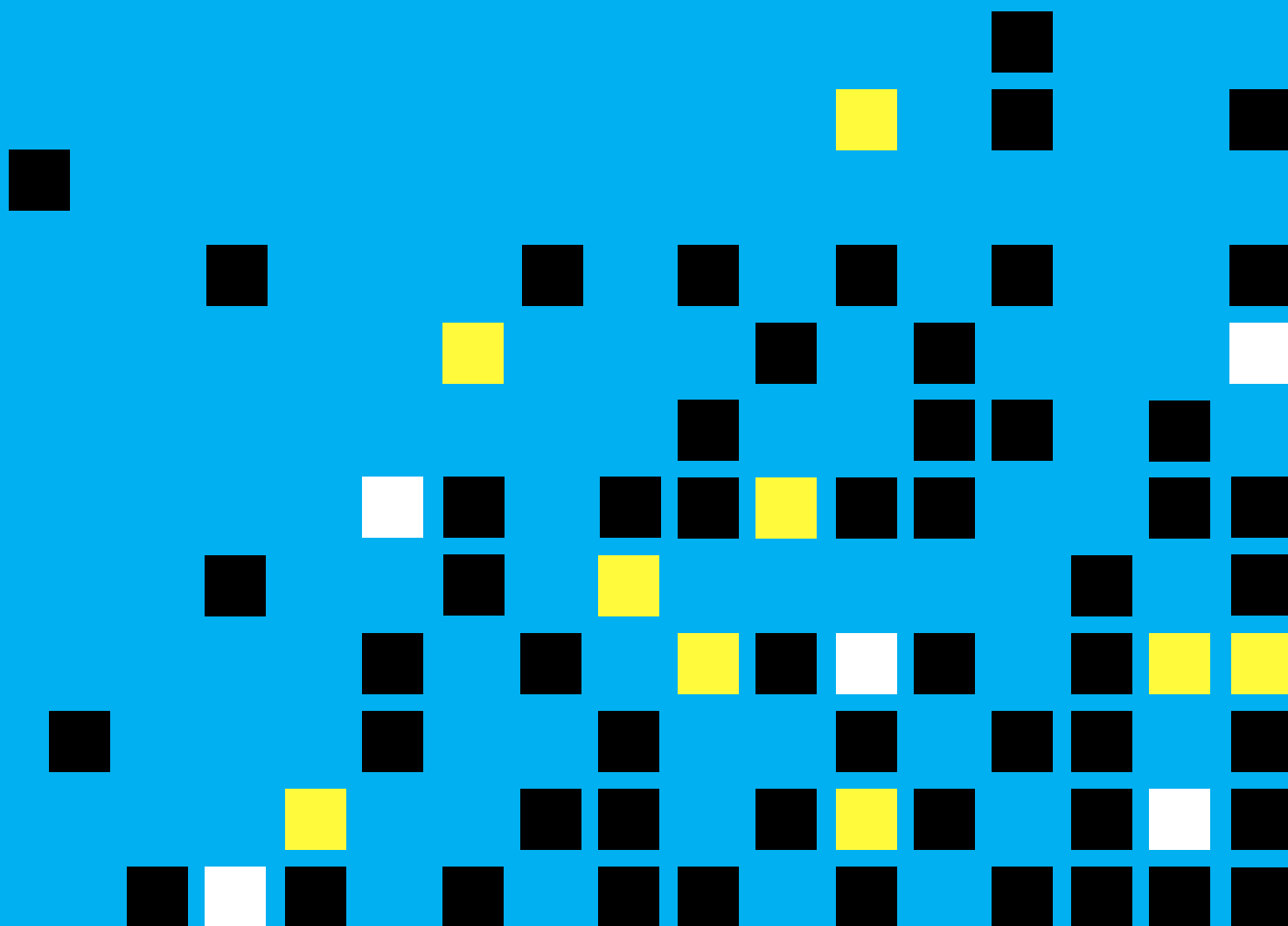


HOW TO MEASURE OUR IMPACT ON LONELINESS



About this Manual

This manual provides an overview of methods used for research and evaluation related to loneliness and social isolation, including a presentation and review of common social connection measures.

About the Social Bubble Project

The Social Bubble project is a growing community of social health researchers, advocates, friends, and neighbours tackling the loneliness pandemic. Join our campaign working to cut loneliness in half by 2030 and end the loneliness pandemic by visiting our website at www.socialbubbleproject.ca.

Need more help?

If you feel like something is missing, let us know and we will include in in the next edition. Also, feel free to email team@socialbubbleproject.ca for expert social consultation in designing or implementing your research or evaluation project.



SOCIALBUBBLE
P R O J E C T

Table of Contents

Introduction.....	2	Getting Responses.....	21
Designing Your Study	3	Analyzing Data	23
Evaluation Research.....	3	Preparing Your Data	23
Epidemiological Research.....	5	Descriptive Statistics	23
Measuring your Impact.....	6	Bivariable Statistics	25
UCLA Loneliness Scale	7	Qualitative Evaluations.....	28
De Jong Gierveld Loneliness Scale	9	Importance of Qualitative Approaches..	28
Single-item Loneliness Measures.....	10	Types of Qualitative Studies	29
Stephoe Social Isolation Index.....	11	Interview Guides.....	31
Lubben Social Network Scale.....	12	Analyzing Qualitative Data	32
Oslo Social Support Scale	13	Organizing Data.....	32
MOS Social Support Scale	13	Organizing Concepts & Building	
Perceived Social Support Scale	14	Themes	33
Social Networking Site Usage	15	Choosing a coding technique	35
Quantitative Evaluations.....	17	Codebooks	36
Sampling Methods	17	Ensuring Reliability and Validity	37
Engaging Participants	20	Conclusion & Recommendations.....	38
Ethics Approval	20		

Suggested Citation

Card & Marziali (2021) "Research and Evaluation Manual: How to measure our impact on loneliness." The Social Bubble Project.

Part 1

Introduction

The Social Bubble Project has joined leading social health organizations and advocacy groups across the globe to cut loneliness in half by 2030. There are no one-size-fits all-solution to loneliness. It is likely that hundreds of small projects and programs will be needed to support the social health of our communities, not to mention the everyday positive actions that we need to do to keep our neighbourhoods and communities connected.

If you are a community organization working to reduce loneliness, you don't want to waste time or money spinning your wheels on programming that is not effective. Even if loneliness isn't the chief focus of your programs, it's very likely that it could have some beneficial effect on loneliness – particularly if it is socially engaging and participatory. For these reasons and more, it is so important to evaluate whether we are making progress towards our 2030 goal to cut chronic loneliness in half.

However, not everyone has evaluation experience. This can be a major barrier to implementing evaluations. That's why this manual exists. We believe that with the right tools anyone can conduct an evaluation of their programs. With collaboration from academic institutions, evaluations can typically be conducted for a low cost – ensuring that your valuable programming dollars are not redirected away from the essential work you are doing.

In this manual we provide a simple overview of study designs (Part 2), an overview of quantitative measures of loneliness (Part 3), practical content needed to implement quantitative (Part 4) and qualitative (Part 5) evaluations, and conclude with our recommendations for research and evaluation related to loneliness (Part 6).



With the emergence of COVID-19, it has never been more important to measure loneliness.

Part 2

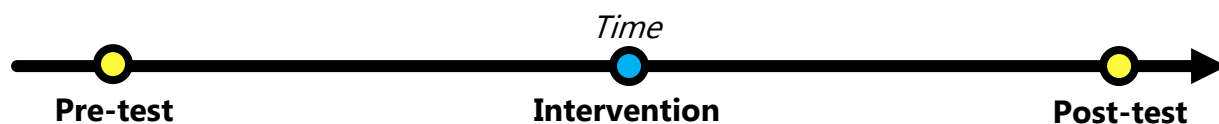
Designing Your Study

Generally speaking, there are two types of research that can be used to study the impact of programs on loneliness and social isolation: (1) Evaluations Research and (2) Epidemiological Research. The first focuses on the **outcomes** of a program among a relatively small number of individuals. The second focuses on the **impact** of a program at the population level. The choice between these two programs is logically related to the program itself. If you are conducting a clinical program (e.g., counselling), an evaluation is likely more appropriate. Whereas, if you are conducting a public health program (e.g., social media campaign), an epidemiological study is probably a better choice.

Evaluation Research

A wide range of programs are needed to address loneliness and social isolation. As such, there are many things to consider when planning an evaluation of your program. This part of the manual provides a brief overview of common and practical study design considerations.

- 1.1. To measure the impact of your program among your participants, you can use a study design known as the **pre-test/post-test design**.



As you can see from the image above, the pre-test/post-test design involves administering a measure – often a validated scale – prior to and after delivery of a program. In Part 2 of this manual, we will review common pre-test/post-test measures used to evaluate the impact of a program on loneliness and social isolation.



A major difficulty of pre-test/post-test designs is that some individuals may be **lost to follow-up**. This is sometimes referred to as **attrition**. To reduce loss-to-follow-up it is recommended you collect multiple pieces of contact information from your participants. These may include:

- phone numbers,
- email addresses,
- social media handles, and/or
- mailing addresses.

Below are some additional considerations that may want to consider as you plan your evaluation:

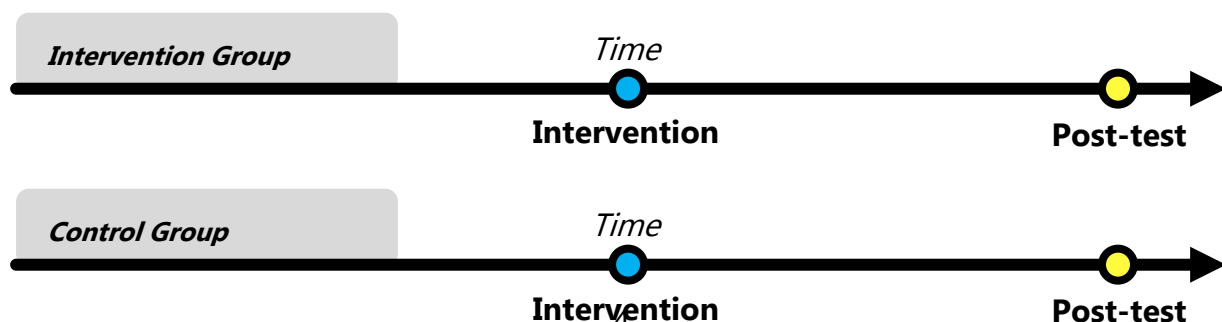
Multiple post-tests. Many pre-test/post-test design studies include multiple post-tests in order to assess whether the desired impact is maintained over time. It is not uncommon for the effect of a program to dissipate as time goes on – especially for behavioral and psychosocial programs. If multiple post-tests are administered, they are typically administered immediately after and at least 3-months following the program.

Control groups. Some pre-test/post-test designs also include a comparison group of individuals that did not receive the program. A comparison group, often called a **control group**, helps you establish that your program was the actual cause of the change in loneliness. To whatever extent possible, the comparison group should be the same as the **program group**, with the exception that they are not exposed to your program – though they may be exposed to a placebo, attention-only program, or the existing gold standard program.

Matched controls. Sometimes, the controls in your study are “matched” to individuals who received the program. Typically matching should be used for variables that are likely to confound your results. For example, you might administer a short questionnaire at the beginning of your study to identify the age group, sex, and income level of your participants and use these characteristics to identify matched-controls with similar demographic characteristics. Matching can be 1:1, meaning that you have one control for every participant enrolled in your study or can be 1:n meaning that you have multiple controls for every participant you enroll. While having more matches increases the statistical power of your study, it is often the case that the first matches are a better match than the second, third, fourth, or fifth matches. As a result, increasing the number of matches is not always preferred if they are not “good” matches.

Randomization. Another tool used to increase the quality of a pre-test/post-test design studies is to randomly assign interested participants to the program and control groups. Individuals in the control group may receive an alternative suitable program or be provided the program at a later date. By randomizing participants to a program, the risk of confounding bias is evenly distributed across each group.

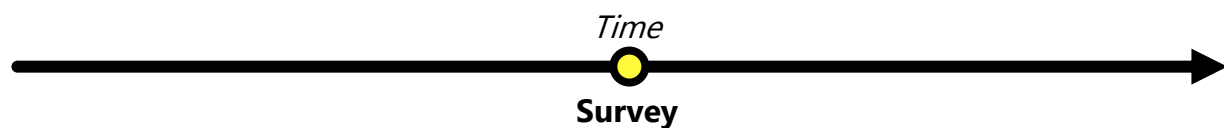
- 1.2. When a control group is used, researchers sometimes choose to conduct a **post-test only design**. This design differs from the previously discussed design because there is no pre-test. The diagram below shows the design diagram for a post-test only design with a control group. However, a control group is not necessarily required – though highly recommended.



Epidemiological Research

You may already be familiar with these, but it is also helpful to be aware of epidemiological research designs that are widely used in the study of loneliness and social isolation.

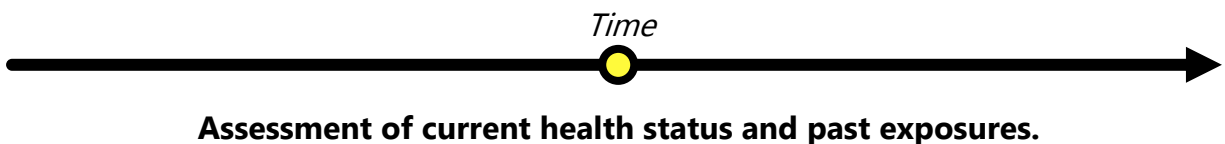
- 1.3.** A **cross-sectional study** typically consists of a survey administered at a single point in time. These designs are widely used to understand the correlation between two concepts or to understand the prevalence of a health related state or event. For example, a cross-sectional study could be used to examine whether the prevalence of loneliness differs between men, women, and non-binary people. With cross-sectional studies there is no implication of causation. In other words, you are not able to say that gender causes loneliness (or that loneliness causes gender, for that matter).



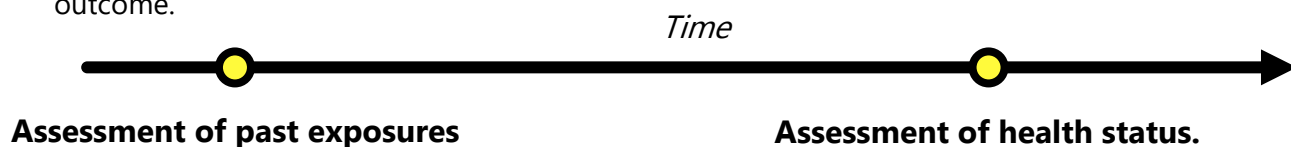
- 1.4.** A **serial cross-sectional study** typically consists of multiple cross-sectional surveys conducted over time. It is typically used to assess the prevalence of a health related state or event and identify whether the prevalence is changing. Each cross-section should be independent of the other cross-sections, meaning participants should only participate once. Frequently cross-sectional surveys ask whether participants have participated before and if they have, they are ineligible to participate. Each cross-section is also asked the same questions as previous cross-sections and recruitment methods should be as consistent as possible between surveys.



- 1.5.** A **case-control study** typically consists of a study in which participants have been recruited based on a specific outcome and are then asked questions about past exposures that might have contributed to their current health status. Differences in exposure are compared between whether they were a "case" (i.e., were affected by the health related state or event being studied) or a "control" (i.e., were not affected by the health related state or event being studied).



- 1.6.** A **cohort study** typically consists of a study in which participants have been recruited based on whether they have been exposed to a potential risk and are followed over time to assess whether they develop an outcome. For example, you might measure loneliness and assess whether it is associated with increased risk of premature death or a specific adverse health outcome.



Part 3

Measuring your Impact

In conducting an evaluation of your program, there are a wide variety of outcomes for you to consider, you may even choose to measure multiple outcomes. In choosing these, it is helpful to recognize that social lives are complex. For our purposes, loneliness is defined as the subjective dysphoric feeling associated with a deficiency in the quantity or quality of social and emotional relationships. It can be chronic (i.e., people experience it for a long time) or transient (i.e., it is merely a passing experience). The literature differentiates loneliness from closely related concepts, including social isolation (i.e., the objective state of being disconnected from or having few ties with a social network), social network size (i.e., the number of social ties with others), social support (i.e., the ability to rely on one's social network for emotional resources that can help one cope with stressful experiences and situations), social inclusion (i.e., the extent to which one is able to participate in social and communal activities), and community connectedness (i.e., a sense that you belong to a community or social group). This section reviews measurement these constructs, however, constructs can be difficult to measure and it is not always clear which concepts are measured b which scales.

In choosing a measure, a variety of considerations should be made, including:

- Is this measure appropriate for my population?
- Does the measure assess subjective experience (e.g., how they feel) or objective descriptions about their daily lives (e.g., what they do)?
- Does the measure assess the quantity of relationships or the quality and function of relationships?



“

When we measure loneliness,
we're communicating that
social connection matters to
us as much as it does to those
experiencing it.

”

UCLA Loneliness Scale

The *UCLA Loneliness Scale* is the most widely used measure for loneliness. There is a 20-item version of the scale and a 3-item version of the scale. The 20-item version is scored on a 4-point Likert Scale (Never, Rarely, Sometimes, Often) and the 3-item version is scores on a 3-point Likert Scale (Hardly Ever, Some of the Time, Often). Higher scores indicate greater loneliness and lower scores indicate less loneliness.

3-item UCLA Loneliness Scale

Please indicate how often each of the statements below is descriptive of you.

		Hardly Ever	Some of the time	Often
2	How often do you feel that you lack companionship?	0	1	2
11	How often do you feel left out?	0	1	2
14	How often do you feel isolated from others?	0	1	2

Scoring for the 3-item Version

The shortened UCLA Loneliness Scale is the sum of items 2, 11, and 14. There are no subscales. Scores range from 0 to 6.

20-item UCLA Loneliness Scale

Please indicate how often each of the statements below is descriptive of you.

		Never	Rarely	Sometimes	Often
1	How often do you feel that you are "in tune" with the people around you?	3	2	1	0
2	How often do you feel that you lack companionship?	0	1	2	3
3	How often do you feel that there is no one you can turn to?	0	1	2	3
4	How often do you feel alone?	0	1	2	3
5	How often do you feel part of a group of friends?	3	2	1	0
6	How often do you feel that you have a lot in common with the people around you?	3	2	1	0
7	How often do you feel that you are no longer close to anyone?	0	1	2	3
8	How often do you feel that your interests and ideas are not shared by those around you?	0	1	2	3
9	How often do you feel outgoing and friendly?	3	2	1	0
10	How often do you feel close to people?	0	1	2	3
11	How often do you feel left out?	0	1	2	3
12	How often do you feel that your relationships with others are not meaningful?	0	1	2	3
13	How often do you feel that no one really knows you well?	0	1	2	3
14	How often do you feel isolated from others?	0	1	2	3
15	How often do you feel that you can find companionship when you want it?	3	2	1	0
16	How often do you feel that there are people who really understand you?	3	2	1	0
17	How often do you feel shy?	0	1	2	3
18	How often do you feel that people are around you but not with you?	0	1	2	3
19	How often do you feel that there are people you can talk to?	3	2	1	0
20	How often do you feel that there are people you can turn to?	3	2	1	0

Scoring for the 20-item Version

The UCLA Loneliness Scale (Version 3) is the sum of items 1 – 20. There are no widely-used subscales. Scores range from 0 to 60.

De Jong Gierveld Loneliness Scale

The *De Jong Gierveld Loneliness Scale* is the second most widely used measure for loneliness. There is an 11-item version of the scale and a 6-item version of the scale. Both scales are typically scored using 3-response options (Yes, More or Less, No). Each of the existing versions consist of two sub-scales: The *Emotional Loneliness Subscale*, which measures the absence of intimate relationships, and *The Social Loneliness Subscale*, which measures the absence of a broader social network. On all versions and subscales, lower scores indicate less loneliness and higher scores indicate greater loneliness.

Please indicate for each of the statements, the extent to which they apply to how you feel.

		Yes	More or Less	No
1	There is always someone I can talk to about my day-to-day problems	0	1	1
2	I miss having a really close friend	1	1	0
3	I experience a general sense of emptiness	1	1	0
4	There are plenty of people I can lean on when I have problems	0	1	1
5	I miss the pleasure of the company of others	1	1	0
6	I find my circle of friends and acquaintances too limited	1	1	0
7	There are many people I can trust completely	0	1	1
8	There are enough people I feel close to	0	1	1
9	I miss having people around me	1	1	0
10	I often feel rejected	1	1	0
11	I can call on my friends whenever I need them	0	1	1

Scoring for the 11-item Version

The *Emotional Loneliness Subscale Score* is the sum of questions 2, 3, 5, 6, 9, and 10. This sub-scale consists of 6 items. Scores range from 0 to 5. The *Social Loneliness Subscale Score* is the sum of questions 1, 4, 7, 8, and 11. This sub-scale consists of 5 items. Scores range from 0 to 6. The *Overall Loneliness Score* is the sum of the *Emotional Loneliness Sub-Scale Score* and the *Social Loneliness Subscale Score*. Scores range from 0 to 11.

Scoring for the 6-item Version

The *Emotional Loneliness Subscale Score* is the sum of questions 3, 9, and 10. This sub-scale consists of 3 items. Scores range from 0 to 3. The *Social Loneliness Subscale Score* is the sum of questions 4, 7, and 8. This sub-scale consists of 3 items. Scores range from 0 to 3. The *Overall Loneliness Score* is the sum of the *Emotional Loneliness Sub Subscale Score* and the *Social Loneliness Subscale Score*. Scores range from 0 to 6.

Single-item Loneliness Measures

A wide range of single-item loneliness measures have been used. If you cannot use a validated measure, using one of the following items is recommended:

- How often do you feel lonely?
[1 = Never, 2 = Seldom, 3 = Sometimes, 4 = Often, 5 = Always]
- How often do you feel lonely?
[1 = Hardly ever or never, 2 = Some of the time, 3 = Often]
- Are you
[0 = Never Lonely, 1 = Lonely at times, 2 = Very Lonely]
- During the past week, have you felt lonely
[0 = Rarely or none of the time (e.g. less than 1 day), 1 = Some or a little of the time (e.g. 1-2 days), 2 = Occasionally or a moderate amount of time (e.g. 3-4 days), 3 = All of the time (e.g. 5-7 days)]

You can see that each of the above differ slightly in what is being measured or how. As such, different measures of loneliness will provide somewhat different pictures of both the prevalence of loneliness and the characteristics of people who are lonely. Choosing the best question to use is often a matter of identifying the specific needs of your study. Consulting with program participants by piloting questionnaires can help you identify whether questions are appropriate.

Retrospective Measures. It is important to note that the loneliness measures described above can be used in pre-test/post-test and post-test only designs. However, if you are conducting a post-test only design without a control group, most measures are uninformative because you cannot compare results to either the pre-test results or the results provided by the control group. In these situations, measures must have a retrospective component that asks the participants to consider how their loneliness has changed. This sort of recall is highly vulnerable to a wide range of biases. However, sometimes using a reflective measure is the only feasible option. In these cases, the following measures are recommended:

- Comparing how you feel now to how you felt _____ ["Six months ago" or "before you participated in the program"], do you feel more or less lonely?
[1 = Much more lonely, 2 = Somewhat more lonely, 3 = About the same, 4 = Somewhat less lonely, 5 = Much less lonely]





Special Considerations for Surveying Youth

Measuring loneliness among children is similar to measuring it in adults. However, the language used should be simpler and should not assume children understand fully the types of social and emotional concepts that adults think of when they think of loneliness. Young people may also be less likely to complete long questionnaires, so shorter instruments are preferred. We recommend the following items for use among children:

- How often do you feel that you have no one to talk to?
[1 = Hardly ever or never, 2 = Some of the time, 3 = Often]
- How often do you feel left out?
[1 = Hardly ever or never, 2 = Some of the time, 3 = Often]
- How often do you feel alone?
[1 = Hardly ever or never, 2 = Some of the time, 3 = Often]
- How often do you feel lonely?
[1 = Never, 2 = Hardly ever, 3 = Occasionally, 4 = Some of the time, 5 = Often/Always]

Steptoe Social Isolation Index

The Steptoe Social Isolation Index measures social isolation on a five-point scale ranging from 0 (Not Socially Isolated) to 5 (Socially Isolated). Scores of two or more are defined as being socially isolated. There are no subscales.

- Are you currently living with a spouse or partner?
[0 = Yes, 1 = No]
- Do you have kids with whom you see, talk to, text, email, or write to at least once a month?
[0 = Yes, 1 = No]
- Do you have other family members with whom you see, talk to, text, email, or write to at least once a month?
[0 = Yes, 1 = No]
- Do you have other friends with whom you see, talk to, text, email, or write to at least once a month?
[0 = Yes, 1 = No]



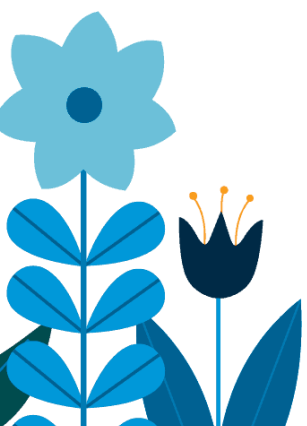
Lubben Social Network Scale

The Lubben Social Network scale measures the quantity and quality of social relationships and is variably considered a measure of social support or social isolation. There are a range of versions of the Lubben Social Network Scale. We present the 6-item measure here as it is an appropriate length and quality for most purposes.

	None	One	Two	Three to four	Five thru eight	Nine or more
<i>Considering the people to whom you are related, either by birth or marriage...</i>						
1 How many relatives do you see or hear from at least once a month?	0	1	2	3	4	5
2 How many relatives do you feel at ease with that you can talk about private matters?	0	1	2	3	4	5
3 How many relatives do you feel close to such that you could call on them for help?	0	1	2	3	4	5
<i>Considering all of your friends, including those who live in your neighborhood...</i>						
4 How many of your friends do you see or hear from at least once a month?	0	1	2	3	4	5
5 How many friends do you feel at ease with that you can talk about private matters?	0	1	2	3	4	5
6 How many friends do you feel close to such that you could call on them for help?	0	1	2	3	4	5

Scoring for the 6-item Lubben Social Network Scale

- The *Lubben Social Network Scale* is the sum of items 1 - 6. Scores range from 0 to 30. Individuals with scores of 11 or lower are considered "at risk for social isolation."
- Two *subscales* can be calculated: Items 1 – 3 comprise a family subscale and items 4 – 6 comprise a friends subscale. Scores range from 0 to 15. Individuals with scores of 6 or lower are considered "at risk for social isolation."



Oslo Social Support Scale

The Oslo Social Support Scale is a short 3-item measure assessing social support. The instrument is scored as the sum of each question. Scores range from 3 (Poor Social Support) to 14 (Strong Social Support). Some authors have used scores of 3 – 8 to represent “poor support”, 9 – 11 to represent “moderate social support”, and 12 – 14 to represent “strong support.”

- How many people are so close to you that you can count on them if you have great personal problems?
[1 = “none”, 2 = “1–2”, 3 = “3–5”, 4 = “5+”]
- How much interest and concern do people show in what you do?
[1 = “none”, 2 = “little”, 3 = “uncertain”, 4 = “some”, 5 = “a lot”]
- How easy is it to get practical help from neighbors if you should need it?
[1 = “very difficult”, 2 = “difficult”, 3 = “possible”, 4 = “easy”, 5 = “very easy”]

Medical Outcomes Study Social Support Scale

The 6-item Medical Outcomes Study Social Support Scale is a measure of social support. Multiple versions of the scale are available. The instrument is scored as the sum of each question. Scores range from 6 (Poor Social Support) to 30 (Strong Social Support).

If you needed it, how often is someone available...

		None of the time	A little of the time	Some of the time	Most of the time	All of the time
1	...to help you if you were confined to bed?	1	2	3	4	5
2	...to take you to the doctor if you need it?	1	2	3	4	5
3	...to share your most private worries and fears?	1	2	3	4	5
5	...to turn to for suggestions about how about personal problems?	1	2	3	4	5
4	...to do something enjoyable with?	1	2	3	4	5
6	...to love and make you feel wanted?	1	2	3	4	5

Multidimensional Scale of Perceived Social Support

The Multidimensional scale of perceived social support consists of 12 items scores on a scale of 1 to 7 (= *Very Strongly Disagree*, 2 = *Strongly Disagree*, 3 = *Mildly Disagree*, 4 = *Neutral*, 5 = *Mildly Agree*, 6 = *Strongly Agree*, 7 = *Very Strongly Agree*). Final Scores are a sum of all items

We are interested in how you feel about the following statements. Read each statement carefully. Indicate how you feel about each statement.

- 1 There is a special person who is around when I am in need.
- 2 There is a special person with whom I can share my joys and sorrows.
- 3 My family really tries to help me.
- 4 I get the emotional help and support I need from my family.
- 5 I have a special person who is a real source of comfort to me.
- 6 My friends really try to help me.
- 7 I can count on my friends when things go wrong.
- 8 I can talk about my problems with my family.
- 9 I have friends with whom I can share my joys and sorrows.
- 10 There is a special person in my life who cares about my feelings.
- 11 My family is willing to help me make decisions.
- 12 I can talk about my problems with my friends.



Social Networking Site Usage

Social media use has become an important measure of social connectedness. The following items are recommended.

How many times per day do you visit social networking websites, on Average?

- [1 = "Less than once per day", 2 = 1-3 "times per day", 3 = 4-8 "times per day", 4 = 9-15 "times per day", 5 = "More than 15 times per day"]

In the past week, on average, approximately how much time PER DAY have you spent actively using social networking websites?

- [1 = "Less than 10 minutes per day", 2 = 10-30 "minutes per day", 3 = 31-60 "minutes per day", 4 = 1-2 "hours per day", 5 = 2-3 "hours per day", 6 = "More than 3 hours per day".]



Key References for Measurement

- Valtorta N.K., Kanaan M., Gilbody S., et al. (2016) Loneliness, social isolation and social relationships: what are we measuring? A novel framework for classifying and comparing tools.
- de Jong-Gierveld, J., & Kamphuls, F. (1985). The Development of a Rasch-Type Loneliness Scale.
- Gierveld, J. D. J., & Tilburg, T. V. (2006). A 6-Item Scale for Overall, Emotional, and Social Loneliness: Confirmatory Tests on Survey Data.
- Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure.
- Hughes, M. E., Waite, L. J., Hawkey, L. C., & Cacioppo, J. T. (2004). A Short Scale for Measuring Loneliness in Large Surveys: Results From Two Population-Based Studies.
- McWhirter, B.T. Factor analysis of the revised UCLA loneliness scale. *Current Psychology*
- Office of National Statistics. (2018). Testing of loneliness questions in surveys Overview of our loneliness question testing, methodology and findings.
- Ahmad, F., Jhajj, A.K., Stewart, D.E. et al. Single item measures of self-rated mental health: a scoping review. *BMC Health Serv Res* 14, 398 (2014).
- Steptoe, A., Shankar, A., Demakakos, P., & Wardle, J. (2013). Social isolation, loneliness, and all-cause mortality in older men and women. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5797–5801.
- Cornwell, E. Y., & Waite, L. J. (2009). Social disconnectedness, perceived isolation, and health among older adults. *Journal of health and social behavior*, 50(1), 31–48.
- Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents.
- Kocalevent, R. D., Berg, L., Beutel, M. E., Hinz, A., Zenger, M., Härter, M., Nater, U., & Brähler, E. (2018). Social support in the general population: standardization of the Oslo social support scale (OSSS-3).
- Lubben, J., Blozik, E., Gillmann, G., Iliffe, S., von Renteln Kruse, W., Beck, J. C., & Stuck, A. E. (2006). Performance of an abbreviated version of the Lubben Social Network Scale among three European community-dwelling older adult populations.
- Holden L, Lee C, Hockey R, Ware RS, Dobson AJ. Validation of the MOS Social Support Survey 6-item (MOS-SSS-6) measure with two large population-based samples of Australian women.
- Coombs, T., Nicholas, A., & Pirkis, J. (2013). A review of social inclusion measures. *Australian & New Zealand*
- Baumgartner JN, Burns JK. Measuring social inclusion--a key outcome in global mental health. *Int J Epidemiol*. 2014 Apr;43(2):354–64.
- Zimet, G.D., Dahlem, N.W., Zimet, S.G. & Farley, G.K. (1988). The Multidimensional Scale of Perceived Social Support.

Part 4

Quantitative Evaluations

Sampling Methods

In epidemiological research there is a lot of emphasis on generalizability and random sampling. Contrary to popular belief, random sampling is not always preferred. In many situations alternative sampling techniques provide more information for less cost. After all, the goal of an evaluation is to make a judgement about a program. You are rarely trying to estimate population parameters with statistics.

- **Quota sampling** uses quotas – often proportional to the population being studied – and recruit's participants to fill these quotas. Quota sampling is particularly useful when you are unable to obtain a probability sample, but you are still trying to create a sample that is as representative as possible of the population being studied.
- **Convenience sampling** involves recruiting convenient participants. For example, you might ask participants waiting in your lobby to complete a survey simply because it is easy to get in touch with them. Convenience sampling is very easy to carry out with few rules governing how the sample should be collected. Furthermore, the relative cost and time required to carry out a convenience sample are small in comparison to probability sampling techniques.
- **Self-selection sampling** is a sampling strategy in which participants opt in to a study – perhaps because they see a flier or social media post. Since the potential research subjects (or organisations) contact you this can reduce the amount of time necessary to search for appropriate units (or cases); that is, those individuals or organisations that meet the selection criteria needed for your sample.
- **Snowball sampling** is a form of chain referral sampling that involves invitations being provided through social networks and word of mouth. Snowball sampling might be useful if you are hoping to understand something about social networks or communities. Further, some populations that we are interested in studying can be hard-to-reach and/or hidden. These include populations such as drug addicts, homeless people, individuals with AIDS/HIV, prostitutes, and so forth.
- **Maximum variation sampling**, also known as heterogeneous sampling, is a purposive sampling technique used to capture a wide range of perspectives relating to the thing that you are interested in studying; that is, maximum variation sampling is a search for variation in perspectives. Participants exhibit a wide range of attributes, behaviors, experiences, incidents, qualities, situations, and so forth – providing you with greater insights from a more diverse population.

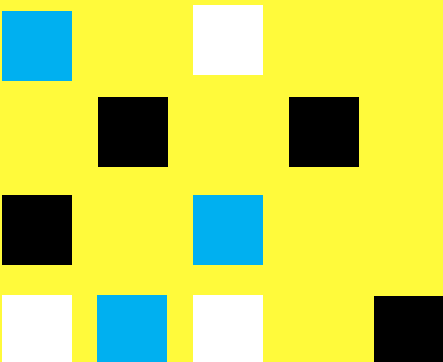


“

Contrary to popular belief, random sampling is not always preferred.

In many situations alternative sampling techniques provide more information for less cost.

”



- **Homogeneous sampling** is a purposive sampling technique that aims to achieve a homogeneous sample; that is, a sample whose units (e.g., people, cases, etc.) share the same (or very similar) characteristics or traits (e.g., a group of people that are similar in terms of age, gender, background, occupation, etc.). A homogeneous sample is often chosen when the research question that is being address is specific to the characteristics of the particular group of interest, which is subsequently examined in detail.
- **Typical case sampling** is a purposive sampling technique used when you are interested in the normality/typicality of the units (e.g., people, cases, events, settings/contexts, places/sites) you are interested, because they are normal/typical. The word typical does not mean that the sample is representative in the sense of probability sampling (i.e., that the sample shares the same/similar characteristics of the population being studied). Rather, the word typical means that the researcher has the ability to compare the findings from a study using typical case sampling with other similar samples (i.e., comparing samples, not generalizing a sample to a population).
- **Extreme (or deviant) case sampling** is a type of purposive sampling that is used to focus on cases that are special or unusual, typically in the sense that the cases highlight notable outcomes, failures or successes. These extreme (or deviant) cases are useful because they often provide significant insight into a particular phenomenon, which can act as lessons (or cases of best practice) that guide future research and practice. In some cases, extreme (or deviant) case sampling is thought to reflect the purest form of insight into the phenomenon being studied.
- **Critical case sampling** involves selecting a small number of important cases - cases that are likely to "yield the most information and have the greatest impact on the development of knowledge" Critical case sampling is a type of purposive sampling technique that is particularly useful in exploratory qualitative research, research with limited resources, as well as research where a single case (or small number of cases) can be decisive in explaining the phenomenon of interest.
- **Total population sampling** is a type of purposive sampling technique where you choose to examine the entire population (i.e., the total population) that have a particular set of characteristics (e.g., specific experience, knowledge, skills, exposure to an event, etc.).
- **Expert sampling** is where you draw your sample from experts – or key informants – in the field you're studying. It's used when you need the opinions or assessment of people with a high degree of knowledge about the study area. "Expert" doesn't necessarily have to mean highly education and skilled in a field. Individuals with lived experience are often the best and most informed *experts* on a variety of issues.

Engaging Participants

In addition to designing your recruitment strategy, you also need to figure out how you will invite participants to participate. Invitations to participate in your study provide a “first impression” of who you are and what you’re trying to do. Saying the wrong thing can thus drive people away. Remember that people like a personal touch (e.g., Dear Kiffer...) and they like to feel special (e.g., “You have been selected!” or “We’d love to hear from you!”), but they are also busy so make sure you provide the key information about what their participation will mean, when the deadline will be, what you’re aiming to show with your results, and what they will get out of the survey. Most ethics review boards will also want you to disclose any potential harms that someone might encounter during a study. How you invite participants will largely depend on how you will be collecting data from them. We will discuss the various data collection methods later in this text.

Ethics Approval

If you are working with researchers at an academic institution, you can typically secure ethics approval from their Research Ethics Board. However, if you are not collaborating with researchers, the Community Research Ethics Office provides ethics reviews for a small fee. Visit communityresearchethics.com for more information.



CENTRE FOR
COMMUNITY
BASED RESEARCH

Getting Responses

Once your survey is finalized, you next have to think about how you will distribute your survey. One of the first questions in this step is to consider when you will distribute it – particularly with regards to holidays, political cycles, life events, vacations, and so on. You can also think about who else is circulating surveys to your target population and consider whether your surveys are sent out to close to one another. Participants of some populations can become fatigued by the number of surveys they are asked to complete – and therefore opt out of completing them altogether. The second question you'll need to consider is how you are administering surveys, are you using in-person interviews? Computer assisted self-interviews (i.e., web or app-based surveys)? iPhone Surveys? Mailers? Each of these are good options with benefits and limitations. Which one you choose will be closely linked to how you are recruiting participants, who your target population is, and what information you are hoping to collect.

- During **in-person surveys**, researchers interact directly with participants and often ask questions verbally. Alternatively, they might just hand out a paper survey or a tablet on which participants can complete the survey (See CASI, below). Interviewers can extend person invites to individuals at a venue, or go door-to-door. These surveys tend to have higher response rates, decrease non-response, can be used to access hard-to reach populations (e.g., seniors), can be done in a variety of settings, allow researchers to make observations to enrich survey data, can involve all 5 senses (e.g., taste testing), and can be used to collect biological data that can accompany survey data (e.g., dried blood spot testing). Unfortunately, they tend to be more expensive, time consuming, and logistically challenging; and of course, interviewer biases can impact how participants respond to surveys.
- **Telephone surveys** are similar to in person surveys, but may be managed either digitally (using recordings and touch tone keypad responses) or manually. These also have better response rates, but are lower cost and less time-intensive than in-person surveys. As with in-person surveys, some populations are not accessible via this method (e.g., unlisted numbers, demographic differences). Further, telemarketers have given phone surveys a bad rap – introducing challenges in response bias.
- **Mailers** can also be used to reach participants with physical mailing addresses. Usually, mailers are accompanied by a letter of explanation, a self-addressed pre-paid envelope for returning the questionnaire, and the surveys themselves. In some cases, an honorarium may be pre-sent along with the survey to encourage participants to participate. Follow-up reminders (usually 2 follow-ups, 2-3 weeks apart) can be sent as postcards to encourage individuals to complete the survey. Mailers allow for larger samples, are cheaper than interviews, and allow respondents to complete the survey on their own time. However, they are more expensive than doing so online, and result in low response rates.

- **Computer-assisted self-interviews (CASI)** are like paper surveys, except they have been digitally recreated. Sometimes they are shared through a web-link or by downloading an app. SurveyMonkey and Qualtrics are the most common platforms used for creating CASI. Often times, participants are recruited on social media or from email list servers. CASI allows for a lot of customization and easily manages skip and display logic. Anonymous links can allow for sensitive data to be collected anonymously. However, if you are offering an incentive, you may want to ensure that participants are not responding more than once. If you have a set group of participants whom you might want to administer multiple surveys to you can create a panel of respondents. Panels allow you to track responses and send reminders for completion to those who have not yet completed the survey. If reminders are sent, they are often sent after about 48 hours – but you can be flexible in choosing the time between reminders. In any case, data collection is inexpensive and automatically entered into a database. Further, you can force responses to certain questions. Of course, these surveys also have lower response rates and require participants to have a computer. Further, technical errors can make responding to surveys difficult.

As discussed in each of these examples a major consideration is the response rate of your survey. Whenever possible it is a good idea to measure the response rate of your survey. This can be done by dividing the number of people who were invited to complete your survey by the number of people who actually completed your survey.

If possible it is also good to identify if there are differences between those who respond to your survey and those who do not. This is partially possible using ads-based delivery available on social media sites. High response rates are desirable because they improve the representativeness of your sample, help you get more diverse opinions, protect against nonresponse bias. When response rates are low, it may introduce bias and confounding if there are important differences in who chooses to respond.

Response rates generally fluctuate with between 20-35% of respondents completing a 10-minute survey. Higher response is influenced by the motivation of participants. Altruism and other intrinsic motivators can be leveraged by distributing the survey through trusted community leaders. Financial incentives can also encourage participation.

Regarding incentives, raffles are generally not as useful as guaranteed small incentives – but don't be afraid to be creative in choosing how you motivate individuals to participate. Note that longer surveys tend to have lower response rates – particularly for individual questions that appear towards the end of the survey. As noted throughout this section, reminders can also be used to increase response rates.

Analyzing Data

Preparing Your Data

Once you have your data – meaning you entered it into an excel sheet or downloaded it from a survey collector – the first thing you will need to do is clean your data. Data cleaning involves removing duplicate entries, removing people who clicked through the survey but provided no usable data, recoding variables into the categories you want, and dealing with extreme values or erroneous data. Well-designed surveys can eliminate the amount of data cleaning required.

As you begin to edit your data it is important that you make a copy of your original data to work from. You should never over-write your original data. You will never know when you will need it. You should also make sure that each question is represented by a single column and that no column contains multiple pieces of information. The levels in each column should match the levels you want to use in your analysis, plus any extra levels you created to account for skip/display logic, missing data, and so on. Your final data set should have no blank cells (use some indicator, such as 9999, to indicate true missingness and other indicators to indicate why observations are missing).

Most survey data will have missing data somewhere. Respondents may not answer a question (either voluntarily or accidentally) or the question may not have been asked. It is important to distinguish between these types of missingness. Participants might also give poor quality data – which you can recode as missing or as “Poor Data Quality.” This is common in text questions. For examples you might ask how many hours of community service somebody has done in the past year and they would report 10,000,000. Clearly this value is too extreme to be taken seriously. Once you have identified poor quality and missing data, you have a few different options:

- You can delete the entire row of data – a method called **list-wise** deletion.
- You can **leave it** missing and report the total number of responses separately for each variable.
- You can **impute** or assign some value (average, median, modeled value).
- You can **randomly assign** a value noting that this adds to random error.

Descriptive Statistics

Once your data is clean you will want to begin to describe your data. For numeric data you should consider where the distribution has its peak (*central location*), how widely dispersed it is on both sides of the peak (*spread*), how tightly the data are dispersed near the peak (kurtosis), and whether the distribution is symmetrically distributed or skewed. You can plot these data as a histogram and assess each thing visually. There are also more formal measures that are used that we call “summary statistics.”

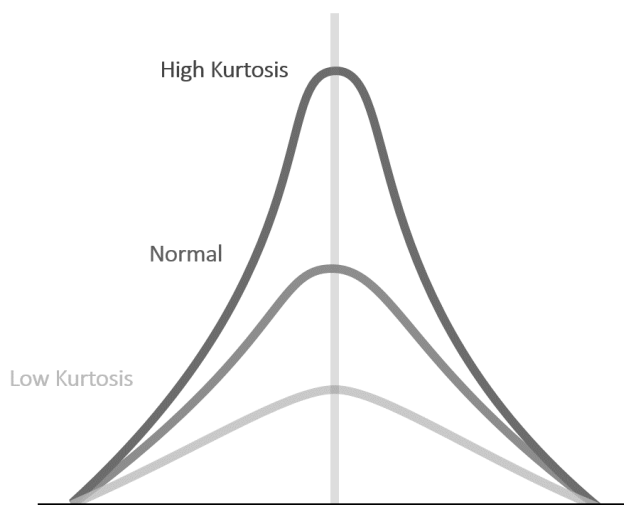


Figure 1. Kurtosis

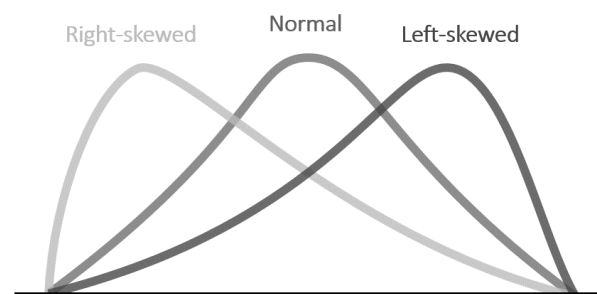
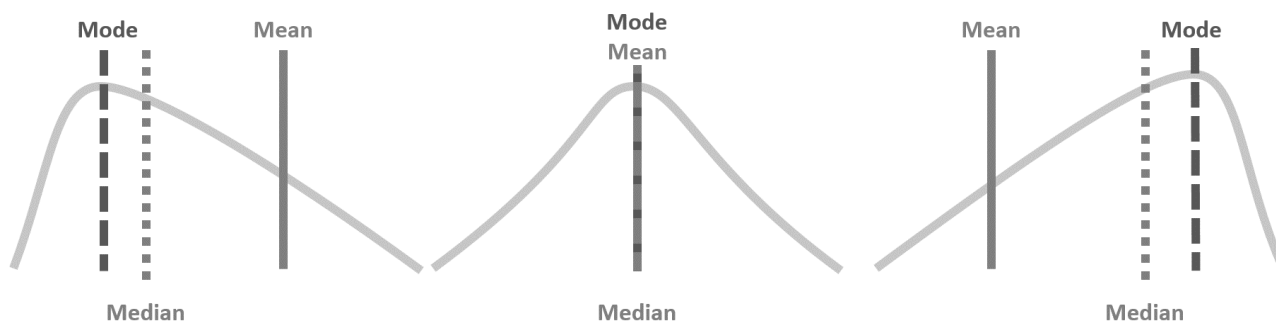
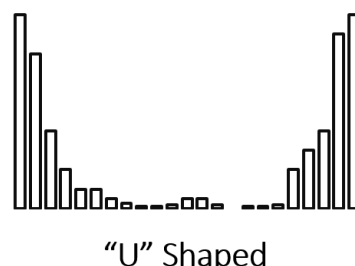
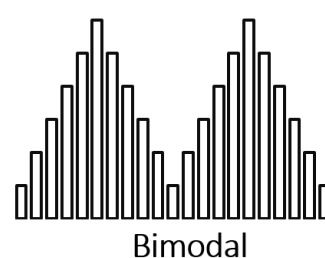
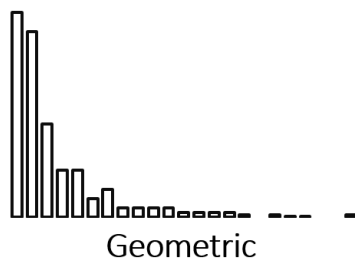
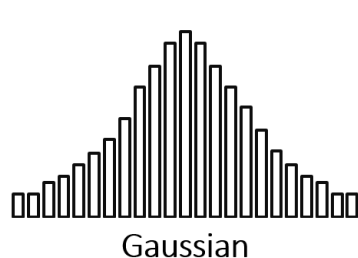


Figure 2. Skewness

Summary statistics for **measures of central location** include the mean, the median, and the mode. The mean is the average value of a dataset, the median is the middle value, and the mode is the most common value. If data are skewed, the median is a preferred measure of central location, if the data is normally distributed the mean is often reported. It is good to calculate each summary statistic for each numeric variable in your dataset.



Creating a histogram of your numeric variables can be very informative. It can help you understand whether your data are normally distributed (i.e., have a Gaussian distribution) or whether they have some other distribution. Depending on the distribution of your data, you may or may not be able to use traditional statistical approaches. You should be aware that many measures in health – particularly measures that count things (e.g., number of days in hospital) – tend to have a distribution that is highly right skewed (e.g., Poisson, binomial, negative binomial, geometric). If your data is not normal, it is wise to consult a statistician prior to conducting your analyses.



Poisson
Binomial
Inverted U
Negative Binomial
Bernoulli
etc. etc.

In addition to measures of central location, measures of spread can help you analyze your data. Common measures of central location include the range, the variance, the standard deviation, and the interquartile range. You should recall the definitions of these from your biostatistics training. Boxplots can also be used to help you visually inspect the spread of your data. It is good to construct box plots and to calculate these statistics for each numeric variable.

Descriptive variables are usually much easier to describe. Traditionally we report the number of individuals in our sample who answered each question, and the number and the proportion of those answering the question that answered the question each way. For example, if you were reporting ethnicity, you would give the total number of respondents in your sample who provided their ethnicity and then report the number and percent of those who chose each response option. Bar charts are a good way to visualize this data.

Bivariable Statistics

In addition to describing your data, you will also likely want to conduct statistical tests to assess whether there is a statistically significant association between two variables. In doing so, you will need to identify whether you are working with numeric or descriptive dependent and independent variables. Numeric variables include factors such as height, weight, BMI, and so forth. Descriptive variables include factors such as ethnicity and HIV-status. If you are working with numeric data you will need to know whether your data is normally distributed and if you are comparing two or more groups, you will need to know if the variances are equal between them.

Two important tests can help you assess normality and the equality of variances:

- The *Shapiro-Wick test* is used to assess normality of a numeric variable.
- The *Levene test* is used to test whether the variance of a numeric variable is equal between two groups.

To decide which tests you should use, you can use online tools such as [The Decision Tree for Statistics](#). As this is not an epidemiology or biostatistics textbook we will not get into all of the many nuances with these measures. Suffice it to say that you should know when to use these tests. In the real world you will probably need to become more familiar with these or higher a statistician to assist you in your evaluation.

In addition to knowing when to apply various tests, it is also important that you are able to provide accurate interpretations of the results from these tests. I find that many tests are easily interpreted incorrectly. Therefore, let's cover some of the basic interpretations for the output of the most common statistical measures:

- The **linear slope** from a linear regression model is interpreted as the unit increase in the dependent variable for each unit increase in the independent variable. Often only the p-value is interpreted and the slope is mistakenly ignored. Do not ignore the slope! The sign (+/-) of the slope indicates whether the association is significantly positive or negative.
- The **Pearson's r** is the strength and direction of an association. Values closer to 0.0 indicate a weak association, those around 0.7 indicate a moderate association, and those near 1.0 indicate a strong relationship. The sign (+/-) of the Pearson's r indicates whether the association is significantly positive or negative.
- The **R²** is the degree to which the observed data fit the modeled line. The R² can only have positive values. The closer the value is to 1.0, the stronger the data fits the line; the closer the value is to 0.0, the less the data fits the line.
- **Odds ratios** are used to compare the relative effect of one group to the "reference" group. Values greater than 1.00 indicate a positive association and values less than 1.00 indicate a negative association.
- **Relative risk** is interpreted the same way as an odds ratio, except it represents risk and not odds. To calculate risk, you need to know the number of people at risk for a disease. This is usually not possible in case-control studies, but is possible in cohort studies. Odds and risks can sometimes be interpreted the same way, however, doing so tends to inflate the association. Take for example a hypothetical group of 129 seeking treatment for loneliness. Imagined that 65 people received a counselling-based program and 64 did not. Of those not receiving the program (n = 64), 18 died. Of those receiving the program (n = 65), 29 died. The overall risk of death in this scenario would be 0.36 (n = 47 died out of 129) while the overall odds of death would be 0.57 (n = 47 died and 87 lived). You can see that if odds are interpreted as risk, the interpretation would give you an inflated estimate of risk.





- **P-values.** To understand how to interpret the p-value it is helpful to know how not to interpret it: The p-value is not the probability that the null hypothesis is true, or the probability that the alternative hypothesis is false; The p-value is not the probability that the observed effects were produced by random chance alone; and the p-value does not indicate the size or importance of the observed effect. A very “significant” p-value could correlate with a minuscule effect. The 0.05 significance level is merely a convention and is nearly indistinguishable from 0.04 or 0.06; yet often times statements regarding “significance” are made based on threshold significance. The p-value is the probability that you are committing a type 1 error. A type 1 error is a false positive – the rejection of a true null hypothesis. The probability of obtaining data as extreme, or more extreme, than those observed if the null hypothesis is correct.
- **Confidence Intervals.** A 95% CI simply means that if the study is conducted multiple times (multiple sampling from the same population) with corresponding 95% CI for the mean constructed, we expect 95% of these CIs to contain the true population mean. A common misunderstanding about CIs is that with a 95% CI there is a 95% probability that the true population mean lies between A and B. This is an incorrect interpretation of 95% CI because the true population mean is a fixed unknown value that is either inside or outside the CI with 100% certainty. In other words, the inclusion of a true population mean is not a probabilistic occurrence. Furthermore, remember that like with the p-value, the choice of whether to use a 90% or 95% CI is somewhat arbitrary, and depends on the level of “confidence” that the investigator wishes to convey in his or her estimate.

Finally, you should be careful to make sure that you are accounting for confounding and bias in your analysis. If you did not use randomization to assign participants to non-participants, multiple regression and other statistical approaches can be used to control for confounding. It is recommended that you work with a statistician on these more advanced approaches. When conducting any of the analyses discussed in this chapter, you can calculate most of them by hand, in excel, or use statistical software such as SPSS, SAS, STATA, or R.



It is not uncommon to feel uncomfortable with analyzing quantitative data. If this is the case, ask your local university if there are researchers, students, or graduates who can help you. Many departments often have many researchers interested in collaborating on community-based evaluations. Practicum students are frequently looking for placements where they can apply their skills.

Part 5

Qualitative Evaluations

Importance of Qualitative Approaches

In many situations, quantitative data is not feasible or appropriate for answering a specific research question. This is particularly so, when not much is known about a topic or when the results of a quantitative study do not actually tell you much about the phenomena of interest. The use of qualitative methods can tell you about lived experiences, meanings, and perspectives and can be used to probe the depths of an issue in a way that is not possible with quantitative data. While some view qualitative methods with suspicion, this worldview is not justified. Qualitative methods are important and necessary for evaluations. The goals of qualitative research can be classified using four general aims:

- **Exploration**, which includes discovering themes and patterns to build an initial understanding of a complex phenomenon, asks questions such as "What kind of things are here or going on?" "How are these things related to one another?" "Do these things fall into natural groups or categories?"
- **Description**, which includes providing an illustration of a phenomenon, asks questions such as "What does this look like?" "Why does it happen?" "What is its purpose?" "Who is involved?" "What are their roles?"
- **Comparison**, which includes looking for differences and similarities in things, asks questions such as "How does X differ from Y?" "How is this group different?" "What factors might drive these differences?"
- **Modelling**, which includes testing conceptualizations developed in the previous steps (or through a priori experience) against observations of new data, asks questions such as "Does this case conform to my general model?" "If not, how is this case different and how should I incorporate it in my model?"



Types of Qualitative Studies

There are many types of qualitative study designs. In this manual we focus on interviews and focus groups, which are some of the most widely used. Under this umbrella there is a gradation of interviewing styles. These interviewing styles may be used to explore ethnographic data (i.e., study of people's experiences) or to illicit insight or opinions about a phenomenon.

- **Informal conversations.** Informal conversations usually are not planned in great detail. While you may have a general idea of what you want to talk about and who you want to ask, informal conversations allow for greater flexibility. These conversations are generally not transcribed word for word – though you might take notes. The conversational nature of these interviews allows interviewers and interviewees to navigate through issues with less social formality. Another important characteristic of these informal interviews is the rapport upon which they are built. While other interview styles may take place between complete strangers, conversational interviews often occur when the evaluator and participant are embedded within the evaluation process together. The dynamic of interviewer-interviewee is thus de-emphasized and the flow of the conversation is less “question-answer” and more “how about this, how about that.”
- **Semi-structured interviews** take things one step farther by creating a somewhat more rigid framework for guiding the conversation. These interviews are often conducted between strangers and are often transcribed word for word. While the topics of various questions have been pre-established and formalized in an ethics guide, the interviews are generally given significant flexibility in how questions are asked, the order in which questions are asked, and the degree to which statements from the respondent are fleshed out. That said, interviewers should still avoid leading questions and design questions that illicit lengthy descriptive answers. The questions should be concise and easily understood as to avoid having to restate the questions. When ordering questions, you generally begin with less sensitive more general questions, but make sure you end on a positive note that's not too emotionally “heavy.” If you are studying a process of life course, it is good to start with earlier events and move on to more recent events. If participants give shorter answers than expected have a few back-up questions and probing questions ready. When participants go in greater depth and answer questions that are intended to appear later – interviewers should adapt by not asking the questions again. That said, an interviewer also needs to be skilled in redirecting discussion back to the topic at hand. While tangents can lead to interesting information – it is important to keep the interview at a reasonable time limit.
- **Structured/standardized open-ended interviews** are the most rigid of the interview styles discussed here. They can technically be administered in person

or by using open ended questions as part of a survey. These interviews are very rigid. The questions are asked the same way for all participants and in the same order. Probing questions, if used, are pre-planned as part of the questionnaire. These are sometimes used because a validated scale or questionnaire requires a specific structure to be accepted as valid. When the interviewer is not an area expert, such as during telephone interviews, these types of surveys can be useful as they do not require an agile and adept individual to conduct the interview.

- **Focus Groups are interviews you conduct with multiple people at the same time.** As with individual interviews, group interviews can be equally flexible or rigid. Nevertheless, they tend to be more similar to semi-structured interviews than anything else. This is because the interviewer will need to allow for the dynamics introduced when multiple people are being interviewed at the same time.

An important distinction between individual interviews and group interviews is that data from focus groups does not represent the viewpoint of an individual. Rather, focus groups will often represent the views of the most vocal individuals or the views of individuals who are delivering answers for the group. This can introduce bias, but can also be considered a strength of the group interview design since it tells you a bit about how individuals express their view points in inherently complex social settings. Given that individuals are not the unit of analysis in a group interview, the questions and research questions generally focus on eliciting diverse viewpoints. The voices of various experts can help you reach new depths by identifying areas that a single participant might not recognize on their own.

Sometimes focus groups can be used to identify a consensus opinion or perhaps identify the key issues of contention. Focus groups are often audio recorded and transcribed. Notes may also be kept by a second interviewer/observer. Body language and other non-verbal speech can also be examined and documented as part of the focus group. Focus group interviews can be expensive and require multiple rounds to get all the issues at hand. Generally speaking, 5-7 people participate in a focus group. However, as many as 10 individuals might participate. The number of participants should be based on the expertise of the participants. In considering the composition of focus group, interviewers should consider power dynamics and how these can limit some individuals from sharing their perspectives. Focus groups also tend to be a bit longer than one-on-one interviews. This allows greater involvement from the larger number of participants. Typically, 2-3 hours is the maximum duration for a focus group, erring towards 2 hours.



Interview Guides

Regardless of which qualitative techniques are selected, it is important that the process of data collection is well-documented. This helps you to account for sources of bias, including interviewer biases that can emerge from inconsistencies in the ways data are collected.

Interviewer guides are not only important for when multiple interviewers are engaged in the interview process, but also for when those collecting the data are not the ones who conducted each of the interviews. Among the various things an interview guide establishes, it identifies the main questions and probing questions that are used to flesh out details or information. Probing is key to successful interviewing as it allows for a more in-depth examination of an issue. There are a variety of probing type questions that can be used:

- **The silent probe** involves waiting for participants to continue talking after they have stopped. It is a small signal to the speaker that they should continue speaking.
- **The echo probe** involves simply repeating back to participants what they last said. It shows that you understand what they said and allows them to add more information if they want to.
- **The Uh-huh probe** is the use of short phrases or sounds to confirm to the speaker that you are engaged and listening. It can help them continue in the current line of thought.
- **The tell-me-more probe** involves asking things like "Can you tell me more about that?" or "Why do you say that?" or "How does that make you feel?"
- **The long question probe** is the use of longer questions to illicit longer responses. For example instead of saying "Why did you start using methamphetamine?" you might say "Some people start using methamphetamine because they are trying to cope with life or to build social connection. Why did you start using methamphetamines?" These sorts of questions are good for sensitive topics because they create comfort and help the hearer to understand where you're coming from.

So, while interviewer guides should include descriptions of the primary and probing questions to be used, they should go beyond this and also identify the

- **characteristics of the interviewer/facilitator** (i.e., credentials, occupation, gender, experience, relationship with participants, how they establish relationships, participant knowledge of interviewer, and other important characteristics);
- **theoretical framework** guiding the development and interpretation of questions;
- **recruitment methods** (i.e., sampling, mode of interview, sample size, response rate);
- **setting** (i.e., presence of non-participants, location and style of interview room);
- **data collection tools** (i.e., number of interviewers, recording, note taking, duration; and transcription methods); and
- information about **how data will be analyzed** (i.e., number of coders, description of codebook; description of how themes will be identified; description of software used in coding process; description of how participants will review findings; description of how quotes will be selected).

All of this information is a helpful part of the interview guide because it helps situate the discrete experience of interviewing into the broader qualitative research framework. Thinking through these aspects will ensure that you have the information you need to successfully execute your study and report on your findings. Further, having information about the interviewer, environment, and other key issues before the interviewer begins will help you to better account for these factors. For example, research has shown that *the deference effect* – which is the effect that differences between the interviewer and participant have on a participant's responses – can play a significant role in shaping how an interview unfolds. By thinking through this issue ahead of time you can better identify strategies to mitigate this during the interview.

As you think about what should be considered in the lead up to an interview or focus group, two helpful resources include (1) the consolidated criteria for reporting qualitative research (COREQ) and (2) the Evaluation Checklist Project's checklist for guiding the development of qualitative research.

It is also important that in designing an interview guide that you consult stakeholders to ensure that they feel the questions being asked are within the scope of your project. Stakeholders with lived experience dealing with the issue at hand will help you flesh out probing questions and identify areas that you might have otherwise ignored. Further, involving stakeholders early in the study design process will encourage buy-in and investment. Invested stakeholders may be more willing to participate in the interpretation of data and in the data analysis process. Having these extra helpers can be of great benefit to you throughout the qualitative research process.

The Process of Analyzing Qualitative Data

Quantitative data analysis involves five primary steps: (1) organizing data, (2) finding and organizing concepts, (3) building overarching themes, (4) ensuring reliability and validity, and (5) rationalizing findings.

Organizing Data

Data from qualitative studies are first transcribed into a format in which they can be analyzed. According to Bailey (2008), "Transcribing appears to be a straightforward technical task, but in fact involves judgements about what level of detail to choose (e.g. omitting non-verbal dimensions of interaction), data interpretation (e.g. distinguishing 'I don't, no' from 'I don't know') and data representation (e.g. representing the verbalization 'hwarryuhh' as 'How are you?'). Representation of audible and visual data into written form is an interpretive process which is therefore the first step in analysing data. Different levels of detail and different representations of data will be required for projects with differing aims and methodological approaches." Once data is transcribed it can be analyzed in specialized software (e.g., NVIVO, ATLAS.ti, QDA Miner, Tams Analyzer, Dedoose, MAXQDA, HyperRESEARCH, AQUAD, Mendeley, and Transana). These software, to varying extents, facilitate the remaining steps in the qualitative data analysis process.


Organizing Concepts & Building Themes

Data from qualitative interviews, focus groups, or any other source can be analyzed using primarily two types of qualitative methods: deductive methods and inductive methods.

- **Deductive methods** are used when a pre-existing theory or framework is used to explore the data. Pre-existing themes and codes, including entire codebooks, can be used to study the data.
- **Inductive methods** are used when there is a lack of previous theory or findings on a topic or when the authors choose to disregard existing theory for the sake of evaluating data from a fresh perspective.

Building off these two general methodological approaches, there are a variety of specific analytic frameworks that can be applied to qualitative data. There are undoubtedly many of these in existence, but the five listed here capture the bulk of qualitative methodology.

- **Content analyses** are either inductive or deductive approaches that examine the content of a data source – often with the goal of quantifying trends and patterns of words used, their frequencies, their relationships, and the structures and discourses of communication.
- **Thematic analyses** are either inductive or deductive approaches that examine the data in order to identify key themes or ideas that are represented by the content of a data source.
- **Discourse analyses** are either inductive or deductive approaches that examine naturally occurring talk and all types of written text to understand how language is used.
- **Narrative analyses** are either inductive or deductive approaches that examine the stories and narratives shared by participants.
- **Grounded analyses** are primarily deductive approaches that examine data sources without respect to pre-existing theory.
- **Phenomenological analyses** are primarily a deductive approach that attempts to describe a phenomenon by articulating its essential nature.



With each of the analytic frameworks listed above, you will need to begin your analysis by creating codes – a process referred to as coding. Codes are themes or short words or phrases that represent a key theme or idea. There are two types of coding used in the coding process:

- **Open coding** involves the organization of raw data to try to understand what has been captured.
- **Axial coding** involves connecting and linking the codes together.
- **Selective coding** involves identifying the core of the issue at hand.

Although coding styles vary between researchers, some of the main questions addressed in open coding include:

- What are the underlying issue and the phenomenon?
- Who are the actors involved and what are their roles?
- How, when, and where does a phenomenon take place?
- How intense or salient is the experience?
- Why and for what reasons does the phenomenon occur?
- Which strategies and tactics facilitate the phenomenon?

When more specific theoretical frameworks are being utilized, specific dimensions of the framework are included. It is important to recognize that in developing codes, you are working to identify both subtle and obvious expressions and themes. Codes can be identified as important based on (1) how frequently they appear, (2) how pervasive they are across different domains, (3) how people conceptualize or interact with the code, and (4) the degree to which the code is influenced by specific contexts or situations. When looking to identify a new code there are several things you can look for:

- Frequent repetitions and circling back to a core issue or element.
- The presence of unfamiliar word or words that are used in unfamiliar ways.
- The use of metaphors or analogies.
- Transitions in tone, thought, or demeanor during a conversation.
- Linguistic connectors that imply causality (e.g., because), conditionality (e.g., if-then), taxonomy (e.g., is a), temporality (e.g., before, after, next), and location (e.g., is close to, by, near).
- Unbalanced qualifiers (e.g., talk about health and women's health, but not men's health).
- What's missing (e.g., is there something important or central that just doesn't get mentioned.)

After the open coding process is complete, axial coding can be used to link specific ideas together; and selective coding can be used to identify the underlying theme that links all of the axial codes together. Again, these processes are highly specific to the exact methods chosen for your analysis and the theories, models, and frameworks that are used to interpret your data. That said, there are a few widely used techniques that can help you process your codes. Among these techniques are:

- **Cutting and sorting** traditionally involves marking each sentence line by line and then cutting each quote out and pasting it to a small index card. The index sorts are then piled based on how similar they are to one another and to what degree the codes relate to one another. Two strategies – splitting and lumping – can be used to either maximize the differences and create more refined themes or minimize the differences and create more general themes.
- **Word lists and key-words-in-context** strategies involve counting the number of times each words appear and the frequencies they appear in which contexts (e.g., Do pretty and handsome appear more frequently for one gender or the other).
- **Word co-occurrence** involves looking at how frequently certain words co-occur. For examine knowing that “Shrouded” occurs frequently with “mystery” or “secrecy” can give you a sense of a common theme. Likewise, crime may appear more frequently with violence than it does with justice – suggesting two potentially distinct themes: criminality (negative) and justice (positive).

Choosing a coding technique

When choosing a technique, it will depend on

- **the kind of data you have** (e.g., Is it text or something else? Do you have verbatim text or just field notes? Are the narratives rich or is your data mostly just short response?),
- **the capacity of your team** (e.g., Do you have enough cultural competence to spot metaphors, connections, and missingness?; Do you have the technical skills to construct co-occurrence data or to use other quantitative analysis methods?; Do you have the man hours to finish the analysis?); and
- **how many themes do you want to identify** (cutting and sorting is flexible, but KWIC and word lists tend to produce a lot of themes while meta-coding and co-occurrence analysis produce fewer themes).

Codebooks

Once you have identified your themes, it is time to begin developing a codebook. Codebooks compile three types of codes: structural codes, theme codes, and memos.

- Structural codes provide information about the reviewer, environment, and interviewee.
- Theme codes capture the themes identified in the last step – these are the substance of what you are studying.
- Memos are field notes about the codes and contain our running commentary as we read through the text.

Codebooks can be developed deductively, but you can also have a predefined inductively developed codebook. For example, you might search the literature in a systematic review for all the psychological models of trauma and then create a codebook that allows you to identify cases in support of each model. Some fields or topic areas have pre-established code books that are available for use or that can be purchased.

Whether you choose to use or build a deductive or inductive code book is largely up to you and your stakeholders. Common features of codebooks included the anonymization of participants, the hierarchical organization of codes and their relationship to one another, and the instructions or criteria used to classify a participant response into the code.

When organizing codes hierarchically, it can be helpful to first identify the general codes and then identifying the sub codes. In some studies, the general codes are identified first and then a separate process or analysis is undertaken to further highlight elements of each constituent code.

“

The complexities of validating qualitative research rest upon their extraordinary power to reflect and conceptualize the nature of the phenomenon, to capture the complexity of the social reality. The validation of qualitative research becomes intrinsically linked to the development of a theory of social reality.”

”

- Steinar Kvale, University of Aarhus

Ensuring Reliability and Validity

Evaluators are often concerned with whether the individuals coding the data are coding the same passages the same way. Using multiple coders for each set of data or training the coders on a smaller test-batch of data can help improve the inter-rater reliability of your analysis. Training coders has all but become a standard approach for qualitative data analysis. Training consists of six steps:

- Give all coders a codebook they can use to code data.
- Review the codebook as a group and have coders code a set of real examples.
- Review their coding as a group and discuss and resolve discrepancies.
- Update the codebook as coders come to agreements about the content of each theme.
- Once coders start coding the bulk of the data, do random spot checking to ensure agreement.
- Repeat the steps above as often as needed to improve the reliability of codes. Agreement of around 80% is the minimum acceptable level for a reliably coded dataset. This means that in a random sample of 10 sections, 8 will have been coded the same way.

One of the main ways that validity of interpretations can be confirmed is through verification. Verification is the process of checking, confirming, and making sure that your results agree with the data and that the claims you make account for each observation. The process of rechecking your data against itself is sometimes called a “constant comparative” approach. During the verification stage, you should systematically review your codes to ensure that coded sections are appropriately classified. Further, you can verify your data by interpreting the data as you collect it. This allows you to probe participants for information that will help you clarify your codes. Further, you can directly ask participants about key aspects of your emerging coding scheme to see if it has obvious face validity. Likewise, you can continue collecting data, until the addition of new respondents does not result in changes to how your data is coding – a condition referred to as **saturation**. Nobel & Smith (2015) provide a thorough list of additional checks:

- Accounting for personal biases which may have influenced findings;
- Acknowledging biases in sampling and data collection;
- Meticulous record keeping, demonstrating a clear decision trail and ensuring interpretations of data are consistent and transparent;
- Seeking out similarities and differences across accounts to ensure different perspectives are represented;
- Including rich and thick verbatim descriptions of participants’ accounts to support findings;
- Demonstrating clarity in terms of thought processes during data analysis and subsequent interpretations;
- Engaging with other researchers to reduce research bias;

Incorporating these sorts of activities as explicit components of your study design will help support the conclusions of your qualitative study and show the rigor inherent in your approach.

Part 6

Conclusion & Recommendations

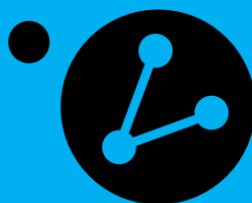
This manual has provided an overview and some practical insight into the design, implementation, and analysis studies. In review, we recommend the following:

- **Loneliness is too important not to measure.** Even if your program is not explicitly designed to address loneliness, monitoring the loneliness of participants is important. Finding a positive impact on loneliness can help justify continued funding and support for a given program.
- **You should use pre-test/post-test evaluations with control groups.** While there are many design considerations in research, pre-test/post-test evaluations with control groups provide strong evidence supporting the effect of a program.
- **We recommend the 3-item UCLA loneliness scale, plus one direct measure of loneliness.** This method provides a fast, easily implementable measure of loneliness that can be used in essentially any setting.
 - How often do you feel that you lack companionship?
 - How often do you feel left out?
 - How often do you feel isolated from others?
 - How often do you feel lonely?

[1 = Hardly ever or never, 2 = Some of the time, 3 = Often]

- **Loneliness is a multidimensional construct.** There are many factors and concepts closely related to loneliness. If your program explicitly aims to address loneliness, it may be important to also measure these factors.
- **A random sample is not always preferred.** There are many established and respective purposive sampling that can provide high quality evidence at low cost.
- **Qualitative interviews and focus groups are important too!** Surveys can only tell you so much about an issue. Engaging participants and program providers in frank discussions about your program is essential to figuring out how it works and how to improve it.
- **Ask for help.** If you are having difficulty designing, implementing, or analyzing your research and evaluation study, reach out to an academic partner and ask for help. You may also contact us at The Social Bubble Project for individualized consultations by emailing team@socialbubble.ca.

.



SOCIALBUBBLE
P R O J E C T