

Methodological issues in measuring the impact of interventions against female genital cutting

IAN ASKEW

Population Council, Nairobi, Kenya

Abstract

With increasing efforts being made to introduce systematic interventions for encouraging abandonment of female genital cutting (FGC) comes the need to better understand how such interventions work and what effects they have. Many interventions are based on theoretical models of behaviour change and so studies to evaluate them should develop indicators appropriate to the type of behaviour change anticipated. Systematic evaluations need also to use some form of quasi-experimental design to be able to attribute change to the intervention and not to any 'natural' change in FGC behaviour or other activities that may be concurrent. A sustained change in the prevalence of FGC is the ultimate indicator and there are several ways this can be measured, although with many limitations given the intimate nature of the practice. Moreover, appropriate sample sizes must be calculated and used to be able to draw valid conclusions. Many of those implementing FGC interventions are not familiar with such basic research principles and so there is an urgent need to ensure that projects are well designed so that valid conclusions concerning their effectiveness can be drawn.

Résumé

Les efforts croissants de mise en place d'interventions systématiques visant à encourager l'abandon de l'excision sont accompagnés du besoin de mieux comprendre le fonctionnement et les effets de ces interventions. Beaucoup de ces interventions s'appuyant sur des modèles théoriques de changement de comportements, leurs évaluations devraient permettre le développement d'indicateurs appropriés aux types de changement de comportements anticipés. Ces évaluations systématiques doivent être élaborées de manière quasi-expérimentale pour permettre de préciser si les changements de comportement sont dus aux interventions, et non à un quelconque facteur «naturel» ou à toute autre activité concomitante à ces interventions.

Une baisse prolongée de la prévalence de l'excision est évidemment le meilleur indicateur, et il existe plusieurs outils pour la mesurer, bien que ceux-ci soient limités par la nature intime de la pratique de l'excision. De plus, pour pouvoir tirer des conclusions valides, il faut utiliser des échantillons de dimensions appropriées. Beaucoup des personnes chargées des interventions sur l'excision ignorent ces principes de base de la recherche. Il est donc urgent de s'assurer que les projets sont correctement élaborés, afin que des conclusions valides sur leur efficacité puissent être tirées.

Resumen

Con los esfuerzos cada vez más numerosos que se están llevando a cabo para introducir intervenciones sistemáticas que convengan a dejar de practicar el corte genital o ablación femenina aparece la necesidad de comprender mejor cómo funcionan esas intervenciones y cuáles son sus efectos. Muchas

Correspondence: Ian Askew is Associate Director for Africa, FRONTIERS in Reproductive Health Program, Population Council, Nairobi, Kenya. Email: iaskew@pcnairobi.org

intervenciones se basan en modelos teóricos de cambio de conducta y, de este modo, los estudios para evaluarlos deberían tener en cuenta indicadores apropiados al tipo de cambio anticipado de comportamiento. Las evaluaciones sistemáticas también deben tener algún tipo de diseño casi experimental para ser capaz de atribuir el cambio a la intervención y no a cualquier cambio ‘natural’ en el comportamiento frente a la ablación femenina u otras actividades que pudiesen ser paralelas. Un indicador principal es el cambio sostenido en el predominio de la ablación del clítoris. Existen varios métodos para medirlo, aunque con muchas limitaciones dada la naturaleza íntima de esta práctica. Además, hay que tener en cuenta y utilizar los tamaños apropiados de muestras para sacar conclusiones válidas. Muchos de los que ponen en práctica las intervenciones frente a la ablación del clítoris no están familiarizados con estos principios básicos de investigación por lo que es necesario y urgente asegurar que los proyectos estén bien diseñados para poder extraer conclusiones válidas con respecto a su eficacia.

Introduction

In recent years, there has been a marked increase in the amount of systematically collected information describing the practice and adverse outcomes of female genital cutting (FGC).¹ This information has been collected through a range of methods, including medical case studies, small-scale surveys and more recently through larger baseline surveys (e.g., GTZ 2003). Since 1989, national Demographic and Health Surveys (DHS) have started including a module with questions on the practice for selected countries, and to date nationally representative data exist for 16 countries where FGC is practised (Carr 1997, MEASURE DHS+ 2003). These descriptive studies have greatly increased understanding about the practice, and have also been used extensively in efforts to advocate for abandonment of FGC.

Many interventions to encourage individuals, families and communities to abandon FGC have been undertaken over the past 70 years by church groups, colonial administrators, government and non-government bodies (World Health Organization 1999; Population Reference Bureau 2001; Toubia and Sharief 2003). These interventions have usually been implemented with little attempt to document how they work, or to evaluate their impact on knowledge, beliefs, attitudes and behaviour, largely because most implementing organizations have been small in size and working with limited budgets. However, several of the larger international development assistance organizations and philanthropic foundations are now supporting interventions with higher levels of funding, which offers the opportunity, in principle, to document and evaluate interventions more systematically. Moreover, with increased levels of funding comes greater expectations that empirical evidence be collected to demonstrate whether or not interventions work, how and why they work, and what effect they have in influencing abandonment of the practice.

A review of FGC-related research issues by the World Health Organization (WHO) notes, however, that ‘the most neglected area is that of applied or operational research on how to design interventions that would convince individuals and communities to stop the practice. Methodologies for monitoring and evaluating different interventions are also lacking’ (WHO 1999). This paper highlights some of the key methodological issues to be considered and offers guidance to those undertaking evaluations of FGC programmes and operations research projects to test FGC interventions.

Behaviour change models and FGC interventions

To evaluate interventions that seek to change behaviour, researchers (as well as those implementing the interventions) need a clear understanding of why and how the

intervention is expected to cause such a change. For example, the ‘diffusion of innovations’ model (Rogers 1962) proposes that new ideas and behaviours are not adopted by all persons at a single point in time, but rather are adopted first by ‘innovators’. Adaptation of the new behaviour diffuses gradually within a community until a critical mass of ‘adopters’ has been reached, at which time the rate of diffusion may accelerate until the new behaviour becomes the norm. This model also recognizes that interventions implemented in a participatory way are likely to lead to more rapid diffusion of behaviour changes. El-Gibaly and colleagues (El-Gibaly *et al.* 2002) have used this model to explain why a decline in FGC has only recently started in Egypt.

The ‘stages of behaviour change’ model builds on the diffusion model to propose a sequence of stages that a person or community needs to pass through for a behaviour change to be made and then sustained. A fuller description of this model and its application to FGC interventions can be found elsewhere (Izett and Toubia 1999).

For those interventions that have been designed explicitly to encourage and enable communities and individuals to move between these stages, an evaluation should use a study design, indicators and data collection methods that can assess the effectiveness of each stage in moving the population on to the next. In most cases, however, FGC interventions have *not* been designed with reference to a theoretical model, but instead developed in response to a particular situation (e.g., uncut girls needing an alternative ritual), or a programmatic experience (e.g., a functional literacy programme leading to women discussing ending FGC), or simply intuition (e.g., converting traditional practitioners). Lack of an underlying theoretical model can make it difficult to identify appropriate indicators for evaluation, however, because the cause and effect relationships between intervention activities and expected outcomes are not always clear or logical. Mackie (2000) provides a helpful analysis of the extent to which the effectiveness of alternative FGC interventions can be explained by sociological theories.

A behaviour change process does not end with a single decision, which could later be reversed, but requires sustaining over time, especially if it is to diffuse widely so that not cutting girls becomes the social norm. Measuring sustained or permanent change can only be done through a longitudinal study spanning several years, and because of resource limitations, most intervention evaluations are obliged to measure an intention to sustain the change (for example, a public declaration or participation in an alternative rite) rather than the sustained change itself.

Designing intervention research studies

Some ‘intervention’ research is purely descriptive, such as case studies of ongoing or completed projects. Studies that analyze and explain the way in which the intervention works have also been undertaken. For example, Abdel-Tawab and Hegazi (2000) synthesized the findings from several case studies of interventions in Egypt, and Chege and colleagues (2001) undertook a study in Kenya that compared the characteristics of those families that did and did not participate in the alternative rites programme.

To evaluate the effectiveness of interventions requires, however, a research design that follows the principles of experimentation. More detailed information about designing experimental research for testing reproductive and sexual health interventions in general can be found elsewhere (Fisher *et al.* 2002, Stephenson *et al.* 2003). but essentially all designs have two basic features:

- there is some control over implementation of the intervention and timing of the data collection;
- the key variables can be measured before as well as after the intervention is introduced.

Measuring the key variables before and after the intervention is introduced into the study communities is the most basic design for testing an FGC intervention, but this type of design can only demonstrate that an FGC intervention *may* have had an effect if changes in attitudes or behaviour are detected between the ‘pre-’ and ‘post-’ measures. This design cannot control for other possible influences that may be happening at the same time, and so is not generally recommended, especially when it is known that a natural change has begun. In situations where it is known absolutely that there is no ‘natural’ change in attitudes towards FGC, or where there is no likelihood of other organizations carrying out activities that may influence attitudes towards FGC, it could be argued that this design is acceptable. Such situations are extremely rare, however, and so studies that can also make a comparison between communities receiving the intervention and similar communities not receiving it are preferable. For example, CARE International used this type of design (known as the ‘pre-post- control group design’) for projects in Ethiopia and Sudan (Chege 2002a), and the Population Council used it for studies in Senegal and Burkina Faso (Diop 2002).

In all four studies, the process of *matching* was used to select the communities forming the comparison group. With matching, the communities are selected because they are (or are believed to be) very similar to the intervention communities in terms of their socio-cultural and other characteristics that may influence FGC attitudes and practices. Matching comparison with intervention communities means that, in principle, the situations are virtually equivalent, and so any attitudinal or behavioural changes found in the intervention communities that are not found in the comparison communities can be attributed to the intervention itself.

There can be difficulties in creating a comparison area in settings where the prevalence of FGC appears to be changing rapidly, because while culturally similar communities once had identical FGC practices and prevalence, change is not necessarily occurring at the same pace, and/ or communities may be at different points in a similar change process. The situation is complicated by different factors related to social change in general or FGC specifically—education, religion, the law against FGC, etc. In such situations it is important to control for these factors statistically so as to adjust for these differences between the comparison communities.

In some studies, a comparison is made between communities receiving one type of FGC intervention with communities receiving another type of FGC intervention. Studies that include more than one intervention can be useful as they compare alternative ways of addressing the same issue among the same population. An example of this design is the study by CARE International in two Somali refugee camps at Dadaab in northern Kenya, which tested and compared a targeted education activity (through which health and social workers discuss FGC in both group and individual interactions) in one camp, with the effect of this intervention plus an advocacy activity (through which community leaders are trained to sensitize community members through public pronouncements and discussions) in the other camp (Chege 2002b). The study tested the hypothesis that the effect of the combined interventions will be greater than the single intervention.²

An alternative way of creating intervention and comparison communities is to firstly select a group of similar communities, and to then *randomly* assign half of the communities to receive the intervention and half to be the control group. In principle this is a stronger design because it ensures that the groups of intervention and comparison communities are completely equivalent. It is extremely difficult to use this design in real life situations,

however, because communities that are the most similar are usually located in close proximity, and so there is the opportunity for interactions between intervention and control communities that would 'contaminate' the influence of the intervention. To date, this design has only been used once in an FGC intervention study, at the Navrongo field station in northern Ghana (Sakeah and Jackson 2002).

The Navrongo study was designed to test two intervention strategies in four paramount chiefdoms in a high prevalence area of the Kassena-Nankana District. The two smaller paramencies were combined into one area to form three areas of approximately equal size. The two interventions are being implemented jointly in one area and separately in the other two, so that the relative effectiveness of each of the three possible combinations of interventions can be tested and compared. Each of the three paramency groups was randomly assigned to receive one of the three intervention strategies. Groups were ranked according to numbers assigned from a conventional random number table to determine which intervention strategy they would receive. A 'stepped wedge design' is being used to implement the intervention in stages, beginning with a pilot area. During analysis, each area is treated as a comparison area prior to the time it begins to receive the intervention.

The power of the study depends on the sample sizes used, which are related to the incidence of the key indicator, female genital cutting. When it was found that the incidence was lower than originally expected, the sample size calculations made before starting the project were too small, and so steps had to be taken to expand the sample size, either by extending the duration of observation or by expanding the population under observation. If the study was not expanded, there was a risk that even if the project showed an impact, the standard errors on estimates would be so high that the impact would be statistically insignificant. For this reason, the study area population was increased by the addition of two neighbouring paramencies, which serve as comparison areas. Because these additional comparison areas were not under observation at the start of the study three years ago, it is necessary to collect retrospective data on FGC incidence in these areas.

Although including comparison communities is strongly recommended, the resources needed for such designs may be greater than some funding agencies are willing to contribute. Traditionally, funding within project budgets for evaluation has been limited to an end of project assessment only, and many interventions are still being introduced without a plan or budget for evaluating their effectiveness through a systematic design. Given that the costs of adding comparison communities also would almost double the evaluation budget, it is understandable why funding agencies that do not normally support strong evaluation components may be reluctant to do so for FGC projects. The onus is on researchers, therefore, to demonstrate that the benefits to be gained by using a stronger design outweigh the additional costs.

One aspect of research into FGC (and particularly in intervention research), that can often be downplayed or sometimes simply ignored, are the ethical principles underlying the way the study is designed and the data collected. Abdel-Tawab (2002) discusses these principles in some depth and analyses how they could be applied in FGC intervention research. Put briefly, she highlights the need to ensure that any intervention study should follow the principles of: beneficence, non-maleficence, respect for autonomy and justice.

Managing unanticipated outcomes

Community level behaviour change interventions cannot be implemented in isolation from other factors that may influence FGC-related attitudes and behaviour, however, and so a

strong intervention study needs to be able to manage and explain outcomes that were unanticipated when the study was originally designed. For studies that include comparison groups, the major challenge is to manage possible 'contamination' of the comparison populations if they also become exposed to the intervention. For example, it was already known that the nomadic lifestyles of the Afar community in Ethiopia and the East Bara communities in Sudan would present challenges to controlling exactly who is and who is not being reached and with what messages by a CARE International project (Chege 2002a). The high level of interaction between the experimental and comparison groups meant that the original design had to be abandoned, although this situation does now present the opportunity to understand *how* the intervention's behaviour change messages are diffused. In addition to monitoring and documenting what has been taking place in both the experimental and control sites, qualitative information was collected to map the migratory patterns and community interactions. Further, questions on exposure to intervention messages and sources of information were included in the endline survey instruments to assess the level of diffusion of FGC behaviour change messages.

Experience from Navrongo in Ghana (Sakeah and Jackson 2002) show that both researchers and implementers should always be prepared to meet unanticipated outcomes that could result in changing the study design. Following the baseline survey, it was found that the prevalence of FGC was considerably lower than anticipated among younger women, and it became necessary to increase the number of girls in the control area by interviewing in an adjacent area to collect information retrospectively about FGC incidence. Researchers have to accept that pre-planned designs can 'breakdown', even after following a participatory planning process.

In many countries, it is becoming increasingly important to control for national legislation or presidential decrees against FGC. In Kenya, there have been a number of presidential decrees banning FGC, and legislation has recently been passed which now outlaws the practice. Since such factors can affect the whole country, and can greatly influence the way in which people answer questions about FGC, it is hard to manage or account for their influence without using a comparison group (Kamau 2002).

Controlling for other possible influences also requires that the organization introducing an intervention must understand as fully as possible the context in which it is operating. Consequently, formative research should always be undertaken first, and this step must be included within a study plan. A vivid example of this is a survey (that collected both quantitative and qualitative data) undertaken by the German Technical Co-operation agency (GTZ) among the Kalenjii community in Rift Valley Province in Kenya (Kamau 2002). This survey indicated an apparently drastic decline in the practice from 65% among adult women to two percent among the youngest generation. Although some decline in the practice could be explained through an apparent change in social norms (which was attributed to the influence of the church and improvements in education), the virtual eradication of the practice within one generation was not credible.

Qualitative data from in-depth interviews and group discussions revealed two alternative explanations. First, that there is underreporting because a proportion of women are cut after completing schooling, and some even after marriage, and so these women would not be reflected in the survey. As the mean age at cutting is 14 years, however, cases of delayed cutting are likely to be few. A more plausible explanation is that the survey was undertaken in the home district of the Kenyan President, who had been a vocal and sustained opponent of FGC since 1982. Consequently, although there is certainly a marked

change in social norms, those continuing the practice are probably unwilling to admit it. However, faced with this information, GTZ could not justify continuing with developing the intervention.

Given the need to understand community values and practices when designing an intervention, and to design an intervention that is fully acceptable to the community whose behaviour it is intended to change, it is strongly recommended that a participatory learning approach (PLA) be used before starting an experimental study design. Using PLA to generate behaviour change makes community members more confident in solving their own problems because they become part of the problem-solving process. This approach has been used successfully in Nigeria (Oduwole 2002), Ghana (Sakeah and Jackson 2002) and in studies implemented by CARE International in Ethiopia, Sudan and northern Kenya (Chege 2002b).

Medicalization of the practice has become an unanticipated and unwanted outcome of education efforts that focus on the health complications in several countries (e.g., Mandara 2000, Shell-Duncan *et al.* 2000), although it can be understood as a rational response by parents who intend to continue the practice but want to minimize potential harm to their daughters. However, medicalizing the procedure still constitutes a violation of a girl's right to bodily integrity and does not address long-term sexual, reproductive and mental health complications that may result from FGC. Medicalization provides a financial incentive for health providers to take up the practice, making it more difficult to eradicate. WHO, the United States Agency for International Development (USAID), and other international organizations are unequivocal in their opposition to medical providers engaging in the practice, and some are working to engage health care providers in opposing medical participation in FGC. It remains, however, a contentious and complex issue (Shell-Duncan 2001).

Indicators for monitoring and evaluating behaviour change

Indicators of change in practice

One obstacle to evaluating specific interventions, and subsequent assessments of characteristics associated with successful interventions, has been the identification of valid and feasible impact indicators; a useful review of indicators has been compiled by RAINBQ through its Review, Evaluation and Monitoring (REM) project (RAINBQ 2002). The ultimate indicators for a reduction in the proportion of girls being cut are the incidence of girls being cut in specific age groups, or the prevalence of the practice in the overall population. Data can be gathered during questionnaire surveys with mothers about recent circumcision status of their daughters or, in situations where FGC is performed at an older age, the girls themselves may be asked for this information (as is the case in Navrongo, Ghana). Clinical examination of the genital area would be required if inaccurate self-reporting is anticipated, but this is generally unacceptable in a survey situation.

If FGC prevalence is used as an indicator, it is essential that age-specific prevalence rates be calculated for measuring the impact of interventions. The reason for this can be seen in an example from Kenya. Using data from 1993 baseline and 1999 endline surveys of their FGC projects in Kenya, PATH found that even though a statistically significant decline from 90% to 82% was found in the overall female population aged 14–60 years, this decline was much more pronounced in the younger age groups, for the simple reason that these are the age groups in which a change in prevalence is possible (PATH 2002). If resources for

evaluation are limited, therefore, only the younger age ranges need be measured at the endline survey to determine changes in the prevalence of cutting.

Comparing age-specific rates over time requires that age at cutting be recorded for both surveys, and that sufficiently large sample sizes be used for each age range and in both rounds of surveys. Controlling for age is now a key feature in analyses of behavioural changes and the role of interventions—for example, Chege and colleagues (Chege *et al.* 2001) did so when comparing the reported cutting of daughters by parents exposed to sensitization activities in Kenya, and El-Gibaly and colleagues (El-Gibaly *et al.* 2002) did the same when comparing self-reported cutting among girls of different ages whose parents had been exposed to media publicity about FGC in Egypt.

Another way of controlling for age can be found in the Navrongo study (Sakeah and Jackson 2002), which recruited and is following a cohort of girls who were aged 12–19 years during the baseline survey in 1999 (12 years being the youngest age for being at risk of cutting). This is an ‘open’ cohort, in that each year of surveillance other girls aged 12 are added. The impact of the two arms of the intervention implemented separately and together will be compared using a discrete time hazard logit model. Any decline in FGC will be measured in terms of the difference between the incidence rates in communities exposed to each intervention compared with the incidence rates in communities not exposed (Akweongo *et al.* 2002).

In ongoing studies to evaluate the Village Empowerment Programme in Senegal and Burkina Faso, the key outcome variable is the proportion of young girls being cut who are currently not cut (Diop *et al.* 2004); the indicator used is the proportion of daughter(s) aged 0–10 years whose mother reports that they are cut (in these countries, 95% of girls are cut by age 10 years). These studies, by necessity, followed panels of women who were exposed and not exposed to the intervention over time, and the change over time evaluated is between the two groups and not within each group.

A life table analysis was used to measure any change in the probabilities of being cut between the baseline and endline surveys for both groups because the experience of some girls in these samples is necessarily ‘censored’, that is, they are not yet circumcised but might be in the future. A life table subdivides the period of observation (i.e., 0–10 years) into smaller time intervals (i.e., individual years) and for each interval all girls who have been observed are included to calculate the probability of being cut occurring during that interval. The probabilities estimated from each of the intervals are then used to estimate the overall probability of the event occurring at different time points.

Indicators of the behaviour change process

Most interventions are implemented for a limited period of time and so may only influence the early stages of behaviour change. Consequently the indicators used during an evaluation must measure the appropriate degree of change that can reasonably be attributed to the intervention in that period of time. For example, statements of intention not to cut future daughters are commonly used indicators of behaviour change, both during individual interviews and as community or group declarations and pledges. Whether these statements should be accepted as indicators of behavioural change (i.e., having reached the ‘action’ stage) or are more appropriate as indicators of attitude change (i.e., the ‘preparation’ stage) depends on the validity of using a public statement as an indicator of personal behaviour. Generally a public pledge is a binding social commitment in most African societies, but making such a statement during an individual and confidential

interview may reflect a courtesy bias to the interviewer rather than a genuine change in the person's intention.

A crucial stage in the behavioural change process is the maintenance of a decision to discontinue FGC (for example a mother deciding not to circumcise a daughter). In situations where whole villages (and in some cases whole intermarrying areas) have publicly declared against the practice, opposition to an individual decision against FGC is expected to be minimized. Holding a public declaration is likely to represent a shift in actual practice, and would thus be a good indicator of reduction in incidence.

Sampling

Calculating the appropriate sample size and developing a strategy for identifying the respondents for a sample are extremely important components of operations research, and can negate the findings of a study if they are not well planned and implemented. The most important issue to bear in mind when developing the sampling plan for intervention studies is that the principles on which the sample size are calculated are different from those used for calculating samples in descriptive studies. This is because the sample size for a descriptive study is based on the need to measure the frequency with which a variable (for example, prevalence of FGC) occurs within a population, whereas the sample size for an intervention study has to enable a researcher to compare the frequency of a variable at two or more times or situations (i.e., in the baseline and endline intervention surveys, and/or in the intervention and control groups), and to be able to detect statistically a change and whether or not there is a meaningful difference between them. More details about these different approaches to calculating sample sizes for intervention studies can be found elsewhere (e.g., Fisher *et al.* 2002), but it is essential that researchers and programme staff fully understand this difference when designing and implementing an FGC intervention study.

In calculating sample sizes for evaluating interventions, it is important to:

- Identify the single most important key variable of behaviour change with which to judge the success of the intervention. It is also necessary to either know or be able to make an accurate estimate of the current level of that variable in the target population.
- Decide on the magnitude of change that the intervention can realistically achieve from this current estimate within the time available, and which would be judged as a success by those interested in this intervention.

Another statistical consideration arises because FGC interventions are usually implemented at the community level, or among groups (or 'clusters') within a community. Sample size calculations must therefore take into account the fact that people within the same community or cluster tend to be more similar than people from different communities or clusters. Because of these possible differences between the clusters, sample size calculations for evaluating community-level interventions should take into account this 'design effect' as it is known. Failure to include this can result in underestimating the sample sizes needed and consequently the results produced may not be statistically valid. Simple methods of including this 'clustering' consideration in calculating sample sizes for community-level interventions have been developed (Hayes and Bennett 1999).

There are other statistical considerations to be considered when calculating the sample size (e.g., confidence and significance levels), but these are the main considerations and ones that will require much discussion and negotiation between those doing the research

and those implementing the intervention to get agreement on them. It is also essential that sample sizes be calculated separately for each of the sub-groups within a population whose behaviour the intervention is intended to change. The decision as to which sub-groups (e.g., adult women, adolescent girls, fathers, etc) to sample will be determined by the causal logic underlying the behaviour change intervention, the indicators chosen and of course by the budget available. If resources are limited, then it will become necessary for these sub-groups to be prioritized in terms of the relative impact of the intervention.

Collecting data to measure indicators

Because of the sensitive nature of FGC, a number of problems can arise when measuring indicators describing the practice. First and foremost is whether or not a woman or girl is actually circumcised, and if so, with what type of cut. Most research into FGC-related behaviour and interventions is undertaken using individual or group interviews, through which respondents are expected to verbally describe their own or their daughter's status. It is impossible, however, to validate the person's response given without actually observing their genitalia, which obviously poses huge ethical and logistical concerns. On the four occasions when this has been done (in Egypt (Huntington *et al.* 1996), the Gambia (Morison *et al.* 2001) and twice in Nigeria (Larsen and Okonofua 2002, Snow *et al.* 2002)), only in urban and peri-urban Nigeria was any difference found between the woman's self-reported status and her observed status, suggesting that questioning individuals may be a valid way of measuring this indicator.

Situations exist, however, in which it is unlikely that self-reported status does accurately reflect actual status. The study undertaken at Navrongo in northern Ghana shows that while longitudinal observation gives researchers a unique opportunity to assess the determinants of FGC on an individual level, repeated observation exposes the fact that some women report their circumcision status inaccurately or inconsistently over time (Sakeah and Jackson 2002, Jackson *et al.* 2003). For example, 50% of women aged 20–24 who reported that they were circumcised in 1995 reported that they were not circumcised in 2000, but this proportion declines steadily with older age groups, so that only 7% of women aged 45–49 denied their status (see figure 1).

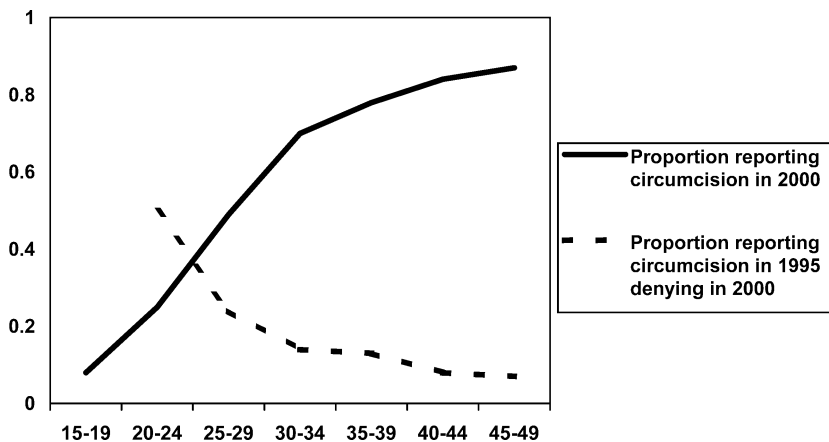


Figure 1. Circumcision and denial rates among Kassena-Nankana District women in the 2000 Panel Survey.

Women who denied being circumcised after earlier indicating that they had been cut were not only younger, but also more likely to be married, educated, and less likely to practice traditional religion than circumcised women who did not subsequently deny their status. Two factors, recent legislation against FGC and rapidly changing attitudes about FGC, may cause women who have previously reported being cut to subsequently report not being cut. Qualitative research suggests this inconsistent reporting is likely to be because of a reluctance to reveal being circumcised to interviewers.

A second set of surveys interviewed adolescent girls aged 12–19 in 1999, 2000 and 2001. Denial among this group appeared to increase over time *and* with exposure to anti-FGC intervention activities. Jackson and colleagues (2003) conclude that in situations where the prevalence of FGC is decreasing due to changes in social support for the practice, where there is legislation against FGC, and/or where an informational campaign to discourage the practice underway, it is particularly important to assess whether or not inaccurate reporting of FGC status biases results. This reason may apply, for example, to the situation in urban Nigeria (Snow *et al.* 2002), where a community-wide reconsideration of the value of FGC may have led women to inaccurately report their status.

In contexts where the practice is illegal, the validity of those respondents who state that their daughters are uncut, or who do not want to cut their daughters in the future is uncertain. The illegal status of the act means that it is not known for sure whether they are replying truthfully or not. Consequently, the proportion of parents who state that their daughters *are cut* is a preferable indicator to those stating they *are not cut*, because it is more likely that these persons are telling the truth, whether or not they know that it is illegal. During analysis of data, however, care needs to be taken to control for factors such as knowledge of the legality of FGC, age of daughter, cut status of other sisters, etc.

Whether self-reporting of FGC status is a valid measure appears to depend, therefore, on the context in which the questions are being asked. If FGC is widespread, socially acceptable and there are no well-publicized interventions causing people to question its acceptability and legality (as was the case in Egypt, the Gambia and rural Nigeria), then self-reporting is likely to be valid. If there are reasons why it would not be attractive for respondents to declare that they are cut (as was the case in northern Ghana, Burkina Faso and parts of Kenya), then self-reported measures should be questioned and ways sought to validate the results.

Clearly it is not normally going to be possible to validate self-reporting through observation of the genitals during administration of a questionnaire survey. The only known example of this was a study in The Gambia (Morison *et al.* 2001); other validation studies have all been in clinic situations. One option may be to collect a sample of observations of actual status (and of type of cut and complications, if appropriate) among clinic clients that is representative of the wider population; the study reported by Jones *et al.* (1999) in Burkina Faso and Mali is an example of this approach. While observing the physical status of adolescent girls is extremely difficult because they rarely attend clinics, such observations could possibly be done during routine antenatal care examinations for first time deliveries, because most women in Africa still have their first delivery at a relatively young age and the vast majority of pregnant women attend for antenatal care. Alternatively, as was done in Kenya and Ghana, qualitative research methods could be used to ascertain whether there are valid reasons why respondents may not be willing to correctly identify their status or accurately describe the type of cut.

Understanding the intervention process

It is critical that sufficient attention and resources be paid to understanding the process through which an intervention is implemented as it happens, including the challenges encountered at each stage and the lessons learnt. Most community-level interventions are never implemented as planned, and descriptions of how activities were actually undertaken in real-life situations is crucial if a successful intervention is to be replicated elsewhere. In particular, being able to understand how and why individuals and families reach decisions about changing their beliefs and behaviours and especially which messages or information most influenced them, can be difficult to do solely through baseline and endline surveys.

Planning for sufficient resources and identifying appropriate research methods for collecting this information are critical when initially designing a study. The Village Empowerment Programme in Senegal, for example, is being implemented in 90 villages and many communities have links through marital and other social relationships to villages that are not in close proximity; consequently, the intervention appears to be having impact beyond the study sites themselves. It has proved essential to have researchers resident in the villages who use ethnographic methods to understand these social networks and communication patterns so as to map and document these ‘ripple’ effects (Diop *et al.* 2004).

Documenting the process of implementation is particularly important when working in unusual or unique situations (such as refugee camps or with nomadic populations), as responses by such communities are often unpredictable. For example, in CARE International’s project in Ethiopia, the traditional *dagu* communication system of the nomadic Afar community has strongly influenced the way in which the messages of the IEC intervention are given (Chege 2002a). Conversely, the implementation of interventions in research field stations, such as Navrongo, are usually extremely well documented to such an extent that the population may suffer from being over-researched and sometimes respondents give the answers they think the researchers want to hear. As a result, although field stations are excellent for testing the effectiveness of medical interventions that are measured through physical indicators of bodily health, they may be too artificial a setting for testing interventions that are measured primarily through reporting of social and behavioural change.

Utilizing research findings

Many countries now have organizations and committees (which may be non-governmental or governmental) that co-ordinate and encourage collaborations between national stakeholders. One of their primary functions is usually sharing information about FGC, and so they are a major audience for research results, and particularly those that are communicated in non-technical language and oriented to provide guidance for information and service programmes.

More needs to be known, however, about the type of evidence and means of communication that are most convincing for programmes and funders. Discussions and negotiations with stakeholders are necessary when writing research proposals to assess the feasibility of the study, interest in the results, and to develop some practical guidelines on how results would be utilized. Moreover, many non-researchers need guidance on interpreting data from operations research studies, and in particular, the validity of findings that each design can and cannot give.

Conclusions

As systematic efforts to develop and implement interventions that encourage the abandonment of FGC increase, so do the needs to ensure that their feasibility and effectiveness are documented and evaluated, and the way in which they function is understood. Programme evaluations and operations research can provide these types of information, but to date few organizations working towards abandonment of the practice, including those undertaking FGC-related research, have embraced the contributions that experimentally-oriented research studies can add to understanding this particular type of behaviour change.

However, there are now sufficient experiences with FGC-related evaluations and operations research to be able to start identifying some general issues that should be considered when designing and implementing such studies, but several challenges remain. These include: ensuring that those implementing and funding anti-FGC interventions appreciate how an operations research approach can benefit their efforts; building organizational and individual capacity to undertake operations research on anti-FGC interventions as well as to use the results from such studies; and communicating widely the findings from operations research so that future activities are evidence-based. The design of concerted behaviour change efforts to encourage abandonment of the practice need to be informed by empirical evidence, and to be evaluated using strong research designs. A better understanding of research methods, and operations research in particular, can contribute to this end.

Acknowledgements

This paper is based on the report of a consultative meeting entitled 'Using Operations Research to Strengthen Programmes for Encouraging Abandonment of Female Genital Cutting' which was convened by the FRONTIERS programme of the Population Council and held in Nairobi, Kenya on 9–11 April 2002. Funding for the meeting was provided by USAID under the terms of Cooperative Agreement number HRN-A-00-98-00012-00. The opinions expressed in this paper are those of the author and do not necessarily reflect the view of USAID or the participants at the meeting. The author is grateful to all of the participants for their insightful presentations and discussions during the meeting, and in particular would like to acknowledge the contributions made by Linda Morison, Elizabeth Jackson, Karin Ringheim, Samson Radeny, Nafissatou Diop and Jane Chege.

Notes

1. This practice is also known as female genital mutilation (FGM), female circumcision and female genital surgeries.
2. The findings showed that this was the case only in relation to increases in knowledge of the harmful effects of FGC. Changes in attitudes and intentions were greater or the same in the site having the single intervention as those observed in the site with combined interventions.

References

- Abdel-Tawab, N. (2002) Ethical Issues in conducting interventions research on FGC. Paper presented at the Conference on Advancing Research on Female Genital Cutting, Bellagio, Italy, 29 April–3 May.
- Abdel-Tawab, N. and Hegazi, S. (2000) *Critical Analysis of Interventions Against FGC in Egypt* (Cairo: Population Council).
- Akweongo, P., Appiah-Yeboah, S., Sakeah, E., Jackson, E. and Phillips, J. (2002) The preliminary impact of a broad-based intervention to decrease the incidence of female genital mutilation: Lessons learned from the

- Navrongo FGM experiment. Paper presented at the Annual Meeting of the Population Association of America, Atlanta, GA.
- Carr, D. (1997) *Female genital cutting: Findings from the Demographic and Health Surveys Program* (Calverton, MD: Macro International Inc.).
- Chege, J. (2002a) Challenges in testing the impact of community-based interventions among refugee and nomadic groups, Paper presented at the Conference on Advancing Research on Female Genital Cutting, Bellagio, Italy, 29 April–3 May.
- Chege, J. (2002b) FGC Operations Research: Challenges in CARE's Multi-Country FGC Abandonment Project in Ethiopia, Kenya and Sudan. Paper presented at the meeting on Using Operations Research to Strengthen Programmes for Encouraging Abandonment of Female Genital Cutting, Population Council, Nairobi, Kenya, 9–11 April.
- Chege, J., Askew, I. and Liku, J. (2001) *An Assessment of the Alternative Rites Approach for Encouraging Abandonment of Female Genital Mutilation in Kenya* (Nairobi: Population Council).
- Diop, N. (2002) Does Changing Reproductive health and human rights knowledge in individuals lead to changing FGC behaviours? Paper presented at the meeting on Using Operations Research to Strengthen Programmes for Encouraging Abandonment of Female Genital Cutting, Population Council, Nairobi, Kenya, 9–11 April.
- Diop, N., Mbacke Faye, M., Moreau, A., Cabral, J., Benga, H., Cissé, B., Mané, B., Baumgarten, I. and Melching, M. (2004) *The TOSTAN Program: An Evaluation of a Community Based Education Program in Senegal* (Dakar: Population Council).
- El-Gibaly, O., Ibrahim, B., Mensch, B. and Clark, W. (2002) The decline of female circumcision in Egypt: evidence and interpretation. *Social Science and Medicine*, 54, 205–220.
- Fisher, A., Foreit, J., with Laing, J., Stoeckel, J. and Townsend, J. (2002) *Designing HIV/AIDS Intervention Studies: An Operations Research Handbook* (New York: Population Council).
- GTZ (2003) *Baseline Data Collection for FGM Programmes: a Review of Selective Approaches and Methods* (Eschborn: GTZ).
- Hayes, R. and Bennett, S. (1999) Simple sample size calculation for cluster-randomized trials. *International Journal of Epidemiology*, 28, 319–326.
- Huntington, D., Nawar, L., Abdel-Tawab, N., Hegazi, S., El Zanaty, F. and Way, A. (1996) *Clinic-based Investigation of the Typology and Self-reporting of FGM in Egypt* (Cairo: Population Council).
- Izett, S. and Toubia, N. (1999) *Learning About Social Change: A Research and Evaluation Guidebook Using Female Circumcision as a Case Study* (New York: Rainbo).
- Jackson, E., Akweongo, P., Sakeah, E., Hodgson, A., Asuru, R. and Phillips, J. (2003) Inconsistent reporting of female genital cutting status in northern Ghana: Explanatory factors and analytical consequences. *Studies in Family Planning*, 34, 200–210.
- Jones, H., Diop, N., Askew, I. and Kaboré, I. (1999) Female genital cutting practices in Burkina Faso and Mali and their negative health outcomes. *Studies in Family Planning*, 30, 219–230.
- Kamau, J. (2002) Implications of a presidential declaration on FGM study, Koibatek District, Rift Valley Province. Paper presented at the meeting on Using Operations Research to Strengthen Programmes for Encouraging Abandonment of Female Genital Cutting, Population Council, Nairobi, Kenya, 9–11 April.
- Larsen, U. and Okonofua, F. (2002) Female circumcision and obstetric complications. *International Journal of Gynaecology and Obstetrics*, 77, 255–265.
- Mackie, G. (2000) Female Genital Cutting: The Beginning of the End. In B. Shell-Duncan and Y. Hernlund (eds) *Female 'Circumcision' in Africa: Culture, Controversy and Change* (Boulder, CO: Lynne Rienner Publishers).
- Mandara, M. (2000) Female Genital Cutting in Nigeria: Views of Nigerian Doctors on the Medicalization Debate. In B. Shell-Duncan and Y. Hernlund (eds) *Female 'Circumcision' in Africa: Culture, Controversy and Change* (Boulder, CO: Lynne Rienner Publishers).
- MEASURE DHS+ (2003) *Female Genital Cutting (FGC): Findings from DHS Surveys, 1990–2002* (Calverton, MD: Macro International Inc.).
- Morison, L., Scherf, C., Ekpo, G., Paine, K., West, B., Coleman, R. and Walraven, G. (2001) The long-term health consequences of female genital cutting in rural Gambia: a community-based survey. *Tropical Medicine and International Health*, 6, 643–653.
- Oduwole, D. (2002) Assessing FGC interventions among populations already exposed to other sensitisation or developmental interventions. Paper presented at the meeting on Using Operations Research to Strengthen Programmes for Encouraging Abandonment of Female Genital Cutting, Population Council, Nairobi, Kenya, 9–11 April.
- PATH (2002) *Evaluating Efforts to Eliminate the Practice of Female Genital Mutilation: Raising Awareness and Changing Harmful Norms in Kenya* (Washington, DC: PATH).

- Population Reference Bureau (2001) *Abandoning Female Genital Cutting: Prevalence, Attitudes, and Efforts to End the Practice* (Washington, DC: Population Reference Bureau).
- RAINBQ (2002) *A Framework for Design, Monitoring and Evaluating Anti-FGM Programs: A Report of the Female Genital Mutilation Review, Evaluation and Monitoring Project 2001–2002* (London: RAINBQ).
- Rogers, E. (1962) *Diffusion of Innovations* (New York: The Free Press).
- Sakeah, E. and Jackson, E. (2002) The Navrongo FGC intervention. Paper presented at the Conference on Advancing Research on Female Genital Cutting, Bellagio, Italy, 29 April–3 May.
- Shell-Duncan, B. (2001) The medicalization of female ‘circumcision’: harm reduction or promotion of a dangerous practice? *Social Science and Medicine*, 52, 1013–1028.
- Shell-Duncan, B., Obiero, W. and Muruli, L. (2000) Women without choices: the debate over medicalization of female genital cutting and its impact on a northern Kenyan community. In B. Shell-Duncan and Y. Hernlund (eds) *Female ‘Circumcision’ in Africa: Culture, Controversy and Change* (Boulder, CO: Lynne Rienner Publishers).
- Snow, R., Slinger, T., Okonofua, F., Oronsaye, F. and Wacker, J. (2002) Female genital cutting in southern urban and peri-urban Nigeria: self reported validity, social determinants and secular decline. *Tropical Medicine and International Health*, 7, 91–100.
- Stephenson, J., Imrie, J. and Bonell, C. (2003) *Effective Sexual Health Interventions: issues in experimental evaluation* (Oxford: Oxford University Press).
- Toubia, N. and Sharief, E. (2003) Female genital mutilation: have we made progress? *International Journal of Gynaecology and Obstetrics*, 82, 251–261.
- World Health Organization (1999) *Female Genital Mutilation Programmes to Date: What Works and What Doesn’t* (Geneva: World Health Organization).