Types of Evaluation

- **Process** - looking at the implementation of strategies and activities
  - What did the program actually do?
  - Is it what it set out to do? If not, why not?
  - What went well?
  - What didn’t go so well
  - What was learnt that will help us do differently/better next time

- **Outputs** – looking at what the program achieved
  - Did we meet our objectives
  - If not, why not? - Unachievable? Wrong strategies? Wrong objectives? Unexpected complications

- **Outcomes** – Longer term - how much did the program contribute to solving the issue/problem it was designed to address

- **Sustainability** – will the solution be enduring
DESIGNING A RIGOROUS EVALUATION

THE TIME TO DESIGN YOUR EVALUATION IS AT THE TIME YOU DESIGN YOUR PROJECT

THEY WILL INFORM EACH OTHER
The Process

1. Design
   - Program Logic
   - Fitness for purpose
   - The question(s)
   - Methodology
     - Resources
     - Indicators & Measures
     - Data Sources and Samples
     - Tools

2. Data: Collection, management and analysis

3. Reporting
Program Logic

Program theory is the set of assumptions about the manner in which the program relates to the social benefits it is expected to produce and the strategy and tactics the program has adopted to achieve its goals. (Rossi et al. 1999:447)

That is, how is it meant to work and what is it meant to achieve (immediately and longer-term)

Exercise 1
A Program Logic

**WHY?**

What is the **VISION** driving the program?

What do we need to change in the world to reach our vision? (**OUTCOMES**)

What will the program do towards bringing about that change? (**GOALS**)

What are the specific program **OUTPUTS** are needed to achieve those goals?

What do we have to do to achieve those outputs (**STRATEGIES**)

**HOW?**

How can we measure whether the program has delivered on its outputs? (**EVALUATION**)

6/05/2008
Fitness for Purpose

To get the right answer you have to ask the right question(s):

- What do we want to use the results to do?
- Who are the end-users and audience(s) for findings?
- What aspects of the program: process, outputs, outcomes do you need to measure?
- What sorts of answers, to what sorts of questions, will allow you to do what you need to with the results?

Exercise 2
Methodology - Resources

- What resources can we bring to this:
  - Time?
    - To spend on the doing the work
    - Before we need the answers
  - Money?
    - To support us while we do the work?
    - To cover the associated costs?
    - To buy in expertise or labour
  - Skills?
    - Program knowledge
    - Design
    - Analysis
    - Writing
  - Available data sources?

Exercise 3
Methodology – Indicators & Measures

You don’t have to measure it just because you can - if you are not going to use it, don’t collect it!

- **Indicators** - markers of what is happening/has happened within the project
  - Are they a genuine reflection of (i.e. strongly correlated directly or indirectly with) the processes or outcomes you are seeking to study?
  - Are changes likely to be measurable over the available time-frame
  - Does it really matter in terms of the reasons for the evaluation?
  - Are they feasibly measurable?
Methodology - Measures

Measuring tools - Ways of assembling information that allows judgements about indicators

Choosing Measures and Tools

- Do the chosen measures provide a valid measure of the indicator?
- Can the measures be operationalised – i.e. can you actually measure them:
  - Within the available resources – time, money, skill?
  - With the available access to subjects and data sources?
- Are the tools we are using appropriate:
  - For our resources?
  - For the sort of data we need for our purposes?
Data Sources

Select based on:

- Usefulness
- Accessibility
- Resource Costs – collection, management and analysis

1. The literature – *you are not alone*
   

2. Secondary Data – *it's never been easier – ABS, AIHW, Reports*

3. Project Documentation – *plan your record management -it’s not much use if you can’t find it.*

4. Primary Data Collection
Quality and Rigour

- Fitness for purpose
  - Answers the question(s)
  - Appropriate to available resources, timeframes and end uses
  - Supports the required generalisability
- Ethical
- Transparency
- Safeguards against Bias
- Validity and Reliability
- Power
- Triangulation/Crystallisation
- Outputs matched to audience
Ethics

When in doubt - DO!

Note: Most journals will not publish results from studies that do not have ethics approval!

Is there any possibility that it could:

- Do harm to, or disadvantage, the subjects –
  - Physically?
  - Emotionally/psychologically?
  - Socially?
  - Economically

- Make claims made on subjects’ time and effort that are not justified by the probable outputs?

- Contravene privacy laws?
Transparency

- Enough detail to inform your audience (and potential critics)
  - What choices you made
  - Why you made those choices
  - What the implications of these choices are for:
    - The validity and reliability of your results
    - The generalisability and application of your findings
Safeguarding against BIAS

- **BIAS**: quality in measurement method (methodology) which leads to misrepresentation of the measure in a particular direction

- **Selection bias** – *addressed by randomisation or highly structured sampling*
  - **Data Source selection** – do all elements have an equal chance of being selected

- **Detection bias** – *addressed by blinding or attention to systematic analysis*
  - Is there a conscious or unconscious tendency to interpret findings in a way that supports a particular outcome or hypothesis
Bias

- **Attrition bias** – are results skewed because of bias in those who refuse to participate or drop out
- **Reporting bias** – is there any bias in the way in which study outcomes are selected for report – successful or unusual/unique cases are interesting but not generalisable.

**CONTROLS**

- Often not practical in evaluations
- Matched to intervention sample in all important variables apart from intervention itself – can be a bold assumption
- Will account for unknown confounding variables if these are evenly distributed in intervention and control groups – best achieved by randomisation
Validity & Reliability

- **Validity:**
  - *Theoretically* – Veracity; is it a ‘true’ measure of the phenomenon under study?
  - *Practically and generally* - Is it logically and intuitively congruent with existing evidence? – that is, does it make sense in terms of what else we know?

- **Reliability** – reproducibility across time, researcher methods and contexts
  - Not usually an issue with evaluations due to non-linear nature of social phenomena

- **Is the effect ‘real’** - or an artefact arising out of methodological weaknesses (bias, inadequate sample) – statistical power

- Applying findings appropriately - *generalisability*
Power – Inferential Stats

- Only applicable when seeking to apply quantitative findings gained from a study sample to a whole population from which the sample is taken.

- **Statistical Significance** (*p value*): level of confidence that the findings reveal a REAL effect not a ‘chance’ finding - a methodological artefact.

- A function of the sample size and the variability within the sample.

- Not usually an issue in evaluations.
Triangulation/Crystallisation
Tools/Instruments

Selected on basis of:
- The sort of data you need
- Your accessible subjects
- Your resources

Validation of Instruments:
- Rigour – is your instrument measuring what you think it is?
- Psychometrics – time-consuming and not justified for evaluation unless the instrument is to be promoted as standard tool
- Piloting – Always a good idea and simple to do
  - Recruit up to 10 people who match your sample in characteristics of interest
  - Administer survey and invite comment (individually or in focus groups)
  - Analyse responses and comments – look for common misinterpretations, missed questions, difficulties with understanding, ambiguous responses
Surveys

Strengths:

- Quantitative or semi-qualitative data – numbers & descriptive statistics can reveal patterns
- Large amount of data for minimal effort and cost
- Ease of data entry and analysis

Limitations:

- High risk of sampling bias – sampling design critical
- ‘Thin’ data - lacks richness
- Good survey design is an under-rated skill
Focus Groups/Workshops

- **Strengths**
  - Rich qualitative data – meaning, values, perspectives
  - Less resource intensive than interview
  - Interaction increases critical-analytical dimension
  - Good reliability and validity if well designed and conducted

- **Limitations**
  - Does not provide numerical data
  - Data transcription and analysis resource hungry
  - Can be difficult to arrange
  - Can get out of control if not well facilitated.
  - Confidentiality and ethical issues

- **Focus Groups** – data collection only
- **Workshops** – data plus collectively achieved outcomes
Interviews

- **Strengths**
  - **Looking for** - In-depth understanding, thick description, rich narratives, detailed examples
  - **Exploring** - Complicated relationships, difficult to express concepts, threatening topics
  - **Dealing with** - Clients who have difficulty communicating, hostile of marginalised subjects

- **Limitations**
  - Very resource intensive at every stage
Sampling

- Less of an issue than in most research
  - Less likely to be using inferential statistics
  - Most samples purposive or opportunistic

Questions to address:
- Access to subjects
- Best approaches and tools given subject characteristics
- Informants versus respondents
- Representativeness, bias and generalisability

Issues:
- Ethics
- Availability
- Confidentiality

*Exercise 3*
Rigour in Data Management

- SYSTEM, SYSTEM, SYSTEM
  - DOCUMENTED - DETAILED AND TRANSPARENT
  - SET IT UP EARLY
  - STICK TO IT
Analysis

- Use the resources that you have to hand
- Quantitative: adding up the numbers
  - Simple descriptive statistics will generally suffice: means, frequencies, percentages 2X2 table
  - Presenting – clear and simple, making it easier to see at a glance
  - Spreadsheet Programs – Excel
  - Specialist statistical packages – SPSS, SAS – some free
    - http://www.psychnet-uk.com/experimental_design/software_packages.htm
- Qualitative (Coding): the re-organisation of text into themes and categories in order to talk about it.
  - Pencil and paper – markers, colour, post-it notes
  - Word processors – files, folders, cut, copy, paste, outline function
  - Specialist analysis programs –
    - http://www.eval.org/Resources/QDA.htm
Reporting Dissemination

If it isn’t disseminated, it wasn’t done

- Who are your intended audience(s)?
- How to reach them?
  - What format?
  - What conduits?

Getting the message across

- Which messages?
- Form – structure, length?
- Language and tone?

- Is there someone who might help with the writing in exchange for joint authorship?

**Exercise 4**
Some Resources

