Monitoring and evaluation

Rachel Jewkes,
Gender & Health Research Unit,
MRC, South Africa
Why M&E?

• The aim of M&E is not to keep your donor happy!
• M&E is the process you need to engage with in order to strengthen your work
How do programs achieve?

“I think you should be more explicit here in Step Two.”
When do you need to do M & E?

– From programme/intervention preparations until the end
Integrated Program Development Cycle

1. ASSESSMENT
   What is the nature of the problem?

2. STRATEGIC PLANNING
   What primary objectives should my organization pursue to address the problem?

3. DESIGN
   What strategy, interventions, and approaches should my organization use to achieve its objectives?

4. IMPLEMENTATION/MONITORING
   How do I know the activities are being implemented according to our design? How do I fine-tune and customize them to our own setting? How do I know if outcomes are being achieved or the program is situated to achieve outcomes?

5. EVALUATION
   How do I know that the strategy is working? How do I judge if we’re making a difference?

6. REPORTING/SHARING FINDINGS
   How are combined results making a difference?

Results Phase

Planning Phase
M&E METHODS AND PROGRAM LOGIC

Findings from M&E

Planning → Implementation → Results

Findings from M&E

- Formative evaluation
- Assessment and planning

- Input/output monitoring
- Process evaluation

- Outcome monitoring
- Outcome evaluation
- Impact monitoring
- Impact evaluation
Research in intervention design

• Needs to start from the conceptualisation stage

• Formative research: this is undertaken to help shape the intervention and enables the intervention to be grounded in
  – the subject being studied (e.g. gender attitudes, masculinities, violence)
  – the target population

• Methods used in formative research: multiple (including desk reviews)
Using research to test interventions

• All interventions need to be tested, if taken from use elsewhere they need to be adapted and tested for local use
• Most testing uses qualitative methods so the problems (and successes) can be studied in detail
• Testing should establish: what works, what doesn’t, how is an intervention understood, how it is used, what are challenges etc
• Testing should be an iterative process – you need to be open to multiple rounds of testing and redesigning and retesting and redesigning etc..
Testing a school curriculum

- Deliver the curriculum to a group of the right age
- Observe lessons/sessions and record problems etc
- Ask the group to discuss each lesson at the end to get feedback
- Ask some to keep diaries of their feelings and reactions to the curriculum and what they found useful, ask for suggestions for improvement
- Revise and repeat…
When to implement?

• Implementation should be done after qualitative evaluation shows that the intervention is well accepted, liked and seen as useful by the target group, and all major problems have been solved

• NB it may not be advisable to always implement !!!

• Once you implement monitoring is essential…
Monitoring

- Monitoring is an on-going process
- Should be based on the collection of data routinely
- Monitoring focuses chiefly on process
- Monitoring may also include collecting data routinely on an ultimate outcome (e.g. prevalence of GBV in Cambodia)
- Monitoring may also focus on adverse outcomes
Key characteristics of monitoring

- **Routine tracking** (collect data on weekly, monthly, quarterly, etc. basis)
- **Systematic collection** (i.e. reporting of number of HIV tests every quarter)
- **Ongoing process** (occurs throughout program life)
- **Fairly simple data collection** (patient registers, reporting form, etc.)
- **Measures current performance** (what are we doing?)
MONITORING ILLUSTRATED
Good monitoring

• Monitoring uses *indicators* to track progress

• Principles of good indicators – need to be SMART!
  – Specific - concrete, detailed, focused and well defined
  – Measurable - tells how many or how much and can be measured with identified measurement sources
  – Agreed upon - stakeholders vested in question should agree that indicator is relevant
  – Relevant - indicator generates data that can answer the question of interest
  – Time bound - specifies time frame of what it is measuring

• Less is more – focus on few telling & robust indicators
Monitoring national action plan effects

- National intervention programmes need to be monitored through national indicators
  - e.g. % women reporting past year physical IPV
- Challenges:
  - Expensive to collect
  - Temptation to piggy back on other surveys e.g. DHS etc - should be done with great caution
  - GBV research highly vulnerable to error thru under-reporting - a lot of under-reporting provides incorrect data
- Monitoring does **not** tell you **what** has brought about change, but just that there was change
Examples of indicators:

• Number of people attending Stepping Stones workshops – process/output

• Proportion of people attending 75% or more of the Stepping Stones programme

• Proportion of participants’ rating programme relevance as 4 or 5 (on a scale from 0-5)

• Proportion of observers rating facilitation quality rated as ‘good’ or ‘very good’

• Proportion of participants who have taken some action (speaking out or otherwise) against GBV six months after the start of the programme
Purpose of monitoring

- Monitoring needs to be used to strengthen programmes so the results need to be used
- Feedback problems identified into modifications in the programme design
- How can the information help us work better?

Monitor delivery

Adjust the intervention or programme
What do we do when monitoring gives information that doesn’t look positive? - reflect on information

- Information quality?
- Are we looking at the right indicators?
- Are we interpreting them correctly?
- Could there be competing explanations?
- Are we doing the wrong interventions? Or achieving too little coverage?
Evaluation

• Research process through which we determine whether a project meets its goals and objectives, i.e. its outcome is achieved

• Evaluation methodology depend on resources, but the choice here influences the confidence you/others have in the results of the evaluation

• Multi-component interventions need complex evaluation strategies
KEY CHARACTERISTICS OF EVALUATION

- **Periodic** (not more than once a year)
- **One off activity** (mid-term evaluation/review)
- **Systematic assessment** (did the intervention work?)
- **Requires rigorous scientific methods** (pilot tested questionnaires, standardized methods for sampling and analysis, etc.)
- **Measures longer term impact** (what have we achieved?)
- **Builds on findings** in monitoring
EVALUATION ILLUSTRATED
Main concern in information quality is bias: aim is an unbiased assessment of outcome

• Bias is a systematic tendency to underestimate (or overestimate) the measure of interest because of a deficiency in the design or execution of the study

• **Main sources in evaluation:**
  – Selection bias
  – Information bias
Selection bias

• Population studied are not representative of those receiving the intervention
  – Does it exclude any of the target population who may differ in important ways from the people studied? (especially important with evaluations of mass communication approaches; volunteers for research differ from selected people)

• Important if only some who receive an intervention or programme are in the evaluation

• If the evaluation had repeated measures (e.g. before and after) it’s a concern if there is loss to follow up (fewer assessed ‘after’)

• Most important if those followed up differ from those enrolled initially (so this needs to be assessed)
Selection bias

• How do we avoid or minimise selection bias?
  – Ensure all the target population is enrolled in research
  – OR, have a large enough randomly selected sample
    and high (aim for 80%+ if possible) response rates
    and show non-responders not to be too different from
    responders

• Random sampling – means that the process of
  selection for the sample is not influenced by
  variables that may be related to the impact of the
  programme
Problem of information bias in evaluations

• Information bias is a systematic tendency to report outcome measures differently in different intervention/programme exposure groups

• Misclassification of outcome can be random or biased

• Random misclassification occurs when outcome measures are not very repeatable irrespective of whether a person has been exposed to the intervention (e.g. too long a time period is used)

• Systematic misclassification (information bias) – problem where interventions teach the research answer e.g. teach its socially unacceptable to use partner violence without stopping IPV
Attributing effect and time to effect

• One of the greatest challenge in evaluation is to ensure that change observed is due to the intervention/ programme and not due to something else
• Need a ‘non-intervention’ comparison group to ensure this
• Need before and after measures
• Preferably both
• Problem as behaviour and attitude change can take a long time
• Many interventions are evaluated too early – ‘after’ means immediately after and not ‘after 2 years’
Designing evaluations

• Population of interest:
  – Everyone who is a recipient of an intervention
  – A randomly selected sample of those receiving the intervention

• Study design:
  – Need baseline and follow up measures (often we have more than one follow up measure)
  – Comparison group (control or non-exposed group)
  – Ideally, a process of random allocation to the intervention and comparison group before the intervention is delivered
  – High follow up rates (80% + if possible)

• Long follow up: 1-2 years (depends on the outcome)
Intervention to be evaluated

- Replication of intervention – evaluation must be of a defined intervention – one that can be replicated (or scaled up) – it is essential to commit to what the intervention ‘is’ before you evaluate
Most rigorous evaluation designs

• Randomised controlled trial: random allocation of people to the intervention and comparison arms, high follow up and long time frame

• Time series evaluation: this is an acceptable alternative if randomisation is impossible, with multiple assessments made in the target population over a long time
Designing the evaluation

- Outcomes: aligned to the programme goals and objectives
- Key outcomes should be few in number
- Should be measurable, in a way that is valid (reflects the truth) and repeatable
- Must standardise measurement (use a questionnaire)
Qualitative methods in evaluation

• These can give information on:
  – how an intervention is received,
  – how the target view the information (what they see as useful, memorable etc),
  – how they made use of it,
  – what barriers they faced in using it,
  – whether they could overcome these and
  – what happened

• Very valuable information, but it does not really give the ultimate outcome
Commissioning evaluations: things to consider

- Expertise in intervention evaluation AND in GBV/gender
- Budget
- Timeframe
- Purpose – ideally evaluation will be rigorous enough to contribute to knowledge regionally/globally BUT this takes time
Monitoring and evaluating national GBV strategies: a public health approach

Four key questions:

a) Are we fully aware of the problem?
b) Are we doing the right things to address the problem?
c) Are we doing them right?
d) Are we doing them on a scale large enough to make a difference?
• **Are we fully aware of the problem?**
  – *Do we ‘know the epidemic’?*
    • What is the prevalence?
    • Are there important focal sub-groups?
      • Conduct a situational analysis on GBV. Analyze by gender (victimisation/perpetration), age, SES, and other relevant stratifiers.
  – *What are the contributing factors?*
    • Describe the risk and protective factors through research

• **Are we doing the right things to address the problem?**
  – *What works? What works for our target groups and risk factors?*
    • Systematic reviews, formative research to adapt and test interventions from other settings or target groups
    • Evaluations of interventions
• **Are we doing them right?**
  – *Do we have enough resources?*
  – *Are we working across all sectors?*
  – *Is the quality of intervention high enough?*
    - Conduct needs, resources, and response analysis as well as input monitoring
    - Conduct process monitoring and evaluation, as well as quality assessments
    - Conduct outcome evaluation studies that measure changes in attitudes, knowledge, practices and behavior related to GBV

• **Are we doing them on a scale large enough to make a difference?**
  - Conduct national household survey or sentinel surveillance, present data disaggregated by gender (victimisation/perpetration status), age, SES, and other relevant stratifiers
TEMPLATE FOR LOGIC MODEL – MORE DETAIL

**Problem Statement**

**Inputs/Resources**
- What we invest
  - Staff
  - Volunteers
  - Time
  - Money
  - Research base
  - Materials
  - Equipment
  - Technology
  - Partners

**Activities**
- What we do
  - Conduct workshops & meetings
  - Train
  - Outreach
  - Deliver services
  - Develop products, curriculums, resources
  - Deliver product
  - Liaison
  - Facilitate
  - Assess

**Outputs**
- Who we reached & what we did
  - Workshops & meetings conducted
  - # of trainings conducted
  - # of participants attending
  - Materials developed

**Outcomes**
- Short term results
  - Learning
    - Awareness
    - Knowledge
    - Attitudes
    - Skills
    - Motivations
  - Action
    - Behavior
    - Practice
    - Decision-making
    - Policies

**Ultimate Result**
- Medium term results
- Conditions
  - Health status
  - Social
  - Economical
  - Environmental

**Assumptions**