



Quantitative Methods for Evaluating Behavioural HIV Prevention Interventions

ICASA Conference Skills Building Workshop: 4-8 December 2011

Files for this workshop can be downloaded from <http://www.whizzkidsunited.org/downloads.php>

Objective: The overall aim of the workshop is to introduce participants to the skills needed to design and implement a quantitative outcome evaluation of an HIV prevention intervention (particularly behaviour change programmes). Specific skills to be imparted include evaluation methodology, questionnaire design, sampling, data capturing, and data analysis.

Description: This workshop is designed to equip practitioners of HIV prevention programmes with the skills needed to carry out a quantitative outcome evaluation. The workshop will include a Powerpoint presentation and an interactive tutorial using Microsoft Excel which will be made available electronically to participants.

About the Facilitator: Mr. Thomas Farrar is a Canadian statistician (M.Math., University of Waterloo, 2007; B.Sc., Brock University, 2006) who has been working with The Africaid Trust, an NGO based in Durban, South Africa, since 2008 as Monitoring and Evaluation Officer. His duties have included designing and implementing a quantitative monitoring and evaluation system for the WhizzKids United soccer-based behavioural HIV prevention intervention for adolescents which is running in schools in two South African provinces (KwaZulu-Natal and North West). For more information about the WhizzKids United programme, see www.whizzkidsunited.org

Acknowledgments: Mr. Farrar would like to acknowledge the support of the ICASA Conference and its sponsors, as well as Canada Africa Prevention Trials Network (CAPTN, <http://captnetwork.org>) for their generous support.



Table of Contents

1. Develop an Evaluation Question.....	3
2. Develop the Evaluation Methodology	3
2.1. Data Tools	3
2.1.1. Questionnaire Design.....	4
3. Collect the Data.....	6
3.1. Sampling.....	6
4. Capture the Data.....	7
4.1. Example: AIDS-Related Stigma Scale	7
5. Analyze the Data	8
5.1. Software.....	8
5.2. Graphical Representation of Results	8
5.2.1. Instructions for Creating Graphs in Microsoft Excel	8
5.3. Descriptive Statistics	9
5.3.1. Mean	9
5.3.2. Median	9
5.3.3. Mode.....	9
5.3.4. Variance and Standard Deviation	10
5.3.5. Cronbach’s Alpha	10
5.4. Hypothesis Tests	10
5.4.1. Test for difference in proportions.....	10
5.4.2. 2x2 Contingency Table – Chi-squared Test for Association	10
5.4.3. 2x2 Contingency Table – Odds Ratio Test.....	11
5.4.4. Paired <i>t</i> -test.....	11
5.4.5. Unpaired <i>t</i> -test.....	11
5.4.6. Kendall’s Rank Correlation Test	11
6. Write the Report	12
7. References	12



1. Develop an Evaluation Question

- What are the key outcomes of the intervention? (E.g. reduction in HIV incidence, delay of sexual debut, increase in condom use, reduction in multiple concurrent partners, increase in HCT, increase in knowledge, shift in gender norms, reduction of AIDS stigma, reduced teenage pregnancy, reduced drop-out from school)
- Formulate the desired outcome(s) into a specific question (multi-part if necessary), e.g. “Has our intervention caused participants to reduce their number of concurrent sexual partners during the following six months?”

2. Develop the Evaluation Methodology

- How can you quantitatively measure the variable(s) of interest in order to answer your research question? The following factors should be taken into consideration:
 - Consider **data tool**: could be questionnaire (e.g. to ask about sexual behaviour, knowledge, attitudes); observation (e.g. count number of condoms distributed, number of HIV test kits used, or school attendance rate); biological sample (e.g. testing people for HIV or HPV)
 - Consider **sampling method**: how will you access the participants to gather data from them?
 - Consider **timeframe**: if you want to show a change, you will need a before/after picture (baseline and one or more follow-ups)
 - Consider **control group**: if you want to rigorously prove causation, you need a control group to compare your participants to
 - Consider **ethics**: do you need consent from the participants, approval from an academic ethics review committee, or permission from a government department?
 - Consider **cost**: do you have the resources to carry out the evaluation? If not, revisit research question.

2.1. Data Tools

- In a very sophisticated evaluation you could take biological samples (e.g. blood tests), but for many behavioural HIV prevention practitioners this is not feasible



- In some cases (especially in a clinical setting) you could track supply of a commodity or service (condoms; HIV tests; MMC's)
- However, the most common data tool for quantitatively measuring the impact of a behavioural HIV prevention programme is the **questionnaire**

2.1.1. Questionnaire Design

- Responses:
 - Have multiple choice, “closed form” responses, not write-in, “open form” responses (exceptions would be when asking quantitative questions like age)
 - Could have “Yes/No”, “True/False”, or a Likert scale with several options, e.g.:
 - Agree strongly, agree slightly, Neither agree nor disagree, Disagree slightly, disagree strongly
 - Always, usually, sometimes, rarely, never
 - Daily, weekly, monthly, less than monthly
 - The crucial thing with a Likert scale is that the choices be **ordinal** (they can be ranked from least to greatest); otherwise analysis is very difficult
- Question Design
 - Keep the wording of the question as brief and simple as possible; ideally the question can only be interpreted in one way (counter-example: I am concerned about getting HIV)
 - Ask only one thing in each question (avoid ‘double barreled questions’)
 - Avoid double-negatives which can confuse the respondent
 - Ensure questions are translated (and back-translated) so they are understandable in the respondents’ native language; have questions read aloud if literacy is an issue
- Scales
 - Scales are a way of comprehensively measuring a particular variable (e.g. HIV & AIDS stigma) by asking several related questions and scoring responses collectively
 - Scales can be measured for internal consistency (e.g. Cronbach’s alpha)
 - You can find some validated and widely used scales in the literature; if you use these you can compare your results to those in other programmes, geographic areas, etc. (but check copyright!)
- Biases
 - Acquiescence bias: the tendency to agree with any statement by default. Can be countered by putting ‘Disagree’ on the far left as the first option seen by respondent
 - Centrality bias: the tendency to remain non-committal rather than agreeing or disagreeing. Can be countered by not having a ‘neutral’ option, forcing respondents to either agree or disagree to some degree



- Social desirability bias: the tendency to choose the option one believes is socially desirable, rather than the truth. Can be countered by the coin flipping technique although this can be complicated. Keep in mind self-reported sexual behaviour data is not too reliable.
 - The **Coin Flipping Technique** is often used by researchers who are asking a potentially embarrassing question, e.g. How often do you wash your hands after using the toilet? Always Sometimes Never
 - The respondent is given the following instructions: (1) Flip a coin before responding (only the respondent sees the result). (2) If the coin came up heads, tick 'Never' (no matter what your true response is). (3) If the coin came up tails, tick your true response.
 - This way, if the respondent ticks 'Never,' the researcher does not know whether they really don't wash their hands, or whether the coin came up heads.
 - However, because the probability of heads is known (50%), the researcher can estimate how many 'Never' responses were due to the coin flip and remove these.
- Piloting
 - It is very important (but often overlooked) to pilot your questionnaire before you do the main sample
 - Choose a group of people who are similar to your target group
 - Run the questionnaire with them interactively, asking them to inform you of any questions that are unclear
 - Advantages of piloting:
 - Allows you to identify poorly worded, confusing or poorly translated questions, or questions that are open to multiple interpretations
 - And Typo's!
 - Allows you to identify irrelevant questions (e.g. a knowledge question that all respondents know the answer to already, leaving no room for improvement)
 - Allows you to tentatively validate your scales
 - Informs you of respondents' feeling about the questionnaire: too long/short, too invasive of privacy



3. Collect the Data

3.1. Sampling

- In order to take a representative sample, you must identify the following as they apply to you:
- **Target Group:** If your intervention had a specific, identifiable group of participants (e.g. a group of youth who participated in an after-school programme, or sex workers operating in a certain neighbourhood) then these participants would be your target population. If your intervention is aimed at a more general group (e.g. reaching an entire community through a local media campaign) then your target population would consist of all members of that community.
- **Control Group:** A group of people who are similar to your target group but do not participate in the intervention; they are surveyed for comparison purposes. Ideally individuals are randomly assigned to target or control group, but this is not always feasible.
- **Sampling Population:** The 'accessible population' from which you actually draw your sample. It could be all people who come through a certain bus stop in the afternoon, or all households listed in the telephone book.
- **Sampling Frame:** The listing from which you draw your sample. In the above examples it could be a bus stop, or a telephone book.
- **Sampling Units:** The units of individuals included in the sample. This could be people, or households, or school classes, for instance.
- **Sample Size:** How many units should you sample? **Many researchers choose the sample size arbitrarily, but there are mathematical formulas for calculating this** (see Excel spreadsheet). If your target population is small and accessible you may want to include everyone instead of random sampling.
- **Follow-up:** You need to think about whether you will want paired or unpaired data (or a population-based sample) in the follow-ups
 - **Paired:** each individual can be tracked over time
 - **Unpaired:** the sample can be collectively tracked over time, but not individuals
 - In either case, you will need to keep track of the sample's personal details and have a cost-effective, non-invasive way of locating them for follow-up. If anonymity is desirable you can have a register or consent form that is signed independently of the questionnaire being filled out
 - If you want paired data, you must capture a unique identifier on the questionnaire itself so that their data can be linked back to them.
 - **Population-based:** You can choose a new random sample from your population for each follow-up. This will mean you do not have to keep track of who was in the sample; but it also reduces the statistical power of your analysis.



4. Capture the Data

- Here are a few tips for efficient data capturing in Microsoft Excel (I am using 2007 version)
 - Make short, descriptive column headers and drag down row so this is always visible
 - Use the numeric keypad (not available on smaller laptops) with Num Lock on
 - Click Windows button (top left), Excel Options, Advanced, you can have cursor move to the right instead of down when you press enter
 - Code the data numerically, so that furthest left answer = 1, next to the right = 2, and so on
 - Under the Data tab, click Data Validation, and you can specify the values that are allowed (according to your coding of the data), e.g. a whole number between 1 and 5
 - Missing values can be left as blank cells, or entered as a nonsense number e.g. 99 (depending what statistical software you are using to analyze the data)
 - If a respondent ticks two answers, e.g. 'Agree' and 'Disagree', treat it as missing data rather than trying to guess which answer they really meant
 - A basic technique for dealing with missing data is to replace the missing value with the median or mode for that variable
 - Or, if there are few rows with missing data you can delete those rows entirely
- Technology tip: Instead of using paper for the questionnaires, another option is to use mobile phones with a survey application. (In South Africa there is a company called Mobenzi Researcher which provides this service.) The data can then be uploaded to a secure server over the mobile network, and no data capturing is necessary! Respondents often enjoy this method of doing a survey more, and it is environmentally friendly (no printing).

4.1. Example: AIDS-Related Stigma Scale

Source: Kalichman, Seth C. et al. Development of a Brief Scale to Measure AIDS-Related Stigma in South Africa. AIDS and Behaviour 9(2): 2005.

Item	Disagree Strongly	Disagree Slightly	No opinion	Agree slightly	Agree strongly
People who have AIDS are dirty		X			
People who have AIDS are cursed			X		
People who have AIDS should be ashamed	X				X
It is safe for people who have AIDS to work with children			X		



People with AIDS must expect some restrictions on their freedom					X
A person with AIDS must have done something wrong and deserves to be punished				X	
People who have HIV should be isolated	X				
I do not want to be friends with someone who has AIDS		X			
People who have AIDS should not be allowed to work	X				

5. Analyze the Data

5.1. Software

- You may use statistical software such as SPSS, SAS, Stata or Minitab to analyze your data, although these software packages tend to be expensive. Personally I use R (free, open source software), but this requires computer programming capabilities you may not have. However, many of the things you would want to do can be done in Microsoft Excel if you know how.

5.2. Graphical Representation of Results

5.2.1. Instructions for Creating Graphs in Microsoft Excel

- Create Frequencies for your data columns: Under each column, put the following functions: Suppose your data is categorical (either 0 or 1). Put =COUNTIF(Range,0) and, in the cell underneath, =COUNTIF(Range,1). The Range is the range of cells where your data are found for that column, e.g. A2:A94. Put a column of labels to describe each frequency, e.g. 'Disagree', 'Agree'
- Click on Insert Tab, and in the Charts section, choose Column and then the top left option.
- Click on outside edge of graph.
- Open "Design" Tab at top of Word window; choose which Chart Layout you prefer.
- Click "Select Data"



6. Click “Add” on Legend Entries (Series). Under Series name, you may leave it blank, or if you are comparing two data series (e.g. Before and After) you could name the series ‘Before’ etc. Under Series values, select the frequency table you created earlier. Click OK.
7. Under Horizontal (Category) Axis Labels, click Edit. Select the column of labels you created earlier. Click OK.
8. Click OK.
9. You should now see the graph in the layout you chose. If you chose a layout that includes a Chart Title, click on the words ‘Chart Title’ and select the title you prefer.
10. Click on the outer edge of graph, right-click and Copy it.
11. Paste it next to the other graph.
12. Repeat from step 3.
13. When all the graphs are finished, you could copy and paste them into a Word document to include them in a report, for example.

5.3. Descriptive Statistics

5.3.1. Mean

- Definition: The arithmetic average of your data
- When to use: With continuous data, e.g. age, CD4 count; or, with categorical data with only two categories, e.g. Yes/No (in this case the mean is called the **proportion**)
- Excel function name: average

5.3.2. Median

- Definition: The middle value of your data when ranked from least to greatest
- When to use: With categorical data, especially when there are more than two categories (e.g. Likert scale data)
- Excel function name: median

5.3.3. Mode

- Definition: The most commonly occurring value in your data
- When to use: With categorical data with more than two categories (e.g. Likert scale data)
- Excel function name: mode



5.3.4. Variance and Standard Deviation

- Definition: A measure of the dispersion or variability within your data set (note that standard deviation is simply the square root of variance)
- When to use: Useful in almost any situation but especially for continuous data or categorical data with two categories
- Excel function name: var, stdev

5.3.5. Cronbach's Alpha

- Definition: A measure of internal consistency of a multi-item scale
- When to use: Whenever you have a scale consisting of multiple questions designed to measure a single variable (e.g. AIDS related stigma, or self-esteem)
- Excel function name: There is no function in Excel. The formula is as follows

5.4. Hypothesis Tests

- See Excel spreadsheet for technical details
- Here are six basic statistical methods that could be useful for quantitatively evaluating a behavioural intervention:

5.4.1. Test for difference in proportions

- **Tests for:** Difference in proportion between two populations
- **When to use:** If you have categorical data with two categories (e.g. Yes/No, Agree/Disagree) and do not have a control group
- **Example:** We want to know whether the proportion of people who agree with the statement 'People who have AIDS are dirty' is different after a stigma reduction campaign than before.
- **Technical Details:** See Excel spreadsheet

5.4.2. 2x2 Contingency Table – Chi-squared Test for Association

- **Tests for:** Relationship between two binary variables
- **When to use:** If you have categorical data with two categories (e.g. Yes/No, Agree/Disagree) and do have a control group
- **Example:** We want to know whether AIDS related stigma has been reduced in people who participated in an intervention, relative to a control group



5.4.3. 2x2 Contingency Table – Odds Ratio Test

- **Tests for:** Relative probability of an outcome between two binary variables
- **When to use:** Same situation as Chi-squared test for association; it is preferable if you need a precise, quantifiable interpretation of the association
- **Example:** You want to know if the probability of accessing sexual health services is higher amongst youth who receive a cash transfer than those who do not

5.4.4. Paired *t*-test

- **Tests for:** Different in means between two populations
- **When to use:**
 - When your data can be assumed to be normally distributed (NOT categorical data) and;
 - You want to test for a difference in means between two populations, and;
 - The two sample sizes are the same and each data point can be paired (e.g. the same individual before and after an intervention)
- **Example:** We want to know whether the average AIDS related stigma (measured on a multi-item scale) is lower after an intervention than before

5.4.5. Unpaired *t*-test

- **Tests for:** Difference in means between two populations
- **When to use:**
 - When your data can be assumed to be normally distributed (NOT categorical data) and;
 - You want to test for a difference in means between two populations, and;
 - The two sample sizes may or may not be the same; data points cannot be paired between the two samples
- **Example:** We want to know whether the mean age of sexual debut in a population has changed after a community-level intervention

5.4.6. Kendall's Rank Correlation Test

- **Tests for:** Correlation of a variable between two populations
- **When to use:**
 - When your data cannot be assumed to be normally distributed (this is a distribution-free, or nonparametric, test)
 - Ideal for categorical data with two OR MORE categories as long as the categories are ordinal (can be ranked from lowest to highest)
 - Do not require equal sample sizes
 - However the calculations are quite involved



- **Example:** We want to know whether the level of agreement with the statement, “Having many sexual partners is a sign of manhood” (with options ‘Agree strongly; Agree slightly; Neither agree nor disagree; Disagree slightly; Disagree strongly’) has declined after an intervention compared with before

6. Write the Report

- Make sure you report the results transparently and do not try to mislead the reader or cover up poor results; mention all known limitations and sources of error or bias.
- Think about your audience when writing the report, e.g. avoid acronyms and technical terms (or at least include a glossary) if it will be read for laypeople who are not experts in your field.
- Be sure to interpret what your analytical results mean and make recommendations on how they can be applied to improve your programme.

7. References

Bartholemew, L. Kay, et al. Planning Health Promotion Programs: An Intervention Mapping Approach. 2nd ed. John Wiley, 2011.

Brace, Ian. Questionnaire Design: How to Plan, Structure and Write Survey Material for Effective Market Research. Kogan Page Publishers, 2008.

Conover, W.J. Practical Nonparametric Statistics. 3rd ed. New York: John Wiley, 1998.

Fuller, Wayne A. Sampling Statistics. New York: John Wiley, 2011.

Kalichman, Seth C. et al. Development of a Brief Scale to Measure AIDS-Related Stigma in South Africa. *AIDS and Behaviour* 9(2): 2005.

McKnight, Patrick E. Missing data: a gentle introduction. Guilford Press, 2007.

Miller, I., Miller, M., and Freund, John E. John E. Freund’s Mathematical Statistics. Prentice Hall, 1999.

Note: You can also find a lot of websites that will automatically calculate statistical tests for you; just Google the name of the test.