Countering Violent Extremism: Data Collection and Analysis Manual

Associate Professor Adrian Cherney, Emma Belton, Ellen Leslie, Dr Jennifer Bell, Lorraine Cherney & Lorraine Mazerolle

(University of Queensland)

Australian and New Zealand Counter-Terrorism Committee, National Countering Violent Extremism Evaluation Framework and Guide

This work was funded by the Commonwealth Attorney-General's Department

Table of Contents

Table of Tables

1. Introduction

This manual provides an overview of various methods that can be used to analyse different types of data collected to evaluate initiatives aimed at countering violent extremism (CVE). Readers should be reminded that the type of data and the methods used to collect that data will determine the forms of analysis that can be carried out. In addition, CVE evaluations face a number of practical, logistical and ethical challenges, which can place limitations on data collection methods (Horgan & Braddock 2010; Koehler, 2017). For each form of data collection and analysis, a hypothetical case study is provided to illustrate how it could potentially be applied to the CVE context. These case studies are generic to ensure they have wide applicability and provide ideas for practitioners and policy-makers who are thinking about conducting their own evaluations. Caution must be taken in assuming a causal link between an intervention and results that show change in a particular indicator or outcome.

The manual is presented in the following order. First, an overview is provided of the different types of evaluation designs and their strengths and weaknesses. The types of data available for conducting evaluations will then be outlined, along with the key factors that can influence data collection. The following four sections explore the use of *quantitative data*, *qualitative data*, *secondary data* and *social media data*. Each section details the particular method, the task of data management and the advantages and implications of each methodology. Hypothetical case studies are provided to highlight the application of these methods to the evaluation of CVE initiatives.

This manual is targeted at policy-makers and practitioners. We have assumed readers will have limited background knowledge in certain forms of data collection and analysis. Hence, we have aimed to use non-technical language as much as possible. Some readers might have certain views of how evaluations of CVE initiatives should be undertaken, as there are diverse opinions relating to the collection and analysis of research data and which methods are valid and rigorous. We have not set out to advocate one particular method or approach over another. Choices relating to data collection and analysis will be determined by a range of factors, such as the resources available and the level of expertise to assist in program evaluation.

2. Evaluation Designs

2.1 Types of evaluation designs

In order to select an appropriate evaluation design, we must first provide some information on the types or forms of evaluations that exist. Each evaluation method will be defined, along with instructions on the feasibility of implementation and associated strengths and weaknesses of the method. The five most common types of evaluation designs are listed, with some regarded as more rigorous than others.

Retrospective post-intervention evaluation

This type of evaluation involves collecting data from program participants at a single point in time, such as asking participants whether their skills, knowledge and behaviour has changed or altered since starting a program. Results can provide insight into change, but attributing causation to program outputs must be done with caution. For CVE evaluations this can be the most straightforward approach to adopt, but there are limitations with this method.

Pre-/post-intervention evaluation

This type of evaluation entails collecting data from a single intervention group at two points in time: before the program starts and after the program finishes. While change may be observed among intervention participants, results cannot be attributed to the intervention alone without a comparison group.

Interrupted time-series analysis

Interrupted time-series analysis uses data to assess changes at multiple time points both before and after a program has been implemented. Results can help determine trends before and after the intervention. However, again results cannot be attributed to the intervention alone without a comparison group.

Pre-/post-intervention evaluation with comparison group

This evaluation design is similar to pre-/post intervention designs except data is also collected from a group similar in background to the program participant group. Differences can then be compared between the groups to understand the impact of an intervention. If there are differences in key outcomes between the groups, this method can provide some confidence in concluding whether the intervention has had an impact or not. However, one must keep in mind that there are still weaknesses with this method, one being whether the chosen comparison group is sufficiently similar to individuals who received the intervention. This can particularly be the case with CVE interventions, in that identifying and accessing a comparison group may be difficult given the intervention can be dealing with a very small number of individuals.

Pre-/post-intervention evaluation with control group

This type of evaluation involves randomly assigning your pool of program participants to either participate in the program (intervention group) or to not participate in the program (control group). This is often referred to as an experimental method or randomised control trial (RCT). Then, data is collected at two points in time: before the program begins and after the program ends. Results provide robust indications of causality. However, in the CVE field there have been very few evaluations based on this method (Feddes & Gallucci, 2015). One reason for this is that many interventions may be dealing with a small number of clients and hence the ability to randomly assign an intervention can be limited, because it will lead to insufficient numbers across the intervention and control groups. Also, agencies may be understandably nervous in randomly allocating individuals at risk of radicalising to violent extremism to a control group (the no intervention group), in case they commit an act of terrorism.

2.2 Identifying issues pertinent to your evaluation design

Several factors will influence the design of a program evaluation. The following section outlines several issues to consider:

Number of program participants

The number of program participants can influence whether one can determine any meaningful impact from a program when conducting an evaluation. Questions you may need to consider include: what changes do you expect to occur as a result of the program, and do previous evaluation studies exist that can be used to estimate possible program outcomes (Fink, Romaniuk & Barakat, 2013)? The problem is that, when it comes to CVE evaluation, there may be few other existing studies to refer to. The number of program participants will influence your ability to find a statistically significant effect (i.e., program outcomes that you are confident did not occur by chance or error). This is an important consideration in CVE evaluations because you may be dealing with a small number of clients.

Selecting a control or comparison group

In pre-/post-intervention evaluations with either a comparison or control group, most evidence is derived from comparing the group participating with a group of non-participants. All eligible participants are randomly assigned either to participate or to not participate – this ensures that the two groups are as similar as possible. If this is not feasible, a comparison group can elicit similar results. The key here is to select a group that is as similar as possible to the intervention group. This can be difficult to do in CVE evaluations, because as noted by a number of researchers, there is no clear or consistent terrorist profile (Horgan, 2008), which means no one violent extremist will necessarily be like another. Hence, identifying a meaningful and similar comparison group can be difficult.

Timing of the evaluation and the intended audience

To generate more robust outcomes, it is better to collect data from participants prior to implementing the program and after the program ends. However, timing of the data collection depends on the length of the program. Funding deadlines and reporting requirements also influence the timing of the evaluation.

Data security and human subject protection

Evaluation studies of CVE programs will involve working with human subjects. Protecting program participants from harm during an evaluation is of the upmost importance. In order to protect participants, three basic principles have been established for human subject research:

1. *Respect for persons*: participants are fully informed of their participation and consent is obtained.
2. *Beneficence*: personal harm is minimised and public benefit is maximised.
3. *Justice*: participation opportunities, benefits and burdens are fair.

These principles are required to gain informed consent, assess risks and benefits, and to select subjects (Helmus et al., 2017). These data security and human subject protections can be particularly salient in CVE evaluations. For example, individuals exhibiting signs of violent extremism can be of interest to police (e.g., young people, and data may be highly sensitive), or they may come from vulnerable groups. The total number of clients in a program may be small, risking participant identification. There may be the risk of emotional or psychological harm to clients when collecting data, such as when asking about traumatic or sensitive topics.

Evaluation expertise

The feasibility of program evaluation can be limited by access to relevant expertise. Knowledge and experience in conducting program evaluations can come from academic partners, non-government organisations, and private sector groups. The level of expertise must be considered before evaluation designs are decided.

Resources available for the evaluation

Program evaluations are resource intensive. Resources include supplies and equipment, staff time, fieldwork costs, financial and organisational resources. It is important to understand the resources you have available in order to consider the type of evaluation design that can be conducted (Fink, Romaniuk & Barakat, 2013; Helmus et al., 2017).

3. Types of data

As discussed in the previous section, the available resources and research objectives will dictate the type of evaluation design chosen. The next step is to decide on the type of data that will be collected.

3.1 Primary and secondary data

Data can be categorised into two types: *primary* and *secondary*.

Primary data can be understood as original data, such as data collected from program participants by the person conducting the evaluation. This could, for example, involve interviewing clients in a CVE program. Secondary data is pre-existing data (i.e., data already collected by a particular source). This, for example, might be data collected by other agencies for another purpose, but could be used to assess relevant outcomes (e.g., welfare and employment data on a population that has been targeted by a CVE intervention). This data is often used for a purpose that it was not originally intended for, or by someone other than the researcher who collected the original data.

3.2 Qualitative, quantitative and administrative data

Primary and secondary data can include *quantitative*, *qualitative*, or *administrative* data. Quantitative data is predominately numeric and can be measured through calculations and systematic methods. Qualitative data is generally textual (e.g., responses to open-ended questions in a questionnaire) that describes, approximates or characterises information.

Administrative data is collected by government departments and other organisations for the purpose of registration, transactions or record keeping (e.g., employment, annual reports) (Leavy, 2017; Walliman, 2017).

The type of data you choose will affect the methods of collecting data (e.g., via survey, observation), how the data is stored and cleaned, and the types of analysis that can be conducted.

4. Quantitative data

4.1 Introduction

Quantitative data refers to data that can be expressed in numeric form or quantified. This includes measures of values or counts (e.g., weight, age, number of attendees at a meeting) that are naturally expressed in a meaningful numeric form. It also includes measurements of characteristics that do not have a meaningful numeric form, but can be assigned numerical values (e.g., gender, country of birth, ethnicity). Quantitative data is collected through structured data collection methods and analysed statistically to examine patterns in the data. This often includes large-scale research using samples selected from a population in order to produce findings that may be generalised to that population. This section provides an overview of the types of quantitative data, and methods for collecting, managing, and analysing quantitative data. A hypothetical scenario is also presented to illustrate how different quantitative methods can be applied to the evaluation of CVE programs.

4.2 Types of quantitative data

Quantitative data can be divided into two broad types:

(1) Data that has a natural, meaningful numeric form, such as counts or measurements (e.g., number of children, height). This type of data is referred to as interval/ratio data. Interval/ratio data can be further categorised as either discrete or continuous. Discrete data refers to data that is measured in whole numbers (e.g., number of program participants), while continuous data includes measurements that may take any value, usually within a certain range (e.g., hours of intervention program attendance). For both discrete and continuous data, the differences between values are numeric and meaningful (e.g., an increase in program attendance from 10 hours to 17 hours).

(2) Data that does not have a natural, meaningful numeric value, but can be assigned a numeric value (e.g., gender, attitudes towards minority groups, beliefs in extremism) to allow for analysis. That is, the names of categories (e.g., gender, the categories may be male and female), are assigned a numeric value (male = 0, female = 1). All categories must be mutually exclusive and exhaustive. This type of data can be broken down into two subcategories: nominal and ordinal. Data are categorised as nominal if there is no natural order of the categories (e.g., gender). As such, the choice of numeric values for categories of nominal variables is arbitrary. In contrast, ordinal data refers to data with categories that can be ranked in a meaningful order (e.g., low to high). The numeric values assigned to each category must preserve their natural order and rank. In social science research, a common type of ordinal data is data collected from a question that uses Likert scale response categories. For example, a survey may ask respondents to rate their agreement with the statement 'I trust the police in my community', selecting from five response options: (1) strongly disagree, (2) disagree, (3) neither agree nor disagree, (4) agree, or (5) strongly agree; or 'It is sometimes appropriate to use violence against other people in order to achieve something I consider very important': (1) strongly disagree, (2) disagree, (3) neither agree nor disagree, (4) agree, or (5) strongly agree (Doojse, Loseman & van den Bos, 2013). The values assigned to these responses would preserve their natural order and rank.

4.3 Sampling

Quantitative research uses a variety of sampling techniques to gather data. Sampling is the process of selecting groups or individuals from a particular population of interest to study, with the aim of drawing conclusions about the larger population. For program evaluations this would involve being able to draw conclusions about the impact of an intervention on its target group. There are several key factors that need to be considered when determining a sampling method. Firstly, the demographic of interest needs to be defined to draw a representative sample. This includes the particular group (e.g., at risk violent extremists), in the geographical area of interest (e.g., schools/neighbourhoods), during the time of interest (e.g., early adolescents born 2000's onwards). The available resources (e.g., time available, financial resources, number of available staff) will also determine what sampling method is used (Saunders, Lewis & Thornhill, 2012). For a CVE evaluation, selecting a representative sample may be challenging because the potential sample size can be very small.

A representative sample is one that resembles the population from which it was drawn. Obtaining a representative sample is important in order to make accurate generalisations about the impact of a program on its target group. Typically, quantitative research will focus on taking a random sample from a given population in order to make a strong conclusion, for example, about an intervention as it relates to a particular demographic or group of people. There are different types of sampling procedures each with their own advantages and disadvantages.

Simple random sampling

Simple random sampling is the most basic and simplest form of probability sampling. Each member of the population (target group) has an equal chance of being selected as a subject. As it relates to evaluation, simple random sampling is an easy way to assemble a representative sample of an intervention's target group. Ideally, a sample size of more than a few hundred is required in order to apply this strategy and be able to make conclusions and generalisations about a certain population (i.e., that an intervention has an impact). However, it may be difficult to administer to a larger population due to the need to have every possible participant listed for selection, or can be a flawed approach when used with very small populations.

Stratified sampling

In stratified sampling, a researcher will divide the study population into relevant subgroups and then draw a random sample for each subgroup. The subgroups are called stratum and are formed based on members' shared attributes or characteristics. A stratified sample is obtained by independently selecting a separate random sample from each subgroup. The most common subgroups used in stratified random sampling are gender, age, socioeconomic status, religion, nationality and educational attainment. This sampling technique is beneficial if a subgroup of interest makes up a relatively small proportion of the overall sample. Hence, it can help ensure subpopulations are represented equally or proportionally within the sample.

Purposive sampling

A purposive sample is a non-probability sample that is selected based on characteristics of a population and objectives of a study. Purposive sampling is acceptable in specific situations, such as when there are a limited number of participants in populations that are difficult to reach (e.g., youth at risk of violent extremism or convicted terrorists). Subjects are selected using judgements about their compliance and representativeness of the target group. A sample is collected at the discretion of someone who is familiar with the relevant characteristics or traits of the population (Tongco, 2007). Purposive sampling is one of the most cost-effective and time-efficient sampling methods. However, purposive sampling is susceptible to low levels of reliability and high levels of bias. That is, the people selected may not characterise the target group of interest. For example, data is collected from program participants that are all of a very similar age. This can limit the types of conclusions that can be drawn from a study's findings, but may still give insights into the impact of an intervention if its target group is quite specific.

4.4 Data collection

There are a variety of different data collection methods used in quantitative research. The type of data collection method you choose for your evaluation will depend on the type of evaluation and its objectives, the target group of your research, and available resources. This section provides an overview of common types of data collection methods for quantitative research, including the strengths and weaknesses of each approach. It should be noted that ideally, regardless of the data collection methodology you choose, pilot testing should be conducted to identify and address potential issues that may arise during data collection.

Surveys

Surveys are a common method of data collection for quantitative research. They involve the development of a standardised questionnaire that is self-completed by survey respondents. The questionnaire may include different types of questions, including closed questions (i.e., a respondent chooses their response from a list of supplied options) or open-ended questions (i.e., a respondent can write their own answer/response).

Surveys are relatively cost-effective and are a good method for collecting data from a large group of people. They can be mailed to respondents, administered face-to-face and on-the-spot at an event or program, or conducted online. As each respondent is asked the same set of questions, the data collected can be compared across the sample as a whole. However, there are a number of important considerations when using surveys. Firstly, questionnaires must be carefully designed for the targeted group to ensure that the questions being asked will be easily understood. For example, surveys targeted at linguistically diverse populations (e.g., groups for whom English is not their first or primary language), should use simple English and avoid colloquial or slang terms. If questions are ambiguous or easily misunderstood, this will impact on the reliability and validity of the data collected. Secondly, questionnaires should be kept to an appropriate length so as to not burden participants.

Surveys can also involve standardised interviews where participants are asked the same questions in the same way that follow a survey-style questions-and-answers format (i.e., quantitative interviews). As such, the questions posed in quantitative interviews are often closed-ended questions – questions that hanged for consistency with beloware easy to answer and have a predetermined set of responses (e.g., 'yes' or 'no'), rather than open-ended questions that allow for a more elaborate response. In quantitative interviews the interviewer follows a schedule, which acts as a guide for the researcher as they pose question and answer options to the participant. There are several benefits of quantitative interviews, such as higher response rates and lower respondent confusion (Saunders, Lewis & Thornhill, 2012). The main methods of administering interviews for quantitative research include: (1) interviewer administered questionnaires, (2) telephone questionnaires, and (3) online/web-based questionnaires.

Interviewer administered questionnaires

Interviewer administered questionnaires are generally conducted face-to-face and are based on a structured questionnaire. The number of questions, sequence and order cannot be modified during the interview. These interviews are traditionally carried out with paper and pencils or using specialised survey software (e.g., Computer Assisted Personal Interviews or CAPI). Interviewer administered questionnaires can be recorded; however, if the majority of questions are close-ended this may not be necessary. If the questionnaire involves open-ended questions, recording is often advised.

Telephone administered questionnaires

Telephone administered questionnaires involve a researcher administering a structured survey to a participant over the phone. One concern with this method of interviewing is that decreasing numbers of people have their numbers listed and more people now only own and use a mobile phone. Random digit dialling programs and mobile phone sampling frames are designed to help to address this problem. Telephone administered questionnaires are also susceptible to reliability issues. They can generate more fence-sitters (someone who supports both sides and cannot make a decision), compared to face-to-face interviewer administered questionnaires. Reponses to sensitive questions or questions perceived as invasive are also generally less accurate when collected over the phone rather than in person (Holbrook, Green & Krosnick, 2003).

Online/web-based questionnaires

Advances in technology allow researchers to utilise web based instruments such as Survey Monkey, Skype, text-based chat rooms, instant messenger protocols and videoconferencing. Access to participants can be achieved via face-to-face, email, and social networking sites. Web-based questionnaires offer a range of advantages over traditional methods of research (Janghorban, Roudsari & Taghipour, 2014). These include reduced cost, ease and speed of distribution, and access to large samples (Fox, Murray & Warm, 2003). In addition, using online surveys allows researchers to target minority or specialised population groups that may not be accessible through traditional means. Despite the reported benefits, necessity of access to high-speed internet, familiarity with online communication, and having digital literacy impact the suitability of online surveys.

4.5 Data management

Creating a database

Once you have collected your data, you will need to create a database (also sometimes called a dataset). A database consists of rows and columns of data, where each row represents a different program participant and each column represents a particular variable. To protect the anonymity and confidentiality of program participants, identifying information (e.g., names, contact information, date of birth) should not be stored in the database with their data. Instead, every participant should be assigned a unique identifier. A unique identifier is a code (usually alphanumeric) that is used to identify a participant without revealing their name or other identifying information. When unique identifiers are assigned to participants, it is necessary to create two separate databases. One database will contain a list of all unique identifiers and information regarding the participant to which each identifier has been assigned (e.g., participant's name, contact details, date of birth). The second database will contain the unique identifiers and the data that has been collected on each participant (e.g., responses to survey questions), thus excluding any identifying information. These databases are kept separate as a data security measure.

Databases can be created in Excel and imported into statistical software packages (e.g., SPSS or Stata), or created directly within statistical software packages. If you have collected your data using specialised software, you may be able to export your data directly into an Excel spreadsheet. However, if you have collected data using a method such as pen-and-paper surveys, you will need to manually enter your data into your database. There are a number of procedures that you can use to help to minimise data-entry errors when manually entering data. Firstly, if entering data into Excel, you can set up parameters within Excel to help reduce data entry errors. For example, you can limit the range of values that can be entered for a particular variable (e.g., values must be between 1 and 5). Secondly, having two or more people working on data entry can help to reduce errors. For example, if there are two people assigned to data entry, the first person can enter the data and the second person can review the data that has been entered to identify errors.

Data storage

Data, particularly data that includes identifying information (e.g., participant names, contact information, date of birth), should be stored securely to protect the anonymity of program participants. Paper documents (e.g., consent forms, pen-and-paper surveys) should be stored in locked filing cabinets and electronic documents (e.g., databases) should be stored on password-protected computer drives. Data should only be accessible to program staff or external contractors who are involved in the evaluation of the program.

It is also important that you keep the data that you generate for an appropriate length of time. The Australian Code for the Responsible Conduct of Research recommends that, in general, research data should be retained for a minimum of 5 years from the date of publication of findings, however this may vary depending on the specifics of the data (NHMRC, 2007). Additionally, once this required period of data retention ends, it is crucial that research data is disposed of in a secure and safe way to protect the anonymity and confidentiality of research participants (e.g., shredding of paper documents).

Data cleaning

Data cleaning refers to the process of identifying and either correcting or removing incorrect, incomplete, or inaccurate data records in a database. Data cleaning generally occurs after a database has been developed. However, issues of incorrect, incomplete, or inaccurate data may be flagged during creation of the database, particularly if manual data entry is used.

Incorrect or inaccurate records may be the result of data entry errors, which can be fixed by referring to the original data record. However, incorrect, incomplete, or inaccurate records may also be the result of errors that occurred when data was originally recorded. For example, a participant completing a survey may decline to answer or accidentally skip a question or set of questions, resulting in incomplete data, or may circle two response options in a pen-and-paper survey when only one response should be given, or even write out their own response in addition to the listed options. It is important that protocols for data cleaning are developed at the beginning of the data cleaning process and are applied consistently. For example, depending on the particular variable (e.g., trust in government, feelings of belonging), it may be decided that if a participant has selected more than one response for a question that necessitates a single response, that particular participant's response will be coded as missing. Alternatively, it may be decided that in this situation the more conservative of the participant's two responses will be chosen as their response (e.g., if the participant was asked to rate how likely they would be to report a crime to police and they circled both 'likely' and 'very likely', their response would be recorded as 'likely').

4.6 Data analysis

There are many different approaches to quantitative data analysis. Broadly, quantitative data analysis involves the use of descriptive and inferential statistics. These two approaches and common analysis techniques for each are discussed in the following sections. While simple data analysis can be conducted in Excel (see Helmus et al., 2017), specialised statistical software is required for more complex analyses.

Descriptive statistics

Descriptive statistics (also referred to as summary statistics) are used to describe, explore or summarise the characteristics of quantitative data. Descriptive statistics cannot be used to draw conclusions or make generalisations beyond the data that has been collected. However, descriptive statistics enable raw quantitative data to be presented in a meaningful way that is easy to interpret. There are three key types of descriptive statistics used to describe quantitative data: (1) measures of central tendency, (2) measures of dispersion, and (3) measures of the distribution.

Measures of central tendency

Measures of central tendency provide summaries of what is 'average' in a dataset. There are three common measures of central tendency: (1) mode, (2) median, and (3) mean. The appropriate measure of central tendency will depend on the level of measurement of your data. An overview of these three measures of central tendency is shown in Table 1, below.

The mode measures the most common value of a variable, that is, the value that has occurred most frequently (e.g., the most frequently rated sense of belonging), and can be used with nominal, ordinal, or interval/ratio variables. As the mode measures the most frequently occurring value, it is possible for a variable to have more than one mode.

The median is the value of the middle score of a variable (i.e., the value that divides the distribution of data in half so that 50% of values are above it and 50% of values are below it). The median can be calculated for ordinal or interval/ratio variables. The median cannot be calculated for nominal variables (i.e., data that has no natural order, see above), as it requires information about order or rank of values. The median is a useful measure of central tendency for skewed data, as it is not as sensitive to outliers (e.g., unusual observations/values that are distant from other observations/values).

The mean is the average value of a variable (i.e., the sum of all values divided by the total number of values). The mean can be used with interval/ratio data and ordinal data in some circumstances. In social science research it is common for ordinal data, such as attitudinal questions measured using Likert scales, to be treated as interval/ratio data in analysis.

Table 1 Measures of central tendency

Measure of central tendency	**What it measures**	**The types of variables with which it can be used**
Mode	The most frequently occurring value of a variable	Nominal, ordinal, or interval/ratio
Median	The value of the middle score of a variable	Ordinal or interval/ratio
Mean	The average value of a variable	Interval/ratio (can also be used with ordinal variables when treated as interval/ratio)

Measures of dispersion

Measures of dispersion produce summaries of the spread or variation around a central point in a dataset. They are very useful in identifying the level of variability within a sample as it relates to a particular variable. Three common measures of dispersion include: (1) range, (2) interquartile range, and (3) standard deviation. Measures of dispersion can only be calculated for interval/ratio data. It is important to be aware that the distribution of your data (e.g., the presence of outliers – extreme responses) will determine which measures of dispersion are appropriate to use. If you are not familiar with statistics, it may be appropriate to consult relevant expertise as to which measures of dispersion are relevant and meaningful for your data. For program evaluations, it is not always necessary to undertake measurements of dispersion, but again this is when consulting a person with relevant expertise in statistical analysis and program evaluation can be helpful.

The range is the difference between the highest and lowest values of a variable and is presented as a single number (not an interval). For example, if a variable measuring age in years has a minimum value of 18 years and a maximum value of 63 years, the range would be 45 years. It is important to note that the range is sensitive to outliers. If a variable has one extreme value, this can make the range very large and not an accurate representation of the data.

The interquartile range is a more appropriate measure of dispersion if your data includes outliers. The interquartile range is the range of the middle 50% of the distribution of the variable (i.e., the 25th percentile subtracted from the 75th percentile). Therefore, the interquartile range excludes the tail ends of the data and potential outliers. Box plots (also called box-and-whisker plots) can be used to provide a visual representation of data dispersion. They present both the interquartile range and the full range of the data.

The standard deviation measures the average distance of values from the mean (i.e., how spread out the data is from the mean). A larger standard deviation means that data is more spread out around the mean, while a smaller standard deviation means that data is more closely clustered around the mean. Similarly to the mean, standard deviation is sensitive to outliers.

Inferential statistics

Inferential statistics are used to make inferences about a particular population on the basis of data from a sample of that population. There is a vast array of different techniques that can be used in inferential statistics. The types of statistical techniques that will be appropriate to use will depend on your evaluation objectives and design (e.g., sampling method, sample size), the types of quantitative data you have collected, and the resources available (e.g., statistical expertise, funding, and time). The following section provides an overview of simpler inferential statistical tests that may be relevant for evaluating CVE programs. This is not an exhaustive list. This section also touches on more advanced statistical techniques that you may wish to explore further.

In this section, the terms ‘dependent’ and ‘independent’ variables will be used. A dependent variable is something that changes (i.e., an indicator, such as your outcome measure or participant attitudes) due to the influence of another variable that is termed the independent variable (this could be components of an intervention, your service, or other types of attitudes or behaviours being measured).

Pearson’s chi-squared test

Pearson’s chi-square test (commonly referred to as just a chi-square test) can be used to test whether there is an association between two categorical variables (e.g., gender, and trust in police or willingness to use violence). It is a non-parametric or ‘distribution free’ test, meaning that variables do not need to be normally distributed for a chi-square test to be performed.

Paired samples t-test

A paired samples t-test (sometimes called a dependent samples t-test) can be used to test whether there is a statistically significant difference in the means of two sets of observations (i.e., two variables) taken from the same group. Paired samples t-tests require repeated measures – that is, each program participant must be measured on a certain characteristic or observation (e.g., sense of belonging) twice, which results in a pair of observations for each participant. Paired samples t-tests are often used in pre/post-intervention designs. For example, you could measure participants’ sense of belonging before their participation in the program and once they had completed the program. You could then use a paired samples t-test to determine whether there is a significant difference in levels of sense of belonging after completion of the program compared to at the beginning of the program.

The paired samples t-test is a parametric test and has several assumptions. Firstly, the dependent variable (e.g., sense of belonging) must be either an interval/ratio variable or an ordinal variable. Secondly, the dependent variable should have an approximately normal distribution. If the dependent variable is not normally distributed, you would need to use an equivalent non-parametric test, such as a Wilcoxon signed-ranks test.

Independent samples t-test

An independent samples t-test (sometimes called a two-sample t-test) is used to measure whether there is a statistically significant difference in the mean of a variable measured across two independent groups or samples (i.e., the individuals in one group are not the same as the individuals in the second group). You would use an independent samples t-test if you had measured one variable (e.g., frequency of discrimination) across two different sets of individuals (e.g., program participants and a comparison group). Similarly to the paired sample t-test, the independent samples t-test is a parametric test and has a number of assumptions. The dependent variable (e.g., frequency of discrimination) must be an interval/ratio or ordinal variable, as a mean must be calculated for each group. The dependent variable should also be approximately normally distributed for each group. If this assumption is not met (i.e., your data is skewed), you would need to use an equivalent non-parametric test, such as a Wilcoxon rank-sum test or a Mann-Whitney U test.

Analysis of variance (ANOVA)

Analysis of variance (ANOVA) is another test that can be used to determine whether there is a significant difference or variation in the mean of a variable across independent groups, similar to an independent samples t-test. However, while an independent samples t-test can only be run with two independent groups, an ANOVA can test for a difference in means across three or more independent groups. Similarly to an independent samples t-test, the groups must be independent (i.e., there is a distinct set of individuals in each group). An ANOVA should only be run with an interval/ratio or ordinal dependent variable that is normally distributed. For example, you could use an ANOVA to test if there is a difference in mean level of trust in the government across three distinct groups (e.g., program participants who completed the entire program, program participants who only completed half or less of the program, and a comparison group of individuals who did not participate in any element of the program), assuming that your data meets the test assumptions for an ANOVA.

Correlation

A correlation test measures both the direction and strength of an association between two variables. There are many different types of correlation tests, however, the most common is a Pearson’s correlation. Pearson’s correlation can be used with data that is either interval/ratio or ordinal and normally distributed. Pearson’s correlation also assumes that the relationship between the two variables is linear (i.e., there is a straight-line relationship). For example, you could use a correlation matrix to examine the association between trust in police and willingness to report suspicious behaviour to police. We may expect to find a moderate positive correlation, meaning that, on average, as trust in police increases, willingness to report suspicious behaviour to police also increases.

More advanced statistics

Regression

Regression analysis is a quantitative research method that involves modelling and analysing two or more variables, where the relationship includes a dependant variable (also referred to as the response variable) and one or more independent variables (or predictor variables). There are several types of regression that can be conducted, the most common being *simple linear regression*, *multinomial regression* and *logistic regression*. A lack of training or familiarity with these tests may require you to seek assistance, if they are deemed appropriate to the aims of the evaluation. We only cover simple linear regression here.

Simple Linear regression

A simple linear regression model attempts to explain the relationship between two variables using a straight line. This regression model is called simple because there is just one independent variable. A simple linear regression predicts the values of the dependent variable (y) as a function of the independent variable (x). For example, an objective of a CVE program may be identifying factors that lead individuals to disengage from society. A simple linear regression model could be used to examine the association between levels of civic participation (dependent variable) and trust in government agencies (independent variable). That is, does an increase of trust in government agencies predict an increase in individual levels of civic participation? If so, then improving levels of trust in government may help improve civic participation. It is important to remember that the larger your sample size the less likely you are to have error in your models. Determining an adequate sample size depends on the confidence levels, estimate type 2 error, and standard deviations of independent and dependent variables.

MANOVA

Multivariate analysis of variance is simply an ANOVA with several dependent variables. Major effects and interactions are assessed on a combination of dependent variables. The number of independent variables involved distinguish a one-way MANOVA from a two-way

MANOVA. When comparing two or more continuous dependent variables by a single independent variable, a one-way MANOVA is appropriate. A two-way MANOVA entails two or more dependent variables and compares them by at least two independent variables.

4.7 Reliability and validity of quantitative data

In quantitative research, key indicators of the quality of a measuring instrument are the validity and reliability of the measures. Validity is defined as the extent to which a concept is accurately measured. For example, a survey that is designed to explore social cohesion but is actually measuring a sense of belonging, would not be considered a valid measure. The second measure of quality in quantitative research is reliability. It evaluates the stability of measures, internal consistency of measurement instruments, and the interrater reliability of instrument scores. In other words, how accurately and consistently does a research instrument produce the same results if used repeatedly in the same situation (e.g., with other populations subject to the intervention in another jurisdiction) (Heale & Twycross, 2015). It is important to consider validity and reliability of data collection tools when conducting program evaluations.

Validity measures in quantitative data collection

It is rare, if not impossible, that an instrument will be 100% valid. The process of validation involves collecting and analysing data to assess the accuracy of an instrument. There are numerous statistical tests and measures to assess the validity of quantitative instruments used to collect data. We will focus on external and content validity.

External validity measures the generalisability of the research. In other words, are the sampling methods and data collected trustworthy, enabling generalisations about the larger population to be made? Establishing external validity follows directly from sampling. Recall that a sample should be a representation of the population from which it was drawn, because the total population may not be available. The sampling technique must be examined to determine factors such as subjects, time, the type of intervention, and measures used that could affect external validity. *Content validity* looks at whether an instrument adequately covers all the content that it should with respect to a particular variable. In other words, do the measures (questions) accurately assess particular indicators?

Reliability measures in quantitative data collection

Reliability can be thought of as consistency, that is, does the instrument consistently measure what it intended to measure? In quantitative research, there are two general estimates of reliability:

1. *Test-retest reliability*: The consistency of a measure evaluated over time. Reliability is assessed when an instrument is given to the same participants more than once under similar circumstances.
2. *Internal consistently reliability*: Cronbach's alpha is a statistical measure to test the consistency of results across time. Strong alpha scores indicate high reliability, while weak scores indicate the instrument may not be reliable.

4.8 Hypothetical quantitative data collection and analysis scenario

One objective of a CVE program may be to increase individuals' willingness to engage in civic activities and assess if their levels of civic participation change over time. Hypothetically, the intervention could have involved the provision of social services, the use of mentors, education campaigns or community outreach in schools.

Selecting the evaluation design and the sample

In this scenario, a pre-/post-intervention evaluation design is used. The sample consists of the participants of the CVE program.

Data collection

To measure change over time, it is necessary to collect data at multiple time points. In this scenario, data is collected from program participants before they begin the intervention and six months after they have completed the intervention, using pen-and-paper questionnaires self-completed by the participants. Participants complete the first questionnaire on their first day of involvement in the program. This measures their baseline levels of civic participation in the last six months. The six-month follow-up questionnaire will be mailed to participants with a pre-addressed return envelope. This questionnaire will measure their civic participation in the six months since they completed the program.

To ensure that matching data is collected at both time points, there are a few issues to consider. Firstly, the measure of civic participation used in the baseline survey should be the same as the measure used in the follow-up survey. For example, you may use a pre-existing measure of civic participation, such as the civic participation scale developed by the South Australian Community Health Unit (see Hodgkin, 2011). This 11-item scale measures frequency of participation in range of individual and group activities (e.g., signing a petition, being involved in a campaign or action to improve social or environmental conditions, being involved with a political party, trade union, or political campaign). Participants are asked this same set of questions in the baseline questionnaire and the follow-up questionnaire to measure their levels of civic participation in the six months prior to the program and the six months after completion of the program.

Secondly, it is important that the participants who complete the follow-up questionnaire are the same participants who completed the baseline questionnaire. In order to measure within-individual change (i.e., changes in civic participation over time), it is necessary to be able to identify who completed each questionnaire in order to match participants' baseline and follow-up questionnaires. To protect participants' anonymity, unique identifiers should be used. Unique identifiers can be assigned in a number of ways. For example, the paper baseline questionnaires may have unique identifiers already assigned. As program staff hand out the baseline questionnaire, they could use a log sheet to record the participant's name and the unique identifier assigned to that participant. This ensures that identifying information is not recorded on the same document as participants' questionnaire responses. When the follow-up surveys are mailed to participants, the unique identifier can also be prefilled on the questionnaire. Alternatively, participants could be asked to record their name on the baseline questionnaire and a unique identifier could be assigned afterwards. Importantly, any paper or electronic documents that contain identifying information should be stored securely.

Data analysis

Before data can be analysed, the information from the paper questionnaires needs to be entered into a database and the data needs to be cleaned. Once data is ready to be analysed, the first step is to examine your data by producing descriptive statistics, such as calculating proportions or means for particular variables. The types of descriptive statistics you choose will be dependent on the type of variable you are examining (e.g., categorical or continuous). For example, if participants were asked how often in the past 6 months they had attended a council meeting ('did not attend a council meeting in the past 6 months', 'once in the past 6 months', 'twice in the past 6 months', '3-4 times in the past 6 months', '5 or more times in the past 6 months'), this would be an ordinal categorical variable and you could calculate the proportion or percentage of participants in each response category at baseline and the 6-month follow-up. This data could be presented in a table or using a histogram.

A paired samples t-test (also known as a dependent samples t-test) could then be used to test for within-individual changes in frequency of council meeting attendance. While paired samples t-tests require a continuous dependent variable, it is common in social science research to treat ordinal variables, such as the variable described above and other variables with Likert scale response categories, as continuous. Paired samples t-tests also assume that the differences between paired variables are normally distributed. If the difference between variables is not normally distributed, you will need to use a non-parametric test instead, such as a Wilcoxon signed-ranks test.

While you can test whether there is a significant difference in frequency of council meeting attendance, this test does not prove that participation in the program necessarily caused this difference (i.e., you cannot claim causation). There may be other factors that could have impacted on frequency of council meeting attendance (e.g., change in employment in the time between the baseline and follow-up survey, recent purchase of a home, birth of a child). If you have measured factors that may be related to civic participation in your questionnaires, you can use a correlation matrix to test for correlations (i.e., relationships) between your variables (e.g., whether having children is associated with more frequent council meeting attendance). You could also use a correlation matrix to examine whether particular factors related to program participation (e.g., percentage of program meetings/activities attended/sessions completed) are associated with the frequency of council meeting attendance after program completion. This information could be drawn from program administrative data. The types of analyses outlined above could be repeated for each variable of interest.

More advanced analysis

Depending on the resources available to you (e.g., staff experienced with statistical analysis or resources to hire someone to conduct the analysis), you may be able to conduct more advanced statistical analysis for your program evaluation. If you have used a scale to measure civic participation, you could calculate a composite score representing civic participation; pre-existing scales will have a particular method for calculating an overall score. Regression analyses could then be conducted to examine associations between your dependent variable (e.g., level of civic participation in the 6 months following completion of the program) and one (univariate regression) or more variables (multivariate regression).

The type of regression you use will depend on how your dependent variable is constructed. For example, if you have a binary or dichotomous dependent variable (e.g., high level of civic participation vs. moderate/low level of civic participation), then you could use logistic regression. However, if your dependent variable has multiple ordered categories (e.g., low level of civic participation, moderate level of civic participation, and high level of civic participation), you would need to use multinomial logistic regression or a similar method.

The key strength of multivariate regression is that it allows you to examine what factors may predict your dependent or outcome variable. In the case of civic participation, you could construct a logistic regression model with level of civic participation in the six months following completion of the program as your dependent variable (e.g., high level of civic engagement vs. moderate/low level of civic engagement). You could include several independent or predictor variables in your model (e.g., initial level of civic participation before participating in the program baseline measure, percentage of program meetings/activities attended, socio-economic status, age, and gender). This model produces odds ratios (i.e., a measure of how strongly the presence or absence of a predictor, e.g., percentage of program meetings/activities attended, is associated with the outcome of the dependent variable, in this example, high level of civic participation), with all odds ratios adjusted for the other independent variables in the model. In other words, this model could examine the association between the percentage of program meetings/activities that participants attended and the likelihood that participants had high levels of civic participation six months after the completion of the program, considering participants' baseline levels of civic participation, as well as their socio-economic status, age, and gender. This can help you determine whether the program had an impact on civic participation and what other factors may influence this.

5. Qualitative data

5.1 Introduction

Qualitative data includes information that can be captured in non-numeric form. Generally, qualitative data is derived from observations and interviews. Qualitative data and analysis can be used in large-scale, rigorous formal evaluations and small-scale pilot studies (Cohen & Crabtree, 2006).

Qualitative data and analysis in program evaluations can enrich understandings of a participant's experiences by using techniques that give voice and articulate participant's perspectives. The use of qualitative analysis techniques also allows for the identification of reoccurring themes and patterns. It is important to keep in mind that, as with any evaluation, time and care must be taken to complete the evaluation in light of responsibilities, time constraints and the resources available. For example, when considering using large data sets (e.g., transcripts from interviews), the evaluation plan should include realistic time frames for conducting interviews, transcribing, coding, and writing.

Qualitative data and analysis can provide useful insights. For example, through interviews and focus groups, qualitative data can be used to measure participant satisfaction with a program, willingness to report suspicious terrorist-related behaviour, and gauge community awareness of violent extremism. Data collection through qualitative methods are most useful when measuring perceptions, attitudes and feelings. Thus, for CVE evaluations, these methods can be utilised to determine different outcomes, such as perceptions of and attitudes towards community engagement, or to examine community awareness of violent extremism and how perceptions of the threat of violent extremism influence people’s behaviour.

This section provides an overview of qualitative data and the common methods used for its collection, management and analysis. A hypothetical scenario that highlights data collection and analysis methods will also be provided to illustrate how various methods can be applied to CVE evaluations. Readers should be aware that this manual provides examples of qualitative methods that can be used when conducting evaluations, but it is not a presentation of all possible methods and techniques (e.g., qualitative case studies).

5.2 Types of qualitative data

Qualitative data is a categorical measurement expressed not in terms of numbers, but rather by means of description. Most qualitative data is textual; however, in some cases textual descriptions can be placed into categories and then assigned a numeric value to allow for quantitative analysis.

5.3 Sampling/recruiting participants

The recruitment of participants in qualitative research is dependent on various factors, which include the nature of the research (e.g., time frame, risk to participants), the infrastructure (e.g., accessibility to potential participants, collaboration with gatekeepers), participant characteristics (e.g., likelihood of giving consent, cooperation) and recruiter characteristics (e.g., experienced researcher) (Newington & Metcalfe, 2014). Recruitment will depend on the audience; for example, if wanting to recruit those who have been influenced by extremist messaging, this will be a relatively small and difficult group to access – this is where qualitative sampling techniques can be useful.

Nonprobability sampling is most commonly used in qualitative data collection; this sampling technique focuses more on understanding rather than on generalisability. Sampling is carried out to include information-rich cases and typically is not done randomly, but by using purposeful sampling techniques. Qualitative samples tend to be smaller, non-random, and do not intend to produce generalisability (Marshall & Rossman, 2011). There are several types of sampling used in qualitative data collection, including purposive sampling, snowball sampling and quota sampling. The explanations below are designed to help you understand the reasons for choosing each method. Sample size is also mentioned.

Purposive sampling

Purposive sampling is the most common sampling strategy used in qualitative research. Participants are chosen based on pre-selected criteria that is determined by the research aims. The main goal of purposive sampling is to include participants who represent a broad range of perspectives (Neuman, 2007). Purposive samples are most successful when data review and analysis are done in conjunction, as purposive sample sizes are often decided on the basis on what is termed theoretical saturation, that is, where new data will no longer give any additional insights.

Snowball sampling

Snowball sampling relies on research participants to recruit other participants or on researchers choosing participants they wish to include. As the name suggests, a snowball sample builds and grows larger as the study continues. This technique is useful when researching a particular behaviour or group that is difficult to access and recruit. For example, a researcher may want to talk to people who view extremist online messaging. They may know someone who does view this material and interview him or her; that interviewee may refer another person to them (e.g., friend or associate) who they know also views such material.

Quota sampling

This technique has a pre-set number of participants with specific characteristics. Characteristics may include age, gender, ethnicity, class, marital status, profession, etc. These set criteria allow the researcher to focus on a set of people considered the most likely to experience, know about, or have insights into a topic. These pre-determined characteristics allow researchers to go into the community and using appropriate recruitment strategies (i.e., applicable to the required location, culture, and study population) to find participants (e.g., members of a program's target group) who best fit the prescribed criteria.

Sample size

Unlike in quantitative research, qualitative sampling has no formula for determining an appropriate sample size. Because statistical power is not the goal, researchers need to consider the trade-offs between breadth and depth, which can be done by considering the purposes of the evaluation (Marshall, 1996). Sample sizes may or may not be fixed prior to data collection depending on the resources, time available and objectives of the evaluation.

5.4 Data collection

There are several ways of collecting data in qualitative research. The type of data collection method you chose will depend on the type of evaluation design, objectives, the targeted population group and resources. Qualitative data collection methods are often time-consuming; therefore, the data collected is often from a smaller sample size than that used for quantitative data. Compared to quantitative data collection methods, qualitative data collection is much less structured and may use a combination of techniques throughout and adjust accordingly to coincide with the research questions and aims. This section provides an overview of the types of qualitative data collection methods.

Interviews

What makes interviews qualitative is the use of open-ended questions that allow participants to express responses in their own words; this gives the benefit of in-depth responses and allows the participant to elaborate further. Open-ended questions should be constructed in a way that allows respondents to use their own words to respond. This means questions should not elicit categorical responses such as “yes”, “no”, “more”, “less” or “often” but should open with questions starting with “who”, “what”, “when”, “how” or “why”, and use prompts such as “can you give me an example”; “why do you see that as significant?”. Due to the less-structured nature of qualitative interviews, participants may feel uncomfortable in sharing particular insights or experiences, and consideration should be given to procedures to build rapport and to ensure participant confidentiality.

Interviews in qualitative research can be structured or semi-structured:

- Semi-structured interviews use an interview guide which includes questions or issues that are required to be discussed, but also allows for flexibility so new issues or topics can be raised. Probing questions may be used to remind the interviewer to follow up on specific issues, but the interviewee is given the flexibility to deviate from the schedule. Semi-structured interviews are methodologically strong, but require experienced staff to know how to properly conduct interviews to ensure that critical areas are adequately covered.
- Structured interviews use a standardised approach, with questions prepared in advance and asked in the same way to each participant. These interviews are the most time-efficient and easiest to administer.

Focus groups

Typically, focus groups include 6-10 participants. Participants are chosen based on a purposive sample, in that they have similar characteristics (i.e., age, ethnicity, socio-economic status) and have something to say about a particular topic of interest (Rabiee, 2004). Based on the group dynamics, information retrieved from focus groups can provide rich data. An interview guide, similar to that used in semi-structured interviews, is generally adopted. Ideally, two appropriately trained and skilled individuals will run the group, one facilitating the group discussion, the other taking notes and dealing with audio recordings.

This approach is much more cost-effective than individual interviews. Participant interaction can improve data quality, as the degree of consensus or variation on viewpoints is much easier to gauge (Cohen & Crabtree, 2006). However, due to the number of questions and range of issues to be explored, participant responses can be limited given the need to allow all group members to be involved. Additionally, issues may be controversial or personal and thus not be suitable to bring up in a group context.

Observations

Observation is another data collection technique that produces qualitative data in the form of detailed field notes. Direct observation yields detailed descriptions of activities, actions and behaviours of individuals in a particular setting or situation. Observations focus on interpersonal interactions, settings, and organisational processes and procedures. Observational data can tap into phenomena that program participants may not express in an interview or be aware of (Patton, 2002). Observations may be particularly useful in assessing program implementation or capturing interpersonal dynamics of participants. The downfall of observations is that they are time and labour intensive. Preparation of field notes to capture what has been observed takes time, in addition to time spent on-site where the observations are taking place.

5.5 Data management

Creating a database

Careful structuring of databases for storage, retrieval, and data management are important issues to consider when collecting qualitative data. Before data collection, choosing and following a clear file naming system, developing a tracking system and establishing document transcription/translation procedures need to be implemented to ensure quality data management (Johnson, Dunlap & Benoit, 2010). During the data collection phase, any unit of work must be recorded – usually written in Microsoft Word – that describes the date, time, location and particular focus (particularly important for observations). Each unit of work needs to have a standard 'header information' which includes the file name, date, approximate time, the researcher's name and participant descriptions. It is important to maintain consistency when creating a file naming system so as not to create duplicates, extensions of names or names that do not specify details correctly (e.g., the use of identifiers that are not easily recognisable). When data is collected and recorded it is critically important that each researcher maintains a copy of every file for subsequent cross-checking. Data can be stored in pen-and-paper records, Microsoft Word or note-taking programs, or using specialised data storage software.

5.6 Data analysis

Qualitative data analysis focuses mainly on understanding and interpreting phenomena, people and situations, through textual analysis. Thus, data analysis is not statistically structured and requires calculated interpretation and coding methods.

Before beginning the process of data analysis and depending on the amount of data you have, (e.g., number of interview transcripts), it may be necessary to go through the process of data reduction, in which data is condensed. This process involves reducing and transforming your raw data (collected from interviews, observations etc.) into an ordered and simplified form (e.g., edited transcripts). The data reduction process allows you to identify and focus on what is meaningful and can be used in data analysis. It is the process of searching through your raw data (data that hasn't been altered or changed) to determine what is significant and then transform the data into a simplified format (i.e., editing down your interview or focus group transcripts or dividing different sections under broad headings/categories). There are risks with doing this: one is that important data can be deleted, another is that the process can be subject to bias (i.e., a researcher only keeps data that confirms their hunches). To avoid this, data reduction should be conducted by two or more people to allow for crosschecking and unanimous decisions on data typologies, categories and removals (Patton, 2002).

For qualitative data to be analysed it must be grouped into meaningful patterns and themes. This process is generally conducted in two ways: content analysis and thematic analysis. It should be mentioned that there are several types of qualitative data analysis methods. The technique of thematic analysis is one of the most common qualitative analysis methods used and will be explained below.

Thematic analysis

This analysis procedure involves viewing the interview or focus data several times (e.g., reading and re-reading manuscripts), identifying patterns and themes (e.g., finding common statements or ideas that appear repeatedly), and categorising the data (e.g., coding the data according to the themes identified). Coding can be based on the concepts or topics followed in the interview guide. Simply, thematic analysis involves three steps:

1. The coding of text;
2. The development of descriptive themes; and

3. Generation of analytical themes.

Thematic analysis can be completed in several ways:

- By hand: If your dataset is relatively small, themes and categories can be organised using the paper-based data via making notes in the margins or highlighting specific themes in the transcripts through the use of a colour-coding scheme.
- Using a MS word document: Similar to handwritten notes, except that the amount of data can often be increased and organising becomes easier.
- Using specialised data analysis software: data analysis software enables you to work with text, images, audios and video data, which is especially useful for qualitative data. Packages such as NVivo and MAXQDA allow you to code qualitative data more quickly and efficiently, search for codes and themes within transcripts and display your data in graphs or charts.

From coding, the researcher identifies descriptive themes, that is, patterns that have emerged from the coded data. Themes can arise from repeated behaviour, frequency of occurrence, or certain incidences that are repeated (Thomas & Harden, 2008). There are limitations with thematic approaches. For example, if you have a large number of interviews this process will be time consuming and alternative analysis techniques should be considered (e.g., through the use of software packages such as Leximancer that allows for automated coding according to the identification of key concepts). Also, thematic coding can be subject to bias and it is advised that it should be conducted by two or more people to ensure there is consistency in how data is coded (see reliability and validity below).

5.7 Reliability and validity of qualitative data

Demonstrating rigour when undertaking qualitative research is challenging. Qualitative research is often criticised for lacking scientific rigour, poor justification of adopted methods, and analytical procedures being subject to researcher bias. Tests and measures used in quantitative research cannot be applied when determining the reliability and validity of qualitative data. Therefore, it is debated whether terms such as reliability and validity are appropriate to evaluate quality in qualitative research. As a result, reliability and validity – terms used most commonly in quantitative research – have been redefined to reflect the ways of establishing accuracy in qualitative research. Terminology such as credibility, transferability, dependability, consistency and trustworthiness are used to describe validity and reliability in qualitative research (Leung, 2015). In terms of testing the reliability of coding methods, there are tests used to determine the accuracy of the methods used. One is to test for inter-coder reliability, which refers to the extent to which two or more independent coders agree on the coding of the content of interest and use of the same coding scheme. Low-level inter-coder reliability may suggest weak coding methods (Cho, 2011).

5.8 Hypothetical qualitative data collection and analysis scenario

An objective of a CVE evaluation could be gauging (positive) perceptions of Australia after the implementation of a school-based program that focuses on civics, Australian culture and the roles of different institutions. As noted in the *Countering Violent Extremism Evaluation Indicator Document*, if people have a positive perception of the country in which they live they will be potentially less vulnerable to extremist messaging or influences.

Selecting the evaluation design

A post-intervention evaluation will be used. Data will be collected from a sample of participants after the intervention has been completed.

Data collection

In this scenario, data will be collected via focus groups. Focus group participants will be recruited from schools that have participated in the intervention (purposive sampling). Participants will be required to provide informed consent to participate in the evaluation. In such circumstances it may be necessary to seek parental consent, given the age of participants. During the focus group interview, an audio recorder will document the interview to ensure no information is missed or lost and will allow for accurate transcription and coding. In addition, notes will be taken on participants' opinions, beliefs, and attitudes. A semi-structured interview guide will be used to keep discussions on track.

Some issues may arise during data collection. Lack of participant interaction will directly influence data quality; to minimise this, it is important to avoid personal questions and build rapport to allow the group discussion to flourish on its own. Discussions must also be guided to ensure subjects and topics are relevant to the evaluation. Given that no data has been collected prior to the initiative, there will be no baseline measure of attitudes to compare to once the intervention is completed. This is where the design of the focus group schedule is important, because it needs to focus on the value and benefits accrued from the program. Therefore, focus group topics could canvass judgements about the most beneficial and interesting component of the initiative, which aspects participants gained the most from, and whether any specific component changed their attitudes about Australian society – and if so, how, and if not, why. Other topics could encompass perceptions of key social institutions, such as the government and legal authorities.

Data analysis

Before any analysis can occur, the wealth of data collected needs to be reduced. Data reduction should always be aligned with the research aims, in this case, any information pertaining or inferring relevance to perceptions of Australia or Australian culture should be kept. This includes any conversations, phrases or nonverbal cues that could give a deeper insight into participants' perceptions of Australia and the benefit of participating in the program. Similarities and differences between transcripts also need to be identified and noted to use later for comparisons (e.g., differences between males and females).

Once the data is formatted for coding, the next step is to identify key themes, form categories, summarise and tabulate the focus group transcripts. This can be done using thematic analysis. This analysis is conducted in various stages. First, focus group content is classified into themes; content can be a sentence, phrase or word. A theme for this scenario could be "reported benefits of the program" or "perceptions of Australia". Next, sub-categories are developed from these broad themes. The theme "perceptions of Australia" could be divided into the sub-categories of positive and negative perceptions, while the theme of "reported benefits of the program" could be divided into positive benefits, and no reported benefits. Once categories and sub-categories are established and segments of the focus data allocated to them, they are compared to determine any links between them. Themes and categories are then reviewed by two coders to ensure inter-coder reliability, consistency and accuracy (i.e., to ensure themes match up and segments of the focus group transcripts have been coded in the same way). Once all categories are established, these are then tabulated to allow for the interpretation of results and conclusions to be drawn.

It must be noted that this method has its limitations, in that it will not definitely provide results as to whether the school-based program changed attitudes towards Australian society and culture. The most it will do is to provide insights into how students perceived the program, what benefits they derived from it, and if any aspect of the intervention needs improvement because the students did not see particular components as useful. It would not assess the primary aim of the intervention, that is, generating positive perceptions of Australian society and culture so youth are less vulnerable to extremist messaging or influences.

6. Secondary data

6.1 Introduction

As previously mentioned in this guide, data can be categorised as either primary or secondary. Previous sections have presented *primary* or 'raw' data collection methods, but data for program evaluations can also come from *secondary* data sources. This section focuses on secondary data and provides details on how such data could be collected, managed and analysed.

Secondary data is pre-existing data that has been collected by someone other than the user, with the data often being used for a purpose that it was not originally intended for. The data may be readily available and have undergone data cleaning and analysis. There are benefits in using secondary data given it saves times and resources. However, there is no control over the quality of the data itself and if there is no associated documentation (e.g., technical reports), it may be hard to identify any potential weaknesses of the data (Cheng & Phillips, 2014).

Readers should be aware of two issues that can arise when using secondary data for the evaluation of CVE programs. Firstly, because secondary data is not used for its original intended purpose, the data may not be aligned with the specific outcomes of your proposed CVE program. That is, while the data could be relevant in providing insight into certain outcomes measures, it may at best only be a proxy measure. For example, Centrelink data can be used to measure welfare outcomes for certain populations. Although not its intended purpose, this data could also be used in the CVE context. Data on welfare payments could be used an as indicator (proxy) for wellbeing or marginalisation. Data relating to school attendance records could be used to assess social participation. The second issue when using data from secondary sources is the data may not capture the appropriate groups targeted by a CVE intervention.

Many of the methods applicable to the analysis of secondary data have been presented in previous sections (see *Qualitative data* and *Quantitative data*). Not all existing methods for secondary data collection, management and analysis are mentioned here. As advised in other sections it may be necessary to consult relevant experts for guidance and advice.

6.2 Types of secondary data

There are various types of secondary data. Administrative data is one of the main forms of secondary data. Such data is collected as part of the day-to-day processes of registration, transaction and record-keeping of organisations, typically relating to the delivery of a service and a particular client group being engaged (Woollard, 2014). Examples include school enrolments, hospital discharge information, Medicare billing records, tax records, prison program attendance and electoral registration. It is possible to link administrative data with other secondary data sources (e.g., social survey data) depending on access and the availability of data linkage infrastructure.

Secondary data may be qualitative or quantitative in nature. Examples of qualitative secondary data sources could include diaries, letters, autobiographies, case notes, newspapers or government reports, historical documents, and annual reports of organisations. Quantitative secondary sources could include national census data, crime rates and unemployment rates.

Secondary data may be used on its own or in conjunction with primary data when conducting an evaluation, depending on the design and aims of the evaluation (Walliman, 2017). The feasibility of using secondary data needs to be considered. Access to some types of data can be restricted or will need to be negotiated. This is particularly relevant when attempting to access sensitive data (e.g., data on known extremists). Hence accessing data, such as case notes or diaries, may require a significant effort in negotiating with gatekeeper organisations (e.g., corrective services, community health providers), given the confidential or sensitive nature of the content of such sources.

An important factor to consider when it comes to administrative data is that the data was not collected for research purposes and therefore may not be documented in a suitable way to support data analysis. To understand this type of data, exploring how the data was originally sourced, gathered and recorded will provide a basis to begin condensing and sorting the data for an evaluation. This stage is important in identifying any inaccuracies and inconsistencies in the data.

6.3 Choosing data sources

When using secondary data as part of a CVE evaluation, you must first locate data sources that could be relevant to your program outcome measures. As mentioned, secondary data may not provide direct measures of particular program indicators but may instead provide proxy measures. The link (program logic) between program outcomes and indicators captured within secondary data sources will need to be considered.

A range of public, private and non-governmental agencies exist which collect and allow access to secondary data sources. In choosing a dataset, consideration needs to be given to its strengths and weaknesses. This requires a detailed understanding of the population that has been targeted by the secondary data (particularly relevant to CVE evaluation), the sampling technique (relevant when using secondary survey data), the timeframe of data collection, how variables were measured and captured, missing data, response rates and quality control measures. In addition, having access to associated documentation such as survey instruments, codebooks or technical notes will help you to judge the validity and reliability of the data (Cheng & Phillips 2014). It may be necessary to consult with the agencies or researchers who collected the data in the first place.

6.4 Data management

The process of secondary data management needs to be considered. The type of data sourced will dictate how the data needs to be stored. Oversight is also an important element of secondary data management. This may need to involve some type of agreement relating to data management, which can be especially important when agencies or organisations are sharing or exchanging data. Time demands associated with secondary data collation and management can be especially burdensome when large volumes of data from multiple administrative systems and agencies are being sourced (Connelly et al., 2016). Those who have little experience with secondary data may struggle with data management.

Data storage

Secondary datasets can be large and as such, computer requirements for storing and manipulating these large files can be demanding. Administrative records may be incomplete or be recorded on an unfamiliar system with ambiguous terminologies. A large proportion of secondary data issues relate to coding and input errors. For example, procedures for data input can change over time; specifically, names may change as can methods for capturing different variables. Security and confidentiality is an essential element of data storage when using secondary data and administrative records, as secondary data can hold individual personal information.

Preparing data for analysis

Due to inconsistencies associated with secondary data, several considerations need to be addressed before performing data analysis. As mentioned above, there can be a lack of systematic data entry procedures. For example, case notes may have no uniform structure, making them subject to data entry errors because there may be no consistency in how case notes are recorded (Rothbard, 2015). Identifying these discrepancies will be necessary. For quantitative secondary data, specific statistical tests can be undertaken to help identify any discrepancies. For example, frequency distributions and cross-tabulations can be used to assess missing values, incorrect codes, and outliers. Duplicate records may also have to be identified and removed. The process of matching, merging and linking data can also be an essential process when dealing with secondary data. When necessary, multiple datasets may need to be linked or merged; to merge accurately, variables intended to match individuals across multiple datasets need to be reviewed carefully (Lin, Maxwell & Forry, 2017). During this process, when using data that has been stored in numerous datasets, it is important to check the accuracy of the identifier variables (e.g., this could be a unique identifier or number), to ensure data from different time-periods or datasets are matched before they are merged (Cheng & Phillips, 2014).

6.5 Data analysis

Secondary and administrative data can be analysed in various ways. The types of secondary data you have will dictate the analysis methods used. Methods used in analysing primary data are also used in secondary data analysis, as mentioned in the previous sections (see *Quantitative data analysis,* pg. 17 and *Qualitative data analysis,* pg. 33). We have not repeated these data analysis methods here and readers should refer to the appropriate data analysis sections if their secondary data is quantitative or qualitative. Secondary analysis can involve the use of single or multiple qualitative datasets, single or multiple quantitative datasets, mixed qualitative and quantitative datasets, as well as administrative datasets (Heaton, 2008).

There are some basic steps that you are recommended to follow when conducting secondary data analysis for program evaluation:

- The relevance of the dataset to the program's target population – does it capture the population of interest?
- A review of the strengths and weaknesses of the dataset to be used – are the necessary codebooks, survey instruments, and any other documentation available?
- A plan needs to be developed that identifies specific variables of interest – how are the specific variables relevant to my program outcomes?
- Identify any inconsistencies or missing data – how will missing values or inconsistencies in the recording of data/information be addressed?
- Generate operational definitions relating to various variables – what are my dependent (outcome) and independent (indicator) variables?
- Identify the types of analysis that are to be completed – what are my most relevant forms of data analysis? (Cheng & Phillips, 2014).

Secondary data analysis can offer numerous benefits. Firstly, it provides large sample sizes which can help to improve the validity of the data and the analysis (see *Reliability and validity of quantitative data*, pg. 23 and *Reliability and validity of qualitative data*, pg. 34). In addition, secondary data analysis can be conducted on populations that were not involved in the original intervention. Hence, it may offer the opportunity for identifying comparison groups who were not the subject of the intervention, but whose data could be compared to the program's target population.

There are also disadvantages of secondary data analysis when used for program evaluations. For example, it may not have the level of detail required to determine a change of behaviour (that is, your outcome of interest). New primary data may need to be collected in addition to using administrative data.

6.6 Hypothetical secondary data collection and analysis scenario

An objective of a CVE evaluation may be to determine any behavioural changes resulting from program attendance as it relates to convicted terrorists or those at risk of radicalising to violent extremism. This would require you to examine administrative correctional data. We recognise that correctional authorities in Australia may collect data on extremist offenders using specific risk assessment tools e.g., VERA 2R (see Pressman et al., 2016). For illustrative purposes, in this hypothetical scenario we aim to highlight how routine correctional data might be drawn on for evaluation purposes. Here the aim is to see if attendance at general in-prison programs, that is, programs not specific to violent extremists, have an impact on any behavioural markers relating to the rehabilitation of extremist offenders (see indicator Outcome Three in the *Countering Violent Extremism Evaluation Indicator Document*). Some of these behavioural markers may not directly relate to reducing the propensity for violent extremism, but could act as proxy measures for rehabilitation. There are limitations with this approach.

Evaluation measures

Various behavioural flags could be monitored as proxy measures for individual behavioural markers resulting from program attendance (this could for example be attendance at an anger management program). Behavioural flags could relate to incidents of anti-social behaviour or violence, ratings of cognitive impairment, consistency of program attendance, and risk assessment ratings.

Secondary data sources

In this scenario, data is collected from correctional services (examples include the Offender Integrated Management System, OIMS, in NSW or the Integrated Offender Management System, IOMS, in Queensland). Offender records are collected during a 12-month period of program attendance. The data is drawn from individuals serving time for a terrorism-related offence. Each offender has a unique identifier that records individual history within the correctional system. These offender IDs monitor and record various behavioural flags and program attendance data. It should be noted that many correctional data management systems do not have an automatic flag for offenders identified as having radicalised. Evidence of this may be collated on correctional intelligence systems or protected data systems, which are not accessible by all staff. One thing to keep in mind is that the sample size relating to convicted terrorists may be too small for calculating statistically significant effects; the data will only be indicative of certain outcomes rather than providing evidence of a causal relationship between particular indicators and program attendance.

Data analysis

Before data can be analysed, quantitative data retrieved from offender ID files needs to be entered into a database and cleaned. Data includes basic demographic information such as gender, age and length of incarceration, as well as the program and behavioural variables of interest. Missing data or data errors will need to be addressed.

As this data is quantitative, the first step is to run descriptive statistics. Frequency tables could be calculated to determine the number of occurrences of any particular behavioural flag over the 12-month period. You could also calculate the proportion or percentage of participants that exhibited this behaviour at baseline and at the end of the 12-months. While you can test a difference in the frequency of, for example, the number of violent incidents, this does not prove that violent incidents are associated with program attendance. To test whether the number of violent incidents is associated with program attendance a Pearson's correlation could be used. A correlation tests both the direction and strength of an association between two variables, with a correlation matrix examining, in this case, the association between violent incidents and program attendance. Results may show that on average as program attendance increases, associated violent incidents decrease. You could also use a correlation matrix to examine whether particular factors relating to program participation (e.g., percentage of program sessions completed, type of program, length of program) are associated with frequency of program attendance. This may give some insight into how program characteristics impact on program attendance. However, there are limitations with this design in that the tests only measure a correlation between program attendance and a behavioural outcome, they do not demonstrate causation. Also, the data is being used for a purpose it was not originally intended for. Lastly, the sample size of offenders is likely to be very small. You could also compare data and results across terrorist and non-terrorist offenders to see if any outcomes are unique to each group.

7. Social Media Data

7.1 Introduction

Social media is a group of web-based and mobile-based Internet-based applications that allow the creation, access and exchange of user-generated content (Batrinca & Treleaven, 2015; Kaplan & Haenlein, 2010). It is typically associated with social networking sites such as Twitter, Instagram or Facebook, but can also include other online platforms such as Really Simple Syndication (RSS) feeds, (a group of web feed formats used to publish frequently updated content, such as blog entries or news headlines), YouTube, blogs, wikis, websites which allow user comments or facilitate online forums, news feeds and sites (Batrinca & Treleaven, 2015; Kaplan & Haenlein, 2010).

Analysis of social media data is of relevance to the CVE context given that violent extremist groups, ideas and ideologies can recruit and attract individuals to their cause through social media platforms and their associated messaging. Social media is a vehicle for the spreading of extremist ideological propaganda, anti-Western rhetoric, messaging and narratives that justify violence, and is particularly attractive to vulnerable youth who may be attracted to the 'sense of communal belonging' that joining online social forums and discussions may provide (Waldman & Verga, 2016). Analysis of social media data is also of relevance to wider CVE initiatives aimed at encouraging social cohesion, community safety and harmony to promote community resilience to violent extremism. Information extracted from social media can be useful to help in the development of CVE programs and strategies as well as understanding outcomes and outputs of initiatives.

Social media research is not yet well developed (Bartlett & Reynold, 2015; Conway, 2017), and fast-changing technologies associated with social media platforms mean that approaches to this form of research are constantly shifting. Furthermore, the use of social media as part of strategies to counter violent extremism is in its infancy, and while some CVE social media campaigns have been launched in recent years, very few have been assessed or evaluated to any great extent (Wilner & Rigato, 2017). Readers need to be aware of these changing dynamics and developments in CVE social media campaigns. This guide provides advice on various elements that need to be considered when carrying out social media research, and some suggestions for how social media research can inform program evaluations, but is not intended to be a comprehensive presentation of all approaches, techniques and methods.

7.2 What are the types of social media data?

In conducting social media research, it is important to understand the range of social media platforms that exist, and which are likely to be relevant to the research or evaluation you are conducting. Social media data can be both quantitative and qualitative, and the analysis of social media data typically involves some component of both.

Social media data sources vary in the amount of information that can be communicated, the level of social presence or influence that the platforms afford users, and the level of self-presentation and self-disclosure that the platform can provide for a user (Kaplan & Haenlein, 2010). These sources include well known social networking sites such as Facebook, Twitter, Instagram, YouTube or LinkedIn, as well as other emerging platforms, such as the discussion website Reddit, or microblogging and social networking website Tumblr. They can also include websites which have an interactive component (e.g., extremist organisation websites or news websites where users can respond to comments, posts or articles). Within Australia, the most popular social media sites from which relevant social media data can be drawn are Facebook, YouTube, Instagram, Wordpress.com (blog site), Tumblr, and Twitter (Cowling, 2017).

Quantitative metrics associated with social media data include frequency counts of individual users or levels of interest in particular social media content. For example, with Facebook, this could mean total page reach (i.e., the number of users who engaged with a Facebook page or viewed an activity on a page), post reach (i.e., the total number of users who viewed a post on a Facebook page) or total page likes (i.e., the number of 'likes' a Facebook page has received). With Twitter, metrics could include impressions (i.e., the number of times users saw a tweet while using Twitter), replies (i.e., the number of times other users referenced a tweet in their own posts), retweets (i.e., the number of times other users reposted or shared a tweet on their own Twitter account), or 'likes' (i.e., the number of times other users indicated support for a tweet). YouTube metrics could include the number of times a video has been viewed or liked (Helmus et al., 2017: 101).

Qualitative social media data relates to the nature of the content of social media information. This is unstructured data such as textual comments and involves analytical approaches such as sentiment analysis. This could include understanding reactions to opinions that are presented online (e.g., reactions to public statements by controversial political figures), understanding attitudes towards social movements (e.g., tweets relating to a particular theme), understanding the nature of comments on extremist organisations websites, or identifying reactions (and the nature of these reactions) to specific events (e.g., posts relating to a terrorist incident). An overview of the various types of quantitative and qualitative data that can be extracted from social media data is provided in Table 2.

Table 2 Overview of social media data forms

Social Media Platform	Example Quantitative Metrics	Example Qualitative Data
Facebook	Via Facebook Page Insights: https://www.facebook.com/business/a/page/page-insights *Page likes:* Total and new likes for a Facebook page. *Post reach:* Number of users who viewed a post on a Facebook page. *Engagement:* Total number of unique people who engaged with a Facebook page, as well as different engagement types (e.g., likes, comments, shares). *Video views:* Number of times a Facebook page's video was viewed. (Also see https://blog.bufferapp.com/facebook-insights)	Sentiment/Opinion[1] – negative or positive comments in response to a Facebook page or post.
YouTube	Via YouTube Analytics Basic: https://support.google.com/youtube/answer/1714323?hl=en *Realtime views:* Estimated channel views over the past 48 hrs. *Views:* Channel views compared with the previous period. *Average view duration:* The length of time that users spent watching a video. *Likes/Dislikes:* How many people liked or disliked a video. *Comments:* Number of negative/positive comments that a video has received.	Search of YouTube videos for tags or videos with a relevant term. Nature of negative/positive comments on a particular propaganda video.
Instagram	Via Facebook Business: https://www.facebook.com/business/help/788388387972460?helpref=related *Impressions:* Number of on-screen posts. *Likes:* Number of likes for a post.	Hashtags – the nature of various Hashtags that could be used and the context in which they are used. Analysis of the nature of comments associated with Hashtags.

[1] Note that Sentiment/Opinion analysis can be considered a combination of both qualitative and quantitative data. That is, while it provides some understanding of the nature of the comments, for example, whether people have a negative or positive sentiment in relation to a particular Facebook or Twitter post, it also provides evidence on the extent of the perception (e.g., the number of people who have a particular sentiment).

	Engagement: Total number of unique Instagram accounts that liked, commented or saved a post. *Reach:* Number of unique accounts that saw a post. *Followers:* Numbers of total and new followers of an Instagram user. *Use of Hashtag:* Number of times that a Hashtag has been used. *Referral traffic:* How much traffic an associated website receives from Instagram.	
Blogs/Websites (including microblogging social media sites such as Tumblr)	<u>Many blog/website metrics can be sourced from Google Analytics; metrics include:</u> *Visits:* Number of blog visitors and visitor frequency. *Traffic source breakdown:* Where visits to the blog are coming from e.g., from other social media, Google searches etc. *Most popular posts:* Which blog post have had the most views. *Location:* Where most visitors are coming from. *Average time on site:* Average amount of time that individuals spending viewing the blog. *Active RSS/Email Subscribers:* Number of subscribers to the blog's RSS feed and active checking of feed/number of email subscribers.	Nature of content on particular websites/themes associated with a particular discourse on any given topic. Comments or discussion posts from users.
Twitter	<u>Via Twitter Analytics:</u> https://analytics.twitter.com/about The Tweet activity dashboard shows detailed analysis of a user's Twitter activity, including: *Impressions:* Number of times users saw the Tweet on Twitter. *Engagements:* Total number of times a user has interacted with a Tweet (e.g., clicks anywhere on the Tweet, retweets, replies, follows and likes). *Followers:* Number of new and total followers.	Semantic analysis can be conducted on social media data (e.g., Twitter) to identify topics or to characterise followers by their profile data.

(Adapted from Helmus et al., 2017)

7.3 What to consider when using social media data

In setting up data collection for social media research, considerations should include the following:

- *Which social media platforms would be the most relevant for the research question or evaluation?* It may be that your research is aiming to understand the coverage of an issue or evaluate a social media campaign that has been launched on a specific social media platform. However, if your research is aiming to gather information on the full coverage of an issue across social media platforms, then a range of social media platforms would need to be included. Decisions on which social media platforms to include will be influenced by factors such as the popularity of the various platforms amongst a target group or population of interest, the issue being investigated, and the accessibility of the social media data itself.
- *What are my main criteria/search parameters for selecting data from this platform?* This includes considerations about timeframes from which the data can be drawn, whether the users are from a particular group of user accounts (e.g., previously identified extremist users, or known extremist group user accounts), and the use of topics and keywords (e.g., use of hashtags if the social media platform is Twitter; other keywords including full text searches). When using topics and keywords you may need to continually review and revise to include new related topics (e.g., associated hashtags) as they arise. Other considerations include whether you are using metadata (e.g., not based on key words or topics but based on location), the timeframe (e.g., all posts on a particular day) and the language (e.g., only those in English).
- *How much data is needed to answer the research question?* You may need to consider data storage, processing infrastructure available to deal with the amount of data being collected, as well as ethical considerations in accessing personal data on-line.
- *What is excluded if data is collected this way?* You need to consider, or at least acknowledge, that social media populations may be biased with little possibility of checking this (i.e., profiling information such as gender, age or the location of users is often not identifiable from the users). Platforms may restrict access and not be completely transparent about what has been restricted (e.g., those who shared their age on their Twitter profile may not be representative of all Twitter users or may even lie about their age) (Mayr & Weller, 2017). Gathering a representative sample is a significant issue when collecting social media data and caution needs to be taken when interpreting the findings.

7.4 How to collect social media data?

The method of collecting social media data will be dependent on the scope of the research you are conducting. For small studies with only basic measurement and analysis requirements (e.g., collection of data related to a specific Facebook page), data can be extracted manually (e.g., number of likes of a Facebook page, number of followers, comments made about a particular post and so on).

However, typically when conducting social media research, the data being collected is very large in size and resource intensive. Thus, collecting social media data can be carried out directly via data analytics programs designed by social media platforms themselves. These tools offer web scraping services (also known as web harvesting or web extraction), which collect online data from social media and other websites in the form of unstructured text (Batrinca & Treleaven, 2015). Or, alternatively, companies that offer data extraction services from social media platforms can be used.

Analysing social media data has become a major research and business activity due to the availability of web-based application programming interfaces (APIs) provided by social media platforms such as Twitter, Facebook and news services (Batrinca & Treleaven, 2015). An API is a software intermediary that allows programmers to gather information from other websites (Pearlman, 2016). APIs can be used to gather content from social media websites. Public APIs provided by social media firms, for example, Twitter REST API or Facebook Graph API, are increasingly used by researchers to collect data for scientific studies and can be used for program evaluation where social media has been used (Janetzko, 2017, pg147). Note, however, that restrictions exist on public API's provided by social media firms, for example, the Twitter REST API only extracts tweets from the recent past (last seven days to last 24 hours) (Janetzko, 2017, pg 156).

A range of other tools developed by commercial organisations exist to collect social media data. These can provide additional access to social media data compared to what free API's can provide, such as sophisticated visualisations, social network analysis, data mining capabilities and linking in with other data sources. A listing of such commercial organisations is provided in Table 3. The tools listed in Table 3 require no prior technical and/or programming skills and provide a number of ways in which to collect social media data, both quantitative and qualitative, from a number of platforms. However, readers should be aware that many other tools exist. The data collected can be social media metrics, for example, the number of tweets from a particular user of interest on Twitter, or the number of Facebook likes for a particular post that may contain extremist messaging. These tools can also be used to extract text and comments for qualitative analysis purposes. Such tools could be used by program evaluators to gather data from social media platforms as part of their evaluation.

Table 3 Overview of social media collection tools

Tool	**Operating System**	**Download and/or access from**	**Platforms**
Archivist (subscription-based service only)	Web-based	http://archivist.visitmix.com	Twitter
Audiense (offers 14 day trial)	Web-based	https://buy.audiense.com/trial/new	Twitter
Boston University Twitter Collection and Analysis Toolkit (BU-TCAT)	Web-based	http://www.bu.edu/com/research/bu-tcat	Twitter
Chorus (free)	Windows (Desktop advisable)	http://chorusanalytics.co.uk/chorus/request_download.php	Twitter
COSMOS Project (free)	Windows MAC OS X	http://socialdatalab.net/software	Twitter
DiscoverText (offers 3 day trial)	Web-based	http://discovertext.com	Twitter Facebook Blogs Forums Online news platforms Ability to import
Echosec	Web-based	https://www.echosec.net	Instagram Twitter Foursquare Panoramio AIS Shipping Sina Weibo

			Flickr YouTube VK
Followthehashtag	Web-based	http://www.followthehashtag.com	Twitter
IBM Bluemix	Web-based	https://www.ibm.com/cloud-computing/bluemix	Twitter
Mozdeh	Windows (Desktop advisable)	http://mozdeh.wlv.ac.uk/installation.html	Twitter
Netlytic	Web-based	https://netlytic.org	Twitter, Facebook, YouTube, Instagram RSS Feed
NodeXL	Windows	http://nodexl.codeplex.com	Twitter, YouTube, Flicker
NVivo and NCapture (NVivo web browser extension)	Windows and MAC	http://www.qsrinternational.com/product	Webpages, social media, online PDFs, Ability to import
Pulsar Social	Web-based	http://www.pulsarplatform.com	Twitter, Facebook topic data, online blogs
SocioViz	Web-based	http://socioviz.net	Twitter
Trendsmap	Web-based	https://www.trendsmap.com	Twitter
Twitonomy	Web-based	http://www.twitonomy.com	Twitter
Twitter Arching Google Spreadsheet (TAGS)	Web-based	https://tags.hawksey.info	Twitter
Visibrain	Web-based	http://www.visibrain.com	Twitter
Webometric Analyst	Windows	http://lexiurl.wlv.ac.uk	Twitter (with image extraction capabilities), YouTube, Flickr, Mendeley, Other web resources

(Adapted from Ahmed, 2017; Kavanaugh et al., 2012).

7.5 How to analyse social media data?

Analysis of social media data will depend on the outcome measures in place and the type of data being collected. Analysis of social media datasets can include approaches that have elements of both quantitative and qualitative outputs.

Quantitative data drawn from social media platforms includes such measures as those presented above (e.g., frequencies of likes, views or followers). These will vary depending on the social media platform being used. Such data is presented numerically and can be analysed like other quantitative data (see *Quantitative data analysis*, pg. 17). That is, quantitative analysis of social media data can involve both descriptive and inferential statistics. Descriptive statistics can involve, for example, numbers of followers of a particular CVE social media campaign. Inferential statistics can measure the direction and strength of an association between two variables (e.g., correlations) or test whether there is an association between two variables (e.g., chi-square test). An example is to test whether there is an association between gender (if it can be identified) and the nature of comments posted around an issue (positive or negative comments towards a particular social media campaign focused on counter narratives to violent extremist messaging).

Social media data can also be qualitative and can be analysed using the same methods as described previously (see *Qualitative data analysis, pg. 33*). As such, secondary qualitative data can be coded into themes or categories using thematic analysis approaches. For example, content analysis of posts against a hashtag that is associated with a CVE campaign can identify themes to allow indexing or categorisation of posts, which then allows greater understanding of the most prominent opinions about the campaign.

Advances in the automation of analytical approaches are applicable to social media data given the large volume of information that is typically associated with social media datasets. This can include sentiment analysis that uses different algorithms to identify and extract subjective information in source materials (Batrinca & Treleaven, 2015: 90; Bartlett & Reynold, 2015). Tools such as MonkeyLearn or Python are packages for performing sentiment analysis, although there are many similar tools that allow you to perform similar functions. Other analytical approaches include Social Network Analysis (SNA), which aims to map and describe the nature, intensity and frequency of networks of social ties. Social network analysis can be conducted on social media datasets for activities including readership or engagement in blogs, news story discussion boards or social media platforms.

7.6 How to manage social media data?

Social media datasets may be very large, and there are considerable issues around data management and how the data will be stored. Data cleaning is also a key consideration. Issues include missing data due to users not providing specific profiling information, or data errors such as typos, or the use of colloquial language that may not be picked up when extracting data on a particular theme. These are often identified during the analysis phase and may require you to go back into the data and perform another wave of processing. Duplication of data is also an issue, particularly when looking at Twitter data when a comment or post may have been re-tweeted hundreds or thousands of times, in which case it is picked up each time during data collection. Other considerations include inconsistency in the units/formats of elements of the data, with these varying across location (e.g., if time is recorded differently in different countries). Finally, data needs to be extracted into a user-friendly format e.g., CSV file (Brown, Soto-Corominas, Suarez & de la Rosa, 2017).

7.7 Hypothetical social media data collection and analysis scenario

In this hypothetical scenario, social media data is used to inform and encourage engagement with a community-based CVE program aimed at addressing intercommunal tensions. The program itself is implemented in the community and includes three community events including community meetings, forums, workshops and activities over a two-month period. The use of social media is an important component to foster engagement with the program.

Selecting the research design

The CVE initiative was developed and implemented by a community-based organisation. The organisation has a website where programs are given a dedicated webpage and contain further details, as well as a social media profile via Facebook and Twitter. The program and its events are advertised via the organisation's various social media platforms and on its website, which aim to inform and encourage engagement with the program. The initiative was advertised via the website and on social media twice prior to the first event. This included a timetable of the three events occurring over the coming two months. One week before each event, a reminder advertising post was sent out via social media, and a final reminder post sent out one day before each event. Following each event, a 'thank-you for your attendance' post was sent out, asking for feedback on the event. Each event post included a directive to the program's website page. As part of the evaluation, social media platforms were used to advertise an online survey to gauge attendees' perceptions of each event, the initiative itself, and public perceptions of intercommunal tensions in their particular community. Data is collected at the conclusion of the initiative, and the sample is those who engaged with the program's events via social media.

Data collection

Data is collected on each event directly from the Facebook and Twitter event reminder and thank-you posts at the conclusion of each event. This can be downloaded directly from the Facebook or Twitter page or can be manually copied and pasted from Facebook or Twitter into your preferred document for analysis, either manually or through qualitative analysis software such as NVivo. Alternatively, NCapture (a web browser extension for NVivo) could be used, which allows you to import posts from a Facebook or Twitter page to then be analysed in NVivo. The comments need to be de-identified as part of the data collection process. This can be done prior to importing the data into NVivo or, once in NVivo, this identifying information can be deleted.

Data analysis

Thematic analysis could be conducted on each event to identify recurring themes and the nature of the positive or negative reactions to the posts. Other measures that could be included are the number of overall comments in response to the program advertising or to each event posting (including the online survey post), the number of people who commented more than once to a particular event posting, the number of people who commented on more than one post, the number of people 'tagged' in comments (thereby sharing the event with other friends), the number of click-throughs to the website via the social media channel, and the number of times that the hashtag (in Twitter) was used. The location and other profiling information of those engaging with the campaign via social media can also be gathered. A comparison of such metrics between Facebook and Twitter could be carried out to assess the effectiveness of each in informing and promoting engagement with the program.

8. References

Ahmed, W. (2017, May 8). Using Twitter as a data source: An overview of social media research tools (updated for 2017) [Blog post]. Retrieved from http://blogs.lse.ac.uk/impactofsocialsciences/2017/05/08/using-twitter-as-a-data-source-an-overview-of-social-media-research-tools-updated-for-2017/.

Bartlett, J., & Reynolds, L. (2015). The State of the Art 2015: A literature review of social media intelligence capabilities for counter-terrorism. Retrieved from Demos website: https://www.demos.co.uk/wp-content/uploads/2015/09/State_of_the_Arts_2015.pdf.

Batrinca, B., & Treleaven, P.C. (2015). Social media analytics: a survey of techniques, tools and platforms. AI & Society, 30 (1), 89-116. doi: https://doi.org/10.1007/s00146-014-0549-4

Brown, D.M., Soto-Corominas, A., Suarez, J.L., & de la Rosa, J. (2017). Overview – The Social Media Data Processing Pipeline, in L. Sloan & A. Quan-Haase. (Eds.), The SAGE handbook of social media research methods (pp125-145). London: SAGE

Cheng, H.G. & Phillips, M.R. (2014). Secondary analysis of existing data: opportunities and implementation, Shanghai Archives of Psychiatry, 26 (6): 371-375. doi: 10.11919/j.issn.1002-0829.214171

Cho, Y. (2011). Intercoder Reliability, in P.L. Lavrakas (Ed.), Encyclopaedia of Survey Research Methods, (pp345-346). Thousand Oaks, CA: Sage Publications doi: http://dx.doi.org/10.4135/9781412963947

Cohen, D., & Crabtree, B. (2006). Qualitative Research Guidelines Project. Retrieved from: www.qualres.org/index.html.

Conway, M. (2017). Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research. Studies in Conflict & Terrorism, 40 (1), 77-98. doi: https://doi.org/10.1080/1057610X.2016.1157408

Cowling, D. (2017). Social Media Statistics Australia – November 2017. Retrieved from https://www.socialmedianews.com.au/social-media-statistics-australia-november-2017/

Doojse, B., Loseman, A., & van den Bos, K. (2013). Determinants of radicalization of Islamic youth in the Netherlands: Personal uncertainty, perceived injustice, and perceived group threat. Journal of Social Issues, 69(3), 586-604. doi: 10.1111/josi.12030

Feddes, A. R., & Gallucci, M. (2015). A literature review on methodology used in evaluating effects of preventive and de-radicalisation interventions. Journal for Deradicalization, 5, 1-27.

Fink, N. C., Romaniuk, P., & Barakat, R. (2013). Evaluating countering violent extremism programming: Practice and progress. In Final Report of Symposium on Measuring the Effectiveness of CVE Programming, Global Counterterrorism Forum.

Fox, J., Murray, C., & Warm, A. (2003). Conducting research using web-based questionnaires: Practical, methodological, and ethical considerations. International Journal of Social Research Methodology, 6(2), 167-180 doi: https://doi.org/10.1080/13645570210142883

Heale, Ra & Twycross, A. (2015). Validity and reliability in quantitative research. Evidence-Based Nursing. 18 (3). 66-67. doi: http://dx.doi.org/10.1136/eb-2015-102129

Heaton, J. (2008). Secondary analysis of qualitative data, in P. Alasuutari, L. Bickman & J. Brannen (Eds). The SAGE handbook of social research methods (pp506-519). London: Sage Publications

Helmus, T., Matthews, M., Ramchand, R., Beaghley, S., Stebbins, D., Kadlec, A., Brown, M.A., Kofner, A., & Acosta, J. (2017). RAND Program Evaluation Toolkit for Countering Violent Extremism. Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/tools/TL243.html.

Hodgkin, S. (2011). Participating in social, civic, and community life: Are we all equal? Australian Social Work, 64(3), 245-265. doi: https://doi.org/10.1080/0312407X.2011.573798

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. Public Opinion Quarterly, 67, 79–125.doi: https://doi.org/10.1086/346010

Horgan, J. (2008). From profiles to pathways and roots to routes: Perspectives from psychology on radicalization into terrorism. The ANNALS of the American Academy of Political and Social Science, 618(1), 80-94. doi: https://doi.org/10.1177/0002716208317539

Horgan, J., & Braddock, K. (2010). Rehabilitating the terrorists?: Challenges in assessing the effectiveness of de-radicalization programs. Terrorism and Political Violence, 22(2), 267-291. doi: https://doi.org/10.1080/09546551003594748

Janetzko, D. (2017). The role of APIs in Data Sampling from Social Media, in L. Sloan & A. Quan-Haase. (Eds.), The SAGE handbook of social media research methods (pp146-160). London: SAGE

Janghorban, R., Roudsari, R. L., & Taghipour, A. (2014). Skype interviewing: The new generation of online synchronous interview in qualitative research. International Journal of Qualitative Studies on Health and Well-Being, 9 (1). doi: https://doi.org/10.3402/qhw.v9.24152

Johnson, B. D., Dunlap, E., & Benoit, E. (2010). Structured Qualitative Research: Organizing "Mountains of Words" for Data Analysis, both Qualitative and Quantitative. Substance Use & Misuse, 45(5), 648–670. doi: https://doi.org/10.3109/10826081003594757

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. Business horizons, 53(1), 59-68. doi: https://doi.org/10.1016/j.bushor.2009.09.003

Kavanaugh, A. L., Fox, E. A., Sheetz, S. D., Yang, S., Li, L. T., Shoemaker, D. J., Natsev, A., & Xie, L. (2012). Social media use by government: From the routine to the critical. Government Information Quarterly, 29(4), 480-491. doi: https://doi.org/10.1016/j.giq.2012.06.002

Koehler, D. (2017). Understanding Deradicalization: Methods, Tools and Programs for Countering Violent Extremism. Oxon: Routledge.

Leavy, P. (2017). Research design: quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches. New York: The Guilford Press

Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. Journal of Family Medicine and Primary Care, 4(3), 324–327. doi: http://doi.org/10.4103/2249-4863.161306

Lin, V., Maxwell, K., & Forry, N. (2017). Considerations in Preparing to Analyze Administrative Data to Address Child Care and Early Education Research Questions. OPRE Research Brief # 2017-18. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Marshall, C., & Rossman, G. B. (2011). *Designing qualitative research* (5th ed.). London: Sage.

Marshall, M. N. (1996). Sampling for qualitative research. Family practice, 13(6), 522-526. doi: https://doi.org/10.1093/fampra/13.6.522

Mayr, P. & Weller, K. (2017). Think before you collect: Setting up a data collection approach for social media studies, in L. Sloan & A. Quan-Haase. (Eds.), The SAGE handbook of social media research methods (pp107-124). London: Sage Publications

National Health and Medical Research Centre (NHMRC) (2007). Australian Code for the Responsible Conduct of Research. Retrieved from the National Health and Medical Research Council website: https://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/r39.pdf

Neuman, W. L. (2007). Basics of social research: Qualitative and quantitative approaches (2nd ed.). Boston, MA: Pearson.

Newington, L., & Metcalfe, A. (2014). Factors influencing recruitment to research: qualitative study of the experiences and perceptions of research teams. BMC medical research methodology, 14(1), 10. doi: https://doi.org/10.1186/1471-2288-14-10

Patton, M. Q. (2002). Qualitative Research and Evaluation Methods (3rd Ed). Thousand Oaks, CA: Sage Publications.

Pearlman, S. (2016, September 7). What are APIs and how do APIs work? [Blog post] Retrieved from: https://blogs.mulesoft.com/biz/tech-ramblings-biz/what-are-apis-how-do-apis-work/

Pressman E., Duits, N., Rinne, T & Flockton, J. (2016). VERA–2R Violence Extremism Risk Assessment – version 2 Revised: A structured professional judgment approach, Nederlands Institut voor Forensische Psychiatrie en Psychologie.

Rabiee, F. (2004). Focus-group interview and data analysis. Proceedings of the Nutrition Society, 63(4), 655-660. doi: https://doi.org/10.1079/PNS2004399

Rothbard, A. (2015). Quality issues in the use of administrative data records, in J. Fantuzzo & D. Culhane (Eds.) Actionable Intelligence (pp. 77-103). New York: Palgrave Macmillan.

Saunders, M., Lewis, P., & Thornhill, A. (2012). Research methods for business students (6th ed.). Harlow: Pearson Education

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. BMC medical research methodology, 8(1), 45. doi: https://doi.org/10.1186/1471-2288-8-45

Tongco, M. D. C. (2007). Purposive sampling as a tool for informant selection. Ethnobotany Research and Applications, 5, 147-158.

Waldman, S., & Verga, S. (2016). Countering violent extremism on social media. Defence Research and Development Canada. Retrieved from http://cradpdf.drdc-rddc.gc.ca/PDFS/unc262/p805091_A1b.pdf

Walliman, N. (2017). *Research methods: The basics*. New York: Routledge.

Wilner, A., & Rigato, B. (2017). The 60 Days of PVE Campaign: Lessons on Organizing an Online, Peer-to-Peer, Counter-radicalization Program. Journal for Deradicalization, (12), 227-268.

Woollard, M. (2014). Administrative data: Problems and benefits. A perspective from the United Kingdom, in A. Duşa, D. Nelle, G. Stock & G. Wagner (Eds), Facing the Future: European Research Infrastructures for the Humanities and Social Sciences, (pp. 49-60) Berlin: SCIVERO Verlag