Evaluation of Behavior Change Communication Programs:

Guidance Note for German Development Cooperation

April 2012
This paper was prepared by **Marianthi Vezertzi** and **Julia Lohmann** for the sector initiative “Applied Research in Health Programs of German Development Cooperation” commissioned by the German Federal Ministry for Economic Cooperation and Development (BMZ).

April 2012

Editor:

Deutsche Gesellschaft für Internationale Zusammenarbeit

Responsible: Stefan Weinmann, stefan.weinmann@giz.de
Table of Contents

Executive Summary ......................................................................................................................................................... 6

1. Part I: Monitoring and Evaluation of Behavior Change Communication Interventions.............................. 7
   1.1. Background Information on Behavior Change Communication ................................................................. 7
       1.1.1. What is Behavior Change Communication? ......................................................................................... 7
       1.1.2. Which are the goals and objectives of Behavior Change Communication? ........................................ 8
       1.1.3. Which are the guiding principles of Behavior Change Communication interventions? ......................... 10
       1.1.4. The role of BCC interventions in HIV prevention .............................................................................. 10
   1.2. Evaluation of Behavior Change Communication Interventions ................................................................. 12
       1.2.1. Why evaluation? ..................................................................................................................................... 12
       1.2.2. Monitoring versus evaluation .............................................................................................................. 13
       1.2.3. The OECD/DAC general monitoring and evaluation framework and its application in German development cooperation ........................................................................................................ 14
   1.3. Different Types of Evaluation and their Application ................................................................................. 16
       1.3.1. Formative evaluations .......................................................................................................................... 16
       1.3.2. Summative evaluations ......................................................................................................................... 18
       1.3.3. Differentiating evaluation efforts: When to perform which evaluation? .............................................. 21
   1.4. Planning a BCC Intervention Evaluation .................................................................................................... 22
       1.4.1. What is an evaluation plan? .................................................................................................................. 22
       1.4.2. Who is to perform the evaluation? ........................................................................................................ 23
       1.4.3. What are the main evaluation questions? .............................................................................................. 23
       1.4.4. Deciding on indicators ........................................................................................................................ 24
       1.4.5. Evaluation project management ......................................................................................................... 24
   1.5. Indicators ....................................................................................................................................................... 25
       1.5.1. What is an indicator? ........................................................................................................................... 25
       1.5.2. How to choose indicators ................................................................................................................... 26
       1.5.3. Clearly defining/stating indicators ..................................................................................................... 28
   1.6. Evaluation Study Design ............................................................................................................................... 29
       1.6.1. Evaluation designs and causal inference ............................................................................................... 29
       1.6.2. Randomized experiments .................................................................................................................... 30
       1.6.3. Quasi-Experiments .............................................................................................................................. 32
       1.6.4. Non-experimental designs .................................................................................................................. 32
       1.6.5. The role of triangulation ...................................................................................................................... 34
   1.7. Data collection .............................................................................................................................................. 35
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7.1.</td>
<td>Quantitative methods</td>
<td>35</td>
</tr>
<tr>
<td>1.7.2.</td>
<td>Qualitative methods</td>
<td>36</td>
</tr>
<tr>
<td>1.7.3.</td>
<td>A closer look at questionnaires</td>
<td>36</td>
</tr>
<tr>
<td>1.7.4.</td>
<td>Sampling</td>
<td>38</td>
</tr>
<tr>
<td>1.8.</td>
<td>Statistical analysis and interpretation of data</td>
<td>42</td>
</tr>
<tr>
<td>1.8.1.</td>
<td>General information</td>
<td>42</td>
</tr>
<tr>
<td>1.8.2.</td>
<td>Common statistics and statistical procedures</td>
<td>42</td>
</tr>
<tr>
<td>1.9.</td>
<td>Dissemination of evaluation results</td>
<td>45</td>
</tr>
<tr>
<td>2.</td>
<td>Part II: Implications on the HIV prevention tools used by GDC</td>
<td>47</td>
</tr>
<tr>
<td>2.1.</td>
<td>Examples of HIV prevention tools</td>
<td>47</td>
</tr>
<tr>
<td>2.1.1.</td>
<td>Join-in Circuit</td>
<td>47</td>
</tr>
<tr>
<td>2.1.2.</td>
<td>Peer education</td>
<td>50</td>
</tr>
<tr>
<td>2.1.3.</td>
<td>Mass media campaigns</td>
<td>53</td>
</tr>
<tr>
<td>2.2.</td>
<td>Final recommendations</td>
<td>55</td>
</tr>
<tr>
<td>2.2.1.</td>
<td>Recommendations to programs</td>
<td>55</td>
</tr>
<tr>
<td>2.2.2.</td>
<td>Recommendations for researchers</td>
<td>57</td>
</tr>
<tr>
<td>2.3.</td>
<td>And some tips...</td>
<td>57</td>
</tr>
<tr>
<td>2.3.1.</td>
<td>When planning an evaluation...</td>
<td>57</td>
</tr>
<tr>
<td>2.3.2.</td>
<td>When implementing an evaluation...</td>
<td>58</td>
</tr>
<tr>
<td>2.3.3.</td>
<td>Before disseminating the results of an evaluation...</td>
<td>58</td>
</tr>
<tr>
<td>2.3.4.</td>
<td>After dissemination and validation of the results</td>
<td>58</td>
</tr>
<tr>
<td>3.</td>
<td>Part III: Country Examples</td>
<td>59</td>
</tr>
<tr>
<td>3.1.</td>
<td>Introduction to case study groups</td>
<td>59</td>
</tr>
<tr>
<td>3.1.1.</td>
<td>Female Genital Mutilation</td>
<td>59</td>
</tr>
<tr>
<td>3.1.2.</td>
<td>Join-in-Circuit</td>
<td>59</td>
</tr>
<tr>
<td>3.1.3.</td>
<td>Theater for change</td>
<td>60</td>
</tr>
<tr>
<td>3.1.4.</td>
<td>Life Choices “3 ½ lives of Philip Wetu”</td>
<td>60</td>
</tr>
<tr>
<td>3.2.</td>
<td>Discussing the M&amp;E system in place</td>
<td>60</td>
</tr>
<tr>
<td>3.2.1.</td>
<td>Female Genital Mutilation</td>
<td>60</td>
</tr>
<tr>
<td>3.2.2.</td>
<td>Join-in-Circuit</td>
<td>62</td>
</tr>
<tr>
<td>3.2.3.</td>
<td>Theater for change</td>
<td>63</td>
</tr>
<tr>
<td>3.2.4.</td>
<td>Life Choices “3 ½ lives of Philip Wetu”</td>
<td>65</td>
</tr>
</tbody>
</table>
3.3. Developing an evaluation plan ................................................................. 66
  3.3.1. Female Genital Mutilation ................................................................. 66
  3.3.2. Join-in-Circuit ................................................................................... 68
  3.3.3. Theater for change ........................................................................... 68

Bibliography .................................................................................................. 71
Appendix 1: Monitoring and evaluation pipeline ............................................ 77
Appendix 2: Cluster Sampling ....................................................................... 78
Appendix 3: Guidance questions ................................................................. 79
Abbreviations ............................................................................................... 82
Glossary ......................................................................................................... 83
Executive Summary

Behavior Change Communication (BCC) interventions have become an integral part of many HIV prevention programs. Monitoring and evaluation is expected to be included in the design of any BCC interventions from the beginning on. However, the know-how on how to demonstrate results and impacts of such interventions in a methodologically sound way often lags behind.

This document aims at enhancing program officials’ understanding of the importance of evaluations as well as the most important steps to take and decisions to make in the course of an evaluation, and at enabling them to oversee the work done by the evaluators.

It is divided into three parts. The first part presents the main general aspects that need to be taken into consideration when monitoring and evaluating BCC interventions with a focus on HIV-related BCC approaches. More specifically, a general framework for monitoring and evaluation is presented, including among others information on the use of appropriate indicators, the development of an evaluation plan, different types of evaluation, study designs of summative evaluations, statistical analysis and dissemination of the evaluation results.

Based on this theoretical background presented in part one, the second part draws conclusions, presents recommendations and gives practical advice on monitoring and evaluating HIV BCC programs. Three types of HIV BCC interventions, relevant to the work of German Development Cooperation (GDC), are also here discussed; these are the tool Join-in-Circuit, peer education programs and mass media campaigns.

The third part presents four country examples of M&E of HIV BCC interventions implemented by GDC programs. These were discussed during a short-course on evaluation of BBC interventions offered by GIZ from 9-10 March, 2012 in Johannesburg, South Africa.
1. Part I: Monitoring and Evaluation of Behavior Change Communication Interventions

1.1. Background Information on Behavior Change Communication

1.1.1. What is Behavior Change Communication?

**Behavior Change Communication (BCC)** is a participatory and interactive process of working with individuals, communities, and societies to develop tailored messages and approaches using a variety of communication channels in order to enable them adopt health-promoting behaviors, as well as develop and maintain individual, community and societal behavior change (FHI, 2002).

Communication interventions can be carried out through training, radio, television, drama, print material, group work, individual counseling or one-to-one education (AIDSCAP, n.d.). Behavior change strategies need to be combinations of approaches sustained over a long time period at multiple levels of influence. The effective mix varies by the health problem addressed (e.g. the HIV transmission dynamics) and depends on the profile of the populations engaging in risky activities (Coates, Richter, & Caceres, 2008).

This guidance uses BCC interventions for HIV prevention as examples. However, most recommendations are valid also for other conditions.

In order to encourage individuals and communities to adopt and sustain behaviors and attitudes that reduce their risk of HIV infection, they must first understand basic facts about HIV, learn and test a set of appropriate life-skills, appreciate the advantages of adopting and maintaining certain behaviors and attitudes and be given access to appropriate health products and quality health and HIV services. An enabling environment in terms of policies, strategies and programs, where individuals have the choice and are able to change individual behaviors which increase HIV risks, is crucial.

Examples of BCC interventions commonly used in German Development Cooperation (GDC) in the areas of HIV and sexual and reproductive health and rights (SRHR) are:

- peer education
- mass media campaigns
- voluntary counseling and testing (VCT) programs and outreaches
- social marketing initiatives\(^1\) (e.g. male and female condom promotion)
- intergenerational dialogue around HIV and SRHR issues\(^2\)
- community dialogue and mobilization

---


- interpersonal communication on HIV and SRHR topics³ (for example through the use of role models)
- training of people living with HIV (PLHIV) or training of religious leaders on HIV and sexuality
- the tool Join-in-Circuit (JiC)⁴
- entertainment education⁵ (like soap operas on HIV)
- advocacy activities⁶ and
- HIV-related information education communication (IEC) material (NAP+, n.d.), and others.

1.1.2. Which are the goals and objectives of Behavior Change Communication?

BCC goals need to derive from the overall goals and the specific behavior change goals of a prevention program. They are related to specific issues identified after “assessing the situation, knowledge, attitudes and skills that may need to be transformed in order to work toward behavior change and program goals” (FHI, 2002, p. 11).

In the case of HIV, effective BCC should aim at (FHI, 2002):

- **Increasing HIV knowledge** through the application of visual or any other appropriate medium and by employing proper language that is age-, gender- and culture-sensitive.

- **Stimulating community dialogue** around the drivers of the HIV epidemic, such as risk behaviors and risk settings, environments and cultural practices related to sex and sexuality and marginalized practices (such as drug use or commercial sex work). It can also stimulate discussions around healthcare-seeking behaviors, HIV prevention, care and support and services available and/or needed etc.


- **Engaging youth in participatory research and evaluation**

  “Although participatory research has been applied by a wide range of disciplines, the engagement of youth as partners in research and evaluation efforts is relatively new. The positive youth development movement has influenced scholars and practitioners to include youth as partners in the design and implementation of research involving issues that affect their lives. Engaging youth in research and evaluation not only generates useful knowledge for communities and individuals but also provides opportunities for the development and empowerment of youth participants, leading to benefits for young people, organizations, the broader community and the research process. However, there has been little systematic study to establish an evidence base for these effects. [Powers and Tiffany] describe four projects that illustrate active youth participation in research. These examples demonstrate opportunities for positive youth development, create a context for intergenerational partnerships, and generate research findings to inform future interventions and organizational improvements, including community mobilization”.

  (Powers & Tiffany, 2006, p. 19)

---

7 Discordant are the couples with different HIV status (one HIV positive and the other HIV negative). Concordant are the couples with the same HIV status (either positive or negative).
1.1.3. Which are the guiding principles of Behavior Change Communication interventions?

The guiding principles of BCC interventions are the following (FHI, 2002):

- During the design phase of a program, BCC should be in line with the program’s logic framework\(^8\) and integrated with its goals. Stakeholders\(^9\) should be actively involved from the design stage of the planned BCC interventions on (see example “Engaging youth in participatory research”).

- Participants’ involvement is considered highly effective in all phases of BCC interventions (FHI/AIDSCAP, 2010). The target population(s) should be included in the development and implementation of BCC, as well as in its monitoring and evaluation.

- A formative BCC evaluation\(^10\) should be conducted prior to program design to promote understanding of the needs of target population(s), as well as of the possible threats to and opportunities for behavior change. Apart from the target group(s), formative evaluations should also include other stakeholders, service providers and the community.

- Having a variety of linked communication channels is generally more effective than relying on a single one.

- Planning for monitoring and evaluation should be an integral part of the design of any BCC program.

- Adaptation to the specific context and pre-testing is essential for developing effective BCC materials.

1.1.4. The role of BCC interventions in HIV prevention

In the 1980s, HIV research was mainly focused on behavioral variations across geographically and socially defined populations. Behavior change was seen as the primary means of controlling HIV epidemics around the world. Since then, there has been a shift in attention towards biological mechanisms, especially other sexually transmitted infections (STI), antiretroviral treatment (ART) and male circumcision, which currently shapes the discussion on HIV epidemiology and prevention (Bingenheimer & Geronimus, 2009). However, advances in scaling up ART, the benefits of male circumcision and the hopes and expectations around pre- and post-exposure prophylaxis and microbicides do not render behavioral strategies obsolete (Coates, Richter, & Caceres, 2008).

---

\(^8\) Management tool used to improve the design of interventions, most often at the program level. It involves identifying strategic elements (inputs, outputs, outcomes, impact) and their causal relationships, indicators and the assumptions or risks that may influence success and failure.

\(^9\) Stakeholder is a person or group with an interest in the outcome of an intervention.

\(^10\) See also chapter 3.4.1.1 on formative evaluation.
Bingenheimer and Geronimus (2009) argue that there is a need to adopt a conceptual model that explicitly recognizes the dynamic interdependence between behavior and other epidemiological and demographic parameters in HIV epidemiology and prevention. This means that although HIV-related behavioral strategies alone are not sufficient in reducing HIV transmission, they are essential in comprehensive HIV prevention strategies.

National HIV programs in partner countries will have to identify the array of behavior change needs and communication strategies throughout the country and within different target groups. Resources can then be devoted to the development of programs specific to each cultural or behavioral group at high risk of HIV infection. Consequently, HIV-related BCC interventions supported by development cooperation must contribute to the national HIV prevention strategies and programs of a country.

As already addressed, HIV behavior change strategies themselves need to be combinations of approaches sustained over a long time period at multiple levels of influence. In terms of intervention form and content, clearly, a “one size fits all” solution is not applicable. The effective mix varies by HIV transmission dynamics and depends on the profile of the populations engaging in risky activities, among which HIV is spreading (Coates, Richter, & Caceres, 2008). It is thus imperative that effective BCC interventions are target group-specific, age-appropriate, as well as culture-and gender-sensitive (UNESCO, 2010). For example, HIV prevention BCC interventions targeting young people might not be appropriate for married couples, BCC interventions that worked for IDUs might not work for MSM, and BCC interventions that were effective in country/region “x” might not be suitable for country/region “y”.

BCC target groups can vary depending on the core mandate and the partners of a specific HIV prevention program. But more importantly, BCC target groups may vary depending on the type and drivers of the HIV epidemic in a given country or region. For example, in concentrated epidemics, HIV prevention BCC interventions should be tailored to the needs of groups that are marginalized and particularly vulnerable to HIV infections, and they should go along with efforts to reduce stigma and discrimination and to promote gender equality. In generalized epidemics, where everyone is at risk, tailored HIV prevention

---

11 An epidemic is characterized as concentrated when HIV prevalence is consistently over 5% in at least one defined sub-population and below 1% in pregnant women in urban areas (UNAIDS, 2005). Asia’s epidemics remain concentrated largely among IDUs, CSWs and their clients and MSM. Almost one in five (18%) surveyed female CSWs in Myanmar tested HIV-positive in the mid-2000s. In southern India, up to 15% of female CSWs were living with HIV. In studies in Myanmar, up to 38% of IDUs have tested HIV-positive; this is estimated to be 30% - 50% in Thailand and more than half in parts of Indonesia. In China, an estimated 7% - 13% of IDUs are living with HIV. High HIV prevalence among MSM has been reported in several countries: 29% in Myanmar, 5% nationally in Indonesia, 6% in the Laotian capital of Vientiane, between 7% and 18% in parts of southern India and 9% in rural parts of Tamil Nadu state in India. The epidemic among MSM in Thailand had been largely ignored until a study uncovered 17% prevalence in Bangkok in 2005. Subsequent studies in 2005 and 2007 found that the infection levels had risen to 28% and 31% and an annual HIV incidence of 5.5% was recorded in 2008 (UNAIDS, 2010).

12 In a generalized epidemic, HIV is firmly established in the general population (>1%). Although sub-populations at high risk may continue to contribute disproportionately to the spread of HIV, sexual networking in the general population is sufficient to sustain an epidemic (UNAIDS, 2005). Sub-Saharan Africa (SSA) still bears an inordinate share of the global HIV burden. Examples of countries in SSA with a HIV generalized epidemic are among others Uganda (6.5%), Malawi (11%), South Africa (17.8%), Lesotho (23.6%) and Botswana (24.8%). However, the trends
efforts should be complemented with BCC interventions that build and strengthen HIV knowledge, skills and health-promoting behaviors based on approaches which are appropriate for the different target groups, their special characteristics, needs and vulnerabilities (UNAIDS, 2007).

At the same time, it makes sense to explore interventions that have the potential to change not only the individual’s behavior but also the social and sexual norms within the wider community. For example, evidence on HIV prevention among youth showed that HIV interventions (including BCC interventions) which target the community as a whole, rather than just young people, are more effective at improving reported sexual risk behavior and impacting biological outcomes (Mavedzenge, Doyle, & Ross, 2010).

Last but not least, BCC can and should be used not only for initiating changes in behavior, but to also for behavior maintenance, as evidence has shown that reductions in high-risk behavior and HIV incidence are very difficult to sustain (Stall, Ekstrand, Pollack, McKusick, & Coates, 1990). To give an example, HIV prevalence in Uganda declined from a peak of 18% in 1992 to 6.1% in 2002. This trend may have been ended accompanied by worsening of behavioral indicators after 2002, especially an increase in multiple concurrent partnerships (UNGASS Country Progress Report Uganda March 2010).

### 1.2. Evaluation of Behavior Change Communication Interventions

#### 1.2.1. Why evaluation?

According to the OECD Development Assistance Committee (DAC), evaluation is the “systematic and objective assessment of an on-going or completed project, program or policy, its design, implementation and results. The aim is to determine the relevance and fulfillment of objectives, development efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision–making process of both recipients and donors.” (OECD-DAC, 2002, p. 21 f.)

As implied by this definition, evaluations may serve a number of different purposes in the life-cycle of a project or program (Rossi, Lipsey, & Freeman, 2003). In a formative sense, evaluations may be used to assess needs for interventions, choose between alternative interventions options, and set realistic goals. They can guide and, if necessary, redirect the implementation process to ensure program success. In a summative sense, evaluations may provide information on whether the program goals were attained in an efficient and cost-effective way. Evaluations might also hint at whether programs have contributed to any higher-order goals (the MDG’s, for instance). Evaluations are thus very valuable tools to project teams as they might facilitate project steering and improvement and increase accountability. Policy-makers and donors profit from evaluation results in policy and budget decision-making processes. Lastly, society as a whole benefits from an increased knowledge base.

In Eastern Europe and Central Asia are alarming. In Russian Federation and Ukraine the HIV prevalence is 1% or higher and experts are concerned that the epidemic continues to outpace the response (UNAIDS, 2010).
BCC interventions base on the idea that increased knowledge and awareness will translate into changes in behavior. Unfortunately, there usually is no linear relationship between knowledge and behavior change. Sexual behavior, for instance, is obviously influenced and shaped by many factors (which are not always clearly evident) and exposure to HIV prevention programs is only one of them. The question of whether behavior changes can be attributed to prevention programs - especially in countries with advanced and generalized HIV epidemics - tends to create some friction between the different stakeholders involved in a BCC intervention.

From a narrow public health perspective, it may not matter which particular intervention caused a desired positive change in health outcomes or health behavior such as safer sexual behavior. In fact, in some cases, when populations are exposed to multiple interventions, it may even be very difficult to identify the influence of single factors. However, from the cost-effectiveness or policy perspective, it is important to determine what caused the observed changes in sexual behavior. Only if favorable changes are attributable to the intervention in question, resource input can be labeled as effective. If the changes would have occurred without a particular intervention, the costs of the intervention could be considered as resources better spent on something more useful. Taking into consideration the current financial situation of many donor countries, prevention programs are under growing pressure to determine which approaches work best for specific target populations in different epidemiologic settings with a given level of inputs in order to (re)allocate resources in a cost-effective manner. In addition, to transfer effective prevention programs to other countries or regions, the success factors have to be well known.

1.2.2. Monitoring versus evaluation

In practical use, monitoring and evaluation are often treated as very different concepts. However, the two techniques are in fact closely related as monitoring data often play a very important role in an evaluation’s attempt to explain how and why programs were effective or not.

As defined above, evaluation is a one-time or periodic “systematic and objective assessment of an on-going or completed project, program or policy, its design, implementation and results” (OECD-DAC, 2002, p. 21).

Monitoring, in contrast, is part of the ongoing management of communication activities and it usually focuses on the process of implementation. It is a “continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds” (OECD-DAC, 2002, p. 27 f.)
So, while monitoring is as a **continuous process** that consists of overseeing the proper execution of the program’s activities, evaluation is a **periodic activity** that involves assessing whether stated objectives have been fully or partially achieved, or not and reveals the reasons for discrepancies between the actual and forecasted levels of achievement of objectives. See **appendix 1** for a graphic display of the relationship between monitoring and evaluation.

The following are factors often monitored in BCC interventions (FHI, 2006):

- **Reach**: Is an adequate part of the target group being reached over time?
- **Coordination**: Are messages adequately coordinated with service and supply delivery and with other communication activities? Are communication activities taking place on schedule and at the planned frequency?
- **Scope**: Is communication effectively integrated with the necessary range of audiences, issues and services?
- **Quality**: What is the quality of communication (messages, media and channels)?
- **Feedback**: Are the changing needs of target populations being captured?

In the following, we will focus primarily on evaluation. However, the principles of social science research discussed apply for monitoring as well. We will first provide a brief overview over the general evaluation framework used in GDC, and describe common types of evaluations and their respective areas of application. For an in-depth discussion of evaluation theory and frameworks, we recommend, among others, the text books by Rossi et al. (2003) and Stockmann and Meyer (2010) that, if not indicated otherwise, we also base our overview on. In addition, the reader might find it helpful to take a look at the OECD-DAC documents referred to in the text.

### 1.2.3. The OECD/DAC general monitoring and evaluation framework and its application in German development cooperation

The most widely applied evaluation framework in development cooperation is the one designed by the OECD Development Assistance Committee (DAC). The OECD-DAC framework clearly distinguishes between one input and three results levels (OECD-DAC, 2010).

- **Program inputs** refer to financial, human and material resources invested in a development program.
- **Program outputs** are the products, capital goods and services which directly result from a development program; for instance, a BCC intervention intended to teach about risky sexual behavior is likely to “produce” increased knowledge on such behavior and its consequences among participants.
Program outcomes refer to the short-term and medium-term effects of a program. In the above example, reduced risky sexual behavior among participants might be one expected outcome of the BCC intervention.

Program impacts are long-term, higher-level effects of a development program. Impacts can be positive or negative, intended or unintended, direct or indirect. In the health sector, long-term results (impact) are expected to be achieved in a period of 5-10 years. For example, a BCC intervention might be expected to contribute positively to HIV incidence rates (via a reduction in risky sexual behavior which, in turn, reduces infection rates). Other common indicators of impact in the area of HIV/AIDS are HIV prevalence, AIDS-related mortality rate, number of AIDS-orphans, poverty alleviation, increased productivity, live expectancy of those with HIV infection etc. (OECD-DAC, 2002).

Depending on the program goals, the distinction between outputs and outcomes is not always entirely clear. Aspects defined as outputs in one intervention logic framework might be regarded as outcomes in another.

The above framework and its elements may seem familiar to you as they are incorporated in the general GDC results model (logic framework):

In addition to inputs, outputs, outcomes and impacts, the GDC results model defines two process-oriented elements: Activities are actions taken or work performed through which inputs are mobilized to produce specific outputs. Use of Outputs refers to the usage of the outputs of a project or program to produce outcomes (GTZ, 2008). While “activities” is an internationally accepted dimension, “use of outputs” is not and should not be used any longer. Most items that have previously been subsumed under “use of outputs” are now being attributed to the outcome level.

It is usually quite difficult to attribute changes on the impact level to single projects or programs. Rather, projects or programs interplay with a variety of other programs as well as numerous other intervening factors. In the GDC evaluation framework, this problem is represented by the so-called “attribution gap”, and there is an on-going debate both on a national and international level as to what extent high-level impact assessment of development interventions can be meaningfully done and interpreted (Caspari & Barbu, 2008).
In addition to the general evaluation framework, the OECD DAC has defined five overall criteria that every program evaluation in development cooperation should address:

- Relevance (“Are we doing the right thing?”)
- Effectiveness (“Are we achieving the project and/or program objectives?”)
- Efficiency (“Is our action cost-effective?”)
- Impact (“Are we contributing to overarching development objectives?”)
- Sustainability (“Are the results durable?”) (OECD-DAC, 2002).

With one single evaluation design, it is impossible to address all five criteria. Most BCC programs supported by GDC so far restrict their evaluation efforts to ensuring high-quality program implementation (activities and output level), and rarely ever address the outcome and impact level. In order to make outcome or even impact attributions possible, methodologically sound evaluation designs are necessary. This guidance note attempts to give you an idea of the most important aspects to consider when planning or commissioning a high-quality evaluation study. To ensure that evaluation results will be strong and useful, expert help in designing and conducting evaluations is generally strongly recommended. Such a study is a good opportunity for capacity development.

It has to be kept in mind that the main task of GDC staff is to manage and commission evaluations and to not directly implement evaluations by themselves. In addition, not every BCC intervention needs a high-quality evaluation study. If there is enough sound evidence on the effectiveness of specific BCC methods in different settings, monitoring by means of valid indicators is sufficient. For innovative BCC interventions, in case of lacking evidence or considerable uncertainty as to its effectiveness in specific settings there are strong arguments if not an obligation to set aside some money for a sound evaluation study with the results being accessible to the wider public.

1.3. Different Types of Evaluation and their Application

1.3.1. Formative evaluations

1.3.1.1. Ex-ante Evaluations / Needs Analysis

Research into intended target populations’ culture, lifestyle, behaviors and motivations, interests, and needs is a key component to a health communication program’s success.

An ex-ante evaluation should thus be conducted during the planning stage of a BCC program to identify and resolve intervention and evaluation challenges before the program is widely implemented. Ex-ante evaluations (sometimes called formative research; usually contain, but are not limited to, needs assessments) help to understand the interests, attributes and motivations of the target populations and other stakeholders in their community. They explore the need for interventions, provide the information necessary to define realistic goals and objectives for the intervention programs, and then help program planners make

---

13 See [http://www.oecd.org/dataoecd/15/21/39119068.pdf](http://www.oecd.org/dataoecd/15/21/39119068.pdf) for an overview over the main questions to be addressed; a German version can be found on the BMZ website: [http://www.bmz.de/de/zentrales_downloadarchiv/erfolg_und_kontrolle/evaluierungskriterien.pdf](http://www.bmz.de/de/zentrales_downloadarchiv/erfolg_und_kontrolle/evaluierungskriterien.pdf)
tentative decisions about effective, feasible intervention strategies and how to carry them out (FHI, 2006) (Commission, 2001).

For example, formative research was conducted in communities in China, India, Peru, Russia and Zimbabwe to obtain information about the social and cultural factors related to health behaviors influencing HIV and STI transmission with the aim of adapting a peer education intervention (Community Popular Opinion Leader) in culturally appropriate ways. The methods used were field observations, focus groups, in-depth interviews with key informants and an observation of community social dynamics (NIMH Collaborative HIV/STD Prevention Trial Group, 2007). These will be explained in more depth later.

Another example comes from the Wits Reproductive Health HIV Institute (WRHI); a Conditional Cash Transfer (CCT) study to keep young girls in school for HIV protection conducted a formative evaluation through a school survey, interviews with the educators as well as interviews with young women who have dropped out of school.\(^\text{14}\)

### 1.3.1.2. Process evaluation

**Process evaluations** concentrate on the input, activities and outputs levels of the GDC results model and are undertaken once the implementation phases of BCC interventions start. **They aim at assessing whether the interventions are actually being carried out as planned.** They monitor inputs and outputs and assess the quality of service delivery. Process evaluation are ideally repeated periodically throughout the course of the program implementation – depending on the scope of the project - and should use different methodological approaches, including reviews of service records and regular monitoring and reporting systems, key informant interviews, exit interviews of service users, direct observations, quantitative population-based surveys to assess program coverage and barriers to service use, and others.

**Process evaluations** can be used as a management tool in steering and possible readjustment of projects and programs. In addition, an understanding of the processes through which intervention activities achieve effects can help to explain the outcomes (or lack thereof) of the intervention. Process evaluations, however, do not reveal whether interventions were in fact effective, but only whether they have been carried out as planned.

---

1.3.2. Summative evaluations

Summative evaluations are mostly done for accountability reasons or to inform decision-making processes as to whether programs are worthwhile to continue.

Effectiveness evaluations pertain to the output, outcome and impact levels of the GDC results model. They assess the extent to which the objectives of a program were achieved and whether or not the interventions implemented did make a difference in terms of the program aims. They might also try to establish whether any other, non-intended results have followed from the program. An effectiveness evaluation is generally based on clearly defined standards of goal achievement (“indicators”) that provide quantitative values to which the output, outcome and impact of interventions can be compared (FHI, 2006). When
judging achieved results, time plays an important role. It is advisable to stratify by **short-term and intermediate program effects** (program output and outcome) and **long-term program effects** (program impact) (FHI, 2006). For example, changes in HIV-related beliefs and attitudes, the reduction of risk behaviors and instead the adoption of protective behaviors, as well as the reduction in sexually transmitted infections (STI) incidence rates (during a defined time period acquired new STI infections as a proxy for new HIV infections) are considered to be the most appropriate short-term or intermediate (also called proximate) outcome measures for interventions designed to reduce sexual transmission of HIV. Long-term effects include impact on HIV trends (e.g. HIV prevalence and incidence), sustainability issues, improved societal response and enabling policy environment (FHI, 2006).

Until now, only a few **BCC interventions with proven effects on HIV incidence and prevalence** are documented. Cowan et al. have conducted one of the first rigorous evaluations of a community-based HIV prevention intervention for young people in southern Africa. The objective of the study was to assess the effectiveness of the intervention in terms of its impact on (i) HIV and Herpes simplex virus type 2 (HSV-2) incidence as well as on rates of unintended pregnancy and (ii) reported sexual behavior, knowledge and attitudes (Cowan, et al., 2008). This study is an excellent example of the evidence-based information which is actually required to prove the likely effectiveness and costs of such programs to policy makers.

**Summative evaluations are closely linked to process evaluations.** The information provided by process evaluations can facilitate understanding on how and why interventions have achieved their effects or failed to do so, and they can make suggestions on what is actually making the difference. By examining outcomes and impacts without looking at the processes underlying these outcomes and impacts, one could be led to erroneous conclusions regarding the effectiveness of the intervention (FHI, 2006).

**Cost-effectiveness analysis (CEA)** is a form of economic analysis that compares the relative costs and outcomes (effects) of two or more courses of action. It is distinct from cost-benefit analysis, which assigns a monetary value to the measure of effect (Bleichrodt & Quiggin, 1999). In other words, CEA also measures program effectiveness, but expands the analysis by adding a measure of program cost per unit of effect (for example, per number of HIV infections averted). By comparing the costs and consequences of various interventions, cost analysis and cost effectiveness estimates can assist in priority setting, resource allocation decisions and program design. The most commonly used outcomes of a CEA are: Cost per unit of behavior or change; cost per HIV infection averted; cost of quality adjusted life years
(QALY) and disability adjusted life years (DALY)\textsuperscript{15} (FHI, 2006). While certainly useful under specific circumstances, CEA’s are rarely found in GDC evaluations for reasons of practicality and resources, and because at least two interventions need to be compared.

\textit{Table 1: Types of evaluations}

<table>
<thead>
<tr>
<th>Type of evaluation</th>
<th>Contribution of different types of evaluation</th>
<th>Sample of questions answered by different types of evaluation</th>
</tr>
</thead>
</table>
| A. Ex-ante evaluations / Needs analysis | Conduct situation analysis and determine the concept and design of the evaluation | ▪ Is an intervention needed/appropriate/cultural sensitive?  
▪ Who is targeted by this intervention?  
▪ What is the profile of the target group?  
▪ Which other stakeholders are involved?  
▪ What aspects should the program take into consideration during implementation? |
| B. Process Evaluations | Monitor inputs and activities and assess service quality | ▪ To what extent are planned activities being carried out?  
▪ Are planned activities implemented within the agreed time frame and budget?  
▪ How well are the services provided?  
▪ To how many people are the services provided? |
| C. Summative evaluations | Assess output, outcome and impact | ▪ Were the objectives of the program achieved?  
▪ What outcomes were observed?  
▪ What do the detected outcomes mean?  
▪ Does the program make a difference?  
▪ Which factors hinder/promote the realization of outcomes? |

\textsuperscript{15} For more information also refer to (FHI, 2006) chapter 17 “Guidelines for Performing Cost-Effectiveness Analysis of HIV/AIDS Prevention and Care Programs”.
1.3.3. Differentiating evaluation efforts: When to perform which evaluation?

Evaluations can be expensive. Due to constraints of time, funds and staff it is thus necessary to balance the ideal against the feasible and useful.

Single BCC interventions (individual projects) do not necessarily require a rigorous research design to judge their effectiveness, unless a new intervention is piloted or an unanswered research question is being addressed.

Individual projects using standard intervention strategies that have already been shown to be effective in other similar settings should focus their evaluation activities on formative evaluation.

New and/or innovative interventions, however, or interventions without well-established evidence base should be evaluated with regard to their outcomes and possibly their impact in addition to formative evaluations.

In a situation in which multiple donors are conducting or supporting multiple interventions with overlapping target groups, evaluations should be coordinated and conducted by country or regional programs. Such comprehensive evaluation approaches are not only efficient in monetary terms, but also make much sense in environments where the effects of individual projects from different donors cannot be determined.

In the context of sector-wide approaches (SWAp), donors have generally also committed to supporting one national health program with one result framework and are increasingly conducting or striving for joint evaluations regarding specific sub-program (e.g. safe motherhood) or health systems strengthening processes.
1.4. Planning a BCC Intervention Evaluation

After being clear on what kind of evaluation is to be done and what purpose it is to serve, the evaluation must be planned in further detail. In the following, we will outline the most important decisions to make and steps to take.

The University of Michigan Evaluation Center has collected and further developed a number of useful and easy-to-use evaluation checklists that might come in handy in planning and coordinating BCC intervention evaluations\(^{16}\). Some of these checklists have been translated into German by Wolfgang Beywl and his colleagues\(^{17}\). Also, we would like to recommend the UNFPA Tool Number 5: Planning and Managing an Evaluation\(^{18}\).

1.4.1. What is an evaluation plan?

An evaluation plan is a written document that describes the overall purpose(s) of the evaluation, the specific evaluation questions to be addressed, the evaluation design and methods to be used, what data is to be collected and how, how the data is to be analyzed, the resources that will be necessary, who will implement the evaluation, and the basic evaluation plan timeline (FHI, 2006). In order to increase perceived ownership of the evaluation results, and thereby their subsequent utilization, feedback and consent should be obtained from all major stakeholders both during the planning phase and after the evaluation plan has been drafted.

Key elements of an evaluation plan are the following:

- Scope of the evaluation
- Methodological approach
- Implementation plan
- Dissemination plan (FHI, 2006)

A list of the most important questions to ask oneself in developing an evaluation plan can be downloaded from the University of Michigan Evaluation Center’s website:  

---

\(^{16}\) http://www.wmich.edu/evalctr/  
\(^{17}\) http://www.univation.org/index.php?class=Calimero_Webpage&id=9023  
\(^{18}\) http://www.unfpa.org/monitoring/toolkit/5communi.pdf
The evaluation plan should be prepared through a joint effort of those who implement a BCC intervention and the evaluators. The evaluation plan is also meant to convince those who fund the evaluation (in general the GDC implementing organization) that the evaluation is going to be methodologically sound and worthwhile to be funded.

1.4.2. Who is to perform the evaluation?

Depending on the type and complexity of the evaluation project, technical capacity and evaluation expertise will be crucial in the personnel decision. As such capacity is usually not present in an HIV prevention program, technical and scientific support should be considered and requested from local universities, research groups etc.

The issue of credibility of the results might be important in certain situations. While formative evaluations might be performed by members of the project team themselves, evaluations with accountability purposes might better be done by “neutral” personnel external to the project team.

When commissioning an evaluation, possible evaluators and researchers applying for performing the evaluation need to demonstrate their methodological and management capacity. This is usually done by submitting CVs, mentioning previous evaluations and experiences and providing a list of publications. Prefer evaluators who provide peer reviewed publications on previously done evaluation studies!

1.4.3. What are the main evaluation questions?

The main questions to be answered by the evaluation vary by the purpose of the evaluation. In summative evaluations that are to investigate whether the program goals have been reached, the main evaluation questions usually derive from the logic framework in a rather straight-forward way. In formative evaluations, questions on target group needs, likelihood of program success, or the quality of implementation might be of interest.

Whatever the research interest, you should keep in mind that evaluations can only answer so many questions due to constraints in time, financial resources, etc. We strongly advise you to focus on only a few, but central questions to decision-making, and to try to answer these few questions in detail using strong research designs.

After the main evaluation questions have been determined, a cursory review of the literature might make sense for three reasons.
• **Firstly**, identifying whether any of the evaluation questions have already been (partly) answered by previous research might save a lot of time and resources. As development programs are usually tailored to the very specific needs of the target group, though, evaluation results obtained in different contexts should be handled with the appropriate care.

• **Secondly**, even if no such results apply to the program to be evaluated, previous research might provide valuable suggestions as to how to perform the evaluation.

• **Lastly**, research results unfortunately tend to depend to some degree on the methodology used (i.e., the study design, sampling strategy, data collection tools, statistical procedures, etc.). In order to be able to compare different research results focusing on similar subjects (e.g., the effectiveness of a BCC intervention in high HIV prevalence vs. low prevalence countries), methodological similarity is thus a very important aspect to keep in mind (see also section 1.8.4 on questionnaires).

We recommend performing a literature research by using scientific search engines, such as PubMed, MEDLINE, Medscape, Google Scholar etc. In addition, other research groups or programs that are possibly planning a similar evaluation should be identified and possibilities of collaboration to avoid duplication should be explored.

1.4.4. **Deciding on indicators**

Having settled on the main evaluation questions that are to be answered, indicators need to be chosen that will signal whether or not a program was relevant, effective, efficient, and had sustainable impact. Ideally, concrete indicators for program success have been defined in or along with the logic framework. More often than not, however, program goals are formulated less clearly, such as “increase in knowledge on HIV transmission”, for instance. What, however, is HIV knowledge, which are the relevant aspects? And what percent increase in relation to a baseline measure is deemed successful? As we feel like the issue of indicators and their appropriate choice and formulation is of central importance, and as there are many indicators that are now used as a global standard, we will elaborate more on it in section 1.5.

1.4.5. **Evaluation project management**

Before proceeding to the methodological design of the evaluation, a preliminary time frame and working schedule should be developed and recorded in the evaluation plan. Ideally, the study design appropriate to answer the evaluation questions should determine time and budget frames. In reality, however, things are usually the other way.

---

around so that one should have an idea of the approximate time and resources available before going into methodological detail. Do not underestimate the time necessary for questionnaire development, data collection, data analysis, and report writing, especially when employing a rather participatory approach where many different stakeholders are to give feedback and consent!

Figure 1 illustrates the major steps in planning an evaluation.

Figure 1: Steps to develop an evaluation plan

1.5. Indicators

1.5.1. What is an indicator?

An indicator is a “quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention or to help assess the performance of a development actor” (OECD-DAC, 2002).

Indicators are approximations of complex processes, events and trends. They can pertain to the tangible results (e.g. service uptake), the intangible results (e.g. community empowerment), the expected as well as the unanticipated results of a BCC intervention. As “operationalization” of program inputs, outputs, outcomes, or impacts, indicators signal as to whether programs were successful with regard to specific aspects.

The most straight-forward way to distinguish different types of indicators is according to their level in the results model:

- **Input indicators** track the means allocated for implementation of the activities such as financial and time resources, personnel, facilities, equipment and supplies.
- **Process indicators** follow up on the implementation of the program; that is, they track the activities in which the inputs are utilized, for instance in training, in establishing a network of NGOs, in planning of the service delivery etc.
- **Output indicators** point out the direct and immediate results of input and processes at the program level. Examples are availability of health services, or improved capacities of supported groups etc.
- **Outcome indicators** refer to the intermediate results at the target population level that are plausibly linked to the work of the program, e.g. increase in HIV knowledge and skills, condom use during last sexual intercourse with non-regular partner, median age at first sex of boys and girls.

- **Impact indicators** refer to the long-term results at the target population level that are plausibly linked to the work of the program, such as incidence and prevalence of HIV and/or STIs among adults (men and women) aged 15-49 in last year, the number of AIDS orphans etc. In national health or specific HIV programs, MDGs are often used to define impact indicators.

Examples of indicators at various levels can be found here:

[Handouts_1.pdf](Handouts_1.pdf)

To evaluate / measure an indicator for which no direct data collection is possible, a **proxy indicator** (an indirect measure) can be used. For example, a proxy indicator for condom use could be the prevalence of Sexually Transmitted Infections (STI).

1.5.2. **How to choose indicators**

In formulating the main evaluation questions, we have already limited the number of aspects of the project or program we’d like to address. From these questions and the underlying logical framework, a set of indicators to answer the evaluation questions needs to be derived.

In choosing indicators, it is important to again **limit the number to a set of critical indicators**. Too many indicators might create problems when analyzing and interpreting the results. Too few indicators might result in an incomplete picture that leaves out crucial aspects. With regard to the acceptance and utilization of the evaluation results, it is very important that all major stakeholders agree on the indicators chosen, and their specific formulation. Formative evaluation could be used to choose indicators.

In addition to the sheer number of indicators, programs should make sure to have a **good balance of different types of indicators**, which give a comprehensive overview of the process. For example, impact indicators only would not be sufficient to track the contributions of the BCC interventions of an HIV prevention program. More background information (provided by process, output and outcome indicators) is needed to be able to plausibly link the results observed to the interventions of the program. In turn, stand-alone output indicators would not be enough to demonstrate the results to the extent required to properly evaluate this program. For that outcome and impact indicators are additionally needed.
Because behavior and social change often take a lot of time to happen, sometimes indicators that indicate progress in reaching the ultimate outcome or impact are needed. In communication programs, for example, “intent to behavior change” has been used as predictor of actual change (UNICEF, 2006).

Table 2 displays a number of commonly used indicators in HIV prevention interventions.

**Table 2: Examples of HIV prevention intervention indicators by results level**

<table>
<thead>
<tr>
<th>Level</th>
<th>Indicator</th>
</tr>
</thead>
</table>
| **Inputs** | ▪ Financial resources allocated  
            ▪ Supplies available to the program (e.g. condoms)  
            ▪ Number of the technical staff available to the program                                                                                   |
| **Outputs**| ▪ Absolute number and proportion of the target group(s) that has been reached by the intervention (in mass media campaigns also: Percentage of target group who recall hearing or seeing a specific message)  
            ▪ Number of condoms distributed to the target groups (on central, regional and/or peripheral level)                                             |
| **Outcomes** | ▪ Percentage of target group who both correctly identify ways of preventing the sexual transmission of HIV and who reject the major misconceptions about HIV transmission  
            ▪ Percentage of target group with correct knowledge of prevention of mother to child transmission of HIV                                                                                     |
| Knowledge  | ▪ Accepting attitudes toward people living with HIV                                                                                         |
| Attitudes  | ▪ Condom use during last act with non-regular partner  
            ▪ Proportion of men and women with non-regular partner during last year  
            ▪ Median age at first sex of boys/ girls  
            ▪ Proportion of young women less than 18 years of age having had sex with                                                                         |

---

20 Surveys assessing such outcomes are often referred to as KAP surveys (knowledge, attitudes, practices).
28 men more than 30 years of age during last year

- Proportion of target group requesting an HIV test, receiving a test and receiving test results

**Impact**

<table>
<thead>
<tr>
<th>HIV incidence/prevalence</th>
<th>HIV prevalence among women less than 25 years of age in antenatal clinics (to approximate incidence)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HIV prevalence among high risk groups, e.g. STI patients, CSW, IDU, MSM</td>
</tr>
<tr>
<td></td>
<td>HIV prevalence among young people aged 15-24 and among adult men and women aged 15-49</td>
</tr>
</tbody>
</table>

**Overall program effort**

- AIDS Program Effort Index (API)\(^{21}\)

*Source: (FHI, 2006), (Global Fund, 2009), (UNAIDS, 2009)*

### 1.5.3. Clearly defining/stating indicators

An indicator should not only define the program aspect which is to be considered in judging program success (e.g. “enhanced knowledge of HIV protection”). In order for little to no room for interpretation to remain, it

- Should very specifically state its frame of reference (“knowledge of the three major ways x, y, and z of protecting oneself from an HIV infection”).
- Should clearly state the minimum target level for program success (e.g., “80% of the participants know the three major ways x, y, and z of protecting oneself from an HIV infection”).
- Should provide any other relevant information (e.g., time frame: “six months after the intervention”).
- Ideally, and if available, baseline information should be reported (e.g., “40% had such correct knowledge before the intervention”).

---

\(^{21}\) AIDS Program Effort Index is a composite index designed to measure political commitment and program effort in the areas of HIV prevention and care. It tries to capture many of the inputs and outputs of a national HIV program. The score is made up of 10 main components of an effective national response: political support, policy formulation, organizational structure, program resources, evaluation and research, legal and regulatory aspects, human rights, prevention programs, care programs and service availability ([http://www.measuredhs.com/hivdata/ind_detail.cfm?ind_id=1&prog_area_id=1](http://www.measuredhs.com/hivdata/ind_detail.cfm?ind_id=1&prog_area_id=1)).
Overall, indicators should be SMART, that is

- **Specific:** They should measure only the program aspect they are intended to measure.
- **Measurable:** They should be (easily) measurable and provide reliable data, independently of who is carrying out the measurement along time.
- **Achievable:** The target level set to indicate program success must be achievable. Establishing adequate goals requires special care. Goals which are too ambitious will eventually not be reached, while goals which are too easy to achieve can suggest apparent success, which nevertheless does not imply that the program made a difference.
- **Relevant:** The information provided by the indicator should be of central importance to decision making.
- **Time-bound:** Definite time periods must be specified that the indicator refers to (GTZ, 2007)

**Other characteristics of indicators** are the following:

- **Validity:** They should measure the condition or event they are intended to measure.
- **Affordability:** The costs of measuring the indicators should be within reason.
- **Feasibility:** It should be possible to carry out the proposed data collection.
- **Operationalizability:** It should be possible to measure or quantify indicators with developed and tested definitions and reference standards.
- **Reliability:** They should produce the same results when used more than once to measure the same condition or event.
- **Sensitivity:** They should be able to detect meaningful changes in the state of the condition or event under observation.

(FHI, 2006)

**1.6. Evaluation Study Design**

**1.6.1. Evaluation designs and causal inference**

Different evaluation study designs reflect different levels of certainty in concluding that observed changes in knowledge, attitudes, behavior, or epidemiological measures are attributable to the intervention program and are not the result of other confounding factors. Experimental designs applying randomized control groups provide the most tangible evidence of such causal relationships. The need for a rigorous study design, however, must be balanced with issues of cost, opportunity and feasibility. Ethical and political issues also often play a role.
In the following, we will give a brief overview over the most common study designs used in summative evaluation studies. We will then introduce different methods of data collection, and talk about sampling, before we get to the analysis and interpretation of the data. For more in-depth information, we recommend the text books by Schnell, Hill and Esser (2008), Kromrey (2009), and Bortz (2005) that we also used as background literature for this overview. In addition, we recommend the very practitioner-friendly, but rather lengthy guide on monitoring and evaluation of adolescent reproductive health programs by Adamchak, Bond, MacLaren, Magnani, Nelson and Seltzer (2000).

1.6.2. Randomized experiments

Of all study designs, randomized experiments provide the highest level of confidence that the observed outcomes are attributable to a specific intervention. In evaluations using this study design, participants are assigned by chance to a group that will receive an intervention (called the intervention group) or to a group that will not receive the intervention (called the control group).

This so-called random assignment is the best way to ensure that the members of the intervention and control groups do not differ systematically from each other in terms of crucial characteristics, such as age, education, family background, attitudes and any other factors that might influence the outcomes that the program.

There are two basic types of randomized experimental designs:

- The pretest-posttest control group design, where one
  - randomly assigns participants to intervention and control group
  - measures indicators both before and after the intervention, and
  - uses the difference between the changes in indicators for the intervention group and the control group ("difference in differences") as estimates for the intervention results.

The pretest-posttest control group design bases on the idea that since the control group is perfectly equal to the intervention group, any kind of influence external to the intervention that might have led to changes in the result parameters would have equally affected both the intervention and control group. In the context of HIV prevention programs, such external influences might be other, parallelly implemented interventions or information campaigns, for instance. By subtracting the changes observed in the control group from those observed in the intervention group, one can thus isolate the net effect of the intervention.

For practical reasons, “after” measurements are often done right after the intervention has ended. However, in order to establish that intervention results are sustainable, evaluations often include additional “after” measurements after appropriate amounts of time have passed.
Although this is the strongest of all designs, it does pose some challenges:

- The assumption of equality of control and intervention group does not always hold true, especially in small samples. In addition, when longer time periods between the intervention and follow-up measurement are involved, there is a chance that confounders have differentially influenced the intervention and control groups. For instance, in response to an intervention teaching HIV knowledge, participants of the intervention group might have sought additional information on the topic. While this is certainly desirable in terms of the program goals, the additional information seeking positively “contaminates” the net effect of the intervention.

- It is not always feasible to randomly assign subjects to intervention and control groups, for example when the project partner determines in which districts, municipalities, schools etc. the BCC interventions should be implemented or when the mandate of the HIV prevention program is restricted to specific geographical areas/regions.

- In order for randomized experiments to produce tangible results, experimental conditions need to be upheld throughout the program implementation. However, HIV prevention programs are implemented into real life conditions. For example, an HIV program might only be able to partially implement the BCC intervention due to political reasons, a shift in the interest of the partner, a budget cut down, or other external factors, such as natural disasters.

- Also, ethical considerations should not be overlooked. Although the program resources are often not sufficient to cover the entire target population, deliberately denying access to interventions to some people (control group) in order to enable a smooth evaluation of the program poses a challenge. A common solution to the problem is to offer the intervention to the control group at a later stage.

The posttest-only control group design, where one

- randomly assigns participants to experimental groups (just as in the pretest-posttest control group design)

- measures indicators only after the intervention and

- defines intervention results as the difference between results indicators for the intervention group and the control group

As the posttest-only control group design does not provide measures before the intervention, it is somewhat weaker than the pretest-posttest control group design in terms of conclusions about program effects as the net effect size cannot reliably be estimated. Again, due to ethical considerations in some cases it is a must to use a posttest-only design. For example, a program cannot deny or delay life-saving treatment, vaccination or access to condoms to the control group to have a stronger study design. Also, in certain situations where the pretest is assumed to have a substantial influence on the intervention effects, a posttest-only design might be favorable.
1.6.3. Quasi-Experiments

If for whatever reason, it is not feasible to randomly assign experimental groups, quasi-experimental designs can be used. Quasi-experiments apply groups similar to the experimental ones, but non-random selection methods are employed. Most often, control groups (then called comparison groups) are selected by matching them with the intervention groups on important characteristics. The non-random selection results in a reduced level of confidence that the observed outcomes are attributable to a specific intervention in comparison to randomized experiments.

In practice, age, sex, geographical location and other socioeconomic characteristics are often used as matching variables. From a methodological point of view, however, matching variables should be chosen on a theoretical and empirical basis in that they should reflect the most important potential external influences on the intervention results.

In comparison to true experiments, quasi-experiments provide weaker evidence that the results are in fact attributable to the intervention as even with the most elaborate matching procedure, the possibility remains that comparison and intervention groups systematically differ in important characteristics. Also, it is often simply impossible to match for all variables considered important. Statistical procedures exist, however, that take into consideration or control for differences in factors that could not be taken into account through matching.

Quasi-experiments can be applied as pretest-posttest or posttest only designs in the same way as randomized experiments.

1.6.4. Non-experimental designs

Non-experimental designs are routinely used when experimental or quasi-experimental designs are not feasible as, for instance, when

- no provisions were made for a control or comparison group during program planning and/or implementation or as part of the evaluation plan
- a program or intervention is expected to reach the entire target population at once (e.g. intensive mass media campaigns, national family life education)
- long-term impacts, such as HIV prevalence, are concerned.
Non-experimental do not assign participants to experimental groups, nor do they necessarily follow participants from before the intervention to after the intervention. Control groups might be constructed retrospectively, but a causal attribution of results to interventions is not possible as potential confounders that might have caused the observed behavior changes cannot be controlled. Nevertheless, non-experimental designs are useful, for instance, when results on highly aggregated levels are concerned that take long time periods to occur, or when the program target population is not clearly definable, such as in mass communication.

Types of observational studies are the following:

**Cohort studies** follow up two or more groups from exposure to outcome. They compare the experience of one group exposed to a factor (exposed group) with that of the other which was not exposed to the factor (control group). If the exposed group has a higher or lower frequency of an outcome than the unexposed, than an association between exposure and outcome is evident. Cohorts look at causes and natural history of diseases and are also useful for examining prognosis of people who already have the disease.\(^\text{22}\)

**Case-control studies** compare individuals who have a specific disease (cases) with a group of individuals without the disease (controls). The proportion of each group having a history of a particular exposure or characteristic of interest is then compared. An association between the hypothesized exposure and the disease being studied will be reflected in a greater proportion of the cases being exposed. It is advantageous for the controls to come from the same population from which the cases were derived, to reduce the chance that some other difference between the groups is accounting for the difference in the exposure that is under investigation. A case-control study generally depends on the collection of retrospective data, thus introducing the possibility of recall bias.\(^\text{23, 24}\)

**Cross-sectional studies** are studies in which disease and exposure status are measured simultaneously in a given population. Cross-sectional studies can be thought of as providing a “snapshot” of the frequency and characteristics of a disease in a population at a particular point in time. This type of data can be used to assess the prevalence of acute or chronic conditions in a population. However, since exposure and disease status are measured at the same point in time, it may not be possible to distinguish whether the exposure preceded or followed the disease and thus cause and effect relationships are not certain.\(^\text{25}\). To give an example, a cross-sectional survey implemented for the program, “Love Life” in South Africa showed that the more young people were exposed to the program, the less was their risk for HIV infection. However, the study design was not “strong” enough to prove a causal relationship. Lower HIV risk was rather associated to than caused by the “Love Life”

---

\(^\text{22}\) From [http://www.drcath.net/toolkit/cohort.html](http://www.drcath.net/toolkit/cohort.html)

\(^\text{23}\) Recall bias occurs when the way a survey respondent answers a question is affected not just by the correct answer, but also by the respondent’s memory.

\(^\text{24}\) From [http://www.ehib.org/faq.jsp?faq_key=34](http://www.ehib.org/faq.jsp?faq_key=34)

\(^\text{25}\) From [http://www.ehib.org/faq.jsp?faq_key=41](http://www.ehib.org/faq.jsp?faq_key=41)
program. It is possible that the young people who participated into the program were more interested, informed and thus less vulnerable to HIV infection anyway.

Unfortunately, the answer to which evaluation design works best is “it depends”, primarily on

- the questions your evaluation is to solve. Not every evaluation questions requires an experiment to be answered.
- the level of the results you are addressing – attributing results on the outcome or impact level to a program or a single intervention will require a methodologically stronger design (e.g. experiments or quasi-experiments) than results on the output level.
- the degree to which your results attributions need to be tangible (i.e., purpose of the evaluation). Methodologically stronger designs produce more tangible results.
- the situation and context in which the evaluation is performed: time, resources, qualified personnel, the political environment, ethical constraints, and so on will influence the degree to which strong designs are feasible and reasonable.

1.6.5. The role of triangulation

Particularly in the absence of an experimental design, triangulation plays an important role in substantiating links between the intervention and observed behavior changes. Triangulation can be achieved by using

- multiple methods to answer an evaluation question,
- multiple data sources, or
- several researchers interpreting the same data.

Results from behavioral quantitative surveys should, for example, be analyzed together with findings from qualitative interviews carried out in sub-samples of surveyed target population.

For instance, self-reports of sexual behaviors in the absence of additional evidence are often considered invalid and unreliable by stakeholders for whom such data are sensitive and run against firmly held cultural norms. Therefore, triangulation of behavioral data with other available and relevant data is required to explain more comprehensively the reasons of behavior change (FHI, 2006).
1.7. Data collection

Having decided on an evaluation design, data collection has to be planned and tools need be developed. In line with the above thoughts on triangulation, **BCC interventions should be evaluated by using both qualitative and quantitative methods.** Data collection may include any of the standard techniques used in social science research. A combination of research methods helps provide more data to work with, helps overcome the individual weaknesses of each method and ultimately leads to a more accurate evaluation. Regardless of the data collection method, programs should strive for sex- and age-disaggregated data.

1.7.1. Quantitative methods

**Quantitative methods** provide numerical information on quantifiable program aspects that can be analyzed with statistical procedures in a rather objective way. When concrete hypotheses on program effects exist, and when the establishment of goal achievement rather than an in-depth understanding is of main concern, quantitative methods are the methods of choice.

The most common quantitative techniques used in program evaluation are:

- **Structured interviews:** A trained interviewer poses a series of pre-established questions to all respondent in the same way and with the same order. The questions are often closed-ended with pre-coded responses. The instrument may also contain some open-ended questions, which are coded after all responses have been obtained.

  The **self-administered interview** is a variation on the structured interviews, in which the respondents complete a **questionnaire** themselves. The instrument consists of a series of questions, which can be closed (pre-coded) or open-ended. Self-administered questionnaires only work among populations that know how to read and write; the wording needs to be simple and the instructions clear. A problem with self-administered questionnaires is that the respondents may leave questions blank or fail to follow the instructions correctly. Furthermore, in many countries, the culture for mail-back questionnaires does not exist (Bertrand & Solís, 2000).

- **Service statistics** or other program data in a broad sense include any type of information that can be obtained from the national health management information system (HMIS), the logistics management information system (LMIS), other national data collection systems or the program records.

---

26 See also chapter 1.7.3. on questionnaires.
1.7.2. Qualitative methods

**Qualitative methods** usually dominate when formative aspects are in the foreground. They examine the reasons for why specific interventions were successful or not by enabling researchers to gain insight into attitudes, beliefs and motives for behavior change. Qualitative procedures usually involve a small number of correspondents who are not chosen on a probability basis. One should thus be careful in drawing conclusions from the results, especially on causal relationships. Three very common qualitative methods are focus group discussions, in-depth interviews and observation:

- **Focus group discussions** capitalize on group dynamics and allow a small group of correspondents to be guided by a skilled moderator into increasing levels of focus and depth on the key issues of the research topic.

- **In-depth interviews** are – like focus group discussions – characterized by extensive probing and open-ended questions, but they are conducted on a one-to-one basis between the correspondent and the interviewer.

- **Observation** includes, among others, direct observation, mystery client or ethnologic techniques such as illness narratives.

### “Mystery clients”

“In 1996, a qualitative study explored the changing dynamics of concurrent sexual partnerships, using a variety of researchers and methods. Although most international ethical research codes prescribe the informed consent of research subjects, the [...] author [...] included the mystery-client method, which omits informants’ consent. Five trained, pilot-tested, and closely supervised male researchers contacted 101 bar girls and “freelancing” women in trading and urban centers to assess the women’s ability to negotiate condom use. The men posed as clients but were instructed not to have sex with their informants. This approach provided important contextualized information to improve HIV transmission-prevention programs, yet it raises ethical concerns.”

*Source: van den Borne (2007, p. 322)*

1.7.3. A closer look at questionnaires

Compiling questionnaires can be a very challenging task, especially when attitudes rather than apparent behaviors are concerned. Generally, it is **recommended to use already developed questions and sets of questions** (‘scales’) which have been professionally developed and tested.

Using such validated instruments not only ensures high-quality data, but also allows for comparison of study results obtained in different contexts and settings. However, as development programs are often tailored very specifically to the target population, an adaptation of previously developed instruments both in content and wording is almost

---

27 Such professionally developed questionnaires can for instance be found in scientific journals (see section 1.4.3. for a list of search engines) or other published literature.
always necessary. The tools used should respect and be appropriate for the social, cultural, political and legal context within which the BCC intervention takes place. Respective background knowledge will help ask the right questions in an appropriate way.

The adaptation process can be challenging, and the support of an expert is recommended. Small variations in wording and the order and way of posing questions can greatly affect the respondents’ answers and consequently the findings of the evaluation. When administering questionnaires in different languages, questionnaires need to be translated and back-translated into the local language(s) to make sure that the real meaning will not be distorted and that appropriate culture-sensitive terminology is used. It is essential to pretest the tools after adaptation to every local setting.

For example, concepts such as “sexuality” or “gender roles” are socially and culturally constructed. A number of implications follow for the way a society or individuals understand and respond to sexuality and young people, for instance, or realize the role of gender in the risk of HIV infection. For example, how do we ask about homosexuality in a country where homosexuality is actually prohibited by law (e.g. in many countries in Sub-Saharan Africa or in the MENA region) or how do we address gender based violence in a country where the majority of women believe that their partner have a good reason to be violent against them (and this is for example the case in Uganda)? In a questionnaire one should start with the less sensitive questions.

Another example: A mission of the German BACKUP Initiative in the Mtwara region in Tanzania revealed that land and inheritance issues, early marriages and early sexual debut, as well as polygamy, traditional dances and initiation ceremonies that are practiced in the region increase the vulnerability of women and girls to HIV infection. In the context of an HIV prevention program that tackles the above mentioned issues, a questionnaire designed to explore the effects of relevant BCC interventions should address these topics.

In addition to culture-specific data collection instruments, it is also useful to develop guidelines for interviewers and supervisors to accompany the instruments. Such documents explain in detail the rationale behind each question and its intended meaning. They also aim at clarifying any ambiguities or misunderstandings that may arise and avoiding the interviewer’s bias.  

---

28 **Interviewer’s bias** is the distortion of the response to an interview, questionnaire etc. due to the personality and/or the personal views of the interviewer or to a particular way of the interviewer posing the questions.
Confidentiality of the respondent and their answers should be ensured and consent should be clearly given by the respondent. No respondent names or other identifying information should be recorded unless absolutely necessary. It has to be clear to respondents that their participation is voluntary and that they can refuse to answer questions or even the entire questionnaire (FHI, 2006). Naturally, the rule of strict confidentiality does not only apply to questionnaires, but also to any other method of data collection.

1.7.4. Sampling

Sampling is the process of selecting elements (e.g. people, households, schools, organizations) from a population\(^{29}\) of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen. Sampling is not to be confused with (random) assignment to experimental groups! While sampling deals with choosing study elements from all elements that the study pertains to (e.g., all potential targets of an intervention), assignment procedures are then used to create intervention and control groups from this sample.

1.7.4.1. Types of sampling methods

Sampling methods are classified as either probability or non-probability methods.

Probability sampling methods make sure that the sample is representative of the population in question. As such, they produce results that do not pertain to the study participants only, but generalize well to the entire target population. Probability sampling methods are mainly used in summative research, and may play an important role in needs analysis.

For instance, when reliable data on prevalence, incidence and such are concerned, probability sampling strategies are the strategies of choice. When trying to establish the effectiveness of an intervention in order to inform decision-making processes over further implementation, it might also be important to have results that generalize beyond the study sample.

However, while probability methods produce generalizable results, they might be costly, time-consuming and demand a high level of expertise.

\(^{29}\) In statistics, the term "population" is used to describe the sum of all elements a study refers to. For instance, the target population of a BCC intervention might be all young people who are potential targets of the intervention. Note that population elements are often human beings, but can be objects or immaterial elements as well.
Probability sampling methods require that:

- each population element has a chance of being chosen (note: chances for different elements do not have to be equal)
- for each population element, it is possible to calculate the probability of being selected, and
- random chance determines which population elements are chosen.

Probability methods include, among others:

- **In random sampling**, each member of the population has an equal and known probability of being selected. Accordingly, when applying a random sampling technique, a list of all population elements must be available or compiled. When sampling from large populations, but also from populations with sensitive characteristics, however, it is often difficult or impossible to identify every member of the population. As a result, the sample might become biased and results might be distorted.

- **Stratified sampling** can be superior to random sampling in certain situations. A stratum is a population subset, the members of which share at least one common characteristic of interest. Examples of strata might be males and females, or people living without and living with HIV. After the relevant strata and their representation in the population are identified, random sampling is used to select a number of elements from each stratum large enough that we can be confident they represent the population (see section 1.5.8.3 for details on effect and sample sizes). Statistical procedures exist to eliminate the bias introduced by stratification. Stratified sampling can reduce the required sample size (and thus time and costs) enormously when subpopulations with low prevalence of the characteristics in question are involved. Just as random sampling, however, “true” stratified sampling requires inclusive lists of elements of each stratum, which are often very difficult or impossible to obtain.

- **Cluster sampling** is a sampling technique where “naturally occurring” groups such as schools or health centers are selected for inclusion in the sample, most often in a two- or three-step procedure. Cluster sampling could be used to introduce HIV vaccines in schools and clinics. As this probability method of sampling is widely used, we will take a closer look at it in appendix 2.

**Non-probability methods** do not choose samples on a random basis and thus do not guarantee that the sample chosen represents the population of interest. As such, they are mostly used with qualitative methods.

Non-probability methods include, among others:
- **Convenience sampling** is often used in exploratory research\(^{10}\) where the researcher is interested in getting time- and cost-efficient estimates of the relationships of interest. As the name implies, convenience samples, also called opportunity samples, are compiled of elements which happen to be available at the time.

- In **judgment sampling**, the sample is selected based on the researcher’s individual expert judgment that it is representative of the population. While there is a risk that the expert is wrong in her or his judgment, judgment sampling is often preferable to convenience sampling when probability sampling is not possible.

- **Quota sampling** is the non-probability equivalent of stratified sampling. Like As in stratified sampling, strata and their representation in the population are identified. However, samples from the strata are not chosen on a probability basis, but by judgment or convenience.

- **Respondent driven sampling** is used for people hard to reach (e.g. sex workers, men who have sex with men). The use of vouchers is very crucial to be able to trace the chains of people (e.g. who was referred by whom to the program, which regions people come from etc.). This sampling method allows generalizing more than snowball sampling. **Snowball sampling** makes no use of vouchers and is a more informal method than the respondent driven.

- **Targeted snowball sampling** is a non-probability method often used when the target population is difficult to identify and reach, or when the characteristic of interest is rare in the general population (e.g. IDUs or CSWs). The data collection process typically begins by interviewing informants and target group members known to the researchers in order to learn of other target group members. The researcher then contacts these persons, collects data and obtains information on where additional target group members might be found. This process of referral is continued until a sample of pre-determined size has been achieved. While the technique can dramatically lower search costs, it comes at the expense of reducing the likelihood that the sample will be a good representation of the population.

**Non-probability methods** are more flexible, less costly and less time-consuming, but give potentially weaker evaluation results as it remains unclear whether and to which degree the results can be generalized to the entire target population. Nevertheless, non-probability methods can be very appropriate in certain situations, for instance, when formative aspects dominate the evaluation.

---

\(^{10}\) In contrast to hypotheses-driven research, exploratory research aims to gain insight into novel problems or situations for which a sound scientific base has yet to be established.
1.7.4.2. Sample size

Sample size is a key design parameter in any study because it is crucial in ensuring that meaningful changes introduced by an intervention will be detected in the course of the study. When anyhow possible, one should thus calculate an appropriate sample size before planning the sampling process. Unless statistical expertise is present in the program we highly recommend to seek external help both with sample size calculations and the statistical data analysis.

Generally, the probability of detecting an effect will increase with increasing sample size. Assembling samples as large as possible, however, is no ideal solution to the “power issue” both for practical and methodological reasons. From a practical point of view, large samples are time-consuming and costly, so that one will generally want to minimize the sample size. From a methodological perspective, the inherent characteristic of most statistical procedures to produce “statistically significant” results when only the sample size is large enough raises the question as to whether such results will also be meaningful in practical interpretation.

Upon starting on the sample size calculation, a statistician will likely ask for information on the following five factors (FHI, 2006):

1. The magnitude of the effect (effect size) that a program expects their intervention to achieve. Estimates can be found on the basis of previous research on similar interventions in similar contexts. Another common practice is to define the minimum intervention effect still large enough to qualify as meaningful in practical terms.

2. The initial or starting level of the indicator(s).

3. The probability with which a program wishes to be certain that an observed change did not occur by chance. Usually, this probability level is set at 95%, but there are situations in which one might want to be more certain (for instance, when potentially harmful medication is concerned).

4. The probability with which a program wishes to be sure that the actual change of the magnitude expected will be detected.

5. The relative frequency with which persons with the characteristics specified in a given indicator may be found in the target group population (see also section 1.7.4.1. on stratified sampling).

Based on the above information, the sample size needed to answer the specific evaluation questions can be calculated.

31 In statistics, the probability of detecting a change given that it has truly occurred is called “statistical power”.
If for whatever reason, it is not possible to estimate the required sample size, as a rule of thumb, go for samples as large as possible. Factors that tend to increase the required sample size are, among others:

- heterogeneity of the sample (in diverse relevant characteristics)
- large variation of the characteristic in question in the population
- intention of subgroup analysis
- number of variables included in simultaneous analysis.

As mentioned before, the sampling strategy chosen will affect the required sampling size, and certain statistical procedures come with minimum sample size requirements. In cluster sampling, a sample of more clusters of smaller size is always preferable to one with fewer clusters of larger size.

1.8. Statistical analysis and interpretation of data

1.8.1. General information

Once data has been collected, it needs to be analyzed to assess possible changes in indicators over time and to explore relationships between indicators.

A multitude of different statistical procedures is available, some of them simple, some more complicated and very specific for certain purposes. In most cases, few basic procedures deliver most of the information needed in development work. The following section attempts to provide a brief overview over the most important statistical terms and procedures, as well as their appropriate application and interpretation. For a deeper understanding, a vast variety of textbooks on statistics exist (Bortz (2005), for instance).

In statistics, the term “variable” is usually used to refer to a certain characteristic or attribute. For instance, variables might be “experimental group”, “age”, “number of condoms used”, or “HIV status”. Variables that are not inherently numerical (such as “experimental group”, as opposed to “age”) need to be coded (i.e., assigned numbers to) before the statistical analysis. Ideally, a coding scheme was already developed along with the questionnaire. Make sure that the coding process is meticulously documented as after the statistical analysis, numbers will have to be translated back into words! After all, statistics have no value in themselves, but only help you in interpreting and making sense of the information collected.

1.8.2. Common statistics and statistical procedures

1.8.2.1. Descriptive statistics

The most often used statistic to describe a variable is the mean. It does not, however, provide any information on how far the data points are spread out to both sides of the mean. The standard deviation should thus be calculated and reported to get an idea of how much “dispersion” from the mean there is in the sample. The t-test is the most commonly used method to test for differences in means between two groups. For example, the t-test can be used to test for a difference in HIV knowledge test scores between a group of young people who received a BCC intervention and a control group who did not receive such an
intervention. When the means of more than two groups are to be compared, analysis of variance (ANOVA) is usually being employed.

**Frequencies** are the numbers of occurrence of a certain states in a given time period, for instance, the number of males and females who participated in the survey. In addition to the absolute numbers, **percentages** should be reported for each state (i.e., each cell in the frequency table).

**Percentages** are superior to absolute values especially when the denominator is not constant (e.g., when population numbers change over time, or when two different populations are to be compared).

**Contingency tables** are a common method of displaying different subgroup frequencies in the variables of interest. As a very basic example, some population might be differentiated by gender and HIV status, resulting in a 2x2 matrix. Testing procedures exist that estimate whether there are statistically meaningful differences in subgroup frequencies (i.e., between the different cells of the contingency table).

**1.8.2.2. Measures of association**

**While the above measures pertain to single variables only, the following measures describe how two or more variables are related. You might have heard of the famous sentence “Correlation does not imply causality”. What it means to say is that when working with such measures of associations, it is of vital importance to remember that no statistical measure or procedure in itself is able to provide proof for a causal relationship! Rather, the study design determines whether statistical measures can be interpreted in causal terms (see section 1.6).**

Measures of **correlation** are most often used to describe the relationship between two or more variables. Correlation coefficients range from -1 to 1, with higher absolute values indicating a stronger (positive or negative) relationship.

Another procedure often used to estimate the association among two or more variables is called **regression** analysis. In contrast to the correlation coefficient, regression explicitly considers the direction of the relationship between variable A (e.g., a BCC intervention) and B (e.g., rate of condom use) as it is assumed in the logic framework. In its basic linear form, given that A is assumed to cause (a change in) B, regression analysis will estimate a linear equation (regression line) that predicts values of B from given values of A. Regression is commonly used with experimental study designs. Unfortunately, the regression coefficient (i.e., the slope of the regression line) cannot be interpreted as easily as a correlation.
coefficient. Rather, a statistical testing procedure will indicate whether variable A is a good predictor of (i.e., is strongly associated with) variable B.

Several complex statistical procedures based on linear regression exist. For instance, multiple or hierarchical regression simultaneously considers several predictor variables (e.g., “availability of condoms” and “gender” in addition the “BCC intervention”) when predicting an outcome variable (e.g. rate of condom use). Often, when you read that some influencing factors have been “controlled for”, multiple regression has been used. Structural equation modeling, or path analysis, is often used with complex theoretical models.

The following websites provide a brief and good overview of basic statistics:

http://www.fao.org/docrep/w7295e/w7295e08.htm
http://www.statsoft.com/textbook/basic-statistics/

A number of statistical analysis computer packages such as EPI-INFO\textsuperscript{32}, SPSS or STATA exist which can be used to analyze behavioral data. These packages ease the analysis process by conducting statistical tests using standardized formulas. However, they should be used by researchers who have a good understanding of these tests and their assumptions, limitations and interpretation. Thus programs should consider to seek technical assistance both in conducting statistical analysis and interpreting the results of the analysis.

1.8.2.3. Relative and absolute measures of effect

In epidemiological studies, the effect of exposure to a BCC intervention can be measured both in relative and absolute terms.

The incidence rate, the risk ratio and the odds ratio are relative measures of effect.

**Incidence rate** is the number of new cases within a specified time period divided by the size of the population initially at risk.

The **risk ratio** (also known as relative risk) is the risk of an event (or of developing a disease) relative to exposure, where RR \text{________________________}. The “exposure” might be a specific intervention such as a BCC intervention.

The **odds ratio** is used to assess the risk of a particular outcome (or disease) if a certain factor (or exposure) is present. The odds ratio tells us how much more likely it is that someone who is exposed to the factor under study will develop the outcome as compared to someone who is not exposed (for example, use of condom (exposure) and HIV infection (outcome)).

There is some confusion about the use of the odds ratio versus the risk ratio; both the odds ratio and the risk ratio compare the likelihood of an event between two groups. However,

\textsuperscript{32} Available at http://www.cdc.gov/epiinfo/downloads.htm
the odds ratio compares the relative odds, while the risk ratio compares probabilities. To better illustrate this, consider the following example on condom use:

<table>
<thead>
<tr>
<th>Sex</th>
<th>Condom use</th>
<th>No condom use</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>308</td>
<td>154</td>
<td>462</td>
</tr>
<tr>
<td>Women</td>
<td>142</td>
<td>709</td>
<td>851</td>
</tr>
<tr>
<td>Total</td>
<td>450</td>
<td>863</td>
<td>1313</td>
</tr>
</tbody>
</table>

Clearly, women are less likely to use a condom than a male; but how much less likely? One can compute the odds ratio or the risk ratio to answer this question.

The odds ratio compares the relative odds of no use of condom among men and women. For men, the odds would be exactly 2 to 1 (154/308=0.5). For women, the odds would be almost 5 to 1 (709/142=4.993). The odds ratio would be 9.986 (4.993/0.5). That means here there is almost a ten-fold greater odds of no condom use for women (i.e., being exposed to being a female) than for men (i.e., being exposed to being a male).

The risk ratio compares the probability of no condom use in each group rather than the odds. For men, the probability would be 33% (154/462=0.3333), for women 83% (709/851=0.8331). The risk ratio of no condom use would be 2.5 (0.8331/0.3333). There is a 2.5 times greater probability of no condom use for women than for men.

Both measurements show that women are less likely to use condoms. But the odds ratio implies that women are much worse off than the risk ratio. The most commonly cited advantage of the risk ratio over the odds ratio is that the former is the more natural interpretation, as it measures events in a way that is interpretable and consistent with the way people really think.

1.9. Dissemination of evaluation results

The process of evaluation should be completed with the dissemination of results. This last step is extremely important and should be taken into account as early as in the design phase of the evaluation.

It ensures that program managers, decision- and policy-makers and other stakeholders are aware of the findings of the evaluation. Think about the different audiences of your evaluation, their background, and their specific information needs. Findings should be translated in a language which is appropriate for the audience, including implications, actions needed and recommendations on the way forward, in order to contribute to better BCC programs with tangible and sustainable results.

Remember that dissemination:
- Is not a one-off activity

– Is an ethical obligation, especially at the community level where the evaluation took place
– Does not have to be a publication; one needs to be creative (e.g. dissemination through SMS).
2. Part II: Implications on the HIV prevention tools used by GDC

Crucial to the success of any effective HIV prevention effort are a number of overarching principles, which programs should be based upon. HIV prevention programs must respect the promotion, protection and respect of human rights and gender equity. They also have to be differentiated and locally adapted to the relevant epidemiological, economic, social and cultural contexts, in which they are planned and implemented. They must be evidence-informed and comprehensive in scope. They require long-term and sustained efforts, as well as coverage, scale and intensity enough to make a difference. Last but not least, community participation is critical for the HIV prevention programs to achieve impact (UNAIDS, 2005). BCC programs are naturally expected to subscribe to these principles.

2.1. Examples of HIV prevention tools

Behavioral HIV prevention programs can target individuals, families, communities, entire societies, or (ideally) a combination of all these. There is therefore a variety of HIV behavior change strategies. These approaches can range from counseling and interpersonal communication to social mobilization, advocacy, behavior and social change communication, social marketing, advertising, film and theatre.

For the purposes of this guidance note, three of the tools used in BCC are presented, as these are particularly often utilized by GDC HIV prevention programs. Four country examples of HIV BCC interventions implemented by GDC programs are presented in part 3.

2.1.1. Join-in Circuit

According to UNAIDS, 45% of new HIV infections in 2007 were among people aged 15–24, yet survey data from 65 countries for 2004–2007 indicate that just 40% of males and 38% of females in this age range had accurate knowledge about the disease and how to avoid transmission (UNAIDS, 2008).

The Join-in Circuit (JiC) was developed by the German Federal Centre for Health Education (BZgA) in the early 90’s. In 2003, after successful field-testing in Germany, the JiC was introduced as a joint BZgA and GTZ cooperation project with the aim to adapt and implement it into different conditions in five partner countries (Ethiopia, El Salvador, Mongolia, Mozambique and the Russian Federation). Based on the positive experiences made, the JiC was adapted in more than 20 countries (BZgA; GTZ, 2005).

The JiC draws on communications and behavioral research to provide an interactive, problem-driven approach for communicating vital information about HIV to large numbers of people belonging to certain groups at risk of HIV – school-children and other young people as well as soldiers and prisoners – in a short period of time. The aim is to strengthen the ability of participants (particularly the youth) to protect themselves from HIV with information that is tailored to the local context, practical and interesting to the audience. The dialogue-oriented facilitation and the playfulness of the approach allow participants to establish an emotional and personal connection, promoting effective learning and creating
an environment in which it is possible to discuss taboo subjects related to sexuality (GTZ, 2008).

The circuit normally consists of five stations through which young people pass in small groups of about ten to fifteen participants. The stations have names such as “HIV Transmission Channels”, “Love, Sexuality and Protection from HIV”, “Contraception” “Living with HIV” or “Sexually Transmitted Infections”. The subject of HIV prevention is embedded within topics which really interest young people or which have a particular significance in certain countries. One of the added values of this prevention tool is that it offers the participants the possibility to incorporate special themes which are particularly important in the context of HIV prevention in a given country (BZgA; GTZ, 2005). It can be used flexibly as a mobile instrument, which can – and should – be adapted to the local situation and cultural preconditions, after consulting the local partners (GTZ, 2010).

2.1.1.1. Successes

Studies indicate that the JiC not only improves the knowledge of participants, it also boosts their tolerance of people living with HIV and raises awareness of risky behavior. Most organizers and facilitators agree that the JiC is an effective tool for HIV prevention among young people. In many places, national stakeholders as well as targeted groups of young people welcomed the JiC with enthusiasm and its educational benefits radiated well beyond the formal sessions.

The JiC has been proven to be a useful tool with the ability to reach big numbers of people. Its flexibility to be integrated and adapted into different settings and with different groups was tested and demonstrated. And though primarily viewed as a vehicle for providing information, subjective evidence suggests that this tool has a positive influence on people’s attitudes (toward PLHIV, for example) and helps people to change their behavior (condom use etc.). Also, evidence indicates that the JiC can add value to national HIV responses (GTZ, 2008).

2.1.1.2. Challenges

Although this prevention tool is highly valued and very well accepted by the participants, one evaluation carried out in 2008, addressed one of the main challenges (Beck R. M., 2008):

“Lack of a framework that enables research, monitoring and evaluation which provides robust evidence of the results achieved by the JiC and promotes quality management”:

Within ten years around 170,000 people had been reached by the German JiC campaign, and in the five adaptation countries 25,000 young people have already been made aware of the issues around HIV within the short period of two years. Moreover, surveys of participants before and after implementation in Ethiopia, Bangladesh, Ecuador, El Salvador, Kazakhstan and Mongolia indicated increased HIV knowledge (GTZ, 2008).

But how one can prove that the JiC really achieves sustainable results in terms of actually initiating and maintaining behavior changes that have the potential to protect the target group(s) from HIV infection? Thus how does the JiC contribute to the achievement of overarching HIV-related development objectives and results (impact)?
Previous experience shows that not all effects of the JiC can be measured, simply because they have not been identified as objectives of observation and reporting during the design and adaptation phase of the JiC (Beck R. M., 2008). Thus, planning for effectiveness evaluation was often overlooked in JiC planning and implementation. How can the results of the JiC be measured and demonstrated?

2.1.1.3. Recommendations

Following the already presented results and challenges in terms of evaluating the effects of JiC, here are some recommendations for future JiC planning, monitoring and evaluation:

i. **Integration of the JiC into existing structures and programs can increase its sustainability and further promote its effects and thus needs to be ensured:** Evaluations so far indicate that the JiC has the potential to be an effective instrument for HIV prevention initiating long-term changes in young people’s attitudes and behavior in regard to HIV. Therefore, it is critical to integrate the JiC into existing (national) prevention programs and structures rather than implementing it as an ad hoc and sporadic activity. Integration of the JiC can promote ownership, increase the chances of its more effective and sustainable use and create a more conducive environment for the evaluation of results.

ii. **Every JiC needs measurable objectives from its onset:** Effectiveness of the JiC can only be measured if clear learning goals are set from the beginning. Therefore it should be clear where the program stands at its outset (for that a baseline study is needed!) and which direction the planned measures should take in order to yield the expected results.

iii. **Monitoring and evaluation of the JiC should be integrated into the routine monitoring and evaluation system of the HIV prevention program it is embedded in to promote a more robust evaluation of its results:** In this way the JiC becomes one of the activities the program has to routinely report on, while effectiveness evaluations can be based on regularly collected relevant and appropriate information that allows objectively verifying and measuring the results of the JiC.

iv. **Effectiveness evaluations should be an integral part of JiC adaptation and implementation until more robust evidence is available:** When adapting or planning the implementation of a JiC, effectiveness evaluation should be not overlooked as to date, there is little robust evidence available on JiC effects. To conduct an evaluation, the program needs to search and/or adapt appropriate tools at the design stage. A set of relevant evaluation tools could be developed. The evaluator could then choose the most appropriate tools depending on the context in which the evaluation will take place. In this way, both quality and comparability of the JiC evaluations across different countries and various settings can be further improved.

v. **Evaluations of the JiC should go beyond simple pre-test and post-test evaluation approach:** Depending on the resources of the HIV prevention program in which JiC is integrated, evaluations should strive to extend the time intervals of follow-ups (e.g. after six months, one year etc.). A pre- and post-test approach might be good to measure immediate changes in HIV knowledge and attitudes. However, it does not necessarily mean that this knowledge will be there one year down the line or even
more that it will be translated into behaviors that reduce the risk of HIV infection. For an effectiveness evaluation, a control group with no or delayed JIC intervention should always be taken into account.

2.1.2. Peer education

Peer education is based on a behavioral theory which proposes that people make changes in their behavior not such much in response to scientific evidence but rather because of trusted peers who have adopted changes and who act as convincing and credible role models for change. By the 1990s peer education was one of the most widely used approaches in HIV prevention initiatives targeting the youth in developed and developing countries (Horizons, 2000). A study from India (Bhardwaj, 2004) has actually revealed that the most dramatic examples of behavior change are even seen among the peer educators themselves.

Peer educators are effective and credible communicators who have inside knowledge of the intended audience and use appropriate language/terminology as well as non-verbal gestures to allow their peers to feel comfortable when talking about issues of sexuality and HIV. They have physical and socio-cultural access to intended audiences in their natural environments without being conspicuous. This is particularly true when working with hard-to-reach populations such as CSWs, IDUs and prison inmates because physical access to such populations can be difficult (UNAIDS, 1999).

Peer education aims at achieving behavior change at the individual level by attempting to modify a person’s knowledge, attitudes, beliefs and/or behaviors. However, peer education may also effect change at the group or societal level by modifying norms and stimulating collective action that leads to changes in programs and policies (UNAIDS, 1999). Peer education is very often implemented by youth reproductive health and HIV prevention programs around the world (UNFPA, 2005). In the GDC context, peer education is one of the methods used to support, for instance, implementation of HIV workplace programs.

2.1.2.1. Successes

According to an evaluation carried out by UNAIDS (UNAIDS, 1999), peer education is a widely utilized HIV prevention strategy that is accepted and valued by program audiences and other stakeholders. It is perceived as an effective tool in promoting the adoption of preventive behavior change with regard to HIV.

Findings from some more rigorous studies that have used HIV risk behavior or STI/HIV incidence as outcome measures and applied experimental or quasi-experimental evaluation designs have found that the interventions that included HIV peer education had a positive impact on STI or HIV incidence and/or risk behavior. Results are of such studies are available for different target groups, including CSWs, communities, industrial workers, MSM, IDUs, PLHIV and youth (UNAIDS, 1999).

A review supported by WHO (Medley, Kennedy, O'Reilly, & Sweat, 2009) concluded that peer education interventions were associated with moderate increase in HIV knowledge,
reduced equipment sharing (needles) among IDUs, and increase condom use (except with youths). There was no consistent effect on STI rates.

2.1.2.2. Challenges

The quality of peer education varies tremendously. Very often, programs are called “peer education programs”, when they are in reality outreach activities. Peer education experts frequently raised the issue of how to standardize peer education (UNFPA, 2005). At the same time, institutionalization of peer education programs might be a challenge because of high peer educators turn over.

Other challenges are related to the design and implementation of peer education programs and include feasibility and sustainability aspects. Further issues of concern are appropriate selection of peer educators, content of their training curriculum and support needs, their supervision and motivation, as well as stakeholder community involvement and acceptance of peer education programs. (UNAIDS, 1999).

Lack of time, funding and technical expertise are seen as possible barriers to measuring the outcomes of peer education programs (UNAIDS, 1999). Also, the evaluation of such programs can be difficult because they are often implemented in conjunction with other activities.

Reviews of the results of peer education programs in youth regarding behavior change are mixed. There is no good evidence on the effectiveness concerning abstinence. Similarly there is only limited evidence on the effects on reducing the number of sexual partners. There were few studies with biological outcomes such as STI infections and none showed a significant effect (Weinmann, 2009).

2.1.2.3. Recommendations

Here are some key recommendations that derive from previous research:

i. **Meaningful participation of the target groups in designing, implementing and monitoring and evaluating peer education programs has to be ensured:** Ownership of peer education programs is crucial for their success and sustainability. Thus such programs should emerge from and reflect the needs of the target groups. Therefore it is recommended to conduct a needs assessment before proceeding with the implementation of peer education programs. Once it is clear that a peer education program is needed, target groups should also be involved in identifying the goals and objectives of the program and should be included throughout its implementation as well a monitoring and evaluation. Such an active participation can increase the sense of ownership and contribute to sustainable and measurable results.

ii. **Peer education programs need clear goals and measurable objectives from their onset:** Effectiveness of peer education program should be assessed on the basis of clear goals and objectives which are defined from the beginning. It should be clear where the program stands at its outset (and for that **a baseline study is needed!**)
and which direction the planned measures should take in order to yield the expected results.

iii. **High quality training of peer educators needs to be provided:** High quality peer educators’ training can build self-confidence and skills, as well as instill clear program objectives to keep peer educators motivated and focused (USAID, 2008). Well trained and motivated peer educators can also support the monitoring and evaluation process, by providing feedback on how the program evolves, how it is perceived by the target group, what the challenges and successes are, and where possibilities for improvement lie.

iv. **Budget allocation should be increased** to promote quality monitoring and evaluation of peer education that provide tangible results.

v. **Staff involved in monitoring and evaluation of program including peer education should be trained** in basic data analysis and the creation of systems for data collection, analysis, and dissemination.

vi. **If resources permit, the program should consider conducting regular KABP surveys** to obtain baseline information on the knowledge, attitude, behavior and practices of the target group and through annual or biannual follow ups document the possible changes that take place as a result of the peer education program. A source of comprehensive information on KAPB surveys, tools, reporting formats etc. is the ACCA toolbox developed by GTZ.

vii. **Within the context of workplace programs, inclusion of peer educators in monitoring and evaluation of the program should be taken into consideration:** Peer educators, as trusted member of their group, are in the position to openly discuss the program with their peers and then give feedback on what the target group thinks about the program (e.g., whether it is widely accepted, perceived as cultural- and/or gender-sensitive; whether it satisfies the expectations of the target group; whether it is seen as successful or not, and which changes should be undertaken to improve the quality of the program). A very useful tool in this context is the “Interactive course: monitoring & evaluation of HIV/AIDS workplace programs” jointly developed and launched by GTZ and the Global Business Coalition on HIV/AIDS (GBC).

---

34 KABP stands for Knowledge, Attitude, Behavior and Practices.


36 The CD-Rom is available in English and French: [acca@gtz.de](mailto:acca@gtz.de)
2.1.3. Mass media campaigns

Mass media campaigns have been used for more than 20 years as a tool in the HIV response worldwide. In many countries, communication programs via mass media have been expanded and now strongly focus on behavior and attitudes change, increasing HIV-related knowledge, decreasing stigma, and producing awareness. In addition, they can be powerful in raising community activism and mobilization of the masses at grass-roots level. Mass media include TV, radio, video, print and the internet. Programs may be presented in the form of shows, songs, soap operas, pamphlets, posters, interactive websites, spots and other. They may be part of social marketing programs, e.g. for improving adolescent reproductive health (Van & Meekers, 2000).

It is well known that economic, social, cultural, legal and political factors shape human behavior and can act as obstacles for behavior change or as enabling forces leading to less risk behavior with regard to HIV acquisition. Against this background, the mass media are expected to influence knowledge, attitudes and self-efficacy with regard to sexual behavior. They are used to increase the perception of risk and to drive behavior change (Bertrand & Anhang, 2006). Mass media campaigns have the potential to generate specific effects in large numbers of individuals at national, regional and local levels.

Based on a large research experience, well-accepted principles of mass media campaigns used for HIV prevention are: (i) conducting research about the target

A. GDC example: HIV Soap Operas

The publication “TV soap operas in HIV education: Reaching out with popular entertainment” looks at cases where those components include soap operas. The examples of the soap operas: “Love as a Test” in Kyrgyzstan, “Amor de Batey” in Dominican Republic, “SIDA dans la Cité” in Côte d’Ivoire. It does not, however, look at the issue of evaluating these programs.


B. Example PSI in Botswana: Evaluation of a Multiple Concurrent Partnerships Mass Media Campaign

The campaign was integrated into community-based interpersonal communication projects in addition to mass media channels. The method was to cite common sayings and then point out the risks inherent in the behaviors the sayings help to normalize or legitimize.

The evaluation involved selecting a random sample of sexually active adults stratified by geographic zone and urbanization. From this sample, respondents were recruited to participate in focus group discussions that were used to assess the reach, recall, and effectiveness of the initiative.


---

37 The chapter on mass media campaigns has been adapted from (Weinmann, 2009).
audience; (ii) using a specific theory as a conceptual foundation; (iii) segmenting the audience into meaningful subgroups; (iv) creating specific and easy messages; (v) utilizing effective channels; (vi) ensuring high message exposure and (vii) using a sensitive outcome evaluation design (Weinmann, 2009).

The results (successes and challenges) presented below are based on systematic reviews performed to assess the impact of mass media campaigns.

### 2.1.3.1. Successes

Mass media interventions showed significant immediate and overall positive effects with testing rates increasing six-fold after the intervention (Vidanapathirana, Abramson, Forbes, & Fairley, 2006). Also, some studies showed positive effects toward less casual sex and lower numbers of sex partners (Bertrand, O’Reilly, Denison, Anhang, & Sweat, 2006).

A more recent review (Noar, Palmgreen, Chabot, Dobransky, & Zimmerman, 2009) revealed that in the last years more and more theory-based interventions have been evaluated and that study designs have improved. 44% of the campaigns were theory-based, 82% used specific slogans, but only 24% contained slogans with direct and clear messages related to behavior change, including HIV testing, safer sex, or condom use. Eight out of ten campaigns that used quasi-experimental evaluation designs showed significant effects on behavior change. In particular, after media exposure, there were significant reductions in the number of sex partners in Tanzania, more youth-initiated condom use, sexual abstinence and less sex partners in Zimbabwe, and an increased AIDS knowledge in China. The authors detected a shift in the purpose of the campaigns from simply aiming at raising HIV awareness to having an impact on safer sexual behaviors and on putting more emphasis on (the integration of) computer- and Internet-based components.

### 2.1.3.2. Challenges

Despite the immediate effects right after the mass media campaigns, no long-term effect was seen (Vidanapathirana, Abramson, Forbes, & Fairley, 2006), while many evaluations of TV or radio programs and of community education campaigns failed to show significant increases in condom use (Bertrand, O’Reilly, Denison, Anhang, & Sweat, 2006).

No study was found that evaluated mass media programs focusing on mobile phones and Information and Communication Technologies (ICT) in low-resource countries, although these media are becoming more and more used to disseminate also health related messages.

### 2.1.3.3. Recommendations

While donor agencies and nations have invested millions of dollars in mass media campaigns relatively few have been subjected to rigorous evaluation. There is no doubt that mass media campaigns in developing countries and countries in transition have been underresearched and require more rigorous evaluation in order to assess their impact and
provide guidance on how to improve their performance (Weinmann, 2009) (Indiana University, 2007).

2.2. Final recommendations

This guidance note has provided background information on the design and implementation evaluations of BCC interventions. It has also presented three practical examples of BCC tools, their successes, challenges and recommendations regarding monitoring and evaluation. In the last section, we will try to distil this knowledge and narrow it down by formulating recommendations, taking into consideration GDC-specific aspects, wherever applicable.

2.2.1. Recommendations to programs

1. **BCC goals need to derive from the overall goals and the specific behavior change goals of the program:** Interventions should be clear about what is being done and what the expected outcomes are. Integration of BCC interventions in the impact chain of the program implies their integration into its routine monitoring and evaluation system. In this way, BCC evaluations and reports on their findings are assured, e.g. through the program annual reports or in the program progress review.

2. **Evaluation is not an ad-hoc activity, but rather an integral part of any BCC program and thus it should be planned during its design phase:** The goals and objectives of BCC evaluations need to be clearly defined from the onset of the program and form an indispensable part of the evaluation plan. Additionally, the required resources have to be timely identified and reserved.

3. **Evaluation should accompany all phases of BCC programs:** Formative evaluation is crucial before the onset of BCC programs to provide a clear picture on where the program starts, as well as on which is the profile and the specific needs of the target group. Process evaluation monitors whether the activities of the program are on the right track, while effectiveness/impact evaluation builds upon formative and process evaluation and goes a step beyond by assessing what the program has actually achieved.

4. **A prerequisite for every sound evaluation is a well-designed study protocol shared and agreed upon by all key stakeholders:** An appropriate evaluation plan is needed that outlines how to measure the results of BCC interventions and interpret the changes in knowledge, attitude, behavior and environment. This concept paper states among others the objectives of the evaluation, the questions that will be answered, the information that will be collected to answer these questions, as well as when collection of information will begin and end.

5. **All evaluations should examine the following questions** (OECD-DAC, 2002):
   - Are we doing the right thing? (relevance)
   - Are we achieving the project and program objectives? (effectiveness)
   - Is our action cost-effective? (efficiency)
   - Are we contributing to the achievement of overarching development objectives and results? (impact)
   - Are the results durable? (sustainability)
– For GDC programs, the **Capacity WORKS** tools used to facilitate the effective and efficient steering of the program should be also taken into consideration\(^{38}\).

6. **Ethical clearance** needs to be obtained or – if there is no official requirement - an explanatory statement for why this is not necessary should be drafted. **Confidentiality** of personal data of the participants needs to be guaranteed, while their right to leave the evaluation process any time has to be clearly communicated.

7. **In all effectiveness evaluations, the evaluation design will define the level of certainty about the program’s results**: Therefore one should first explore whether it is feasible to randomly select an intervention group and a control group. If this is not possible (e.g. due to ethical considerations) one should create, along with the intervention group, a non-randomized comparison group of people who will either be exposed to the BCC interventions at a later stage or will not be exposed for reasons that are outside the control of the program.

8. **SMART indicators have to be used** to provide a simple and reliable means to measure achievement, to reflect the changes connected to an intervention or to help assess the performance of a development actor.

9. **Every effectiveness evaluation is based on a sound baseline assessment**: A baseline study simply defines the “pre-intervention exposure” condition for the set of indicators that will be used to assess achievement of the outcomes and impact expressed in the program’s logical framework. When compared with the condition of the same indicators at some point during implementation (mid-term evaluation) and post-operation implementation (final evaluation), the baseline study forms the basis for a “before and after” assessment or a “change over time” assessment. Without baseline data to establish pre-operation conditions for outcome and impact indicators it is difficult to establish whether change at the outcome level has in fact occurred (WFP, n.d.).

10. **A participatory approach is imperative for successful BCC evaluations**: A participatory approach allows all stakeholders - particularly the target audience - to take part in the design of a BCC project, its ongoing assessment and the response to the findings. It thus increases ownership of the evaluation along with the chances that these will be translated into action for improvement of BCC programs and/or their environment. However, a big number of stakeholders might sometimes be difficult to coordinate, stretching the resources of a program and delaying its implementation. One needs to strike a balance between the participation and input of various stakeholders and the resources (human, financial and time) available to complete an evaluation. In the context of German development cooperation, Capacity WORKS\(^{39}\) can provide useful tools on how to prioritize the involvement of different stakeholders at the different stages of an evaluation.

11. **Technical “know-how” should be obtained to ensure sound evaluations**: Evaluation can be a complicated process demanding technical knowledge and expertise at all stages. The program should consider obtaining support, e.g. from local academic institutions with expertise in this area (Beck R. M., 2008). Greater collaboration is needed between

---

\(^{38}\) For more information on GTZ monitoring and evaluation system visit [http://www.gtz.de/en/unternehmen/31522.htm](http://www.gtz.de/en/unternehmen/31522.htm)

\(^{39}\) For more information visit [http://www.gtz.de/en/leistungsangebote/28379.htm](http://www.gtz.de/en/leistungsangebote/28379.htm)
program managers and researchers to facilitate effective monitoring and evaluation design. In addition to obtaining their technical expertise, the involvement of an “outsider” has the potential to increase the credibility and assure independence of an evaluation.

12. **In all evaluation reports** the goals and objectives of the evaluation, the type and the methodology used, as well as all relevant results (also the unexpected ones) should be described. Finally, conclusions and recommendations should be clearly stated. The report should always be shared with relevant stakeholders and target groups in the partner country.

13. A **(peer-reviewed) internationally accessible publication should be taken into consideration**: Peer review methods should generally be employed to maintain standards, improve performance and provide credibility. Publication of one’s results makes the knowledge generated accessible to others who might have similar research questions.

2.2.2. **Recommendations for researchers**

The first systematic review of what works to prevent HIV infection among young people in developing countries was conducted by WHO (2006). Here are some of the recommendations formulated for the researchers:

1. There is a critical need to strengthen research and program monitoring and evaluation capacity in developing countries
2. High-quality evaluations and monitoring of the impact of HIV prevention interventions among in developing countries are urgently required
3. Operations research is needed to better understand the mechanisms of action of interventions
4. Clarity is needed about the specific vulnerabilities of the target group(s)
5. Standardization of outcome indicators would greatly facilitate comparisons of results across studies
6. Costing and cost-effectiveness studies should be built into evaluation studies
7. Research is needed to better understand the relationship between reported effects on behaviors and biomedical impacts.

2.3. **And some tips...**

2.3.1. **When planning an evaluation...**

- Plan and start the process on time
- Inform/include all relevant stakeholders
- Identify which BCC interventions will be evaluated
- Become clear on the goals of the evaluation, and the major questions to be addressed
- Develop an evaluation plan

---

2.3.2. When implementing an evaluation...
- Ensure collection and unhindered flow of quality data (preferably a combination quantitative and qualitative)
- Collect not more than the data you need
- Analyze data using appropriate (statistical) tools
- Record lessons learned, successes, challenges
- Formulate recommendations

2.3.3. Before disseminating the results of an evaluation...
- “Translate” the results into a language appropriate for policy- and decision-makers, partners, key stakeholders, the news media and funding agencies
- Go through a peer review process, if possible
- Explore possibilities to disseminate the results both on national and international level

2.3.4. After dissemination and validation of the results
- Revise or redesign program based on evaluation findings

All along the way, do not forget to engage people with technical and scientific expertise to support the evaluation.
3. Part III: Country Examples

To further facilitate the understanding of such programs and the application of monitoring and evaluation (M&E) methods in day-to-day work, a short course tailored around interventions supported by programs of GDC was offered from 9-10 March, 2012 in Johannesburg, South Africa.

The training was based on the Guidance Note elaborated by the Program to Foster Innovation, Learning and Evidence in Health and HIV programs of German development cooperation (PROFILE) and built directly on the personal experience of the participants.

The content of the training included:

- Principles, concepts and frameworks of M&E of behavioral interventions including indicators
- Introduction to different study types and designs of evaluations and data collection
- Planning an evaluation: Workgroups, case presentations and examples of good practices
- Research evidence available for typical BCC interventions supported by GDC
- Managing an evaluation of a BCC intervention: Terms of Reference (ToR), supervising consultants/researchers, time frame and budget.

Part 3 presents the country case studies and the results of the groups work. The four groups worked step by step on developing / improving the evaluation plan of the country examples. Starting point was the M&E structure of the respective GDC program already in place (or not) to monitor and evaluate the results of their HIV BCC interventions.

3.1. Introduction to case study groups

Before embarking on group work the case studies were presented to the participants. The four case studies follow.

3.1.1. Female Genital Mutilation

The first case study presented comes from a program implemented in Burkina Faso. The program works on female genital mutilation (FGM). It was presented by Zakari Congo. The presentation is available here.

3.1.2. Join-in-Circuit

The second case study presented was from an HIV project in Zimbabwe. Theresa Ndikudze presented how Join-in-Circuit (JiC) is being implemented, monitored and evaluated by the project. Her presentation is available here.
Some additional supporting documents were also submitted by the project. These were: a) a result chain and b) the facilitator’s manual in Zimbabwe.

3.1.3. Theater for change
The case study on Theatre for a Change (TfaC) comes from Malawi and was presented by Claire Walsh. Her presentation is available here.

3.1.4. Life Choices “3 ½ lives of Philip Wetu”
This is an example from Namibia on a mass media campaign presented by Luise Haunit. Life choice “3 ½ lives of Philip Wetu” is an interactive DVD. The presentation is available here.

3.2. Discussing the M&E system in place
The participants joined one of the four groups (FGM/JiC/TfaC/Philip Wetu). The guidance questions, developed by Paola Frattola, to support the group discussions around evaluation of the case studies can be found in appendix 3.

The first task was to discuss assets and needs of our behavioural intervention’s monitoring and evaluation system and develop ideas how our behavioural intervention’s monitoring and evaluation system could be improved.

3.2.1. Female Genital Mutilation
The group prepared a power point presentation which is available here.
Summary of the presented M&E System:

- The M&E System was well presented.
- It was suggested to consider other intervening factors that could also influence outcomes.

Indicators:

- Impact indicator is derived from BMZ prescribed indicators.
- MoH, MoSA, and the project share one common indicator in the national M&E system.
- Partners including MoH, MoSA, MoW, use the family planning indicator derived from GIZ, as a process indicator.

Data collection

- For the two intervention regions, the program has implemented a specific baseline study in one of them. DHS data (FGM prevalence) was used as the baseline for the other region.
- Program has however no written M&E Plan in place, but a number of individual elements of them.
- The overall impact has been shown to be difficult to attribute to the intervention solely.

Types of evaluation used

A) Outcome evaluation

- A yearly outcome evaluation to measure the knowledge and attitudes concerning FGM is conducted. This for internal GIZ purposes. (Increase by 25% knowledge of men, women, primary and secondary students)
- MoSA and MoEducation benefits from this evaluation, as the main partner. Other ministries however benefit- MoW, MoH.
- The workshops on advocacy, capacity building, where data is discussed and disseminated are successful.
- Positive unintended outcomes: other ministries are benefiting from the data collected.
- Collaboration with stakeholders and ministries for data collection is difficult.

B) Impact Evaluation

The impact indicator is reduction of the prevalence of FGM.

- At this point, it is difficult to attribute all changes to the program.
- Net effects are the decrease in prevalence of FGM.
- It is hoped that at the end, the general reproductive health of women would be improved.
- The strong political commitment will to reduce the prevalence of FGM helps to avoid unintended effects of the intervention.
**Evaluation indicator**

An impact evaluation would look into indicators that measure a reduction in the prevalence of FGM:

- Decreasing FGM prevalence of girls aged 0-14 years by 50% at the end of the program in 2015.
- Outcome indicators measure the increased knowledge of and attitudes towards FGM.

**Evaluation design**

A future impact evaluation will use a cross sectional collection of data in a follow-up period with the same data set that was used for the baseline study.

- Quantitative evidence will be collected through questionnaires and structured interview.
- Qualitative evidence will be collected through in-depth interviews & focus group discussion.
- Mothers and children are interviewed and results are correlated.

**Sampling methods**

- A non-probability sampling method would be used.

**3.2.2. Join-in-Circuit**

The group started its work by reviewing and commenting on the result chain of the JiC in Zimbabwe. A lot of time was dedicated in **reaching a common understanding on terminology**. The comments on the result chain were the following:

- The **impact** was “number of HIV new infections, STIs and teenage pregnancies in the project regions decrease”. It was felt that young people should be included for the number of HIV new infections and STIs.
- The **objective** was “young people (aged 15-24 years) in the project regions engage increasingly in protective behavior and make use of adequate health services”. The groups thought that it was actually two objectives: 1) “young people (aged 15-24 years) in the project regions engage increasingly in protective behavior” and 2) “young people (aged 15-24 years) in the project regions make use of adequate health services”.
- Instead of “use of outputs” it was proposed to use the term “immediate outcome”. This was “youth is informed on HIV/AIDS and SRH and has access to youth-friendly services”.
- **Outputs**: 1) people trained 2) manual developed etc.
- **Activities**: 1) Training of trainers in JiC 2) training of JiC facilitators (and all other activities of the program).

The group proposed the following **indicator**:

Percentage of trained youth that accessed the services increased by x%
After in detail discussions, the group agreed that the method used by the program for data collection is meaningful. Sex-disaggregated data is collected on a daily basis. This data is then compiled into quarterly reports. Additionally, feedback is regularly provided by the trained NGOs. The group also addressed the necessity of collecting data from different entry points, e.g. from the service providers and from the beneficiaries.

The groups underlined the need to adapt and evaluate even already widely used and evaluated tools such as JiC, if these are applied in a totally new context.

The methods of evaluation that are being used by the program are a variety of qualitative and quantitative data. With regard to the qualitative methods, the program (among others) randomly selects participants and based on a set of guiding questions, conducts its FGDs.

The group also recommended compiling a good practice report and sharing this with other stakeholders in the country.

### 3.2.3. Theater for change

The group started its work by reviewing the goal and objective of the TfaC Education program and discussing whether these are SMART.

- **Impact (Goal):** To reduce the risk of HIV infection and promote HIV prevention and communication strategies among primary school teachers, learners and parents.
- **Objective:** Target groups demonstrate the knowledge, attitudes and abilities to protect themselves and others from HIV infection.

Feedback provided by the group on the stated goal and objective was that both formulations were too broad and should be rephrased to be more concise.

The group further discussed terminology and that it is critical to have a common understanding on what is meant by attitude, what distinguished attitude from ability, and ability from practice.

Whereas HIV-related transmission and knowledge on protection is rather easy to capture and most surveys include standardized questions on this, measuring HIV-related attitudes is more difficult. Questions (here of medical personnel) could be for example:

*White, married women, age 40-45, would you test her for HIV? Would you treat someone who is HIV positive?*

TfaC uses observation of theatre to capture abilities / assertiveness / resilience. See assessment tool below used by facilitators to assess participants. Standardization workshops to address the problem of subjectivity bias were discussed. The tool is promising but not a validated method yet.

---

41 Notes by Jasmin Dirinpur.
The underlying theory of change would require changes in knowledge first, before shifts in attitudes and abilities/behavior/practices could happen. However, observed and persisting knowledge-do gaps pose a challenge to evaluation of BCC strategies.

The group then reviewed TfC indicators used to capture behavior change (and preceding changes in knowledge and attitudes). These are as follows:

- Pre-service teachers have comprehensive knowledge of HIV transmission, prevention and treatment
- % of pre-service teachers reporting increased knowledge on malaria prevention
- Pre-service teachers have a positive attitude to HIV prevention strategies
- Pre-service teachers demonstrate the ability to say no to sex
- Pre-service teachers report they know how to use a female and male condom
- % of female and male pre-service teachers reporting condom use
- Number of pre-service teachers going for HTC
- Learners have comprehensive knowledge of HIV transmission and prevention and treatment
- Learners report that they speak to an adult about HIV
- Learners have a positive attitude to HIV prevention methods

A spontaneous reaction was that these are far too many indicators to observe and report on which makes a proper analysis a daunting task. The group therefore thought a prioritization of indicators to be useful. However, expect one indicator on pre-service teachers ability to resist to sex which is an internal “nice-to-have” indicator, all others are “have-to-need indicators” given the donor requirements.

The “nice-to-have” indicator was further reflected upon as an example for a very sensitive, hard to capture / measure aspect of behavior change (see also assessment tool above).

TfC beyond behavior change aims to capture changes in the fields of advocacy and capacity building. Indicators here were briefly discussed and were found to be rather at the output (e.g. Numbers of trainings, Number of schools with fully functioning procedures etc.) than at the outcome level.

The group further exchanged on data collection and the social desirability bias in data on sexual and HIV-related behaviors. Parts of the problem could be circumvented by use of data from health facilities, or use of proxy indicators (e.g. teenage pregnancy or STI prevalence as proxies for unprotected The challenge to measure the results was that this was sex).
3.2.4. Life Choices “3 ½ lives of Philip Wetu”

- Not a concrete intervention with a clearly defined beginning and end (e.g. there was no control over the amount of DVDs distributed and the share of “facilitated” presentations of the film).
- The group proposed a cross-sectional representative population survey that asks questions of exposure (whether the respondent has watched the film, and if yes, facilitated?) and questions with regard to knowledge, attitudes and behavior. By doing so, one could possibly come up with knowledge about “cove Claire Walsh rage” (e.g. have men having sex with men (MSM) been exposed) and with some kind of correlations, e.g. have those exposed a better knowledge on HIV/AIDS and other attitudes. In addition a dose-response-effect relationship could be assessed. However, this design does not allow establishing a causal relationship.
- Another option could be in-depth interviews and FGDs at the workplaces where the DVD was shown. In this way one could possibly draw some conclusions and adjust the training accompanying the DVD accordingly.
- In addition, one could examine whether the uptake in Voluntary Counselling and Testing (VCT) increased (or not) at the services where the film was shown.
- It would be also interesting to ask the people accessing the services whether they have been exposed to the film.
- Unfortunately, no baseline data can be collected any more as the film has already been distributed and is continued to be distributed. There are already many people that have been exposed.
- The experts from WITS also proposed the following:
  - The program could find out whether other larger surveys are planned and use the opportunity to integrate exposure questions into the survey (this approach to add questions to an ongoing survey is also called “piggybacking”, for an example see Working Paper by Pascaline Dupas: “Education and HIV/AIDS Prevention: Evidence from a randomized evaluation in Western Kenya”, 2006, with Esther Duflo, Michael Kremer and Samuel Sinei. World Bank Policy Research Working Paper Series 4024).
  - Find, if possible, a population where the DVD was not shown and do a pre-and post-test. Moreover, the program could follow-up six months afterwards.
  - Conduct a formative research to see where the DVD was used and where not and which were the results. This does not have to necessarily be a big and costly survey.
  - Embed the tool into a program, so that its M&E will be integrated into the larger M&E system of the program.
3.3. Developing an evaluation plan
The second part of the group work was to discuss the steps to an evaluation plan. What are the reasons to conduct an evaluation? Does the intervention require an evaluation? Why to evaluate?

3.3.1. Female Genital Mutilation
The group came up with the following work plan for evaluation:

M&E work plan and budget to implement an impact evaluation of the FGM project in Burkina Faso

Indicator: Reduction of FGM Prevalence by 50% in girls aged 0-14 years in Burkina Faso

Methodology: External evaluation (8-10 months)

<table>
<thead>
<tr>
<th>Activity</th>
<th>Timeline</th>
<th>Responsible person</th>
<th>Budget or Number of Working Days</th>
<th>Expected outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial planning stages</td>
<td>4-6 months</td>
<td>Project coordinator and staff</td>
<td>According to current costs</td>
<td>Authorisation of evaluation plan</td>
</tr>
<tr>
<td>1. Partners and stakeholders will be informed about the intended evaluation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Development of ToR - If the tender is above 20,000 Euros, a tender process will be used. Below this amount, the best consultant out of three best offers received, will be selected.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. A meeting with the successful consultants to discuss the ToR-methodology, budget, design of evaluation, meeting schedules &amp; any other important items in the ToR before signing the contract.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Authorisation from the ethical committee, for the proposal, sample population and study is required.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. A letter from the relevant Ministry is obtained and shared to relevant partners-regional directors, other ministries and community members.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Implementation of the evaluation

1. **Implementation of the evaluation by the consultant.**
2. **Supervision of the consultants in order to ensure quality.**
3. **First draft of the evaluation from consultants is received, and feedback is provided.**

<table>
<thead>
<tr>
<th>Time Frame</th>
<th>Roles and Responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 weeks</td>
<td>Consultant (s) and M&amp;E officer</td>
</tr>
<tr>
<td></td>
<td>According to costs in the contract</td>
</tr>
<tr>
<td></td>
<td>A commented draft evaluation report</td>
</tr>
</tbody>
</table>

### Dissemination of results

1. **Dissemination of final draft to partners for their comments.**
2. **Workshop to disseminate results for partners and stakeholders.**
3. **Dissemination workshops in regions**
4. **After receiving the final comments from relevant stakeholders, an inclusive report will be printed.**
5. **Distribution of final report.**

<table>
<thead>
<tr>
<th>Time Frame</th>
<th>Roles and Responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 weeks</td>
<td>Consultant, M&amp;E officer, key partner</td>
</tr>
<tr>
<td></td>
<td>According to costs in the contract</td>
</tr>
<tr>
<td></td>
<td>A printed final report</td>
</tr>
</tbody>
</table>

The ToR of the consultant should include the following:

- Review relevant documents and conduct a literature review
- Submit an evaluation plan, including details on the study design and the budget needed
- Implementation:
  - Training interviewers
  - Pre-test of questionnaire
  - Revision of questionnaire
  - Conduct study
- Data entry and analysis
- Draft report
- Revise the draft report
- Final draft report
- Validation workshop with stakeholders / partners
- Final evaluation report
### 3.3.2. Join-in-Circuit

<table>
<thead>
<tr>
<th>Activity</th>
<th>Timeline</th>
<th>Responsible</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening for a consultant</td>
<td>Within two weeks (3 working days)</td>
<td>Team leader: overall responsibility Programme officer(s): quality assurance Consultant: conduct the evaluation, analyze and present the results Partner(s): facilitation, coordination and ownership</td>
<td></td>
</tr>
<tr>
<td>Inform the partners about the upcoming evaluation – (involving the partners during the whole evaluation process)</td>
<td>1 month (5 days)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop the ToR for the evaluator / consultant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis of the programme’s already existing data and monitoring tools of the programme and these of other programmes / countries (e.g. pre- and post-test questionnaires, FGDs etc.)</td>
<td>1 month (5 days)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop a concept note on the evaluation process, the tools and data sources to be used (e.g. FGDs, health centre data etc.)</td>
<td>1 month (5 days)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selection and training of the implementation team</td>
<td>1 week (3 days)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-testing of the chosen tools and necessary adaptation</td>
<td>2 weeks (5 days)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data collection and analysis</td>
<td>1 month (20 days)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compiling, reviewing and validating the report</td>
<td>3 weeks (10 days)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dissemination (e.g. workshop, media etc.)</td>
<td>Continuous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback of the partner on the way forward</td>
<td>1 week (1 day)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 3.3.3. Theater for change*42*

1. **What are the key activities in the operational plan of our evaluation?**

Theatre for a Change (TfC) has a very limited budget to conduct an evaluation for the Education Programme for the 2011 – 2012 programmatic year. TfC currently works in ten Teacher Training Colleges with approximately 1,000 pre service teachers and with 4,000 primary school learners in approximately 150 primary schools. There are TfC staff members based at the Teacher Training Colleges, so evaluating that population using a representative sample is achievable and low cost. TfC will conduct an evaluation at this level using the existing plans.

The challenge with evaluation lies at the primary school level. The primary schools TfC works in are distributed nationwide in rural areas – the schools are not close together and difficult to get to. Most schools take two hours to drive to from the closest city or town. In

*42 Notes by Claire Walsh*
addition there is a fuel crisis in Malawi. Fuel cannot be easily sourced, meaning if TfaC sends teams to the field there is a strong possibility they will be stranded until the next fuel delivery. This is a risk to the budget as TfaC has to continue paying for the stranded staff’s accommodation or resort to black market fuel. This presents a logistic and, thus, financial problem.

Therefore, the main focus of TfaC’s planning is the primary school section of the Education Programme because it presents the most challenges. The main activity for the primary school evaluation would be a small evaluation of five schools across the region per year. The same questionnaire used at the baseline will be used at the endline. Qualitative interviews and focus groups will be conducted where the budget allows. If TfaC conducts an endline evaluation of five schools over the 3-year period of the grant, then at the end have surveyed 10% of total schools will have been surveyed. If an external consultant is hired this data can contribute to his final evaluation.

2. What are the time frames for the operational plan of our evaluation?

The evaluation will be done at the end of the school year in June. Data collection in the form of the questionnaire will occur simultaneously in one day and interviews will be conducted over the space of two weeks. Data analysis and the first draft of the report will take two-three weeks because it will be done internally and TfaC has to take into consideration the staff’s other responsibilities.

3. Who is to perform the evaluation?

TfaC cannot afford an external evaluator, so TfaC will have to use its staff members. However, TfaC thought that it could train primary school mentors to conduct the evaluation and interview so there would be capacity for subsequent years. At the end of the project cycle TfaC would have an independent evaluator who would be able to use the data gathered over the years to inform his/her evaluation.

4. What resources will we need for our evaluation?

TfaC can likely complete this evaluation with short, photocopied questionnaires that will be given to staff members when they return to the office in April. They will then administer them during their monitoring visits in the last semester and provide a brief capacity building training on evaluation to the mentor. Depending on budget and time, the TfaC staff members will complete some focus group discussions as well.

43 Currently, the black market is selling fuel at 600 / L. The price at the pump is 290 / L.
5. **What shall be included in the ToR of our evaluators?**

Even though TfaC plans on conducting an internal evaluation, it is important to involve the stakeholders – mentors and primary school administration. It would be good to write up a short overview of the evaluation plan and inform them (as well as ask them to participate) of any additional responsibilities.

6. **How will the results be translated and disseminated?**

The group proposed some creative ways to disseminate our results to the participants and stakeholders. TfaC has a radio show that is broadcasted nation-wide and could dedicate a radio show to the results of the evaluation.
Bibliography


Appendix 1: Monitoring and evaluation pipeline
Appendix 2: Cluster Sampling

In the context of large scale HIV interventions, cluster sampling is the most widely used type of probability sampling. For a two-stage cluster sampling one needs to:

1. **Define clusters** (also known as primary sampling units (PSU)): A cluster is a clearly defined group of sampling elements from which a smaller sub-sample can be selected (e.g. geographic areas, schools, health centers etc.).

2. **Develop a sampling frame**: A sampling frame is a list of all existing clusters in the target population. It is important that sampling frames cover the entire geographic universe defined for a given study. To achieve this, community visits within the program’s target area are often required to compile lists of schools, health facilities, youth organizations, etc. If creating a list of all elements within the entire target population is financially or logistically not feasible, study results are prone to bias as the behaviors of target group members excluded from the possibility of selection for the study might somehow systematically differ from those who were included.

3. **Select a sample of clusters**: Once a sampling frame of relevant PSUs has been created, a sample has to be chosen. The recommended procedure for doing so will depend on whether any information on the size of clusters is available before the selection of sample clusters (also known as measures of size (MOS)). In case of a notable size variance among the clusters, systematic sampling with sampling probabilities proportional to cluster size (PPS) is the most efficient method at the first stage. When MOS for clusters are not available, sample clusters will have to be chosen with equal probability.

*Select a sample of elements from the clusters that have been chosen*: This is often done using simple random or systematic sampling, but one could also use other sampling schemes, depending on the context. Two alternative approaches to second stage sample selection are proposed. The first option, a quota sampling approach, entails including target group members as they come into contact with sample sites or locations until a target sample size has been reached. Alternatively, a “take-all” strategy could be adopted in which all target group members who come into contact with a sample site during a specified data collection interval (for example, on a particular day or night) would be included in the sample irrespective of their number (FHI, 2006) (Bertrand, Magnani, & Rutenberg, 1996).
Appendix 3: Guidance questions

Purpose of the group work is to discuss interactively. Start for each discussion item with the intervention which has been the basis of the case study of your working group. At the end of each session, the results of the four working groups shall be presented to the plenum.

1) Discuss assets and needs of our behavioural intervention’s monitoring and evaluation system. Develop ideas how our behavioural intervention’s monitoring and evaluation system could be improved.

   - Do we have “smart” goals and objectives?
     o Specific
     o Measurable
     o Achievable
     o Relevant
     o Time bound

   - Has the results chain been properly developed?
     o Inputs
     o Activities
     o Outputs
     o Outcomes
     o Impacts

   - Is the “Logic Model” of our intervention working? Does it cover all goals and objectives?

   - Do the indicators follow the RAVESS model? Yes, if they are
     o Reliable
     o Appropriate
     o Valid
     o Easy
     o Sensitive
     o Specific

   - Is the data collection appropriate for our monitoring and evaluation system?
     o Data Sources
     o Base Line
     o Frequency of data collection

   - In which way our data will be analyzed?
     o Regular
     o Management response system to monitoring and evaluation results

2) Discuss steps to your evaluation

   a) Guiding Questions Cluster 1: What are the reasons to conduct an evaluation? Does the intervention require an evaluation? Why to evaluate?
- What kind of evaluation would be appropriate?
  - Activity evaluation
    - How is the programme implemented?
    - Are activities delivered as intended?
    - Fidelity of implementation?
    - Are participants reached as intended?
    - What are participants’ reactions?
  - Outcome evaluation
    - To what extent are desired changes occurring?
    - Are goals met?
    - Who is benefiting / not benefiting?
    - How benefiting / not benefiting?
    - What seems to work / not to work?
    - What are unintended outcomes?
  - Impact evaluation
    - To what extent can changes be attributed to the programme?
    - What are the net effects?
    - What are final consequences?
    - Cost effectiveness: Is the programme worth resources it costs?
    - Cost per unit of effect?

- Which questions do we want to evaluate and how do we want to prioritize these questions?

- Which indicators are appropriate for our evaluation?

- Which data does already exist / is easily accessible?

- Do we have any guidance on what works (e.g. research results)?
  b) Guiding Questions Cluster 2: How to evaluate?

- How shall our evaluation design look like?
  - Quantitative evidence / methods
    - Structured Interview
    - Self-administered Interview
    - Questionnaire design
  - Qualitative evidence / methods
    - In-depth Interviews
    - Focus group discussions
    - Participant observation
    - Cognitive mapping
    - Participatory research
  - Sampling Methods
Discuss necessary steps to manage the evaluation of a BCC intervention

1. What are the key activities in our operational plan of our evaluation?
2. Which time frames has our operational plan of our evaluation?
3. How much will our evaluation cost (working days)?
4. Who is to perform the evaluation?
5. What shall be included in the ToR of our evaluator/s?
6. How the results of the evaluation will be translated and disseminated?
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>AIDS Program Effort Index</td>
</tr>
<tr>
<td>ART</td>
<td>Antiretroviral Treatment</td>
</tr>
<tr>
<td>BCC</td>
<td>Behavioral Change Communication</td>
</tr>
<tr>
<td>BZgA</td>
<td>German Federal Centre for Health Education</td>
</tr>
<tr>
<td>CCT</td>
<td>Conditional Cash Transfer</td>
</tr>
<tr>
<td>CEA</td>
<td>Cost-effectiveness analysis</td>
</tr>
<tr>
<td>CSW</td>
<td>Commercial Sex Workers</td>
</tr>
<tr>
<td>DALY</td>
<td>Disability Adjusted Life Years</td>
</tr>
<tr>
<td>FGM</td>
<td>Female Genital Mutilation</td>
</tr>
<tr>
<td>GBC</td>
<td>Global Business Coalition on HIV/AIDS</td>
</tr>
<tr>
<td>GDC</td>
<td>German Development Cooperation</td>
</tr>
<tr>
<td>GTZ</td>
<td>German Technical Cooperation Agency</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>ICT</td>
<td>Information and Communication Technologies</td>
</tr>
<tr>
<td>IDU</td>
<td>Injecting Drug User</td>
</tr>
<tr>
<td>IEC</td>
<td>Information Education Communication</td>
</tr>
<tr>
<td>JiC</td>
<td>Join-in-Circuit</td>
</tr>
<tr>
<td>MENA</td>
<td>Middle East and North Africa</td>
</tr>
<tr>
<td>MOS</td>
<td>Measures of Size</td>
</tr>
<tr>
<td>MSM</td>
<td>Men having Sex with Men</td>
</tr>
<tr>
<td>PLHIV</td>
<td>People Living with HIV</td>
</tr>
<tr>
<td>PPS</td>
<td>Probability Proportional- to-Size</td>
</tr>
<tr>
<td>PROFILE</td>
<td>Program to Foster Innovation, Learning and Evidence in Health and HIV programs of German development cooperation</td>
</tr>
<tr>
<td>PSU</td>
<td>Primary Sampling Unit</td>
</tr>
<tr>
<td>QALY</td>
<td>Quality Adjusted Life Years</td>
</tr>
<tr>
<td>SRHR</td>
<td>Sexual and Reproductive Health and Rights</td>
</tr>
<tr>
<td>SSA</td>
<td>Sub-Saharan Africa</td>
</tr>
<tr>
<td>STI</td>
<td>Sexually Transmitted Infection</td>
</tr>
<tr>
<td>TfaC</td>
<td>Theater for a Change</td>
</tr>
<tr>
<td>ToR</td>
<td>Terms of Reference</td>
</tr>
<tr>
<td>VCT</td>
<td>Voluntary Counseling and Testing</td>
</tr>
<tr>
<td>WRHI</td>
<td>Wits Reproductive Health HIV Institute</td>
</tr>
</tbody>
</table>
Glossary

**AIDS program effort index** is a composite index designed to measure political commitment and program effort in the areas of HIV prevention and care.

**Behavior change communication** is a participatory and interactive process with communities to develop tailored messages and approaches using a variety of communication channels in order to enable them adopt health-promoting behaviors, as well as develop and maintain individual, community and societal behavior change.

**Cluster** is a clearly defined group of sampling elements from which one can select a smaller sub-sample.

**Concentrated HIV epidemic** is when HIV prevalence is consistently over 5% in at least one defined sub-population and below 1% in pregnant women in urban areas.

**Confounder** is a factor that is prognostically linked to the outcome of interest.

**Cost-effectiveness analysis** is a type of economic evaluation that examines both the costs and health outcomes of alternative intervention strategies.

**Effectiveness evaluation** assesses the extent to which the objectives of a program were achieved and whether the interventions implemented do make a difference.

**Efficacy evaluation** refers to determining whether an intervention could work if implemented optimally.

**Evaluation** is the systematic and objective assessment of an on-going or completed project, program or policy, its design, implementation and results. The aim is to determine the relevance and fulfillment of objectives, development efficiency, effectiveness, impact and sustainability.

**Evaluation plans** describe the overall purpose(s) of the evaluation, the specific evaluation questions to be addressed, the evaluation design and methods to be used, what data is to be collected and how, how the data is to be analyzed, the resources that will be necessary, who will implement the evaluation and the basic evaluation plan timeline.

**Formative evaluation** is research conducted during the planning (or re-planning) stage of a behavior change communication program to identify and resolve intervention and evaluation issues before the program is widely implemented.

**Generalized epidemic** is when HIV is firmly established in the general population (>1%). Although sub-populations at high risk may continue to contribute disproportionately to the spread of HIV, sexual networking in the general population is sufficient to sustain the HIV epidemic.

**Impact indicators** refer to the long-term results at the target population level that are plausibly linked to the work of the program.

**Impacts** are positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.

**Incidence rate** is the number of new cases within a specified time period divided by the size of the population initially at risk.
**Indicator** is a quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention or to help assess the performance of a development actor.

**Input indicators** track the means allocated for implementation of the activities such as financial and time resources, personnel, facilities, equipment and supplies.

**Inputs** refer to financial, human and material resources used for the implementation of a development intervention.

**Logic framework** is a management tool used to improve the design of interventions, most often at the program level. It involves identifying strategic elements (inputs, outputs, outcomes, impact) and their causal relationships, indicators and the assumptions or risks that may influence success and failure. It thus facilitates planning, execution and evaluation of a development intervention.

**Monitoring** is a continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds.

**Non-probability sampling method** is a sampling method where members are selected from the population in some non-random manner.

**Odds ratio** is used to assess the risk of a particular outcome (or disease) if a certain factor (or exposure) is present.

**Outcome evaluation** assesses if the desired outcomes are observed in the target population and if these changes are likely to be the result of the intervention.

**Outcome indicators** refer to the intermediate results at the target population level that are plausibly linked to the work of the program.

**Outcomes** are the likely or achieved short-term and medium-term effects of an intervention’s outputs.

**Output indicators** point out the direct and immediate results of input and processes at the program level.

**Outputs** are the products, capital goods and services which result from a development intervention; may also include changes resulting from the intervention which are relevant to the achievement of outcomes.

**Probability sampling method** is a sampling method where each member of the population has a known non-zero probability of being selected.

**Process evaluation** aims at assessing whether the interventions of a program are actually being carried out as initially planned. It monitors inputs and outputs and assesses the service quality.

**Process indicators** follow up on the implementation of the program.
Recall bias occurs when the way a survey respondent answers a question is affected not just by the correct answer, but also by the respondent's memory.

Risk difference is an absolute measure of effect and it is calculated by subtracting the risk of the outcome in exposed individuals from that of unexposed.

Risk ratio (also known as relative risk) is the risk of an event (or of developing a disease) relative to exposure.

Sampling error is the degree to which a sample might differ from the population.

Sampling frame is the actual set of clusters from which a sample has been drawn.

Sampling is the process of selecting units (e.g. people, schools, organizations) from a population of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen.

Stakeholder is a person or group with an interest in the outcome of an intervention.

Statistical power is the probability of detecting a change given that a change has truly occurred.

Stratum is a subset of the population the members of which share at least one common characteristic.

Target population is a group within a population who share similar characteristics and behaviors and upon whom BCC activities are focused.

Triangulation is the use of multiple theories, sources or types of information, or types of analysis to verify and substantiate an assessment.

Validity is the extent to which the data collection strategies and instruments measure what they purport to measure.