research describes social phenomena by using the breadth of data to facilitate broad and valid generalizations about populations *(15)*. In contrast, qualitative research aims to develop understanding of social phenomena through exploring, describing, troubling or explaining them. Qualitative research is based on an interpretivist (as opposed to a positivist) paradigm *(19–20)*. Qualitative approaches focus on in-depth analysis of data, the findings of which highlight the underpinning factors that explain the social world. Although qualitative designs can in some cases complement quantitative methods, as is it the case in mixed methods studies (Chapter 4.13), in general, qualitative methods generate different types of data, which enable researchers to answer different types of questions that quantitative designs are not suitable for *(20–21).*

Qualitative and quantitative forms of research correspond respectively to inductive and deductive approaches to inquiry. Inductive research, which is favoured in qualitative research, is a 'bottom-up' approach that involves reaching a conclusion based on observation and analysis of data gathered in the field. Inductive research builds theories based upon data collected in the process of doing research *(22)*. Deductive research, which is favoured in quantitative research, is a 'top-down' approach to theory and research that means finding a solution to a problem based upon evidence *(22)*. Deductive research tests theories which are developed through what is known in the existing literature and validated or troubled through the process of doing research *(4)*. It is common for researchers to use both inductive field-based theories and deductive literature-based theories in the analysis of qualitative research.

Table 4.12.1 summarizes common differences between qualitative and quantitative research methods. This list is adapted from and combines lists presented by Creswell *(15, p.15)* and Denzin and Lincoln *(10),* who have summarized the differences to help researchers decide which approach to use.

Table 4.12.1 Common differences between qualitative and quantitative research (adapted from *(10)* and *(15)*)

	Qualitative Research	Quantitative Research
Purpose	Understand and explore behaviour, opinions, experiences from participants' perspectives	Describe social phenomena; Discover facts
Design	Emerging and flexible	Standard and fixed
Paradigm	Multiple interpretations of reality exist (subjective)	Reality is fixed (objective)
Setting	Naturalistic (contextual)	Controlled (empirical)
Sample size	Small	Large
Data Collection	Open-ended Observation, interviews, focus groups, narratives, document analysis, artifacts	Closed-ended Objective measurements Questionnaires and surveys
Data analysis	Inductive Themes, text, images	Deductive Numerical comparisons and statistical inferences
Biases	Acknowledged and assumed to influence findings	Reduced or eliminated
Standards for Quality	Dependability, Credibility and Authenticity, Auditability, Transferability, Confirmability	Internal Validity, External Validity, Reliability, Objectivity

4.12.5 Assumptions in qualitative research

A key point to remember is that qualitative research and quantitative research are based on different assumptions. Much of the debate about the differences between the two approaches concerns paradigms, which are sets of beliefs or worldviews *(23)*. In quantitative research, it is assumed that bias must be reduced and eliminated (Chapter 4.1). In qualitative research, bias is acknowledged and assumed to influence the interpretation of the findings. When reports of qualitative studies are peer reviewed, it is not uncommon for critiques from inexperienced reviewers to include the need to eliminate bias. However, bias is inherent in any research project and is part of the underlying assumption in qualitative designs *(17)*. It is important when reading reports of qualitative studies to understand this underlying assumption and focus on how rigour is managed in the study.

Methods for enhancing rigour in qualitative research are built into the study design in order to ensure interpretations are accurate representations of the data generated. Although researchers have identified as many as 60 ways to think about research *(22)*, this section focuses on five worldviews that frequently inform qualitative disaster research: social constructionism (*24*), post-positivism, advocacy or participatory approaches, and pragmatism *(23)* as well as the importance of reflexivity in research *(25)*.

Post-positivism is based on the assumption that findings cannot be proven beyond doubt, but that confidence is improved through robust measures of reliability and validity. Researchers should remain neutral and reduce bias through attempting to both verify and falsify their hypotheses *(26)*. Post-positive approaches are suited to research that attempts to predict how people will act in a given situation.

Social constructionism recognizes that knowledge is not disinterested or apolitical, and that understandings and meanings are constructed and sustained through social interaction. Multiple realities co-exist, foreclosing the notion that there is one universal truth *(26)*. Social constructionism is particularly useful in the study of identity as well as of experience. In contrast, social constructivism has many similarities with social constructionism, but tends to focus on how individuals learn through social interaction within their peer group. As an example, research knowledge in this view is co-created by researcher and participant *(10)*. An example of the social constructivism approach may be found in the '7 Up' documentary series *(27)*.

The advocacy or participatory approach recognizes that lay people have their own knowledge systems and are able to act and solve local problems. Participatory research is community based, empowering and transformative *(28).* Participatory research is particularly useful when working with communities or marginalized groups.

Pragmatism is the belief that the meaning of actions and beliefs are found in their consequences. Actions are situational, depend on shared sets of beliefs, and linked to consequences that are subject to change based on new experiences. Pragmatism, for example, is implicated in the choice of research method as assumptions are made about the research outcomes that may result from each method. Pragmatic inquiry is particularly suited to research in the area of decision making as well as in relation to novel events *(21).*

In qualitative research, reflexivity involves 'understanding the role of self in the creation of knowledge' *(25*, p. 220) through attention to how the situated knowledge of the researcher impacts on their research (for example, their choice of research design, disciplinary background, beliefs, personal experiences and demographic characteristics) *(25, 29)*. It is therefore important to be transparent with the reader about the researcher's worldview because it will have practical implications for the study, including theoretical frameworks, methodologies, and methods *(23)*.

4.12.6 Subjects versus participants

In qualitative studies, people who contribute to the research by being interviewed or completing arts-based activities as part of data generation are referred to as participants or co-researchers, rather than subjects. This discourse is reflective of a paradigm where research is not done 'on' subjects, but 'with' people. In many participatory methods, there are strong relationships between the researchers and participants or community organizations. These relationships and projects can span many years, and there is joint ownership and direction of the projects. In a disaster context, this point is extremely important given the nature of projects where citizens and communities may be in vulnerable settings following a disaster. The term 'participants' conveys voluntary engagement in the research and reflects the relationships in partner-based projects.

4.12

4.12.7 Five common approaches to qualitative research

Reports of qualitative research should provide a detailed description of the approach, reference seminal authors and justify why the approach was chosen and how the approach informs the procedures of the study (for example, interview type, focus group, observation and so on *(30)*). Outlined below are five common research methodologies used in qualitative research, as described by Creswell *(30)* – narrative research, phenomenology, grounded theory, ethnography and the case study. Also included are a brief description of ethnomethodology and a case study that highlights its application in New Zealand, in order to illustrate the potential of this approach for disaster research.

Narrative research

Narrative research explores people's experiences, as told in the form of stories from one or more individuals of interest *(30)*. Ideally, this leads to an exploration of an individual's life, their identity and how they situate themselves in the world. Storytelling, giving an account of events or actions, predominantly uses interviews and documents to collect the data, but can also rely on observation, use of pictures and group conversations as data collection methods. Several strategies for data analysis can be chosen, depending on the purpose of the research, including thematic analysis, structural analysis and dialogic/performance analysis. It is common for researchers to 're-story' or reconstruct a story told by a participant, so that the report presents the story chronologically, highlighting 'turning points', and important contextual information. With this restructuring, the researcher is often seen as a collaborator in the storytelling process and thus requires much reflexivity on the part of the researcher to reflect on their own assumptions and experiences and how that might affect the way they re-story the data. An additional challenge to this type of research is the amount of data collection that must occur to capture a full and clear picture of the context surrounding the story.

Phenomenology

The purpose of phenomenology is to understand the universal 'essence' of the experience of a phenomenon *(30)*. This approach differs markedly from narrative research as phenomenology goes beyond the individual experience to describe the common meaning for several individuals. In this case, the unit of analysis is 3 to 25 individuals who have all experienced the same phenomenon (such as grief). Just as for narrative research, individual interviews are the most common method of data collection. However, documents, observation and art have also been used. The researcher's stance in phenomenology is to bracket themselves out of the study by reflecting on personal experiences with the phenomenon and setting those aside to focus on the experiences of their participants. Textual and structural analysis of the data summarizes what and how the phenomenon is experienced, ending with a descriptive report of the universal essence. Challenges to this approach include discussing philosophical assumptions of abstract concepts (such as grief), careful selection of participants so that they have all experienced the same phenomenon, and the difficulty that researchers often find in trying to bracket their personal experiences with the concept under study.

Grounded Theory

The purpose of grounded theory is to generate a theory that is grounded in the data to explain a process (for example, the process of dying) *(30)*. Grounded theory uses theoretical sampling to collect data from 20 to 60 participants who have all experienced a process. As is the case with narrative research and phenomenology, one-on-one interviews are the most common method of data collection. The grounded theory researcher constantly compares data across interviews with their memos on the researcher's emerging ideas for a theory. The data analysis strategy used depends on the grounded theory approach a researcher chooses. Glaserian grounded theory uses active codes (see Charmaz *(31)* for more on this approach), while Straussian grounded theory uses open coding, axial coding and selective coding (see Corbin & Strauss *(32)*). It is important not to confuse literature on the two distinct approaches. Straussian grounded theory presents a more structured approach than the Glaserian methodology. In their final report, the researchers will produce a diagram, hypothesis or both to accompany the discussion of their results. A negative attribute of this approach is that it tends to be reductive.

Ethnography

Ethnography describes the social behaviours of a culture-sharing group *(30)*. Here the researcher is tasked with both describing and interpreting topics such as group values, behaviours, beliefs and languages learned. In this case, the unit of analysis is an entire – or subset of a – large culture-sharing group. This approach requires extensive fieldwork using a variety of data collection methods, such as observation, interviews, symbols and artifacts. Most often, researchers are participant observers in which they become immersed in the day-to-day lives of the group they are researching, both observing and participating in the world around them. Data analysis in ethnography typically begins with an insider emic perspective of the data through verbatim quotes, which then gets moved into an etic scientific perspective to develop the overall interpretation of social behaviours of the group. There are several approaches to ethnography including, but not limited to autoethnography, critical ethnography, participatory action research and realist ethnography.

Case study

The purpose of a case study is to develop an in-depth understanding of a single case, or multiple cases *(30)*. Cases can be one person, several people, a group, a programme, an activity, a setting and so on. It is important that the case be clearly defined within a bounded system. A distinct characteristic of the case study is the use of multiple sources of data or multiple forms of data collection methods in a single study to develop an in-depth understanding of the case. Data collection methods can include interviews, observations, documents and artifacts. Data analysis can be explanatory, exploratory, or descriptive using themes and cross-case themes. Approaches to a case study differ depending on whether the researcher wishes to look at the case itself (intrinsic), the wider purpose of the case (instrumental), or look at comparing cases (collective case). It can be difficult to successfully identify and bound a case, and to keep a case study focused, because the more cases are studied, the more the overall analysis will be diluted. For this reason, it is recommended to include no more than four or five cases in a multiple case

4.12

study design. It is also important that, whatever decision is made, a rationale is provided for these choices.

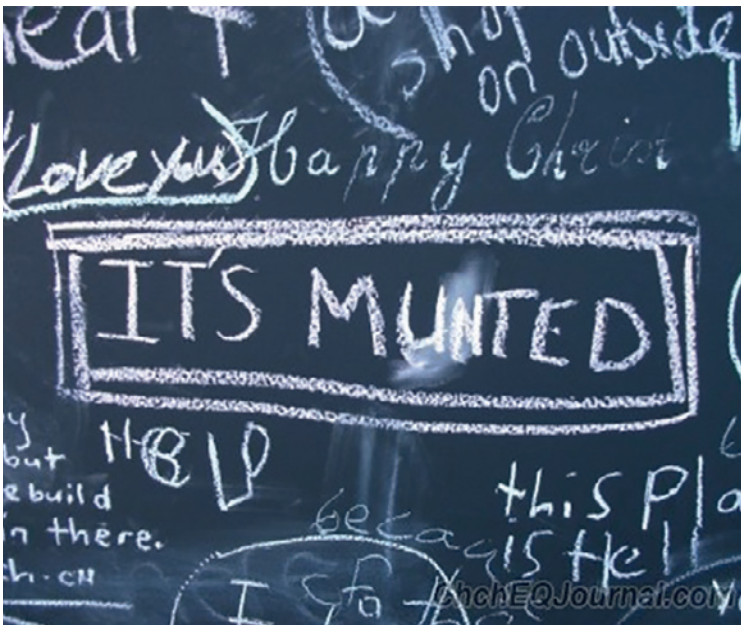
Ethnomethodology

Ethnomethodology has its origins in sociology and a focus upon disruption, which makes it a particularly useful research method in Health EDRM and disasters. This research methodology explores singular events, including how people interact and make sense of occurrences. Although similar to ethnography, ethnomethodology differs through its focus on the knowledge and methods employed by people in their everyday lives (for a discussion of differences between ethnomethodology and ethnography, see Pollner and Emerson *(33)*). Ethnomethodology considers the context of language and meaning through attention to the work of the streets. In an ethnomethodological approach, disruption enables consideration of the process through which the stable features of social organized environments are created and sustained *(34–35)*. Ethnomethodology may be used to look at the everyday micro processes of social interaction, as well as how people cope with and make sense of large scale events *(36)*.

As an example of paying attention to the work of the streets, people in Canterbury New Zealand used the term 'munted' to create a shared language around the 2010-2011 earthquakes, as the images in Figure 4.12.1 illustrate.

Figure 4.12.1 The creation of a shared language in relation to the Canterbury Earthquakes

Source: Outside the Square Creative, https://www.outsidethesquare.net.nz/portfolio/munted/



Source: ChchEQJournal.com. Written on a community blackboard in which people shared their feelings, Christchurch, February 2011.

Use of the term 'munted' was also evident in the following sample narrative texts from qualitative studies conducted in Christchurch by Phibbs and Kenney, following the Canterbury earthquake sequence:

I was getting text after text... don't go down Fitzgerald Ave, no bridge. Dallington is munted, no access. St Albans is closed
(NL, Female, 2012, Māori community research).

P1: The house next door to us... that's triple brick so it's got no internal... timber framing,... it's basically just cracked right through... P2: It's munted
(P1 Male, P2 Elderly Female 2012, disability and disaster research).

We didn't open the marae (Māori community centre) because we had no toilet facilities... so we weren't able to operate, we had all our ablution block, piping, our plumbing that was all totally munted so it couldn't happen for us
(ML, Female, 2012, Māori community research).

The term 'munted', which before the earthquakes had referred to an intoxicated person or something that was broken, came to symbolize the way in which individuals were interpreting and expressing their experiences of the post-disaster city.

A fundamental premise of ethnomethodology is that social reality and social order are accomplished through the ongoing actions of individuals who 'make meaning' out of the practices of everyday life *(26, 33)*. Disasters, as disruptors of everyday life, lend themselves to ethnomethodological analysis because they bring into view the taken-for-granted 'sense-making' processes through which social life is experienced, ordered and sustained.

4.12

4.12.8 Community-based participatory research (CBPR)

This chapter highlights one type of participatory approach to research, which is not limited to but frequently employs qualitative methods – community-based participatory research (CBPR). Because many disaster studies focus on the community level, it is often desirable and necessary to adopt participatory designs. This type of research has different names: participatory action research, advocacy research, CBPR, or community-led research (CLR), which is more commonly conducted by Indigenous researchers in partnership with Indigenous communities (see also Chapter 5.4). However, the premise is the same, with a focus on creating social change with a community through collaborative partnerships and shared decision-making. Regardless of whether a study is being done before a disaster (for example, prevention, mitigation or preparedness) or during and after (response and recovery), there are important guidelines for working with communities. When the research focuses on post-disaster impacts, special consideration must be given to avoiding unintentional harm in the community. Guidance from community partners is essential in order to understand the context and conduct the research in a way that is appropriate for the community circumstances. The following table summarizes the principles of CBPR outlined by Israel and colleagues *(37)*.

Table 4.12.2 Principles of Community-based participatory research (CBPR) (37)

Principle	Explanation
Acknowledge the community as a unit of identity	Community is not necessarily geographic. One of the defining characteristics of a community is identifying with it. People who identify with a community feel a sense of belonging based on certain attributes.
Research initiatives build on the strengths and resources within the community	This is what is referred to as a strengths-based or asset-based approach. In collaboration with community partners, researchers identify what assets (see Chapter 3.1) or resources are in the community and build on those strengths.
Facilitate collaborative, equitable, empowering partnerships where power is shared and inequalities are addressed through the research	The emphasis on collaborative, equitable partnerships is central to participatory projects. Within these partnerships, power is shared through collaborative decision-making. The research focuses on inequalities and social change to disrupt power differentials.
Co-learning and capacity-building for all partners	CBPR projects emphasize capacity-building within the community and within partnerships. The focus is for people to learn from one another and build capacity within themselves, their organizations and their communities.
Knowledge generation is balanced with intervention activities so everyone benefits	The knowledge that is generated from research processes must be balanced with intervention activities so that it is mutually beneficial for everyone involved and the community.

Principle	Explanation
Locally relevant projects which address public health problems and consider ecological perspectives related to determinants of health	Partnerships shape the research agenda by identifying locally relevant issues. CBPR projects can be directed toward understanding and acting on determinants of health which contribute to locally relevant health issues.
Foster systems development using a cyclic, iterative process	By understanding the context of a community and working in partnership with local citizens and organizations, CBPR projects can contribute to systems development or change through an iterative process. As new knowledge is generated, it can be integrated to improve systems within the community.
Sharing the findings and involving community partners in wider dissemination of knowledge	Knowledge which is generated from the project is shared and partners are involved in knowledge mobilization activities for wider dissemination of the findings. This ensures lessons learned from the project are shared with people who can benefit.
Long-term commitment with consideration of sustainability	CBPR projects involve long-term processes starting with partnership development, identification of the issues, designing the project components, securing funding, implementing the research activities, analyzing data, and dissemination of the knowledge generated. Partnerships often continue beyond individual projects in the interest of ensuring sustainability and long term system change.

According to Phibbs and colleagues *(38)*, there is a distinction between top-down participatory approaches and bottom-up community development. The different approaches to working with communities influence working partnerships and relationships between DRR organizations and communities. They write:

"In community-based health promotion, problems, targets and actions are defined by the sponsoring body. The notion of community is relatively unproblematic, with community settings being viewed as venues for interventions that largely target the individual. In these top-down community-based interventions, activities are mainly health, or in this case disaster preparedness, oriented. Community-based initiatives tend to be single issue focused and time-limited, discontinuing once the sponsoring body has withdrawn."

In contrast, an approach focused on community development ensures the identification of priorities, problems and appropriate actions that are determined by the community. Potential power differentials are recognized, empowerment is a priority, and actions focus on capacity building in the community. In a community development initiative,

> ...the target of the intervention may be the community itself or structures, services or policies that impact negatively upon the community by creating vulnerabilities. Activities may be broad-based, targeting wider factors which are associated with negative social outcomes, such as discrimination, poverty or crime, thereby providing

indirect disaster resilience outcomes such as facilitating community empowerment and enhancing social capital *(38)*.

The following case studies provide examples of qualitative research of particular relevance to Health EDRM.

Case Study 4.12.1
The EnRiCH Youth Research Team Photovoice Project, Canada

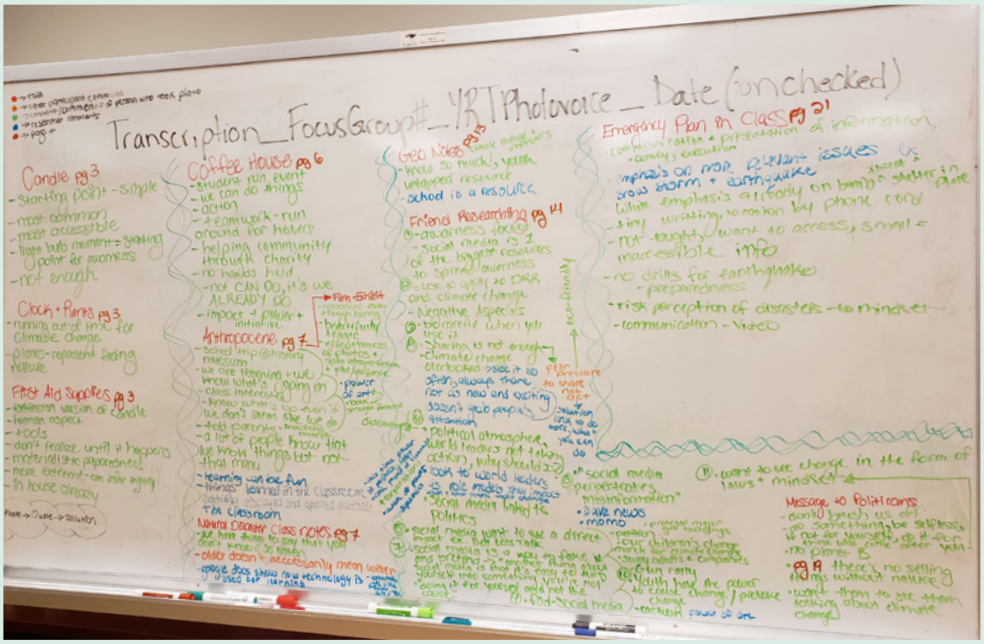
The EnRiCH Youth Research Team is a grass-roots initiative to engage youth in DRR research and action for social change to promote resilient communities. The team has met monthly since 2016, with an annual one-week mini-enrichment course, through a specialty programme offered for high school students by the University of Ottawa. The youth who are members of the team range from 13 to 17 years of age, and they are mentored by undergraduate and graduate university students who are part of the EnRiCH research team. As part of the regular meetings, the youth team members learn about the research team's projects and design and work on projects, including a series of education modules to teach youth about disaster preparedness. When the youth team members heard about the research team's Photovoice project, they asked if they could do their own Photovoice project to express their views about youth engagement in DRR and climate change action.

Photovoice is a qualitative participatory action-based research method used to engage and empower community members to reflect and co-create knowledge with researchers *(39)*. Participants are invited to take pictures of their personal experiences, and express their ideas through picture narrations *(40)*. Participants are actively involved in each of the five steps:

1. Identifying objectives and intended outcomes;
2. Deciding on Photovoice assignments;
3. Taking photographs about the topic;
4. Identifying themes; and
5. Planning a photo exhibition to connect with influential stakeholders in the community *(39)*.

The first Photovoice session for the EnRiCH Youth Research Team was held in March 2019. Over a one-year period, they participated in eight sessions, each lasting two hours, where they shared photos related to youth engagement in DRR, discussed issues and solutions for change, and identified concepts they would like to take photos about for subsequent meetings. The youth team will invite influential stakeholders related to DRR and climate change action to attend their exhibition in 2020, along with leaders and decision-makers in the education system and youth in the area. Figure 4.12.1 shows how the qualitative data were analysed from each Photovoice meeting to bring back preliminary themes for the youth participants to confirm or revise. This is referred to as member-checking in qualitative research and ensures rigour in confirming the themes are representative of the data.

Figure 4.12.2. Preliminary theme analysis in preparation to report back at the next Photovoice session (EnRiCH Research Lab – University of Ottawa)



This project underscores many principles of participatory research including:

- Pproject design shaped by the needs and preferences of the community members;
- Focus on existing strengths and resources in the community (see also Chapter 3.1);
- Advocacy and emphasis on social change for a locally relevant issue;
- Co-learning and capacity-building for all partners;
- Collaborative and equitable partnerships where decision-making is shared;
- Sharing of knowledge and involvement of everyone in planning the exhibition and dissemination of the findings.

This initiative has been long-term (supported by two research grants, 2012-2017; 2016-2019) and has required sustained effort to maintain resources and continuity for the youth involved in this team.

4.12

Case Study 4.12.2
Experiences of the 2010-2011 earthquakes in New Zealand

As community-based participatory research encourages trust on the part of community research partners and has been shown to promote the wellbeing of indigenous communities *(41)*, Kenney and Phibbs *(42)* conducted qualitative research using a participatory approach that encompassed similar principles, in New Zealand. The researchers' partnership with an indigenous Māori tribe, Ngāi Tahu, facilitated collaborative design and implementation of research that explored the earthquake experiences of local Māori following the 2010-2011 Canterbury earthquakes. Project aims included identification and documentation of cultural factors that facilitate Māori health and well-being, and development of recommendations for improving responders' approaches to addressing the psychosocial and health needs of communities, during disasters.

A point of difference with traditional CBPR was that an indigenous Kaupapa Māori research methodology *(43)* shaped the design and conduct of the research. The research was designed by and for Māori, as well as conducted by Māori researchers under the oversight of the local tribe and in accordance with Māori ethical principles *(44)*. Themes arising from the research were confirmed by the community, with the local tribe Te Rūngana o Ngāi Tahu retaining intellectual property rights over the findings and acting as co-authors of publications arising from the research. This process ensured the research became community-led rather than community-based or centred, and strengthened community engagement.

Historically, Māori like most indigenous peoples have used stories to create and ensure the intergenerational transmission of knowledge *(45)*. Contextually relevant narrative research methods which accommodated this process *(46–47)*, were therefore applied to gathering and analysing participant's stories.

Researchers specifically drew on dialogical interviewing techniques to capture participants' viewpoints. Dialogical interviewing *(48)* is an approach that is effective for disrupting power differentials between researchers and participants *(49)*. It is therefore particularly useful for gathering data when researching with marginalized individuals and communities, including, as in this instance, indigenous collectives.

The researchers used thematic analysis to identify discrete stories nested in participants' interviews and analysed the stories using whole narrative unit analysis. Participants' stories were examined to identify contextually complete blocks of texts which were analysed in paragraph format rather than line by line in order to retain the narrative quality of each participant's story. This approach also ensured that analytical findings did not become decontextualized. Narrative analysis highlighted how a nationalized Māori Recovery Network mobilized resources and support to the culturally diverse communities of Christchurch following the earthquakes. Findings showcased ways in which cultural attributes, Māori knowledges, values and practices, interwove to create moral and relational technologies, that when operationalized, addressed the immediate needs and facilitated the health and wellbeing of Māori.

Participants' talk also documented how cultural attributes could be used to enhance the recovery and resilience of the wider Christchurch community.

The prompt and effective disaster risk management approach implemented by Māori, aligns with key recommendations in the Sendai Framework for Disaster Risk Reduction (2015) *(50)*. Māori implementation of best practice in DRR, has generated increased willingness on the part of regional civil authorities, and government to engage and collaborate with local Māori tribes in strategizing for national resilience. To that end, research findings have shaped the development of New Zealand's recently released National Disaster Resilience Strategy *(51)*, as well as informing United Nations disaster science initiatives *(52)*.

4.12.9 Conclusions

This chapter presented an overview of different qualitative methods, as well as some of the issues to consider when designing and implementing studies using these designs. We also highlighted the importance of participatory design and collaborative partnerships, distinguishing between community-based and community-development approaches. In Health EDRM, qualitative methods can be used alone, or in conjunction with quantitative methods in a mixed method methodology (Chapter 4.13). Regardless of the overarching design, it is important to be aware of the differences in paradigms for each method and to consider how to minimize power differentials and maximize empowerment when conducting research with communities.

4.12.10 Key messages

- **Qualitative research design and methods has an important role in Health EDRM.**
- **Rigour is needed to ensure qualitative research contributions are of high quality and credible.**
- **Community-based research is based on partnerships and shared ownership of projects, where the voice of citizens in a community are valued and represented in the research findings.**
- **Qualitative research design has the potential to fill many research gaps in DRR, building on the fact that everybody experiences disasters differently and that disasters affect everybody in different ways.**
- **The emergent design of qualitative research offers the flexibility to address these complex and differing experiences.**

4.12

4.12.11 Further reading

Phillips BD. Qualitative disaster research. In Leavy P, editor. The Oxford handbook of qualitative research. Oxford: Oxford University Press; 2014.

Creswell JW. Qualitative inquiry and research design: Choosing among five approaches (3rd edition). Thousand Oaks, CA: Sage; 2013.

Saldaña J. The coding manual for qualitative researchers (3rd edition). London: Sage Publications; 2016.

Emmel N. Sampling and choosing cases in qualitative research: A realist approach. London: Sage Publications; 2013.

Curtis B, Curtis C. Social research: A practical introduction. London: Sage Publications; 2011.

4.12.12 References

1. Philips B. Qualitative Disaster Research. In Leavy P, editor. The Oxford Handbook of Qualitative Research; 2014.
2. Emmel N. Sampling and choosing cases in qualitative research: A realist approach. London: Sage Publications; 2013.
3. Saldaña J. The coding manual for qualitative researchers (3rd edition). London: Sage Publications; 2016.
4. Curtis B, Curtis C. Social research: A practical introduction. London: Sage Publications; 2011.
5. Vojinovic Z, Abbott M. Flood risk and social justice: From quantitative to qualitative flood risk assessment and mitigation. London: IWA Publishing; 2012.
6. Aitsi-Selmi, Amina & Blanchard, Kevin. Ensuring science is useful, usable and used in global disaster risk reduction and sustainable development: A view through the Sendai framework lens. Palgrave Communications; 2016: 2. 16016. doi: 10.1057/palcomms.2016.16.
7. Alexander D. Natural disasters. London: Routledge; 1993.
8. Murray V. Evidence for disaster risk management – information and knowledge needs for policy makers and field practitioners; 2011. https://www.preventionweb.net/files/18197_205murray.evidencefordisasterriskma.pdf (accessed 22 January 2020).
9. Willig C. Introducing qualitative research in psychology. Maidenhead, United Kingdom and New York, NY: Open University Press; 2008.
10. Denzin NK, Lincoln YS. The Sage handbook of qualitative research (4th edition). Thousand Oaks: SAGE Publications; 2011.
11. O'Reilly M, Kiyimba N. Advanced qualitative research: A guide to using theory. London: Sage Publications; 2015.
12. Jack S. Utility of qualitative research findings in evidence-based public health practice. Public health nursing; 2006: 23(3): 277-83.
13. Grynszpan D, Murray V, Llosa S. Value of case studies in disaster assessment. Prehospital and Disaster Medicine; 2011: 26(3): 202-5.

14. Merriam S, Tisdell E. Qualitative research: A guide to design and implementation. San Francisco, CA: John Wiley & Sons; 2016.

15. Creswell JW. 30 Essential skills for the qualitative researcher. Thousand Oaks, CA: SAGE Publications; 2016.

16. Patton MQ. Qualitative research & evaluation methods, Fourth Edition. Thousand Oaks, CA: SAGE Publications; 2015.

17. Giardina M, Newman J. The politics of research. In Leavy P, editor. The Oxford handbook of qualitative research. New York: Oxford University Press; 2014.

18. Stewart M. Whose evidence counts? An exploration of health professionals' perceptions of evidence-based practice, focusing on the maternity services. Midwifery; 2001: 17(4): 279-88.

19. Angen M. Evaluating interpretive inquiry: reviewing the validity debate and opening the dialogue. Qualitative health research; 2000: 10(3): 378-95.

20. Leavy P. Research design: Quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches. New York, NY: The Guildford Press; 2017.

21. Morgan D. Integrating qualitative and quantitative methods: a pragmatic approach. Thousand Oaks, California. Sage Publications; 2014.

22. Saldaña J. Thinking qualitatively: methods of mind. Los Angeles: Sage Publications; 2015.

23. Creswell JW. Qualitative inquiry and research design: Choosing among five approaches (2nd edition). Thousand Oaks, CA: SAGE; 2007.

24. Crotty M. Constructionism: The making of meaning. In: The foundations of social research. London: SAGE Publications; 1998: pp. 42-63.

25. Dodgson J. Reflexivity in qualitative research. Journal of Human Lactation; 2019: 35(2): 220-2.

26. Leavy P, editor. Oxford library of psychology. The Oxford handbook of qualitative research. Oxford University Press; 2014. doi: 10.1093/oxfordhb/9780199811755.001.0001.

27. Apted M. The UP series. The definitive collection of all the original Up series of films. Granada Television (1964 to the present); 2007.

28. Bowd R, Özerdem A, Kassa D. A theoretical and practical exposition of 'Participatory' Research Methods. In: Özerdem A, Bowd R, editors. Participatory research methodologies: Development and post-disaster/conflict reconstruction. London and New York: Routledge; 2010.

29. Bourdieu P. Vive la crise! For heterodoxy in social science. Theory and Society 17: 773-87; 1988.

30. Creswell JW. Qualitative inquiry and research design: Choosing among five approaches (3rd edition). Thousand Oaks, CA: SAGE Publications; 2013.

4.12

31. Charmaz K. Constructing grounded theory. London: SAGE Publications; 2006.
32. Corbin J, Strauss A. Basics of qualitative research: Techniques and procedures for developing grounded theory (3rd edition). Thousand Oaks, CA: SAGE Publications; 2007.
33. Pollner M, Emerson R. Ethnomethodology and ethnography. In Atkinson P, Coffey A, Delamont S, Lofland J, Lofland L, editors. Handbook of ethnography. London and New York: Sage; 2001.
34. Garfinkel H. Evidence for locally produced, naturally accountable phenomena of order, logic, reason, meaning, method, etc. in and as of the essential quiddity of immortal ordinary society, (I of IV): An Announcement of Studies. Sociological Theory; 1988: 6: 103-9.
35. Pollner M. Left of ethnomethodology: The rise and decline of radical reflexivity. American Sociological Review; 1991: 56: 370-80.
36. Phibbs S, Curtis B, Kenney C . "It's munted": Disasters and precariousness. Sociological Association of New Zealand Conference. Victoria University, Wellington; 4-7 December 2018.
37. Israel BA, Eng E, Schulz AJ, Parker EA. In: Israel BA, Eng E, Schulz AJ, Parker EA, editors. Introduction to methods for CBPR for health, methods for community-based participatory research for health, 2nd Edition, San Francisco: Jossey-Bass; 2012: pp.3-38.
38. Phibbs S, Kenney C, Severinsen C, Mitchell J, Hughes R. Synergising public health concepts with the Sendai Framework for Disaster Risk Reduction: A conceptual glossary. International Journal of Environmental Research and Public Health; 2016: 13: 1241.
39. Wang C, Burris MA. Empowerment through photo novella: Portraits of participation. Health Education & Behavior; 1994: 21: 171–86.
40. Latz AO. Photovoice research in education and beyond: A practical guide from theory to exhibition. New York, NY: Routledge; 2017.
41. Kwiatkowski R. Indigenous community based participatory research and health impact assessment: A Canadian example. Environmental Impact Assessment Review; 2011: 31(4): 445-50.
42. Kenney C, Phibbs S. A Māori love story: Community-led disaster management in response to the Ōtautahi (Christchurch) earthquakes as a framework for action. International Journal of Disaster Risk Reduction; 2015: 14(1): 46-55.
43. Smith L. Decolonizing methodologies research and Indigenous peoples (2nd edition). Dunedin, New Zealand: University of Otago Press; 1999.
44. Irwin K. Māori research methods and processes: An exploration. Sites; 1994: 28: 25-43.
45. King D, Goff J, Skipper A. Māori environmental knowledge and natural hazards in Aotearoa-New Zealand. Journal of the Royal Society of New Zealand; 2007: 37(2): 59-73.

46. Frank A. Why study people's stories? The dialogical ethics of narrative analysis. International Journal of Qualitative Methods; 2002: 1(1): 109-17.

47. Somers M. The narrative construction of identity: A relational and network approach. Theory and Society; 1994: 23(4): 605-49.

48. Frank A. What is dialogical research and why should we do it? Qualitative Health Research; 2005: 15(7): 964-74.

49. Sonn C, Green M. Disrupting the dynamics of oppression in intercultural research and practice. Journal of Community & Applied Social Psychology; 2006: 16(5): 337-46.

50. United Nations Office for Disaster Risk Reduction (UNISDR). Sendai Framework for Disaster Risk Reduction 2015-2030; 2015 http://www.unisdr.org/files/43291_sendaiframeworkfordrren.pdf (accessed 15 January 2020).

51. Department of Prime Minister and Cabinet. National Disaster Resilience Strategy Rautaki ā-Motu Manawaroa Aituā. Wellington, New Zealand: New Zealand Government; 2019.

52. Kenney C, Solomon MW. Māori Community-led Disaster Risk Management: An effective response to the 2010-2011 Christchurch Earthquakes. UNISDR Scientific and Technical Advisory Group Case Studies – 2014. http://www.preventionweb.net/files/workspace/7935_kenneyandsolomon.pdf (accessed 20 June 2019).

4.13 Addressing complexity through mixed methods

Authors

Tracey O'Sullivan, Faculty of Health Sciences, University of Ottawa, Ontario, Canada.

Yasmin Khan, Public Health Ontario and the University of Toronto, Ontario, Canada.

4.13.1 Learning objectives

To understand key factors to consider when developing a mixed methods study for research in health emergency and disaster risk management (Health EDRM) , including:

1. The basic principles of mixed methods research.
2. The relevance of mixed methods design for disaster research.
3. Systems thinking for use in disaster research.
4. The basic tenets of complexity theory and their relevance for disaster research.

4.13.2 Introduction

The timing, characteristics and non-linear impacts of different types of disasters contribute to the complexity of prevention, preparedness, response and recovery – as well as to the challenges for designing research relevant to disaster health and Health EDRM more generally. Although warning systems make it possible to anticipate some weather-related events, other types of disasters such as wildfires, tornadoes and pandemics typically provide little warning. For disaster health research, it is rare to see a simple, single research design that can capture the complexity needed for disaster studies, given the dynamic nature of the context around risks, hazards and events leading to a disaster. Mixed methods and a systems approach provide additional options to address some of these issues.

While mixed methods research is typically described in terms of its evolution over the past 30 years, some argue that this approach has been around for at least a century *(1)*. Nevertheless, it is recognized as a third methodology, with its own set of assumptions and criteria for quality *(2–3)* and not surprisingly, its own set of critiques *(1)*.

Mixed methods research combines qualitative and quantitative methods, and grew from the recognition that some research questions require both quantitative and qualitative methods to provide comprehensive answers. This approach is often used with complex problems, when quantitative or

qualitative methods are not sufficient on their own *(3)*. Given the challenges of defining mixed methods research, and the historical evolution of this approach, Johnson et al *(4)* asked leaders in research methods to provide a definition. Integrating the 19 definitions they received, they presented this definition:

"Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (such as the use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration."

While mixed methods research is common, it is not without critique and practical considerations. Flick *(1)* provides an excellent overview of the myths and mantras, and perhaps the most salient critique is that mixed methods research is somehow regarded as superior to quantitative and qualitative methodologies, despite the limited scope of methods used in mixed methods research (surveys, interviews, for example). Other challenges include defining what is meant by mixed methods research, and how to manage conflicting assumptions, paradigms and values. Criticisms of these studies often include lack of integration of the findings from the quantitative and qualitative arms of the design *(4)* and the need for more comprehensive triangulation *(1)*.

4.13.3 Mixed Methods Research Designs

When deciding whether or not to use a mixed methods research design, the first step is to consider the research question *(1)*, which as with all studies should drive the methodological approach (Chapter 3.5). Generally, complex questions require more complex methods. Simple research questions are characterized by having one concept or variable of interest, and one type of data needed to answer it. Complex questions have multiple concepts or groups, and changing trajectories. More than one type of data is needed to answer the question *(5)*. Once the research questions are in place, the most appropriate and feasible methodologies can be identified. In doing so, it is essential to be aware of theoretical and epistemological differences between the quantitative and qualitative methods being considered *(1)*.

In developing the design, researchers must decide whether the main method needed is quantitative or qualitative and how the supplementary method will support their analyses *(6)*. Palinkas and colleagues *(7)* and Creswell and Plano Clark *(8)* provide excellent overviews of different types of designs. They use capital letters (QUAN or QUAL) to indicate the weighting of the main method, and lower case letters to indicate the weighting of the other method (quan or qual), and the → or + symbols to indicate whether the methods will be implemented sequentially or simultaneously *(8)*. Table 4.13.1 summarizes different types of designs using this notation, and provides examples relevant to Health EDRM research.

Table 4.13.1 Overview of mixed methods designs and applications to Health EDRM research

Design	Structural description	Applied example	Data collection and analysis
QUAN + QUAL	Quantitative and qualitative methods are implemented simultaneously and have the same weighting	Randomized trial measuring behavioural outcomes following a disaster preparedness campaign using a community survey and telephone interviews with a subset of the survey sample	QUAN: Analysis of numerical survey responses on a Likert scale QUAL: Thematic analysis of interview data
QUAL + quan	Main method is qualitative, implemented simultaneously with quantitative method which is weighted less	Focus groups with citizens who have experienced flooding of their homes, supplemented with a short survey related to accessing mental health services	QUAL: Iterative content analysis of focus group data quan: Analysis of binary responses (for example, yes and no).
QUAN + qual	Main method is quantitative, implemented simultaneously with qualitative method which is weighted less	Exit survey with citizens attending influenza vaccination clinics, supplemented with field observations of crowd control and dynamics of the flow of service delivery	QUAN: Analysis of Likert scale ratings from survey qual: thematic analysis from 2 observers field notes
quan → QUAL	Supplemental method is quantitative, implemented before the main qualitative method	Questionnaire sent to participants before a table top exercise to identify priority topics for discussion, followed by field observations and thematic analysis of the discussions during the exercise and debriefing sessions	quan: Analysis of ranking of topics. QUAL: Thematic analysis and triangulation of field notes taken by observers with transcripts of discussions.
qual → QUAN	Supplemental method is qualitative, implemented before the main quantitative method	Key informant interviews to pilot test items for a health risk perception survey being administered to first responders following a prolonged response to wildfires	Interview data analysed deductively according to a coding grid based on topics from the risk perception survey
QUAL → QUAN	Quantitative and qualitative methods are weighted equally, but the qualitative method is implemented first	Community consultation focus groups followed by a community survey to set priorities for a public health action plan	Inductive thematic analysis of focus group data to identify priorities, followed by ranking of priorities

Design	Structural description	Applied example	Data collection and analysis
QUAN → QUAL	Quantitative and qualitative methods are weighted equally, but the quantitative method is implemented first	Rapid needs assessment survey distributed to households impacted by a chemical spill, followed by focus groups with a subset of the population who are at heightened risk	Standardized post-disaster survey analysed to identify neighbourhoods disproportionately impacted using GIS mapping; subsequent outreach and exploratory data collection and analysis to understand short-term health impacts

In the case of intervention research, Minary et al *(9)* provide a framework to guide evaluation design for complex interventions, as well as considerations for evaluating effect, implementation and how mechanisms and context interact to determine intervention outcomes. Further information on the practical considerations for designing mixed methods research studies is available elsewhere *(7, 10–11)*. An important decision to consider in the design stage is how best to ensure comprehensive triangulation, which goes beyond integration of different types of data *(1)*.

4.13.4 Considerations for design and implementation

When combining qualitative and quantitative methods, it is important to consider how the different paradigms will be reconciled *(2)*, and to be certain of the rationale for using mixed methods research to answer the research questions. The mixed methods research design is often methods-centric, with the focus on combining specific methods (for example, quantitative surveys with qualitative interviews) at different timepoints in the project *(1)*. Typical design decisions include determining whether the data will be collected and analysed at the same time, or separately and integrated later.

The decision to do mixed methods research should come after the research questions are identified. It is often described as the obvious choice, under the assumption that two methods are better than a single method *(1)*. This assumption has infiltrated funding programs where mixed methods research projects are deemed more comprehensive. However, as Flick *(1)* emphasizes, most such studies use limited qualitative methods (such as interviews) and don't explore the range of qualitative methods. When applied to a disaster health research context, this can limit creativity in addressing complex issues.

It is important to consider how theory and epistemological differences will be managed in mixed methods research *(1)*. Morgan *(12)* defined research paradigms as "systems of beliefs and practices that influence how researchers select both the questions they study and methods that they use to study them". Paradigms are guides for researchers to determine how to approach a research topic, including the research questions, design, methods and analyses. A pragmatic paradigm is most widely used

in mixed methods research *(13)*. When doing multiple methods (separate studies not mixed), this task is simpler because the paradigms are acknowledged for each method and presented separately.

The function of the mixed method study will determine how the qualitative and quantitative methods are combined at the interface point *(3, 6)*. When determining function, it is important to consider whether the methods need to be combined to answer the same research question, or whether a series of research questions related to the topic require mixed methods. The need for triangulation is a common reason for choosing mixed method design. Flick *(1)* provides a good overview of how triangulation has evolved in recent decades, beyond confirming, disconfirming and expanding findings.

The findings of mixed methods research can be integrated in different ways and at different times. O'Cathain and colleagues *(14)* provide practical suggestions for how to follow a thread and move back and forth between the quantitative and qualitative datasets to confirm or expand the analyses. Use of a mixed method matrix is another option for identifying patterns in the different datasets. Publication of triangulation protocols is an important contribution for the literature, because these enable readers to understand at what points the datasets were integrated and what steps were followed.

Johnson and Schoonenboom *(11)* present a series of tables suggesting different ways qualitative methods can be combined with quantitative methods to enhance randomized trials (Chapters 4.1 and 4.3). In the early stages of designing a randomized trial, qualitative approaches can be used to establish the fit of a conceptual framework or theory (Chapter 4.10) for the intervention. They can also be used to gather information about the context in which the intervention will be implemented and engage stakeholders. For complex interventions, mixed methods research can be used to evaluate outcomes and implementation *(9, 14)*. Qualitative methods are often used in the process of constructing or piloting surveys or other data collection instruments. They are also frequently used to add depth to quantitative designs, such as the example in Case Study 4.13.1 where interviews supplemented survey responses following an earthquake to provide more in-depth understanding of survivor perceptions.

Case Study 4.13.1
Perceptions of earthquake survivors in Amatrice, Italy *(15)*

A series of devastating earthquakes occurred in Central Italy in 2015-2016. In the town of Amatrice, 238 people died out of a population of 2500 people. Massazza et al. *(15)* conducted a mixed methods research study with earthquake survivors in the town, publishing their results in 2019. They explored how survivors perceived the damage from the earthquake and how those perceptions aligned with the concept of natural versus human-made disasters.

Massazza and colleagues *(15)* used a mixed methods, longitudinal design which included quantitative surveys and interviews conducted at two time points, 16 months apart. At the first time point, they received 127 responses to the survey and recruited 52 of the survey respondents to participate in one-to-one interviews. The follow-up survey was completed by 112 of the original respondents. The mixed method design allowed the

researchers to triangulate the findings from the quantitative data with narrative data from the qualitative interviews. The qualitative data was also used to corroborate and expand the analyses for indepth understanding of the complexity of perceptions and understandings of natural versus human-made disasters.

As an excellent example of how mixed methods can be presented together, Massazza and colleagues *(15)* present a summary of the quantitative results in text, tables and graphs, followed by a detailed explanation of the emergent themes from the qualitative data. The discussion includes points of convergence, divergence and how the indepth thematic analysis expanded understanding of the quantitative results.

As an intervention is rolled out, qualitative methods can be used to assess fidelity of the implementation, to determine the extent to which the protocol is being completed as intended *(11)*. Context is important for understanding the mechanisms of why an intervention works and in what circumstances *(16)*. Qualitative approaches provide distinct options for generating process-related data, which can be used in the interpretation of the success of an intervention.

4.13.5 Systems thinking and complexity

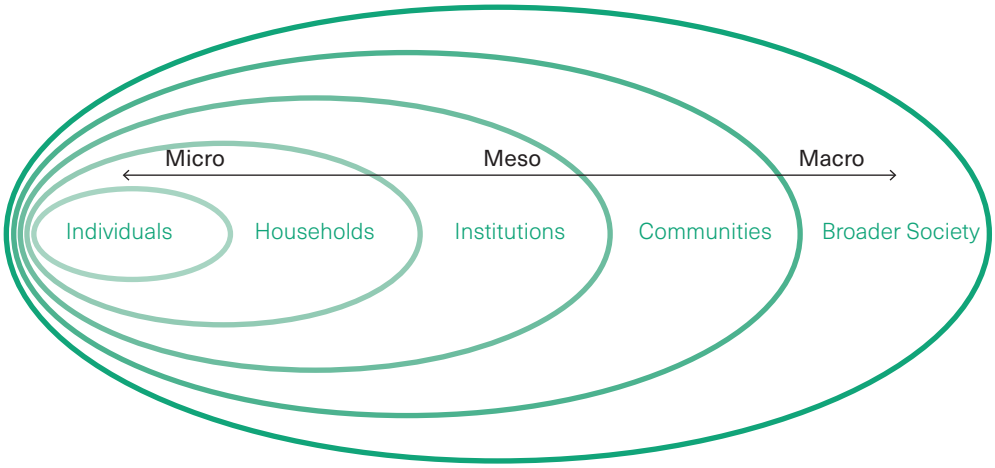
Most disasters are complex and involve collaboration across different sectors, organizations and jurisdictional boundaries. The type of disaster will determine which organizations and jurisdictions must be involved in planning for Health EDRM, including prevention, preparedness, response and recovery, and, therefore, in disaster research. For example, in research related to influenza pandemics, it is essential to consider the roles and impacts on the health and social services sectors, and also on essential services sectors (for example, hydro, transportation) which are likely to experience operational disruptions when absenteeism is high *(17)*. With this in mind, it is useful to look at disaster health research questions through a systems lens, and to acknowledge the complexity in the design of research projects, particularly the interventions (Chapter 3.3).

Systems are made up of different interdependent components and actors or stakeholders. They can be complex, depending on how tightly-coupled the interdependencies are *(18)*. Systems thinking has gained recognition in various fields, because it can be used to understand context, mechanisms and outcomes. It is a way of examining how things are connected within a whole and how the parts within the whole interact in complex ways *(19)*.

Berry and colleagues *(16)* defined systems thinking as “a set of ‘synergistic analytic skills’ used to help describe a complex set of interacting factors that produce outcomes, to predict their behaviour and to formulate interventions to achieve desired (and avoid pernicious) results”. It enables disaster researchers to examine an issue in terms of a dynamic, interconnected collection of components; recognizing how macro, meso and micro level factors influence its operation *(20-22)*. Micro level factors are associated with individuals or households, whereas meso and macro levels refer to factors at the organizational or community and societal levels, respectively (Figure 4.13.1).

4.13

Figure 4.13.1 Multiple Levels for a Systems Approach



Complex research questions require methods that will unpack various influences that interact across multiple levels of society. For example, to achieve a comprehensive understanding of how intervention strategies promote influenza vaccine uptake, it is necessary to have knowledge about people, organizations, communities, health care policy and media; and how different variables intersect across micro, meso and macro levels. The complexity of vaccine uptake across a population includes how messaging influences preventive health behaviours, how social factors influence access and awareness, how mass vaccination is coordinated at the organizational level to increase accessibility, political climate, availability of subsidized health care, and social media threads circulating at the macro societal level *(17)*. The complexity of interdependencies across different levels of the system is the essence of why this problem requires systems thinking.

Complexity has been discussed in the literature for many fields. Cilliers *(18)* outlines different tenets of complexity that are characteristic of complex adaptive systems, including dynamic context, interconnectivity, emergence, self-organization, adaptability, feedback loops and non-linearity. Because complex adaptive systems are open and interact with their environments, the environmental context is inherently dynamic. Systems are composed of different parts and actors which are interconnected, meaning that actions within individual components of a system lead to changes which emerge in other components and the whole system *(19)*. The changes at different levels of the system are non-linear and are, therefore, unpredictable and it is difficult to trace the original causes *(23)*. Nonlinearity is one of the reasons that mixed methods research is important for Health EDRM.

As described by Cilliers *(18)*, complex systems have the capacity for self-organization "... which enables them to develop or change internal structure spontaneously and adaptively in order to cope with, or manipulate, their environment". In the absence of structure or protocols, self-organization naturally follows change in social systems, with people within the system creating structure or strategies to adapt and preserve system functioning. The impacts of changing context within a system are non-linear and feedback loops provide important information about operational functioning *(18)*. In the example of pandemic vaccines, social

media provides a salient example of how self-organizing works. When accurate, timely information is missing, people will look for information and share what they find. This has implications for the rapid spread of misinformation, which can influence beliefs and vaccine uptake.

Table 4.13.2 shows an example of how complexity theory can be applied to understand or map out issues within a pandemic context. Social networks are used as an example, but the same table could be created to examine other issues related to the complexity of pandemics (for example, vaccination or supply chain management). This technique can be used for integration in the analysis phase of mixed methods studies, to understand how complexity manifests within a given research topic and needs to be considered in intervention design (see Case Study 4.13.2).

Table 4.13.2 Application of complexity theory to social networks in pandemic prevention, preparedness, response and recovery

Tenet of complexity	Application
Interconnectivity	Interconnectivity is inherent in relationships, partnerships and strong social networks. Effective pandemic response is dependent on actors from different parts of the health system working together; communication, which is a connective activity is central in pandemic prevention, preparedness, response and recovery.
Dynamic context	Social networks are dynamic. People change positions, retire or meet new people, and the relationships within the network change. People also develop new expertise and experience, which contributes to the dynamic nature of the entire system.
Emergence	Knowledge and ideas emerge within social networks. Behaviours also emerge and influence social norms within networks – both positive and counter-productive. Emergence can spark innovation and contribute to different intervention strategies.
Feedback loops	Social networks provide opportunities for feedback from different parts of the system. This feedback loop creates opportunities for networking, relationship building, and co-learning.
Self-organization	Networks contribute to self-organization in the absence of clear policies or plans which outline roles and responsibilities. When structure and information are needed, but missing, people self-organize to create structure and fill the gaps. Self-organization can support pandemic response and recovery, but in vaccination campaigns, it can also hinder formal processes and awareness campaigns if not managed.
Non-linearity	Social networks are non-linear. Social media is a good example of how social networks do not develop in linear patterns. Communication and influence within social networks are dependent on the relationships and connections of each actor. Non-linearity prohibits cause-effect relationships from being established.
Adaptability	Networks contribute to adaptability. They create opportunities for learning and innovation. Actors within social networks provide different sources of information to enhance situational awareness.

4.13

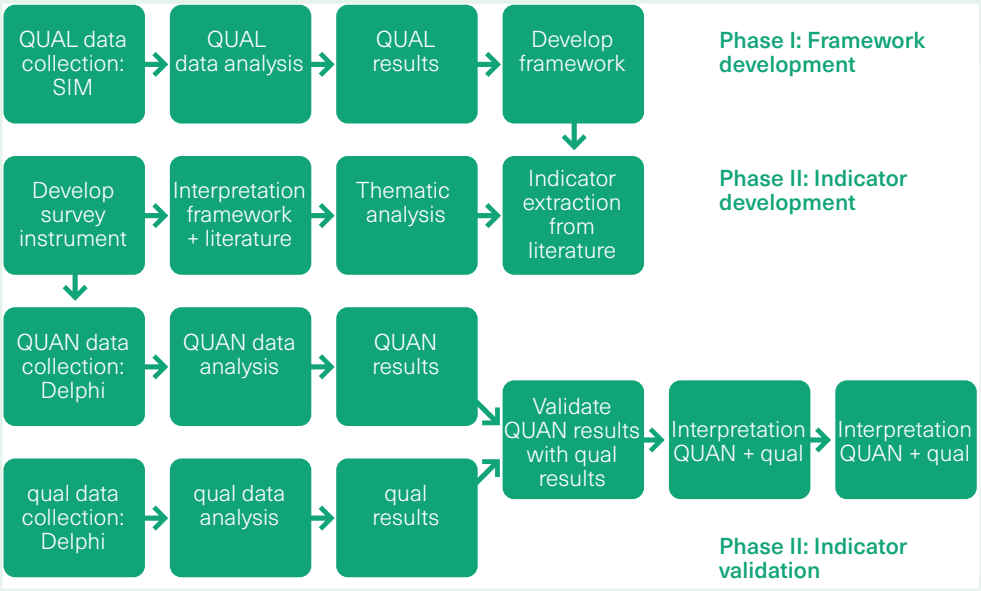
Case Study 4.13.2
Advancing performance measurement for public health emergency preparedness (24–25)

An important knowledge gap in Health EDRM is understanding levels of preparedness or readiness in advance of a disaster. This is a challenging topic, but one of important policy relevance, given the increasing frequency of emergencies and the value of defining and measuring preparedness to guide improvement. This topic was well-suited to a two phase mixed methods study to address the dual objectives: "how do we know if we are prepared?" and "how do we measure it?". Furthermore, using mixed methods enables a consideration for complexity, which is seen as increasingly important for public health systems research *(22)*.

The initial exploration aimed at defining emergency preparedness for the public health system in Canada was achieved using a qualitative study design. Rich qualitative data was analysed using a complex adaptive systems lens to develop a framework defining the essential elements of a resilient public health system *(24)*. The framework reflects the complexity of the role of the public health sector in emergencies and was used to ensure that the approach to measurement considered what the system is aiming to do.

The framework elements informed a mixed methods Delphi survey to develop indicators for public health emergency preparedness (PHEP) *(25)*. The Delphi is a mixed methods research technique well-suited to fields where there is a paucity of evidence, such as PHEP research, and is a structured and rigorous approach to collecting data on expert opinion *(26)*. Its use for developing indicators for clinical contexts such as cancer care also rendered it appropriate for developing PHEP performance indicators *(27)*. In the Delphi process, the combination of deductive thematic analysis of the literature, open ended questions for comments on indicators and suggestions for new indicators, and quantitative rating of indicators enabled the development of a list of preparedness indicators *(25)*. The sequence and combination of mixed methods approaches for the two phases is displayed in Figure 4.13.2.

Figure 4.13.2 Mixed methods used in defining a framework and developing performance indicators for public health emergency preparedness (24–25)



Interdependencies are the norm in modern society and are the reason systems thinking is useful for disaster health research. Lechner and colleagues *(28)* provide a salient example of the interdependencies between the digital society and financial institutions, which has substantial implications in a disaster context. A digital crisis can trigger a financial disaster as the cascading impacts jump between sectors. This complexity underscores the need for collaboration across disciplines and sectors to support situational awareness *(24)*. Expansive, diverse expertise on collaborative teams can also support integrated knowledge translation, which facilitates diffusion and uptake of research findings *(11, 24)*.

The need to understand context is widely acknowledged in the evaluation literature. In fact, the literature base on realist evaluation underscores the importance of understanding context and how it interacts with a mechanism to influence particular outcomes *(29)*. In supporting this point, Johnson and Schoonenboom *(11)* emphasize the utility of qualitative methods to support quantitative methods in process evaluation, with context being a critical consideration. The integration of concepts of complexity, disaster health research and mixed methods approaches are described in the above example of Case Study 4.13.2.

4.13.6 Conclusions

This chapter has introduced mixed methods research design, systems thinking, and shown how complexity can be addressed in Health EDRM research. When conducting mixed method research, it is essential to consider the theoretical and epistemological differences of the methodologies being combined. It is also important to develop the research questions before making the assumption that mixed methods research is the most appropriate methodology for the study.

4.13

When the decision has been taken to use mixed method research as a methodology, careful planning must be done to plan how best to ensure there is comprehensive triangulation, which includes (but is not limited to) integration of data from different methods. Examples provided in this chapter illustrate some of the different strategies that can be used to approach complex questions with mixed methods.

4.13.7 Key messages

- **Mixed methods, which combines quantitative and qualitative methods, has evolved into a third type of methodology which can provide a more comprehensive explanation for the complexity inherent in disaster research.**
- **Systems thinking in disaster health research focuses on the interactions of factors across macro, meso and micro levels of society.**
- **Integration of data, analysis and findings in mixed methods studies is central to the methodology. Many mixed methods studies fall short in the integration process, but this is one of the defining features of mixed methods.**
- **Challenges and practical considerations for designing and implementing mixed method research include theoretical and epistemological differences between methodologies.**

4.13.8 Further reading

Creswell JW. A Concise Introduction to Mixed Methods Research. Thousand Oaks, CA: SAGE; 2015.

Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research (2nd edition). Los Angeles: Sage Publishing; 2011.

Gunderson LH, Holling CS, editors. Panarchy, Washington, US: Island Press; 2002

Johnson RB, Schoonenboom J. Adding Qualitative and Mixed Methods Research to Health Intervention Studies: Interacting With Differences, Qualitative Health Research; 2016: 26(5): 587-602.

O'Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies BMJ; 2010: 341: c4587.

Palinkas L, Aarons G, Horwitz S, Chamberlain P, Hurlburt M, Landsverk J. Mixed Method Designs in Implementation Research. Adminstration and Policy in Mental Health; 2011: 38: 44–53.

4.13.9 References

1. Flick U. Mantras and Myths: The Disenchantment of Mixed-Methods Research and Revisiting Triangulation as a Perspective. Qualitative Inquiry; 2017: 23(1): 46-57.
2. Denzin NK. Moments, Mixed Methods, and Paradigm Dialogs. Qualitative Inquiry; 2010: 16(6): 419–27.
3. Fetter D. Six Equations to Help Conceptualize the Field of Mixed Methods, Journal of Mixed Methods Research; 2018: 12(3): 262–7.
4. Johnson RB, Onwuegbuzie AJ, Turner LA. Toward a Definition of Mixed Methods Research. Journal of Mixed Methods Research; 2007: 1(2): 112–33.
5. Morse J, Niehaus L. Mixed Method Design: Principles and Procedures, Walnut Creek, CA: Left Coast Press; 2009.
6. Creswell JW. A Concise Introduction to Mixed Methods Research. Thousand Oaks, CA: SAGE: 2015.
7. Palinkas L, Aarons G, Horwitz S, Chamberlain P, Hurlburt M, Landsverk J. Mixed Method Designs in Implementation Research. Adminstration and Policy in Mental Health 2011: 38: 44–53.
8. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research (2nd Edition). Los Angeles: Sage Publications Inc.; 2011.
9. Minary L, Trompette J, Kivits J, Cambon L, Tarquinio C, Alla F. Which design to evaluate complex interventions? Toward a methodological framework through a systematic review. BMC Medical Research Methodology; 2019: 19: 92.
10. O'Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. BMJ: 2010: 341: c4587.
11. Johnson RB, Schoonenboom J. Adding Qualitative and Mixed Methods Research to Health Intervention Studies: Interacting With Differences. Qualitative Health Research; 2016: 26(5): 587-602.
12. Morgan D. Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. Journal of Mixed Methods Research; 2007: i(1), 48-76. doi: 10.1177/2345678906292462.
13. Shannon-Baker P. Making paradigms meaningful in mixed methods research. Journal of Mixed Methods Research; 2016: 10(4): 319–34.
14. Drabble S, O'Cathain A, in Hesse-Biber SN, Johnson RB, editors. Moving From Randomized Controlled Trials to Mixed Methods Intervention Evaluations, The Oxford Handbook of Multimethods and Mixed Methods Research Inquiry, Oxford Handbooks Online; 2016.
15. Massazza A, Brewin CR, Joffe H. The Nature of ''Natural Disasters'': Survivors' Explanations of Earthquake Damage. International Journal of Disaster Risk Science; 2019: 10(3): 293-305.

16. Berry HL, Waite TD, Dear KBG, Capon AG, Murray V. The case for systems thinking about climate change and mental health. Nature Climate Change; 2018: 8: 282-90.

17. O'Sullivan T, Phillips K. From SARS to Pandemic Influenza - The Framing of High-Risk Populations. Natural Hazards; 2019: 98(1): 103-17.

18. Cilliers P. Complexity and postmodernism: Understanding complex systems. London: Routledge; 1998.

19. Peters DH. The application of systems thinking in health: why use systems thinking? Health Research Policy and Systems; 2014: 12: 51.

20. Bergström J, Dekker SWA. Bridging the Macro and the Micro by Considering the Meso: Reflections on the Fractal Nature of Resilience. Ecology and Society; 2014: 19(4): 22.

21. Luhmann N. Introduction to Systems Theory, English Edition, Cambridge, UK: Polity Press; 2013.

22. Rutter H, Savona N, Glonti K, Bibby J, Cummins S, Finegood DT, et al. The need for a complex systems model of evidence for public health. Lancet; 2016: 390, 2602-4.

23. Sturmberg JP, Martin CM, editors. Handbook of Systems and Complexity in Health, New York: Springer; 2013.

24. Khan Y, O'Sullivan T, Brown A, Tracey S, Gibson J, Généreux M, et al. Public health emergency preparedness: A framework to promote resilience. BMC Public Health; 2018: 18(1): 1344.

25. Khan Y, Brown A, Gagliardi A, O'Sullivan T, Lacarte S, Henry B, Schwartz B. Are we prepared? The development of performance indicators for public health emergency preparedness using a modified Delphi approach. PLoS ONE; 2019: 14(12): e0226489.

26. Keeney S, Hasson F, McKenna H. Delphi technique in nursing and health research (1st edition). Ames, Iowa: Wiley-Blackwell; 2011.

27. Gagliardi A, Simunovic M, Langer B, Stern H, Brown AD. Development of quality indicators for colorectal cancer surgery, using a 3-step modified Delphi approach. Canadian Journal of Surgery 48: 441-52; 2005.

28. Lechner S, Jacometti J, McBean G, Mitchison N. Resilience in a complex world – Avoiding cross-sector collapse. International Journal of Disaster Risk Reduction; 2016: 19: 84–91.

29. Pawson R, Tilley N. How to design a realist evaluation. In: Pawson R, Tilley N. Realistic Evaluation, London: Sage Publications; 1997: pp. 83–114.

4.14 Natural experiments in a hazard context

Authors

Hyun M. Kim, School of Health Sciences, University of Canterbury - Te Whare Wānanga o Waitaha, Christchurch, New Zealand.

Alex G. Stewart, College of Life and Environmental Science, University of Exeter, Exeter, United Kingdom.

Philip J. Schluter, School of Health Sciences, University of Canterbury - Te Whare Wānanga o Waitaha, Christchurch, New Zealand.

4.14.1 Learning objectives

To understand the potential utility of natural experiments in health emergency and disaster risk management (Health EDRM), including:

1. Process of conducting a natural experiment in a disaster context.
2. Framework for, and outcomes of, natural experiments.
3. Important strengths and limitations of natural experiments.

4.14.2 Introduction

Health researchers are often interested in understanding the effects of certain conditions on health risk or disease outcomes. Typically, constructed and controlled experiments are the cornerstone of studying such causal relationships between exposures and outcomes. An exposure can be any type of condition that is associated with an outcome of interest. For example, the efficacy of influenza vaccine (exposure) can be analysed in relation to the frequency of influenza illness (outcome). In the context of traditionally designed medical experiments, such as randomized trials, exposures are manipulated and are often termed 'treatment'. By contrast, natural experiments are characterized by exposures that are unexpected and cannot be controlled nor manipulated. This exposure may still be referred to as 'treatment' since it essentially performs the same role as the treatment in a randomized trial. Chapter 4.1 explains how to design, conduct and interpret randomized trials in the context of Health EDRM. This chapter discusses natural experiments, an alternative method for studying causal associations. The key components of a causal framework for natural experiments are briefly described in table 4.14.1.

Table 4.14.1 Main components of natural experiments

Component	Description
Exposure/ treatment	'Exposure' broadly refers to any factors (biological, behavioural, lifestyle, environmental) that are being studied in relation to an outcome of interest. 'Treatment' is a technical term that embraces a variety of exposures that differ across experimental groups. In natural experiments, exposures are often disasters caused by natural hazards or anthropogenic (human-instigated) hazards that are typically outside the researchers' control (for example, earthquake, weather shocks and conflicts), and may still be referred to as 'treatment'.
Outcome of interest	'Outcome' is a generic term for the various results that are being investigated in relation to a particular exposure or treatment. In epidemiological and health research, outcomes usually refer to incidences of diseases and health risks. In natural experiments, the hypothesized or observed effects of natural and anthropogenic hazards can be studied as outcome variables. For example, cancer (outcome) can be studied among the population exposed to radiation as a result of breach in a nuclear power plant.
Treatment group	The treatment group describes those people who are assigned to receiving the experimental treatment. In natural experiments, treatment groups are exposed to natural or anthropogenic hazards not by design or deliberate random assignment, but by chance. The treatment unit may be individuals or clusters of people according to affected geographical or jurisdictional borders, regional policies or household units.
Control group	The control or comparison group serves as a reference group in an experiment. In randomized trials, people in the control group might be given the existing best treatment or a placebo, instead of the treatment being tested. In natural experiments, the control or comparison group may be less exposed (or unexposed) to a hazard than the exposed or treatment group since there may be a range of exposure types or concentrations.
Instrumental variables	Instrumental variables are a proxy measure for the independent variable of interest. In the natural experiment context, instrumental variables are often used when the exposure or treatment is difficult to directly measure or quantify (see Case Study 4.14.3). Alternatively, instrumental variables may be related to other variables that could independently influence the outcome (for example, unobserved factors or factors that are not directly included in the model), but may still influence the outcome (see below for an example using *(1)*).
Confounding factors	The exposure-outcome relationship can be influenced by factors that are associated with both the exposure and the outcome. For example, when studying the efficacy of influenza vaccine on reducing the occurrences of flu related illnesses, chronic medical conditions in patients can be a potential confounder (example adapted from *(2)*). Patients with chronic medical conditions or compromised immune system are more likely to be vaccinated (association with the exposure) and more likely to contract influenza viruses (association with the outcome). However, the chronic conditions are unlikely to be directly on the causal path (that is, influenza vaccination can cause chronic illnesses, which in turn, can cause influenza illnesses), and not being directly on the causal pathway is an important condition for a confounding variable *(3)*. In observational studies, any presence and effects of confounding factors need to be taken into account when analysing causal relationships.

A traditional randomized trial design exhibits at least the following three characteristics: 1) random assignment of people into the exposure/treatment and control/comparison groups; 2) researchers' having and exercising control over exposure/treatment assignments; and 3) comparison of outcomes between exposed and control groups. The mechanism of randomly assigning people into exposure/treatment and control groups is of fundamental importance, as it implies that, on average, people across these groups are similar to each other in both known and unknown pre-exposure characteristics *(3)*. This pre-exposure equivalence ensures that any confounding effects from factors that are related to both the exposure and the outcome of interest are balanced across the groups and removes the need for including confounding variables in models and explicitly analysing their effects. Despite this appealing feature, the traditional experimental design is not always a feasible or a practical option. For instance, it would be impossible to control and unethical to simulate a disaster.

Disasters and hazards of various kinds are occurring more frequently and in greater severity. With the world's rapidly expanding and dispersing population together with the impacts of global environmental change, these disasters have greater potential to significantly impact our planet's environmental integrity and its people's health and wellbeing *(4)*. Such occurrences alter the way people live and respond in the affected areas on a scale that would often be logistically or ethically implausible to implement a study using traditional experimental designs *(4)*. Yet, robust evidence-based and informed strategies are needed to serve the affected populous and their environments, together with those experiencing similar events elsewhere or in the future. Natural experiments are, by design, adaptations of conventional approaches or novel methods in providing this evidence-base for Health EDRM. Concordantly, there has been a dramatic increase in the implementation and publication of studies purporting to use natural experiment designs, although their internal validity varies greatly *(5)*.

The randomized trial design is often posited as the minimal standard in considering causation of an effect. However, conventional random assignment, which is the hallmark of randomized trials may be impossible in the disaster context. Nonetheless, it is still possible to have populations that can be demarcated into exposure (treatment) and control groups via a mechanism that is (nearly) as good as random assignment *(6)*. When there is a well-defined exposure that can be contained within a sub-population, and this sub-population is exposed as if in a random assignment, then the natural (or quasi-natural) experimental framework can be used as an alternative to the randomized trial design to infer cause and effect. This chapter follows the convention of Dunning *(5)* and refers to the assignment mechanism that results from an accidental exposure of certain groups of people and which is as good as random as being an 'as if' random assignment. An exposure to a natural or human-instigated (anthropogenic) hazard is an example where natural experiments have been used to understand their impacts on a subject population. This chapter looks at how natural experiments can be used in a hazard/disaster context and the strengths and limitations associated with the framework.

4.14

4.14.3 Natural experimental framework

The natural experimental framework has embedded in its structure many of the elements that characterize randomized trials. These include the exposure, control and outcomes that are outlined in table 4.14.1. When a disaster occurs, such as an earthquake, a well-defined exposure can then be defined which is known to affect all people within a particular perimeter. These people can be considered as comprising the exposed group. Those outside this perimeter remain unaffected and can be categorized as the unexposed, control or comparison group. In some cases, the level of exposure may vary across people and those who are less directly affected may also need to be considered in the analysis.

The focus of using and reporting a natural experiment should be on establishing validity and making a plausible argument for a treatment assignment that is as good as random, or for the difference in exposure of two or more groups. The onus is on the researcher to make a compelling argument for the credibility of 'as if' random assignment by providing both quantitative and qualitative evidence. In a natural experiment with a persuasive 'as if' random assignment argument, the groups are assumed to be similar in all pre-exposure characteristics including any confounding factors, as in the case for true randomization. However, natural experiments are in fact observational studies as the manipulation of 'treatment' cannot be controlled by the researcher as in a true experiment. It is important to distinguish natural experiments from other observational studies, such as quasi-experiments and matching designs (Chapter 4.5), where assignment is neither random nor 'as if' random and hence confounding (both observed and unobserved) becomes an issue to the validity of causal inference *(5)*. In such cases, the effects from confounding factors may need to be explicitly taken into account by adding the confounding variables to the outcome-exposure model and analysing their effects on the association.

In a natural experiment with convincing 'as if' randomization, the data analysis is often simple and interpretable. It usually involves comparing the estimated outcome means between differently exposed groups. For instance, when analysing the level of anxiety after an earthquake, the average effect can be estimated by the average level of anxiety (measured using some form of testing) for all those who experienced the earthquake (by some definition) compared to those who were unexposed to the earthquake. In some natural experiments, exposure/treatment assignment happens at the cluster level (for example, policy implementation in cities, jurisdictional borders or natural boundaries) related to the exposure under consideration. The simplest approach to analyse the average causal effect is to use the average cluster means (that is analyse at the level of random assignment). For example, when analysing the efficacy of a district-wide policy roll-out which affects everyone within the district but not those outside the district boundaries, the average effect is estimated by comparing the average outcomes across different districts rather than across individuals. Sometimes, this is not possible and more sophisticated approaches are needed (see *(5)*).

Three key elements are considered in a typical process for implementing a natural experiment: study design, statistical analysis and validation.

4.14.4 Study design

Exposure-outcome causal model is defined and causal parameters of interest are determined. The 'as if' random assignment argument is also validated using suitable quantitative and qualitative methods. At this point, research hypotheses around the effects of exposures can be considered and formalized.

4.14.5 Statistical analysis

When assumptions around 'as if' random assignment and other model assumptions related to analysing experiments are met, the Neyman–Rubin potential outcomes model is often applied (*7*). One important model assumption is the 'non-interference' assumption: the independence of the effects of exposures across participants, that is, the effects of exposure on one individual do not influence the effects of exposures on other participants and vice versa. Another key assumption is the 'excludability': the effects of exposure on the outcome depend only on the exposure itself and not on other features of the experiment. In a strong natural experimental design, the average exposure/treatment effect is estimated by the difference between average values of observed outcomes for all participants in the exposed groups compared to those in the control/comparison group.

4.14.6 Validation

Quantitative methods are available to test the assumptions about similarities in pre-exposure characteristics between the participant groups. Hence, before the exposure, numbers of participants in each sex, demographic, and other socioeconomic backgrounds are balanced across the exposed and comparison groups almost as if they were randomly assigned to these groups.

Qualitative knowledge about context and process is equally crucial for establishing internal validity in treatment assignment, the integrity of exposure-outcome causal model and the assessment of model assumptions such as non-interference and excludability. Qualitative knowledge is also essential for reporting and assessing external validity such as in replicability and generalisability of results.

4.14.7 Natural experiment designs and their applications

Disasters due to natural hazards often strike with little or no warning and can impact on any population regardless of their attributes, which render disasters persuasive circumstances for implementing a natural experimental design. Perhaps not surprisingly, the natural experimental framework has increasingly been used in broad natural/anthropogenic hazard contexts. For example, the framework has been extended to analyse the impact of arguably one of the most critical natural and anthropogenic hazards that we face today: climate change. Case Study 4.14.1 illustrates a study where children's wellbeing outcomes (measured by undernourishment, labour force participation, and adequacy of medical attention) were analysed in the aftermath of devastating Hurricane Mitch in Central America (October-November 1998).

4.14

Case Study 4.14.1
Children's vulnerability to weather shocks: A natural experiment *(8)*

Agricultural societies are often more vulnerable to weather shocks such as severe storms and hurricanes. Hurricane Mitch hit the Republic of Nicaragua in the last week of October of 1998, and was one of the most destructive storms ever to strike Central America. It left behind more than 50 inches (1.27 metres) of rain and more than 20% of the population was in need of new housing. But, not all municipalities within Nicaragua were directly affected. Fortuitously, a household-level survey had been initiated before the hurricane, the Living Standards Measurement Study (LSMS), which collected data in 1998, 1999 and 2001. Exploiting the LSMS, wellbeing outcomes of Nicaraguan children residing in areas affected by Hurricane Mitch were compared to their unaffected counterparts using a 'double difference' analysis.

The assumption about 'as if' random assignment was made based on the unpredictability of the location of the impact, and that any region was as likely to be on the path of the hurricane as any other regions nearby. The children from households in the municipalities severely affected by the hurricane were analysed as the exposed group. The children from households located outside these areas were used as the comparison group. Validity checks were performed using both quantitative and qualitative methods. The characteristics between the exposed and less exposed households were analysed to validate the 'as if' random assignment argument. Rural areas were more directly hit by the hurricane and the differences in median income and parental educational attainment were detected between the exposed and less exposed groups. These differences were controlled once the treatment effect was conditioned on location. The households were used as the instrument for assigning children into exposure groups. This implies that, after conditioning on location, the outcome of interest (demand for education and health services) was only influenced by whether the households were directly exposed to Hurricane Mitch or not, and not by other underlying household characteristics or other unobserved factors.

Qualitative checks were also performed to analyse the disruption in the supply of school and health services due to the hurricane, as this was considered a potential confounding factor for the demand for those services. The study found that children living in the regions affected by Hurricane Mitch were 30% less likely to be taken for medical consultation when sick, experienced 8.7% increase in the probability of being undernourished, and had 8.5% increase in labour force participation. Although the randomization unit was at the household level, the analyses were performed at the individual child level. The correlation between children within the same household needs to be taken into account when computing variability estimates. However, the extensive validity checks performed in this study to assess the 'as if' random assignment argument were exemplary.

Novel ways of adapting natural experimental designs are continuously being devised. One illustration is a study looking at the application of natural experiment to evaluate cyber security policies *(9)*. Digital hacking is

a relatively new type of man-made security hazard that could place huge cost and burden on people and systems at a global level. Much investment has been made by many countries and organizations on building capacity to deal with any potential breach in cyber security and yet, testing such systems is challenging. Natural experiments are proposed as an alternative to costly and, in some cases, unethical application of traditional experimental design in evaluating the integrity of such programmes.

4.14.8 Regression-discontinuity design

Regression-discontinuity designs are natural experiments where treatment assignment depends on a certain threshold value of a variable (Chapter 4.5). For example, patients may receive a new type of drug depending on their measure of blood pressure being above a certain cut-off value. Around the levels very close to this cut-off, the patient characteristics may not differ greatly even though they are assigned into two distinct groups: those who receive the new drug and those who do not. It can be graphically characterized by a jump or break in the trend for the probability of receiving a treatment versus control around this value of the variable. The 'as if' random assignment argument is only plausible for cases around the near neighbourhood of this threshold as observations farther apart are likely to differ more systematically. In the above example, patients with blood pressure much higher than the cut-off value are likely to have very different lifestyle characteristics than those with values much lower than the cut-off used. So any observed differences between the outcomes being studied may be due to these lifestyle differences rather than the new drug. Case Study 4.14.2 is an application of a regression-discontinuity design for studying the changes in people's lifestyle choices and provision of healthcare services as a result of the 2011 Great East Japan Earthquake (Chapter 1.3) *(10, 11)*.

Case Study 4.14.2
Residential relocation and obesity after a disaster: A natural experiment from the 2011 Great East Japan Earthquake and tsunami (adapted from (11))

Residents in a neighbourhood typically share common demographic characteristics or lifestyle patterns. However, when the east coast of Japan was hit by a massive earthquake and tsunami in 2011, a large-scale exodus ensued that could not have been foreseen nor planned. Approximately 345 000 people were displaced from their homes, disrupting their normal way of life and possibly their long-term wellbeing. This disaster was used as the 'as if' random assignment mechanism where the outcomes of survivors before and after the earthquake were compared. Coincidentally, a nation-wide cohort study of ageing population, the Japan Gerontological Evaluation Study, had been established seven months before the earthquake, allowing the researchers to investigate the impact of disaster in comparison to the extensive pre-disaster information available on the cohort.

For example, the cohort was followed up about 2.5 years after the disaster to study the impact of relocation on 3594 participating survivors' weight gain measured using the Body Mass Index (BMI). The change in the

4.14

distance to the nearest food outlets, bars, supermarkets was used as an explanatory variable in a fixed effects multinomial logistic regression model. Various covariates such as age, socioeconomic status and mental health and behaviours were also added in the analysis. Adjustments for confounding variables are necessary if systematic differences between the survivors pre- and post-disaster are to be assumed. The study found that moving 1km closer to supermarkets, bars or fast food outlets increased the odds of BMI change from normal to the obese range by 1.46 (95% confidence interval (CI): 1.15 to 1.86), 1.43 (95% CI: 1.11 to 1.86), and 1.44 (95% CI: 1.12 to 1.86), respectively. Such findings suggest that the impact of a disaster on survivors' lifestyle choices is of pervasive nature, and could have long-term health and wellbeing implications.

The discontinuity in Case Study 4.14.2 is at the point of the disaster, when changes can occur and differentiate people's post-event characteristics from those of before. Around the time of event, the probability of being exposed to certain risks can be higher for people within the vicinity of the disaster compared to those further away. Some changes, such as the residence displacement, will likely be irrevocable, and the consequences of those can be analysed as illustrated in Case Study 4.14.2.

In Chapter 2.4, Case Study 2.4.1 described an example in which the impact of moving toward a more integrated health system on emergency room attendance and acute admission rates was analysed for the population affected by the 2011 Christchurch, New Zealand earthquake *(12)*. Figure 2.4.1 in Chapter 2.4 shows visible breaks in the trends for emergency room attendance and admission rates before and after the earthquake.

4.14.9 Instrumental variables design

Instrumental variables are proxy measures for the actual variable of interest that may be difficult to measure or could lead to biased estimation. In instrumental variables design, people are assigned at random (or 'as if' random) to this proxy for the variable of interest. For example, Angrist *(1)* sought to measure the long-term labour market consequences of those veterans who served in the military during the Viet Nam era compared to their nonveteran counterparts. Here, military draft eligibility was used instead of actual military service in a natural experiment design which produced robust unbiased estimates. Using the military draft eligibility as an instrumental variable ensured that all those who were subject to randomization were analysed rather than those who complied with the assigned treatment (that is, completed military service). Furthermore, those who volunteered to serve or those who did not pass the health tests after being randomly selected for draft eligibility would have had different characteristics to those who did not serve in the military. So, it was important to use the draft eligibility, which was closely associated with actually serving in the military, but also would not otherwise influence an individual's lifetime earnings.

Case Study 4.14.3 *(13–14)* explores prevalence of iodine deficiency disorders that are endemic to areas characterized by subducting plates in the Himalayan region. Iodine deficiency is a disaster that is not sudden, but is easily preventable. It is a devastating issue in many communities due to

its link with high levels of infertility and miscarriages, cretinism and lowered cognition, as well as the usually harmless but visible goitre. The research described in the case study focuses on the Baltistan region, northeast Pakistan, before any long-term iodization programmes, where clear regional differences in prevalence of goitre were found north and south of the Main Karakoram Thrust, where Asia and India meet geologically, giving a natural experiment. This experiment is characterized by exposure ('north-south goitre prevalence') that is unexpected and cannot be controlled nor manipulated. It can be argued that selection of individuals was "controlled" precisely on the basis of the north-south goitre prevalence, leading to the outcome of the incidence of iodine deficiency. The geological separation was used as an instrumental variable in categorising two communities by their environmental differences, which were otherwise difficult to quantify (for example, geological and geochemical differences).

Case Study 4.14.3
Differences in endemic goitre prevalence in the Karakoram mountains, north Pakistan: a natural experiment suggesting an unrecognized cause (adapted from *(13)*)

Environmental iodine deficiency, of which endemic goitre is one manifestation, causes several disorders, none of which were seen as related to goitre by the local community in Baltistan, northern Pakistan in the 1980s. However, the community noted that more people living in the north of the region suffered from goitre than did those living in the south. Furthermore, goitre was accepted as normal, not triggering clinic visits. Careful qualitative investigation of the communities on both sides of the rivers did not show any north-south differences in ethnicity, diet, farming practices or other obvious causes of the difference. Residence village was used as the 'as if' random assignment mechanism.

New patients attending a clinic fell naturally into four groups: northerners with goitre, northerners without goitre, southerners with goitre, southerners without goitre. There was a significant difference in prevalence due to age-sex and, independently, to residence north or south of the Main Karakoram Thrust. This plate tectonic boundary divides the region into two clearly distinct geologies, and increased the prevalence in the north by 15-18%. A literature search revealed two other studies by another team more than 100 miles to the west, straddling the Thrust. Findings were similar: villagers on the northern plate had consistently more goitre. The geology was the explanatory variable, and indicates that the distribution of iodine deficiency disorders in this and other mountain ranges are likely related to plate tectonics in addition to iodine deficiency.

The study shows most of the strengths in Table 4.14.2. This robust observation allows prediction of the distribution of iodine deficiency disorders which can be tested by further observational studies, with a stronger hypothesis than many of the standard explanations for the occurrence of iodine deficiency disorders, such as leaching of soil iodine by rain or removal by glaciation.

4.14

Table 4.14.2 Strengths and limitations associated with natural experiments for Health EDRM

Strengths	Limitations
No ethical constraints about exposure.	No control over baseline differences in the exposed and less or unexposed groups.
Can infer cause-effect when 'as if' randomization can be validated.	There is no random assignment in the traditional sense, which may restrict causative assertions if 'as if' randomization cannot be established.
Obviates confounding typical in an observational study.	May be difficult to contain the treatment and control groups within certain temporal and spatial perimeters.
Quantitative analysis can be simple and transparent.	May be difficult to isolate an effect of an exposure.
Statistical results often easy to interpret.	Exposure/treatment may not be of research relevance or interest.
Can be less costly than the randomized trials or quasi-experiments if data already available.	Internal and external validity may be difficult to analyse.
Can be tailored to the hazard or disaster.	
Possible to analyse the effect of a slow onset hazard.	
Possible to plan a prospective study.	

Countries and jurisdictional borders can form natural clusters. In some cases, they can be used as instrumental variables for studying various social, political, environmental and health related differences across groups. Historical borders and policy differences across countries are usually outside the control of the researchers (that is, exogenous to the model). The administrative and structural differences also mean that the countries are 'as if' assigned to different types of treatments. When applying instrumental variables, it is important to check that the outcome of interest is influenced mainly through the association between the instrument used and the explanatory variable being studied, and not through other factors unexplained by the model. For example, in Case Study 4.14.3, demographic characteristics between two communities were analysed to ensure that the instrument used, which was related to the geology of the region, was what explained the observed difference in prevalence of goitre, and not the demographics.

In another example, the extent of food insecurity across 21 countries was analysed in relation to the economic hardship, measured using the unemployment rate and decrease in wages, experienced during the 2004-2012 European recessions *(15)*. The country-level analyses revealed that both measures of economic hardship were associated with an increased sense of food insecurity. Also taking advantage of jurisdictional and policy differences, the association was further analysed using the level of social protection in each country. The risks of food insecurity associated with economic hardship were mitigated in countries that spent more on provision of social protection.

Similar designs have also been applied in studies looking at the effects of environmental policies implemented at the prefecture- or city-level of governance. Environmental regulations on sulphur dioxide emission and

acid rain were put in place across different provinces in China in order to reduce air pollution *(16)*. The resulting changes in the volume of industrial activities in the regulated cities were compared to those of unregulated cities. Similarly, gains in energy efficiency following the roll out of 'Smart City' policies in China (aimed at integrating government services and achieving low carbon emitting and ecologically sound urbanization plans) were analysed and compared across the 'Smart Cities' and control cities *(17)*.

Other examples where country-level policy differences have been used to analyse human-instigated hazards can be found in studies of health risk control policies. The impact of tobacco control policy on cardiovascular morbidity and mortality in the Russian Federation was analysed in relation to other countries without such control *(18)*. Similarly, the implementation of trans fatty acid control policy in the Republic of Austria was used as the setting for a natural experiment where the cardiovascular and coronary heart disease mortality was compared between the population under the regulation and the international control population from countries without the regulation *(19)*.

4.14.10 Conclusions

Natural experiments provide researchers with opportunities to investigate some topics of relevance to Health EDRM that are not amenable to designs, such as randomized trials. They have important strengths and limitations for hazard and disaster epidemiology, which are listed in Table 4.14.2.

4.14.11 Key messages

- **In natural or human-instigated hazard contexts, implementing the traditional experimental design to study cause-effect relationship can be unfeasible or unethical.**
- **When people are assigned into exposure/treatment and control groups by chance, but in a way that resembles true randomization, natural experiments can be used to infer relationships between exposures and outcomes, just as in a traditional experiment.**
- **The credibility and validity of natural experiments depend on the persuasiveness of the 'as if' random assignment argument. The randomization ensures that the exposed and control groups are similar in their pre-exposure characteristics and hence mitigates the effects of observed and unobserved confounders.**
- **Quantitative analyses of pre-exposure characteristics and qualitative evidence around context and processes are useful for establishing the credibility of natural experiment design.**
- **If the assumption of random, or 'as if' random, assignment is persuasive, then the estimation of causal (or treatment) effect is as simple as taking the difference between the means of outcome from the treatment and control groups.**

4.14

4.14.12 Further reading

Craig P, Cooper C, Gunnell D, Hawk S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. Journal of Epidemiology and Community Health; 2012: 66: 1182-6.

Scottish Government Social Research Group (2009). Social Science Methods Series Guide 3: Natural experiments; 2009 (https://www2.gov.scot/Resource/Doc/175356/0091395.pdf, accessed 19 January 2020).

4.14.13 References

1. Angrist JD. Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records. The American Economic Review; 1990: 80(3): 313-6.
2. Centers for Disease Control and Prevention, National Center for Immunization and Respiratory Diseases. How Flu Vaccine Effectiveness and Efficacy are Measured; 2016 (https://www.cdc.gov/flu/vaccines-work/effectivenessqa.htm, accessed 19 January 2020).
3. Webb P, Bain C, Page A, editors. Essential Epidemiology. An Introduction for Students and Health Professionals (4th edition). Cambridge, United Kingdom: Cambridge University Press; 2020.
4. Loevinsohn ME. Never let a disaster go to waste: Using natural experiments to understand vulnerability and resilience; 2016 (https://www.ids.ac.uk/events/never-let-a-disaster-go-to-waste-using-natural-experiments-to-understand-vulnerability-and-resilience, accessed 19 January 2020).
5. Dunning T. Natural Experiments in the Social Sciences. A Design-Based Approach. Cambridge, United Kingdom: Cambridge University Press; 2012.
6. Dinardo J. Natural Experiments and Quasi-Natural Experiments. In: Durlauf SN, Blume LE, editors. Microeconometrics. London, United Kingdom: Palgrave Macmillan; 2010.
7. Sekhon J. The Neyman–Rubin Model of Causal Inference and Estimation Via Matching Methods. Oxford, United Kingdom: Oxford University Press; 2008.
8. Baez JE, Santos IV. Children's vulnerability to weather shocks: A natural disaster as a natural experiment; 2007 (https://www.alnap.org/help-library/childrens-vulnerability-to-weather-shocks-a-natural-disaster-as-a-natural-experiment, accessed 19 January 2020).
9. Dean B. Natural and quasi-natural experiments to evaluate cybersecurity policies. Journal of International Affairs; 2016: 70(1): 140-60
10. Hikichi H, Tsuboya T, Aida J, Matsuyama Y, Kondo K, Subramanian SV, et al. Social capital and cognitive decline in the aftermath of a natural disaster: a natural experiment from the 2011 Great East Japan Earthquake and Tsunami. Lancet Planetary Health; 2017: 1(3): e105-3.

11. Hikichi H, Kondo AK, Tsuboya T, Kawachi I. Residential relocation and obesity after a natural disaster: A natural experiment from the 2011 Japan Earthquake and Tsunami. Scientific Reports; 2019: 9: 374.

12. Schluter PJ, Hamilton GJ, Deely JM, Ardagh MW. Impact of integrated health system changes, accelerated due to an earthquake, on emergency department attendances and acute admissions: a Bayesian change-point analysis. BMJ Open; 2016: 6: e010709.

13. Stewart AG. For debate: Drifting continents and endemic goitre in northern Pakistan. BMJ; 1990: 300: 1507-12.

14. Stewart AG, Pharoah POD. Clinical and epidemiological correlates of iodine deficiency disorders. Geological Society Special Publications; 1996: 113: 223-30.

15. Loopstra R, Reeves A, McKee M, Stuckler D. Food insecurity and social protection in Europe: Quasi-natural experiment of Europe's great recessions 2004-2012. Preventive Medicine; 2016: 89: 44-50.

16. Chen B, Cheng YS. The impacts of environmental regulation on industrial activities: Evidence from a quasi-natural experiment in Chinese prefectures. Sustainability; 2017: 9(4): 571.

17. Yu Y, Zhang N. Does smart city policy improve energy efficiency? Evidence from a quasi-natural experiment in China. Journal of Cleaner Production; 2019: 229: 501-12.

18. Gambaryan M, Reeves A, Deev A, Popovich M, Drapkina O, Snell A, et al. Effects of tobacco control policy on cardiovascular morbidity and mortality in Russia. European Journal of Public Health; 2018: 28(2): 14-6.

19. Grabovac I, Hochfellner L, Rieger M, Jewell J, Snell A, Weber A, et al. Impact of Austria's 2009 trans fatty acids regulation on all-cause, cardiovascular and coronary heart disease mortality. European Journal of Public Health; 2018: 28(2): 4-9.

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

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 United Kingdom. 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 Heath 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

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

- **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.**
- **The M&E framework chosen by a researcher will determine the study focus during data collection, analysis and interpretation of its findings.**
- **Randomized trials might not be practical for some Health EDRM M&E studies and quasi-experimental designs are increasingly used.**
- **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.**
- **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

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. Brightion, 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. 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.ijdrr.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

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.

4.15

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