

Epidemiological approaches to evaluating health care interventions

Rachel Jewkes,
MRC Gender & Health Research
Unit, Pretoria, South Africa



Why do we need evaluation?

- We need to have confidence that what we are doing is of value and to learn how to do it better
- Evaluation has shown us that some of our 'good ideas' for interventions don't work or are counter productive
- Evidence of effectiveness is becoming increasingly important for getting new interventions accepted and resources allocated to them
- Evaluation is crucial if we are to advance the services we provide for survivors and to convince policy makers and funders of the value of our work



What needs to be evaluated?

- We need to understand the value and effectiveness of all of the current components of health care after sexual violence
- We need to evaluate / have evaluated new interventions
- What is a new intervention?
 - Changes to policy, training programmes, advocacy/education programmes, information leaflets, changes in staffing, counselling approaches, treatments such as post-exposure prophylaxis etc..



Using evaluation to develop an evidence base

- Evaluations can be thought of as falling into two classes: formative and summative
- Formative evaluation is the term used for what is usually a set of small studies that are used for assessing an intervention in the development stages. The findings tell us not ‘whether it works’ but rather ‘what is happening and what needs to be fixed’
- Summative evaluation is the term used to refer to evaluations of interventions that are established. These projects answer questions of how well an intervention works.



Formative evaluation

- This will usually involve multiple iterations of testing and often the use of a range of methods
- Successful interventions, whether behavioural or biomedical, always have theoretical bases and are built on previous research
- This needs to be articulated and a mapping exercise should be undertaken to address the question ‘what do I need to know in order to do this well?’
- If the knowledge base has gaps it may be important to conduct more basic research before developing interventions
- Sometimes you won’t know everything when you develop interventions, but it is very helpful to have mapped out your gaps as you may be able to the addressing then the course of the formative research or in parallel studies



Formative evaluation

- Qualitative research is particularly valuable in the first stages of formative evaluation because it enables you to learn the unexpected
- Very often we initially test information and behavioural type interventions by exposing a small group of people to them and then gathering feelings, reactions, responses, initial feedback etc using qualitative methods
- Usually this qualitative information can be collected by skilled note taking, processed rapidly and used to inform a next draft of the intervention
- In very well resourced settings (such as the US) all this data would be taped and transcribed etc whether this should be done depends on the resources available, the skill gap between researchers & intervention developers (are they the same people), and the local circumstances



When do we start looking at impact

- This is little point in assessing impact until an intervention has the confidence of those using it and those receiving it, and there is agreement on how it should be used, all demonstrated in qualitative research
- At this point the question ‘what does this intervention do?’ can be quantitatively assessed
- This is usually a small pilot study, the design of which depends on the intervention
- Control arms are usually not necessary but ensuring impact is assessed at a meaningful point post-intervention is important
- You may develop a questionnaire that explores a wide range of possible impacts and to test this with people receiving the intervention
- Interviewing those providing an intervention can also be helpful as they may observe impact that is useful too



Understanding the impact of an intervention

- The findings of a pilot need to be carefully assessed
- It is important to ask ‘what changed’ and if it wasn’t what was expected, ‘why not?’
- Sometimes you need some more qualitative research to answer this
- It may point to a further need to refine the intervention or some aspect of its delivery and if so you may need another pilot
- You can be confident that piloting is finished when you have demonstrated change of a reasonable magnitude in indicators that you had reasonably hoped would change. Then you use your information to:
 - Refine the proposal for a summative evaluation so that objectives are realistic & a focused dataset on impact collected
 - Understand the sample size needed to determine impact. This is based on effect size seen in the initial study (i.e. not just ‘what changed’ but ‘how much did it change by?’)



Decisions around summative evaluation design: some options

- **Before & after studies** – weakness is that its impossible to attribute changes observed to that intervention. Could they have been due to something else?
- **Non-randomised controlled trials** – where there are two arms but allocation is not random (e.g. one may have volunteers for a particular treatment). Main problem is the potential for selection bias.
- **Randomised controlled trial** – this is the evaluative gold-standard



Key features of a RCT

- Two or more study arms – a comparison arm and (1+) treatment group
- Allocation to the study arm is done on the basis of randomisation (chance)
- Ideally neither the allocator or the study participant knows who is in which arm at the point of allocation (double blind trial)
- Ideally, the person doing the impact assessment should not know which arm a person is allocated to at the point of impact assessment
- Where possible biological outcomes that are not subject to reporting biases should be used



Choice of comparison treatment arm

- Crucial decision must be made as to whether to use a placebo (anticipated to be inactive with respect to main study impacts) or an alternative intervention (can be current practice or alternative treatment)
- Usually an ethical decision: it is not ethical to 'do nothing' if there is evidence that something else has worked or that something is current practice.
- But, if you use an efficacious control intervention you only learn relative impact of an intervention, not absolute impact – is that what you want to know?



Allocation to intervention and control arms

- The use of random allocation has the effect of also randomly allocating all unmeasured risk factors (and ones you can measure) between the two arms
- However, it can only be relied on to do this if a large number of individuals or clusters are randomised
- If there are variables which it is essential to have equally allocated between the study arms these can be rendered 'strata' and a stratified random allocation performed in which equal numbers of subjects are allocated to each study arm in each strata
- Baseline data by treatment arm must be collected to ensure equal allocation between arms



RCT design

- RCTs should be designed to test one hypothesis, which articulates the relationship between the intervention and the primary outcome
- The primary outcome needs to be explicitly defined
- There may be a limited number of secondary hypotheses addressing the relationship between the intervention and secondary outcomes
- The success of the intervention must be determined against the primary outcome, it cannot be determined to be successful if one or more secondary outcomes are achieved
- The primary outcome and the anticipated degree of impact are needed for the sample size calculation, and this must take into account possible loss to follow up



Reducing measurement bias

- Outcomes must be very precisely measured to minimise measurement bias
- Often there is a subjective element to outcome assessment, bias can be introduced if the assessor knows which group a study subject is in
- Ideally the study subject reporting e.g. satisfaction, behaviour change or symptoms should also not know this
- Bias can operate in two directions: it can lead to exaggeration of impact, as well as understatement of impact



Collecting outcome data

- Data must be collected to measure the primary and secondary outcomes and on key confounders (usually social & demographic characteristics)
- Minimise the number of observers to reduce inter-observer variation
- Both non-compliers and people lost to follow up normally have different risk of adverse health outcomes from compliers and those retained in a study, so you need to be really careful to minimise loss to follow up
- All people invited to enter the trial and those dropping out at any stage should be counted and the details of their numbers and reasons for loss to follow up (refusal, untraceable, death etc) presented in the publication (see CONSORT guidelines for reporting RCTs)



Types of randomised controlled trials

- Double-blind randomised controlled trials – subjects are randomly allocated to a treatment or placebo and neither subjects or investigators know which is received
- Cluster randomised trials – groups of individuals are randomised rather than single individuals
- Multi-factorial trials – two or more interventions are compared
- Cross-over trials – each study subject is his/her own control by receiving two different treatments at different times in the trial. They are randomly allocated to treatment order.



Ethical issues

- The normal concerns of informed consent, lack of coercion and confidentiality apply
- **Safety:** the new intervention must have been proven safe before the trial
- **Choice of comparison treatment:**
- **Interim analysis:** Some trials with long follow up have their results examined at fixed intervals to look for evidence of effect; if a statistically significant difference between study arms is determined the trial will be stopped. Aim is to reduce the period over which the control group are denied an efficacious treatment.



Ethical issues in sexual violence research: lessons from South Africa

- **Recruitment in post-rape care:** survivors are highly stressed and vulnerable in post-rape care so great caution needs to be exerted before recruiting them for research (i.e. ask if you need to do this).
- Providing information on the research, asking for a decision to be made and for attention to be focused on research logistics can be very stressful.
- Fieldworkers need to be very well trained in both understanding rape as well as research



Sensitivity in follow up:

- Follow up of rape survivors is difficult:
 - Stress affects memory: they may not remember have been asked to consent to the research
 - Contact is difficult: survivors often move to increase their sense of safety after rape and cell phones etc are often stolen in rapes
 - Secrecy: many people you would imagine would be close to the rape victim will not have been told about the rape (especially husbands & boyfriends)

Offer survivors something they perceive they need (such as support and advice on how to deal with police after rape) – even if this is not relevant to your immediate needs. Just do it equally to people in both study arms

Payment of incentives for research participation



Data analysis

- Method of analysis – usually ‘intention to treat’ so that all participants are kept in their study arm for the analysis even if they did not take all (or any) of the intervention
- This is a conservative approach to data analysis which avoids the potential biases which could arise from different levels of participation in the intervention and control arms, or due to the fact that non-participants differ from participants
- Because confounders are allocated between arms you do not need to adjust for possible confounders in analysis, but you may need to adjust for the level of the variable at baseline

Cluster RCTs much have cluster level analyses



Interpretation of the study results

- Clinical importance of the effect, as well as statistical significance should be assessed, both are essential
- Consider alternative explanations for results:
 - Could they be due to bias?
 - Was randomisation (and blinding) effective?
 - Did the study have enough power to detect an effect?
- Generalisability of results – depends on study population – you need to be careful not to define this too narrowly as doing so limits generalisability
- Adverse events – must be monitored and reported
- Costs per unit of benefit?
- Are there indirect effects which should be considered?



Efficacy v. effectiveness?

- Efficacy trials aim to estimate the maximum possible benefit that can be derived in an ideal setting from the intervention (with 100% compliance)
- Effectiveness trials estimate the degree of benefit that can be derived in a 'real world' setting where compliance will usually not be 100% and there may be all sorts of constraints on the treatment delivery
- These both have value but they are different



Limitations of RCTs

- The need to interpret success based on a primary outcome constrains interpretation and this can feel frustrating (but there are no free lunches!)
- A limited range of impacts are illustrated
- Sub-group analysis is usually not pursued – depends on study design & sample size
- Measuring relative effect is some times ethically necessary but doesn't provide the information most needed for policy decisions
- Problem of assessing impact too soon (can over or under-estimate long term impact)
- Inadequacies of measurement can appear as inadequacies of the intervention

