Interventions to reduce the prevalence of female genital mutilation/cutting in African countries

PROTOCOL

Eva Denison, Rigmor C. Berg, Simon Lewin & Jan Odgaard-Jensen

THE CAMPBELL COLLABORATION

First published: 18 Feburay, 2011
Approval date: 07 Feburary, 2011

Contact information:
Eva Denison
The Norwegian Knowledge Centre for the Health Services
Po.Box 7004 St. Olavsplass,
0130 Oslo
Norway
Cell phone: +47 464 00 496 Switchboard: +47 23 25 50 00
Eva.Denison@kunnskapssenteret.no

Funding: 3IE

1. Background

1.1 Definition and classification of female genital mutilation/cutting[1]

Female genital mutilation/cutting (FGM/C) is a traditional practice that involves the partial or total removal or other injury to the female genital organs for cultural or other non-therapeutic reasons (WHO, 2008). The current classification describes four types of FGM/C: Type 1, *clitoridectomy* involves partial or total removal of the clitoris and/or the prepuce. Type 2, *excision*, involves partial or total removal of the clitoris and the labia minora, with or without excision of the labia majora. Type 3, *infibulation*, involves narrowing of the vaginal orifice with creation of a covering seal by cutting and appositioning the labia minora and/or the labia majora, with or without excision of the clitoris. Infibulation is considered the most invasive type of FGM/C. Defibulation, opening of the covering seal, is often necessary prior to childbirth. Reinfibulation refers to the recreation of an infibulation after defibulation. Type 4, *other*, involves all other harmful procedures to the female genitalia for non-medical purposes, for example: pricking, piercing, incising, scraping and cauterizing (WHO, 1997).

1.2 Prevalence of FGM/C

FGM/C is practised in more than 28 countries in Africa, usually on girls under the age of 15 years, and in some countries in the Middle East and Asia. FGM/C is also practised by immigrant communities in a number of other countries, including Australia, Canada, France, New Zealand, Norway, Sweden, Switzerland, the United Kingdom, and the United States (HRP, 2006).

Recent figures for African countries show a prevalence of FGM/C of more than 70 percent in Burkina Faso, Djibouti, Egypt, Eritrea, Ethiopia, Guinea, Mali, Mauritania, Northern Sudan, and Somalia (Yoder & Kahn, 2008). However, there is great variation in prevalence between and within countries, reflecting ethnicity and tradition. Therefore, UNICEF (2005) has proposed that countries be categorized in three groups according to FGM/C prevalence rates: *Group 1*, 80 percent or higher prevalence, e.g. Ethiopia and Somalia, *Group 2*, 25-79 percent, e.g. Senegal and Kenya, and *Group 3*, 1-24 percent, e.g. Nigeria (Table 1).

[1] We adopt the official terminology used by UNICEF and UNFPA 'female genital mutilation/cutting' (FGM/C) throughout this protocol. When referring to published studies we adopt the terminology used by the authors.

Table 1: Prevalence of FGM/C between and within selected countries

Country	Total prevalence [1]	2 lowest [2]	2 highest [2]	Group
Burkina Faso	76.6	41.5	86.9	1
Egypt	97.0	-	-	1
Ethiopia	79.9	0.8	100	1
Kenya	32.2	0.7	96.3	2
Mali	91.6	53.1	98.0	1
Nigeria	19.0	0.5	52.9	3
Senegal	28.2	1.6	78.2	2
Somalia[3]	88.0 [3]	-	-	1

Legend: 1= data from UNICEF (2005). 2= "2 lowest" refers to the two ethnic groups in the country with the lowest FGM/C prevalence; "2 highest" refers to the two ethnic groups in the country with the highest FGM/C prevalence. 3= data from Yoder & Kahn (2008).

1.3 Reasons for the practice of FGM/C

The practice of FGM/C is rooted in religious, personal and societal beliefs within a frame of psycho-sexual and social reasons such as control of women's sexuality and family honour, which is enforced by community mechanisms (WHO, 1999). While reasons for the practice vary across cultural groups, social reasons may include FGM/C as an initiation act for girls into womanhood, as an act of social integration and for the maintenance of social cohesion. Socio-economic reasons include beliefs that FGM/C is a prerequisite for marriage or an economic necessity in cases where women are largely dependent on men. Religious reasons rest on the belief that it is a religious requirement. Hygienic and aesthetic reasons for FGM/C include beliefs that the female genitalia are dirty and unsightly, and health reasons include beliefs that FGM/C enhances fertility and child survival. FGM/C may also be an important source of income for circumcisers (UNFPA, 2007).

1.4 Consequences of FGM/C

1.4.1 Physical consequences

Girls exposed to FGM/C are at risk of immediate physical consequences, such as severe pain, bleeding, and shock, difficulty in passing urine and faeces, and infections. Long term consequences can include chronic pain and infections (WHO, 2008). In general, the consequences are similar for FGM/C Type I, II, and III, but they tend to be more severe and more prevalent the more extensive the procedure (WHO, 2008).

A systematic review of the health complications of FGM/C (WHO, 2000) identified a range of obstetrical problems, the most common being prolonged labour and/or obstruction, episiotomies and perineal tears, post partum haemorrhage, and maternal and foetal death. A recent study investigating 28,393 women attending 28 obstetric

centres in several African countries (Banks et al, 2006) concluded that women with FGM/C are significantly more likely than those without to have adverse obstetric outcomes such as a caesarean, postpartum blood loss ≥500 mL, extended maternal hospital stay, birth weight <2500 g, infant resuscitation, and inpatient perinatal death. The authors also concluded that the risks seemed to be greater with more extensive FGM/C.

1.4.2 Psychological consequences

For many girls and women, undergoing FGM/C is a traumatic experience that leaves a lasting psychological mark and may adversely affect their mental health. In fact, several psychological and psychosomatic disorders such as disordered eating and sleeping habits have been attributed to FGM/C (HRP, 2006). There are also reports of posttraumatic stress disorder, anxiety, and depression associated with FGM/C (WHO, 2008).

1.4.3 Social consequences

FGM/C is a deeply entrenched social convention among some ethnic groups and as such carries consequences both when it is and when it is not practised. When girls and families conform to the practice they acquire social status and respect. For girls, undergoing FGM/C promotes honour and her full acceptance in the community, as well as imparts a sense of pride and of coming of age (UNICEF, 2005). In some societies, the link between FGM/C and value is explicit: girls who undergo FGM/C often receive rewards in the form of celebrations and gifts, and the bride price for a girl who has been cut is much higher than that for one who has not (Wheeler, 2003). For families, fulfilling the cultural expectation that girls should be cut assigns status and community membership. Conversely, failure to conform leads to difficulty in finding a husband for the girl, shame, stigmatization, as well as loss of social status, honour and protection, resulting in the family's social exclusion in the community (UNICEF, 2005).

1.4.4 Sexual consequences

Sexual consequences of FGM/C were summarized in a non-systematic literature review (Obermayer, 2005), which concluded that the available evidence does not support the notion that FGM/C automatically precludes sexual activity or the enjoyment of sexual relations. The results from a recent systematic review by the authors of the present project proposal suggest that women with FGM/C experience pain and reduction in sexual satisfaction and desire compared to women without FGM/C. The evidence base is however insufficient to draw causal conclusions about the consequences of FGM/C (Berg et al, 2010a).

1.5 Interventions to reduce the prevalence of FGM/C

Efforts to abandon the practice of FGM/C in Africa have used several different approaches which, in turn, have had implication for interventions. These approaches include those based on human rights frameworks, legal mechanisms, health risks, alternative rites, positive deviance, training health workers as change agents, training and converting circumcisers, and the use of comprehensive social development processes. Interventions based on these approaches have targeted stakeholders at individual, interpersonal, community and national levels (Muteshi & Sass, 2005).

In 2007, the Population Reference Bureau (PRB) published their results of an extensive survey of current intervention projects taking place in African countries (Feldman-Jacobs & Ryniak, 2007). In total, PRB identified 92 projects, 27 of which were evaluated, mostly by observational designs. Only four of the 27 evaluated projects (15%) used a controlled before-and-after design, and about a dozen of the evaluations used cross-sectional or pre-post intervention questionnaires or interviews without a comparison group. While contributing great understanding about the range of interventions initiated to curb the prevalence of FGM/C, the overview did not reach any conclusions about the effectiveness of interventions.

More recently, the authors of the proposed project specifically examined the effectiveness of interventions to reduce the prevalence of FGM/C in a systematic review (Denison et al, 2009). Through our literature search of February 2009 we identified a total of seven controlled studies, six of which could be obtained in full text. All six studies were controlled before-and-after studies carried out in African countries. In contrast to the PRB overview (Feldman-Jacobs & Ryniak, 2007), we included only controlled studies, i.e. studies with reference to a non-intervention comparison group, and we concluded that while the evidence base is insufficient to draw definite conclusions, there are possible advantageous developments as a result of interventions. Notably, the review highlighted the uncertainties regarding relevance of the interventions (e.g. regarding objectives, intervention targets, activities). That is, since it was not a focus of the systematic review, we were unable to provide any assessment of the degree to which the interventions were appropriate responses to the populations' needs with respect to FGM/C, including the degree to which factors that contribute to the perpetuation of the practice were taken into account in the interventions. It is apparent that the degree of relevance of the intervention exerts a considerable influence on an intervention's effectiveness in reaching its designated goals, and may to a large extent help explain variation in behavioral and other outcomes among members of groups.

In sum, two recent publications have examined aspects of interventions designed to reduce the prevalence of FGM/C. However, the effectiveness of interventions in the context of relevance has yet to be explored.

1.6 Contextual factors related to the continuance or discontinuance of FGM/C

FGM/C is a long-standing tradition that has become inseparable from ethnic and social identity among many groups (UNICEF, 2005). Disaggregation of data from the Demographic and Health Surveys (DHS)[2] shows that the practice of FGM/C varies by demographic variables such as age, urban-rural residence, and region or province, and also by variables such as education, ethnicity and religion (Yoder et al, 2004). Further analysis of DHS data by UNICEF suggests that educational attainment, a woman's own circumcision status and ethnicity have the greatest influence in explaining support or opposition to the practice (UNICEF, 2005). Thus, programmes designed to reduce the prevalence of FGM/C should be country specific and adapted to reflect regional, ethnic and socio-economic variances while also taking into account the diverse reasons why FGM/C is practised among a given ethnic or cultural group (UNICEF, 2005).

[2] MEASURE *DHS+*, ORC Macro, Calverton, Maryland, USA. http://www.measuredhs.com/

2. Objectives

In line with the goals of the Campbell Collaboration we aim to review 1) the empirical research on the effectiveness of interventions to reduce the prevalence of FGM/C in African countries, and 2) the empirical research on contextual factors that may help explain the effectiveness, or lack thereof, of such interventions. Specifically, the proposed review will answer the following research questions:

1. What is the effectiveness of interventions designed to reduce the prevalence of FGM/C compared to no or other active intervention?

2. How do factors related to the continuance and discontinuance of FGM/C help explain the effectiveness of interventions designed to reduce the prevalence of FGM/C?

The review will summarize data relating to 1) key intervention program features, targeted participants, main outcomes, and estimates of intervention effectiveness and 2) factors related to the continuance and discontinuance of FGM/C, such as demographic factors, the frequency and strength of various stakeholders' cognitions and behaviours related to FGM/C, and stakeholders' lived understanding of the persistence of the practice. This will allow us to analyse not only effectiveness of interventions but also their relevance, i.e. the extent to which intervention programs have heeded and built upon factors related to the continuance and discontinuance of FGM/C, the extent to which interventions have been provided to the most appropriate stakeholder groups, and which forces have been overlooked as critical program elements.

3. Methods

Overall, we will conduct the review according to the guidelines in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins & Green, 2009) and guidelines for systematic reviews in the social sciences (Petticrew & Roberts, 2006). Answering research question 2 will involve synthesizing quantitative and qualitative evidence and we will follow guidance by Pope et al (2007). Briefly, we will proceed with the review in the following manner: search for empirical studies, screen studies, extract data, summarize data, perform analyses, and write up results.

3.1 Criteria for inclusion and exclusion in the review

3.1.1 Review question 1: What is the effectiveness of interventions designed to reduce the prevalence of FGM/C compared to no or any other intervention?

Eligible for inclusion:

- *Population*: Girls and/or young women at risk of FGM/C; Members of communities practicing FGM/C such as women, men, traditional circumcisers, religious leaders, educators, community elders, youth, government officials, health workers.
- *Intervention*: Any intervention intended to prevent, or reduce the prevalence of, FGM/C including: Legislation against FGM/C; Education about health risks associated with FGM/C; Training health workers as change agents; Training and

converting circumcisers; Alternative rites; Positive deviance; Comprehensive social development.

- *Comparison*: No FGM/C intervention, wait list, or other active FGM/C intervention.
- *Primary outcomes*: Rates of FGM/C; Behaviours related to FGM/C, such as encouraging others not to cut their daughters.
- *Secondary outcomes*: Intentions regarding FGM/C; Attitudes towards FGM/C; Beliefs related to FGM/C; Knowledge of adverse effects of FGM/C; Awareness of rights.
- *Study designs*: Randomized controlled trials, quasi-randomized trials, controlled before-and-after studies, and interrupted time series designs.
- *Languages*: Publications in all languages will be included and publications in languages not mastered by the review team will be translated into English when considered to meet inclusion criteria.

3.1.2 Review question 2: How do factors related to the continuance and discontinuance of FGM/C help explain the effectiveness of interventions designed to reduce the prevalence of FGM/C?

Eligible for inclusion:

- *Population*: Members of communities practicing FGM/C such as women, men, traditional circumcisers, religious leaders, educators, community elders, youth, government officials, health workers.
- *Interest*: Factors related to the continuance and discontinuance of FGM/C, such as demographic factors, the frequency and strength of various stakeholders' cognitions related to FGM/C, and stakeholders' lived understanding of the practice.
- *Context*: African countries where controlled studies of interventions to reduce the prevalence of FGM/C have been carried out. So far we have identified that controlled interventions have been carried out in Burkina Faso, Egypt, Ethiopia, Kenya, Mali, Nigeria, and Senegal.
- *Study designs*: Cross-sectional quantitative study designs, qualitative study designs, or a combination of the two (mixed-methods studies). Specifically, any type of cross-sectional study design reporting quantitative data. Qualitatively-based studies must have used either individual interviews or focus group interviews to collect data about FGM/C and used qualitative data analysis methods, such as thematic analysis, to be eligible for inclusion.[3] Mixed-methods studies that incorporate both quantitative and qualitative components where the research design matches the nominated study designs will be included. Both the quantitative and the qualitative component of the study will be subjected to the same inclusion criteria as the mono-methods studies and the study will only be included when the inclusion criteria are met.

[3] We define qualitative evidence as "papers had to report results of qualitative (i.e. text-based and interpretive) analysis based on qualitative methods of data collection" (ref Smith 2006, Lancet p826).

- *Study quality*: Study quality will be assessed by checklists appropriate to study design, i.e. cross-sectional and qualitative. Qualitative studies of the highest level of quality will be given priority in the analysis of qualitative studies, and studies with weak methodological quality (as determined by quality assessment) will contribute less to our conclusions. Studies that do not meet any quality criteria at all will be excluded.
- *Languages*: Publications in all languages will be included and publications in languages not mastered by the review team will be translated into English when considered to meet inclusion criteria.

3.1.3 Screening and selection of literature

Selection of primary studies will be based on the inclusion criteria described above. The Reference Manager database containing the search results will be used to keep track of references identified through the electronic database search (to be screened for inclusion).

Screening of literature will proceed at two levels. For level 1 screening, two reviewers (RB and ED) will independently perform an assessment of the identified records by reading the title, and when available, abstract. The pre-developed inclusion questions for level 1 are based on the inclusion criteria described in 3.1.1 and 3.1.2. They are listed in Appendix 1. The reviewers answer each question "Yes" (=promote), or "Can't tell" (=promote), or "No" (=exclude, do not promote). The reviewers then compare and discuss their judgments. Differences in opinion at level 1 screening will be resolved by promoting the record to level 2 screening. Records that unmistakably fail to meet the inclusion criteria will be excluded, such as editorials and commentaries. Records not excluded at level 1 are promoted to level 2 screening, and ordered in full text.

At level 2 screening, two reviewers (RB and ED) will independently evaluate the full text of each record promoted from level 1 screening for inclusion, in accordance with Cochrane guidelines. They will use a pre-developed inclusion form (Appendix 1) based on the inclusion criteria described in 3.1.1 and 3.1.2. There will be a separate set of screening questions for records describing the effect of interventions and records reporting on reasons for the perpetuation of FGM/C (Appendix 1)). The reviewers answer each question "Yes" (=Include), "No" (=Exclude) or "Can't tell" (=Discuss). The reviewers then compare and discuss their assessments. Differences in assessment at level 2 screening will be discussed until consensus is reached. If consensus can’t be reached, a third review member (SL or JOJ) will be asked to resolve disagreements. The reference is included when the reviewers agree to score "Yes" to all questions. The reference is excluded when the reviewers agree to score "No" to any one question. If the reviewers score "Can't tell" to any one question, the inclusion question will be resolved by re-reading of the text, discussion and consensus (or resolved by a third person if consensus can’t be reached). The main reason for exclusion at this stage will be recorded for each record, and a list of excluded records (with reasons) will be created. These steps are in accordance with the Cochrane Handbook (Higgins & Green, 2009)

In the interest of time, the reviewers will not at any screening level be blinded to the authors or other information about the record when assessing the studies. When there is more than one record of the same study, we will include all records meeting the inclusion criteria, but use the most relevant one, i.e. the publication containing the most complete data set, as the main record. Once included, we will group all included studies according

to their methodological focus into three main study types: 1) effectiveness studies, 2) quantitative views studies, 3) qualitative views studies.

3.2 Search strategy for identification of relevant studies

The primary method of study identification will be electronic searches, as advised by the Cochrane Handbook (Higgins & Green, 2009)). For the recently completed systematic review of the effectiveness of interventions designed to prevent the prevalence of FGM/C we searched systematically for relevant literature up to February 2009 in 13 international databases: African Index Medicus, Anthropology Plus, British Nursing Index and Archive, The Cochrane Library (CENTRAL, Cochrane Database of systematic Reviews, Database of Abstracts of Reviews of Effects), EMBASE, EPOC, MEDLINE, PILOTS, POPLINE, PsychINFO, Social Services Abstracts, Sociological Abstracts, and WHOLIS. Under the guidance of one reviewer (ED) a research librarian performed the searches using a strategy incorporating subject headings (for example MeSH terms in MEDLINE, see http://www.nlm.nih.gov/mesh/) and text words (in title and abstract) relating to FGM/C and the four classifications thereof. No method filters were applied as we were more concerned about sensitivity than specificity and prepared to screen a large number of references. We did not restrict our searches by country or language. The MEDLINE search strategy served as the model for the other database searches using appropriate controlled vocabulary as applicable. Our search strategy for MEDLINE is shown:

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1950 to Present

1. Circumcision, Female/
2. ((female$ or wom#n or girl$1) adj3 (mutilation$ or circumcis$ or cutting$)).tw.
3. "fgm/c".tw.
4. ((removal$ or alteration$ or excision$) adj6 female genital$).tw.
5. pharaonic circumcision$.tw.
6. sunna.tw.
7. (clitoridectom$ or clitorectom$).tw.
8. (infibulat$ or reinfibulat$ or deinfibulat$).tw.
9. or/1-8

We will include databases for 'grey' literature (defined here as "reports that are produced by all levels of government, academics, business and industry in print and electronic formats but that are not controlled by commercial publishers" [Higgins & Green 2009]): Demographic and Health Surveys , British Library for Development Studies, IDEAS, JOLIS, Google Scholar, and Google. Theses and dissertations will be included. In addition to the electronic database search, we will perform bibliographic back-referencing to identify new leads. We will also supplement the electronic search with searches in databases of six international organizations that are engaged in projects regarding FGM/C: Centre for Development and Population Activities (CEDPA), Population Council, Population Reference Bureau (PRB), The United Nations Children's Fund (UNICEF), The United Nations Population Fund, and The World Health Organization (WHO). Notably, Popline contains ministerial reports on FGM/C. We will also ask for suggestions for literature we may have missed from FGM/C experts, e. g. the research director of Population Council. Forward citation tracking will be performed through the ISI Web of Knowledge in order to identify further studies. Lastly, we will conduct a hand search of relevant journals (African Journal of Reproductive Health, Social Science & Medicine).

3.3 Details of study coding categories

3.3.1 Data extraction

Two authors (RB and ED) will independently extract data from the published sources using a pre-designed data extraction form, as recommended in the Cochrane Handbook.

Data to be extracted from studies included for answering *Research question 1*: publication identification details (author, year, and source), study design, study setting, population, intervention details, comparisons, outcomes, follow-up details, attrition and missing data details, outcome data (for dichotomous data: number of events, number of persons in the groups, and p-values; for continuous data: means, standard deviations, and p-values), and effect estimates. Intervention details will be classified according to a taxonomy of behaviour change techniques used in interventions proposed by Abraham and Michie (2008). The data extraction form for RQ 1 appears in Appendix 4.

Data to be extracted from studies included for answering *Research question 2*: publication identification details (author, year, and source), study design, study setting, population, quantitative or qualitative data related to the continuance and discontinuance of FGM/C, i.e. demographic factors, the frequency and strength of various stakeholders' cognitions and behaviours related to FGM/C, and stakeholders' lived understanding of the practice. With respect to qualitative-based studies, the reviewers will read the texts independently of each other and extract all text data related to views about FGM/C in light of stakeholders' reasoning for its continuance and discontinuance. Our extraction of data will be inclusive (Briggs et al 2007). We will copy all findings in the form of sentences, phrases or text units appearing to deal with reasons for FGM/C's continuance and discontinuance verbatim onto our pre-designed data extraction form. The data extraction form for RQ 2 appears in Appendix 5.

After independent extraction of data, the two reviewers will compare their forms, and resolve any discrepancies by discussion and consulting the text.

3.3.2 Quality assessment

Two authors (RB and ED) will independently assess the quality of studies, using appropriate checklists for included study designs (see below). Next, they will compare and discuss their assessments. If consensus can't be reached, a third person will be asked to resolve disagreements.

Studies included for answering *Research question 1* will be assessed by The McMaster University, Effective Public Health Practice Project, Quality Assessment Tool for Quantitative Studies (Thomas, nd) (Appendix 2). The Cochrane Handbook recommends this tool for reviews in the public health domain (Higgins & Green, 2009: Chapter 21 Reviews in public health and health promotion). The tool includes the following sections: A) selection bias, B) study design, C) confounders, D) blinding, E) data collection methods, F) withdrawals and drop-outs, G) intervention integrity, and H) analyses. Sections A – F are each given a rating of strong, moderate or weak according to pre-specified criteria. Finally, a global rating of strong, moderate or weak is given, according to the following criteria: strong (four strong ratings with no weak ratings); moderate (less than four strong ratings and one weak rating); weak (two or more weak ratings). A final decision of strong, moderate or weak methodological quality will be agreed upon by the reviewers after discussing whether there was a discrepancy with respect to the

component (A – F) ratings. Component ratings and global rating will be reported in table format.

Quantitative studies included for answering *Research question 2* will be assessed according to guidelines for evaluating prevalence studies (Boyle, 1998). The guidelines cover sampling-, measurement-, and analysis issues which have been compiled into a checklist comprising seven quality criteria by researchers at The Norwegian Knowledge Centre for the Health Services (2009) (Appendix 3). A global rating of strong, moderate or weak study quality is given according to the following: high quality (all/almost all criteria are met, potential weaknesses cannot change the conclusion of the study); moderate quality (some criteria are not met or not sufficiently described, potential weaknesses may change conclusions); low quality (few criteria are met or sufficiently described, the conclusion may be wrong); exclude (no criteria met). The reviewers will agree on a final decision of high, moderate or low methodological quality (or exclude), after discussing whether there was a discrepancy with respect to the seven criteria.

Qualitative studies included for answering Research question 2 will be assessed by a tool designed by the Critical Appraisal Skills Programme (CASP, 2006). The tool comprises 10 questions addressing the rigour, credibility, and relevance of qualitative studies. Two questions are screening questions to determine whether it is worth proceeding with the remaining questions. The remaining eight questions cover research design, sampling, data collection, reflexivity (research partnership relations/recognition of researcher bias), ethical issues, data analysis, findings, and value of the research. The tool does not contain explicit guidance as how to judge the quality of studies. We will apply the same procedure as for the assessment of prevalence studies.

3.4 Statistical procedures and conventions

For the synthesis of effectiveness evidence (Research question 1; Synthesis 1, Figure 1, Appendix 6), the studies will be grouped by intervention, and key intervention features, outcomes and effect estimates will be described in tables. Sensitivity analyses will be used to examine the stability of the effect estimates in relation to quality of studies.

With respect to statistical analyses, we will present dichotomous data for the outcomes listed in the inclusion criteria in results tables when pre- and post scores for both intervention and comparison groups are reported by study authors, allowing comparison. We will estimate effects of interventions in two ways. First, we estimate effect by the adjusted absolute risk difference (ARD) in which the pre-post change score (in percentage points) in the comparison group will be subtracted from the pre-post change score (in percentage points) in the intervention group. Whether ARD is deemed to be large is a judgment whereby we also take baseline difference into account. Second, we will estimate effect by the relative risk (RR) and 95 percent confidence interval (95%CI) based on post-intervention data. The benefit of using this approach is that effect can be estimated even though prognostic similarity at baseline cannot be assumed due to non-randomization. Because none of the effect studies we have identified so far is randomized we will need to use this approach. Should randomized studies be identified in the update search and included in the review we will synthesize these separately. An obvious drawback is that some element of judgment of baseline similarity will be present when conclusions are drawn from the results. We will present continuous data with mean difference and 95%CI.

Forest plots of effect sizes for each grouping of outcomes will be presented. Subgroup analyses or moderator analyses will be used to explore sources of heterogeneity. If sufficient numbers of studies are identified (a minimum of ten studies with one dichotomous moderating variable) meta-regression will be used to perform moderator analyses. Potential moderating variables are intervention content and underlying theory of action (e.g. educational only, training health workers as change agents, comprehensive social development), stakeholder group/s involved, and setting.

3.5 Treatment of quantitative and qualitative research

The contribution of quantitative and qualitative data describing contextual factors will pertain primarily to the understanding of heterogeneous results from the intervention studies, and to help define interventions more precisely in relation to their contexts (Research question 2).

We will utilize an integrative evidence approach, whereby quantitative evidence has methodological dominance. Data extraction and analyses of quantitative and qualitative evidence will largely be completed in separate streams (Figure 1, Appendix 6). Data from cross-sectional survey studies (quantitative data) will be combined with data from studies which examined various stakeholders' perspective of factors influencing the continuance and discontinuance of FGM/C (qualitative 'views' studies). Our integrative evidence approach is largely based on published examples and guidelines from the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI) (e.g., Harden et al., 2004; Shepherd et al., 2006; Thomas & Harden, 2008). The synthesis will be aggregative (Dixon-Woods et al., 2006) and focus on summarizing data by pooling conceptually similar data from the quantitative studies and the qualitative studies. First, we will analyze the two sets of evidence separately (each step is delineated below). Results from the quantitative data set will be used as organizing principles for the qualitative data analysis. Throughout the analysis we will place most weight on the quantitative results, such that the qualitative results are subsumed under the quantitative results and the qualitative results extend the results from the quantitative analysis.

With respect to the quantitative studies (Synthesis 2, Figure 1, Appendix 6), we will use a generic inverse variance approach (Higgins, 2009) to synthesize the extracted quantitative data across studies. Further, for each study and each group of stakeholders, we will review the units extracted and categorize them according to our predefined categories, such as expressed reasons for continuing FGM/C. We will then determine the frequencies of these categories in order to create a ranked list of factors, ending up with one list for each stakeholder group in each country/local region. Thus, we will be able to both synthesize data across studies (settings) and describe local variation.

With respect to the qualitative evidence (Synthesis 3, Figure 1, Appendix 6), our analysis is thematic and draws on published EPPI studies (e.g., Harden et al., 2004; Shepherd et al., 2006). Thematic analysis involves identifying prominent or recurring themes in the literature and summarizing the findings of the different studies under thematic headings (Dixon-Woods, 2005). The data for synthesis of qualitative studies are in text form. In order to synthesize these, we will copy all findings verbatim into a word processing program. These data extracts include both extracts from research participants and extracts of the interpretations made by the researchers. We will organize the findings

from each study according to whether the factors are continuance or discontinuance factors, for each stakeholder group separately. Next, we will examine the findings of each study in turn and assign descriptive codes to describe the findings. Codes will be created without prejudging the meaning of the data and inductively to capture meaning and content of each sentence or phrase. For example, we will code the finding "People perform FGC to reduce a girl's sexual desire" under continuance factors as 'It curbs women's sexuality'. During the coding process, the reviewers will look for similarities and differences between codes in order to start grouping them. Next, we will group findings into thematic categories. This will be based on commonality of meaning as well as frequency and strength of various stakeholders' cognitions about FGM/C, separately for each stakeholder group. Consistent with thematic analysis, it involves taking concepts from one study and recognizing the same concept in another study, though they may not have been expressed using identical words. As explained by Dixon-Woods (2005), thematic analysis can be data driven, i.e. driven by the themes identified in the studies that are included, or theory driven, i.e. driven by themes identified through assessment of the literature. For this systematic review, given quantitative evidence will be given methodological dominance, we will work by using both a priori codes from the included quantitative studies to seek out evidence from the qualitative findings, as well as allowing themes to emerge from the qualitative findings. Text units appearing to deal with related content will be identified and sorted into categories, to which we assign thematic headings, such as ‘reduced sexual interest’ and ‘affirming community membership’. We will do this separately for each stakeholder group, thus developing broader concepts that captured similar themes from different papers. The organization of findings into related areas will first be conducted individually by two reviewers, who then discuss and agree on a set of categories. We will select a set of quotations for each category that represents views that appear frequently, thereby describing the stakeholders' understandings of the factors related to the continuance and discontinuance of FGM/C.

In the last qualitative analysis step, having created categories that represent descriptive themes of the findings, we will then combine categories to create synthesized themes. This will involve reflecting on the thematic categories as a whole and looking for similarities and differences among the categories. The two main reviewers will work together and examine the themes and their corresponding codes in light of the review question, inferring continuance and discontinuance factors from the views stakeholders were expressing about FGM/C. Through discussion and reflection, the main authors will come to a consensus on overall understanding and develop a comprehensive set of analytic themes, across stakeholder groups, such as ‘sexual restraint’ and ‘community allegiance’. For each analytic theme, we will select a set of quotations that capture the essence of each theme.

Once both the quantitative and qualitative sets of data are analyzed, they will be integrated (Synthesis 4, Figure 1, Appendix 6). The integration will involve creating a matrix (table) in which we juxtapose the list of quantitative factors and thematic categories for each stakeholder group. We will work from the quantitative results and seek out evidence from the qualitative results, thus working "down" from pre-existing quantitative understandings. We will generate a set of statements that represents the aggregation of the underlying forces of the continuation and discontinuation of FGM/C from different papers.

The procedure described above was used by Berg and Denison in a systematic review of factors promoting and hindering the practice of female genital mutilation/cutting (FGM/C) from the viewpoints of stakeholders residing in Western countries (Berg 2010b) In the final step of the analysis (Synthesis 5, Figure 1, Appendix 6) the results from Synthesis 1 (effectiveness of interventions) will be integrated with the results from Synthesis 4 (factors related to the continuance and discontinuance of FGM/C) in a realist synthesis model that aims to explain the extent to which intervention programs have heeded and built upon factors related to the continuance and discontinuance of FGM/C, the extent to which interventions have been provided to the most appropriate stakeholder groups, and which forces have been overlooked as critical program elements.

Realist synthesis is concerned with review of complex social interventions (Pawson, 2006). It follows the same "standard" sequence as systematic reviews, i.e. identifying the research question, searching for primary studies, quality appraisal, extracting the data, synthesizing the data, and dissemination of the results. The basic components of realist explanation are outcome patterns, generative mechanisms, and contextual conditions. The result of a realist synthesis would be described in an abstract model explaining how efficacy (outcome patterns) of an intervention varies depending on the particular configuration of its constituent mechanisms and contexts (Pawson, 2006). In our proposed synthesis (Synthesis 5, Figure 1, Appendix 6) the effect estimates extracted from the effectiveness studies will provide the outcome patterns. The data describing the interventions will be used to identify, or generate hypotheses about, the change theories underlying the interventions. The results from Synthesis 4 will provide data concerning contextual conditions (e.g. factors related to the continuance and discontinuance of FGM/C in different stakeholder groups). Recent examples of realist syntheses are Greenhalg et al. (2007) and Kane et al. (2010).

4. Review team

Rigmor ("Rimo") C Berg, Ph.D, adjunct assistant professor, is a researcher at the Norwegian Knowledge Centre for the Health Services. She has been the principal investigator on several projects related to sexual health, primarily HIV prevention among sexual minorities, during the past seven years. Rimo has also collaborated on related projects. For the past few years, she has conducted several systematic reviews, both as project leader and as collaborating reviewer. All systematic reviews that she has been involved in have covered some health aspect and several of them a sexual health topic. For example, she recently secured a grant from the European Centre for Disease Prevention and Control to write a systematic review on the effectiveness of HIV/STI prevention interventions for men who have sex with men. She is the author and co-author of several systematic reviews on FGM/C, three of which were recently completed and one that is ongoing.

Eva Denison, Ph.D, associate professor, is a researcher at the Norwegian Knowledge Centre for the Health Services. She has several years of primary research experience in behavioural medicine, including assessment of risk factors for persistent pain and disability, and development and evaluation of comprehensive interventions for secondary

and tertiary prevention of persistent pain and disability. She is the project leader for three completed systematic reviews on FGM/Ca.

Simon Lewin, Ph.D, is a researcher at the Norwegian Knowledge Centre for the Health Services and editor for the Cochrane Effective Practice and Organization of Care Review Group and the Cochrane Consumers and Communication Review Group. He has contributed to several published reviews and also has experience in conducting syntheses of qualitative studies.

Jan Odgaard-Jensen, M.Sc, is a Statistician/Senior Advisor at the Norwegian Knowledge Centre for the Health Services. Jan is also a Statistics Advisor/Editor with two Cochrane Review Groups: Methodology, and EPOC Norwegian Satellite. Jan has worked as a statistician on a number of Systematic Reviews within healthcare. He has also experience in conducting and analysing data from primary research (clinical Phase III trials and effect studies). Jan has a broad theoretical background and training in mathematical statistics, making him qualified to perform a wide variety of statistical analyses, including meta-analysis, meta-regression, time series analysis, and survival analysis.

The team recently completed a systematic review on the effectiveness of interventions to reduce the prevalence of FGM/C (Denison et al, 2009).

5. Timeline

	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct
C2 title registration	X										
Search for literature		X	X								
Selection of literature			X	X							
Revision of protocol draft (internal)			X								
Data extraction				X	X						
Revision of protocol draft (external)				X	X						
Analysis and summarisation						X	X				
Write review							X	X			
Peer review									X	X	
Revision, publication and dissemination											X

6. Plans for updating the review

The appropriate timing of an update will be discussed in the review team and also with our appointed consultants, Dr Elise R Johansen at WHO and Dr Ian Askew at the Population Council, who have a good overview of evaluations of ongoing projects aiming to reduce the prevalence of FGM/C. Update at 24 or 36 months will probably be a relevant time frame.

7. Acknowledgements

We would like to thank 3ie for funding this review.

8. Statement concerning conflict of interest

The authors are not aware of any conflict of interest, financial or otherwise, that may influence the objectivity of the review.

9. References

Abou Shabana K, el Shiek M, el Nazer M and Samir N. (2003) Women's perceptions and practices regarding their rights to reproductive health. *East Mediterr Health J*, 9:296-308.

Abraham C, Michie S. (2008) A taxonomy of behaviour change techniques used in interventions. *Health Psychol*, 27:379-387.

Asekun-Olarinmoye EO, Amusan OA. (2008) The impact of health education on attitudes towards female genital mutilation (FGM) in a rural Nigerian community. *Eur J Contracept Reprod Health Care*,13:289-297.

Babalola S, Brasington A, Agbasimalo A, Helland A, Nwanguma E and Onah N. (2006) Impact of a communication programme on female genital cutting in eastern Nigeria. *Trop Med Int Health*, 11:1594-1603.

Banks E, Meirik O, Farley T, Akande O, Bathija H et al. (2006) WHO study group on female genital mutilation and obstetric outcome. Female genital mutilation and obstetric outcome: WHO collaborative prospective study in six African countries. *Lancet,* 367:1835-1841.

Berg R, Denison E, Fretheim A. (2010a) Psychological, social and sexual consequences of female genital mutilation/cutting (FGM/C): a systematic review of quantitative studies. Report from Kunnskapssenteret nr 13–2010. Oslo: Nasjonalt kunnskapssenter for helsetjenesten.

Berg R, Denison E, Fretheim A. (2010b) Factors promoting and hindering the practice of female genital mutilation/cutting (FGM/C).Report from Kunnskapssenteret nr 23-2010. Oslo: Nasjonalt kunnskapssenter for helsetjenesten.

Boyle M. (1998) Guidelines for evaluating prevalence studies. *Evidence Based Mental Health,* 1:37-39.

Briggs LA. (2002) Male and female viewpoints on female circumcision in Ekpeye, Rivers State, Nigeria. *Afr J Reprod Health*, 6:44-52.

Bodunrin, HO. (1999) Female genital mutilation: perceptions and beliefs in a Nigerian rural community. *J Pan-African Anthropological Association*, 6:72-81.

Critical Appraisal Skills Programme (CASP). (2006) *10 questions to help you make sense of qualitative research*. National Health Services, Public Health Resource Unit. http://www.phru.nhs.uk/Doc_Links/Qualitative%20Appraisal%20Tool.pdf

Chege JN, Askew I, Igras S and Mutesh JK. (2004) *Testing the effectiveness of integrating community-based approaches for encouraging abandonment of female*

genital cutting into CARE's reproductive health programs in Ethiopia and Kenya. Population Council. Frontiers in Reproductive Health. Washington D.C.

Denison E, Berg RC, Lewin S and Fretheim, A. (2009) *Effectiveness of interventions designed to reduce the prevalence of female genital mutilation/cutting*. Report from Kunnskapssenteret nr 25–2009. Oslo: Nasjonalt kunnskapssenter for helsetjenesten.

Diop N, Sangaré M, Tandia F and Touré K. (1998) *Etude de l'efficacité de la formation du personnel socio-sanitaire dans l'educacion des client(e)s sur l'excision et dans le traitemaent de ses complications au Mali*. Mali: Population Council.

Diop N, Faye MM, Moreau A, Cabral J and Benga H. (2004) *The TOSTAN Program. Evaluation of a community-based education program in Senegal*. New York: Population Council.

Diop N, Moreau A and Benga H. (2008) *Evaluation of the long-term impact of the TOSTAN programme on the abandonment of FGM/C and early marriage: results from a qualitative study in Senegal*. Dakar, Senegal: Population Council, Frontiers in Reproductive Health.

Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. (2005) Synthesising qualitative and quantitative evidence: a review of possible methods. Journal of Health Services Research Policy, 10(1): 45-53.

Dixon-Woods M, Bonas S, Booth A, Jones D, Miller T, Sutton A, Shaw R, Smith J, Young B. (2006) How can systematic reviews incorporate qualitative research? A critical perspective. Qualitative Research 6(1): 27-44.

Ekenze SO, Ezegwui HU and Adiri CO. (2007) Genital lesions complicating female genital cutting in infancy: a hospital-based study in south-east Nigeria. *Ann. Trop. Paediatr*, 27:285-290.

Elnashar A and Abdelhady R. (2007) The impact of female genital cutting on health of newly married women. *Int J Gynaecol Obstet*, 97:238-244.

El-Zanaty F, Hussein EM, Shawky GA, Way AA and Kishor S. (1996) *Egypt Demographic and Health Survey 1995*. Cairo, Egypt: National Population Council, Sep. xxiv.

Ericksen KP (1995) Female circumcision among Egyptian women. *Women's Health*, 1:309-328.

Feldman_Jacobs C and Ryniak S. (2007) *Abandoning Female Genital Mutilation/Cutting: an in-depth look at promising practices*. Washington D.C.: Population Reference Bureau.

Freymeyer RH and Johnson BE. (2007) An exploration of attitudes toward female genital cutting in Nigeria. *Population Research and Policy Review*, 26:69-83.

Greenhalg T, Kristjansson E, Robinson V. (2007) Realist review to understand the efficacy of school feeding programmes. *BMJ*, 335:858-861.

Harden A, Garcia J, Oliver S, Rees R, Shepherd J, Brunton G, Oakley A. (2004) Applying systematic review methods to studies of people's views: an example from public health research. *Journal of Epidemiology and Community Health, 58,* 794-800.

Higgins JPT and Green S. (2009) *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.2.* The Cochrane Collaboration.

HRP. The UNDP/UNFPA/WHO/World bank special programme of research, development and research training in human reproduction. (2006) Progress in sexual and reproductive health, Newsletter 72. Geneva: World Health Organization.

Jaldesa GW, Askew I, Njue C and Wanjiru M. (2005) *Female genital cutting among the Somali of Kenya and management of its complications*. New York: Population Council.

Kane S, Gerretsen B, Scherpbier R, Dal Poz M, Dieleman M. (2010) A realist synthesis of randomised controlled studies involving use of community health workers for delivering child health interventions in low and middle income countries. *BMC Health Services Research*, 10:286.

Mounir GM, Mahdy NH and Fatohy IM. (2003) Impact of health education program about reproductive health on knowledge and attitude of female Alexandria University students. *J Egypt Public Health Assoc*, 78:433-466.

Muteshi J and Sass J. (2005) *Female genital mutilation in Africa: an analysis of current abandonment approaches*. Nairobi: PATH.

Obermayer M. (2005) The consequences of female circumcision for health and sexuality: An update on the evidence. *Culture, Health & Sexuality*, 7:443-461.

Odimegwu C, Okemgbo C and Bisiriyu L. (1998) Female genital mutilation in Nigeria: What do adult and adolescent men know*? J Nigerian Anthropological and Sociological Association*, 1-2:97-108.

Ouoba D, Congo Z, Diop N, Melching M and Banza B. (2004) *Experience from a community based education program in Burkina Faso, the Tostan Program. (Expérience d'un programme d'éducation à base communautaire au Burkina Faso, le Programme de Tostan).* Washington D.C.: Population Council, Frontiers in Reproductive Health.

Pawson R. (2006) *Evidence-based policy. A realist perspective*. Sage Publications.

Petticrew M, Roberts H. *Systematic reviews in the social sciences*. A practical guide. Blackwell Publishing. 2006.

Pope C, Mays N and Popay J. (2007) *Synthesizing qualitative and quantitative health research. A guide to methods*. McGraw-Hill, Open University Press.

Shepherd J, Harden A, Rees R, Brunton G, Garcia J, Oliver S, Oakley A. (2006) Young people and healthy eating: a systematic review of research on barriers and facilitators. Health Education Research, 21(2): 239-257.

The Norwegian Knowledge Centre for the Health Services. Nasjonalt kunnskapssenter for helsetjenesten. (2009) *Slik oppsummerer vi forskning*. Håndbok for Nasjonalt kunnskapssenter for helsetjenesten. 2nd ed. Oslo: Nasjonalt kunnskapssenter for helsetjenesten.

Thomas H. (No Date) *Quality assessment tool for quantitative studies*. Effective Public Health Practice Project. Toronto: McMaster University.

Thomas J and Harden A. (2008) Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8:45.

United Nations Population Fund. (2007) *A Holistic Approach to the Abandonment of Female Genital Mutilation/Cutting*. New York: United Nations.

United Nations Children's Fund. (2005) Female genital mutilation/female genital cutting: a statistical report. New York: United Nations.

Wheeler P. (2003) The role of the law. *Int J Children's Rights*, 11:257-271.

UNAIDS, UNDP, UNECA, UNESCO, UNFPA, UNHCHR, UNHCR, UNICEF, UNIFEM, WHO. (2008) *Eliminating female genital mutilation: an interagency statement*. Geneva: World Health Organization.

WHO. (2000) *A systematic review of the health complications of female genital mutilation including sequelae in childbirth*. Geneva: World Health Organization.

WHO. (1999) *Female genital mutilation. Programmes to date: what works and what doesn't? A review*. Geneva: World Health Organization.

A Joint WHO/UNICEF/UNFPA Statement. (1997) *Female genital mutilation*. Geneva: World Health Organization.

Yoder S, Abderrahim N and Zhuzhuni, A. (2004) *Female genital cutting in the Demographic and Health Surveys: A critical and comparative analysis*. DHS Comparative Reports No. 7. USA, Maryland: ORC Macro, Calverton.

Yoder S and Kahn S. (2008) *Numbers of women circumcised in Africa: the production of a total*. DHS working papers. USA: United States Agency for International Development.

APPENDIX 1

INCLUSION FORM LEVEL 1 SCREENING
Questions for promoting / excluding records from title or abstract reading

Ref number: _____ First author, year_____

Reviewer decision (after completing the form):
Promote ☐ (All questions are answered "YES" or "Can't tell")
Exclude ☐ (Some questions are answered "No")

Final decision (after discussion):
Promote ☐ (All questions are answered "YES" or "Can't tell")
Exclude ☐ (Some questions are answered "No")

1. Is it a study pertaining to FGM/C?
Yes ☐ No ☐ Can't tell ☐

2. Are there reported data?
Yes ☐ No ☐ Can't tell ☐

3. Are the reported data either regarding effectiveness of an anti-FGM/C intervention, or, continuance and discontinuance of FGM/C?
Yes ☐ No ☐ Can't tell ☐

4. Are the study participants members of a community practising FGM/C?
Yes ☐ No ☐ Can't tell ☐

INCLUSION FORM LEVEL 2 SCREENING
Questions for including / excluding records from full text reading

REVIEW 1: "EFFECTIVENESS OF INTERVENTIONS"

Ref number: _____ First author, year_____

Reviewer decision (after completing the form):
Include ☐ (All questions are answered "YES")
Discuss ☐ (Some questions are answered "CAN'T TELL")
Exclude ☐ (Some questions are "NO")

Final decision (after discussion):
Include ☐ (All questions are answered "Yes")
Exclude ☐ (Some questions are answered "No") → Reason /#: __

---

1. Are there reported data from a primary study with at least two groups that received a pretest and a posttest? Design = RCT, quasi-RCT, CBA study, interrupted time series [Mixed methods studies that incorporate both quantitative and qualitative components where the research design matches the nominated study designs will be included, but only the quant component, which is subject to the same inclusion criteria as the mono-methods studies]
Yes ☐ No ☐ Can't tell ☐

2. Are the participants females at risk of FGM/C or other members of a community practising FGM/C?
Yes ☐ No ☐ Can't tell ☐

3. Is an intervention given that is intended to prevent or reduce the prevalence of FGM/C? [including but not limited to: legislation, education, training, alternative rites, positive deviance, social development]
Yes ☐ No ☐ Can't tell ☐

4. Is the intervention compared with no intervention, wait list, or (an)other intervention?
Yes ☐ No ☐ Can't tell ☐

5. Is an index of the status of FGM/C an outcome? [Rates of FGM/C; Behaviours related to FGM/C; Attitudes towards FGM/C; Beliefs related to FGM/C; Knowledge of adverse consequences of FGM/C; Awareness of rights]
Yes ☐ No ☐ Can't tell ☐

Comments: _____

_____

INCLUSION FORM LEVEL 2 SCREENING
Questions for including / excluding records from full text reading

REVIEW 2: "FACTORS RELATED TO THE CONTINUANCE AND DISCONTINUANCE OF FGM/C"

Ref number: _____ First author, year_____

Reviewer decision (after completing the form):
Include ☐ (All questions are answered "YES")
Discuss ☐ (Some questions are answered "CAN'T TELL")
Exclude ☐ (Some questions are answered "NO")

Final decision (after discussion):
Include ☐ (All questions are answered "Yes")
Exclude ☐ (Some questions are answered "No")→ Reason /#: __

---

1. Are there reported empirical data from a primary study? [X-sectional quant, qual or mixed] [Mixed methods studies that incorporate both quantitative and qualitative components where the research design matches the nominated study designs will be included. Both the qualitative component and the quantitative component of the study will be subject to the same inclusion criteria as the mono-methods studies]
Yes ☐ No ☐ Can't tell ☐

2. Are the data provided by participants who reside in a country from which we have included an effectiveness study (review 1)? [e.g., Burkina Faso, Egypt, Ethiopia, Kenya, Mali, Nigeria or Senegal]
Yes ☐ No ☐ Can't tell ☐

3. Are the participants stakeholders in FGM/C? [including but not limited to: females at risk of FGM/C, males from a community practicing FGM/C, traditional circumcisers, religious or community leaders, educators, health workers, child- or social welfare workers, legislators and govt. officials]
Yes ☐ No ☐ Can't tell ☐

4. Is part of the study focus factors related to the continuance or discontinuance of FGM/C? [including but not limited to: knowledge, attitudes, beliefs, perceptions, awareness, understandings, demographics]
Yes ☐ No ☐ Can't tell ☐

5. IF THE DATA ARE QUALITATIVE, did the researchers ascertain first hand accounts about people's views about FGM/C by asking directly about FGM/C?
Yes ☐ No ☐ Can't tell ☐

6. IF THE DATA ARE QUALITATIVE, were the data collected through interviews or focus groups?
Yes ☐ No ☐ Can't tell ☐

7. IF THE DATA ARE QUALITATIVE, were the data analyzed in a qualitative way [e.g. phenomenology, narrative, hermeneutics, thematic, etc]
Yes ☐ No ☐ Can't tell ☐

Comments: _____

_____

APPENDIX 2

MCMASTER QUALITY ASSESSMENT TOOL FOR QUANTITATIVE STUDIES (THOMAS, N.D.)

A) SELECTION BIAS

(Q1) Are the individuals selected to participate in the study likely to be representative of the target population?
1. Very likely
2. Somewhat likely
3. Not likely
4. Can't tell

(Q2) What percentage of selected individuals agreed to participate?
1. 80-100% agreement
2. 60-79% agreement
3. Less than 60% agreement
4. Not applicable
5. Can't tell

RATE THIS SECTION:
1. Strong
2. Moderate
3. Weak
See dictionary

B) STUDY DESIGN
Indicate the study design
1 Randomized controlled trial
2 Controlled clinical trial
3 Cohort analytic (two groups pre + post)
4 Case-control
5 Cohort (one group pre + post (before and after))
6 Interrupted time series
7 Other specify _____
8 Can't tell

Was the study described as randomized? If NO, go to Component C.
No Yes
If Yes, was the method of randomization described? (See dictionary)
No Yes
If Yes, was the method appropriate? (See dictionary)
No Yes

RATE THIS SECTION:
1. Strong
2. Moderate
3. Weak
See dictionary

C) CONFOUNDERS
(Q1) Were there important differences between groups prior to the intervention?
1 Yes
2 No
3 Can't tell
The following are examples of confounders:

1 Race
2 Sex
3 Marital status/family
4 Age
5 Socio-economic status (SES) (income or class)
6 Education
7 Health status
8 Pre-intervention score on outcome measure

(Q2) If yes, indicate the percentage of relevant confounders that were controlled (either in the design (e.g. stratification, matching) or analysis)?
1 80 – 100%
2 60 – 79%
3 Less than 60%
4 Can't Tell

RATE THIS SECTION:
1. Strong
2. Moderate
3. Weak
See dictionary

D) BLINDING
(Q1) Was (were) the outcome assessor(s) aware of the intervention or exposure status of participants?
1 Yes
2 No
3 Can't tell

(Q2) Were the study participants aware of the research question?
1 Yes
2 No
3 Can't tell

RATE THIS SECTION:
1. Strong
2. Moderate
3. Weak
See dictionary

E) DATA COLLECTION METHODS
(Q1) Were data collection tools shown to be valid?
1 Yes
2 No
3 Can't tell

(Q2) Were data collection tools shown to be reliable?
1 Yes
2 No
3 Can't tell

RATE THIS SECTION:
1. Strong
2. Moderate
3. Weak
See dictionary

F) WITHDRAWALS AND DROP-OUTS

(Q1) Were withdrawals and drop-outs reported in terms of numbers and/or reasons per group?
1 Yes
2 No
3 Can't tell

(Q2) Indicate the percentage of participants completing the study. (If the percentage differs by groups, record the lowest).
1 80 -100%
2 60 - 79%
3 less than 60%
4 Can't tell

RATE THIS SECTION:
1. Strong
2. Moderate
3. Weak
See dictionary

G) INTERVENTION INTEGRITY

(Q1) What percentage of participants received the allocated intervention or exposure of interest?
1 80 -100%
2 60 - 79%
3 less than 60%
4 Can't tell

(Q2) Was the consistency of the intervention measured?
1 Yes
2 No
3 Can't tell

(Q3) Is it likely that subjects received an unintended intervention (contamination or co-intervention) that may influence the results?
1 Yes
2 No
3 Can't tell

H) ANALYSES

(Q1) Indicate the unit of allocation (circle one)
community organization/institution practice/office individual

(Q2) Indicate the unit of analysis (circle one)
community organization/institution practice/office individual

(Q3) Are the statistical methods appropriate for the study design?
1 Yes
2 No
3 Can't tell

(Q4) Is the analysis performed by intervention allocation status (i.e. intention to treat) rather than the actual intervention received?
1 Yes
2 No
3 Can't tell

GLOBAL RATING
COMPONENT RATINGS
Please transcribe the information from the gray boxes on pages 1-4 onto this page.

	Rate this section (see dictionary)
A SELECTION BIAS	1. Strong 2. Moderate 3. Weak
B STUDY DESIGN	1. Strong 2. Moderate 3. Weak
C CONFOUNDERS	1. Strong 2. Moderate 3. Weak
D BLINDING	1. Strong 2. Moderate 3. Weak
E DATA COLLECTION METHODS	1. Strong 2. Moderate 3. Weak
F WITHDRAWALS AND DROP-OUTS	1. Strong 2. Moderate 3. Weak

GLOBAL RATING FOR THIS PAPER (circle one):
1 STRONG (four STRONG ratings with no WEAK ratings)
2 MODERATE (less than four STRONG ratings and one WEAK rating)
3 WEAK (two or more WEAK ratings)

With both reviewers discussing the ratings:
Is there a discrepancy between the two reviewers with respect to the component (A-F) ratings?
No Yes

If yes, indicate the reason for the discrepancy
1 Oversight
2 Differences in interpretation of criteria
3 Differences in interpretation of study

Final decision of both reviewers (circle one):
1 STRONG
2 MODERATE
3 WEAK

Quality Assessment Tool for Quantitative Studies Dictionary
The purpose of this dictionary is to describe items in the tool thereby assisting raters to score study quality. Due to under-reporting or lack of clarity in the primary study, raters will need to make judgments about the extent that bias may be present. When making judgments about each component, raters should form their opinion based upon information contained in the study rather than making inferences about what the authors intended.

A) SELECTION BIAS
(Q1) Participants are more likely to be representative of the target population if they are randomly selected from a comprehensive list of individuals in the target population (score very likely). They may not be representative if they are referred from a source (e.g. clinic) in a systematic manner (score somewhat likely) or self-referred (score not likely).
(Q2) Refers to the percentage of subjects in the control and intervention groups that agreed to participate in the study before they were assigned to intervention or control groups.

B) STUDY DESIGN
In this section, raters assess the likelihood of bias due to the allocation process in an experimental study. For observational studies, raters assess the extent that assessments of exposure and outcome are likely to be independent. Generally, the type of design is a good indicator of the extent of bias. In stronger designs, an equivalent control group is present and the allocation process is such that the investigators are unable to predict the sequence.

Randomized Controlled Trial (RCT)
An experimental design where investigators randomly allocate eligible people to an intervention or control group. A rater should describe a study as an RCT if the randomization sequence allows each study participant to have the same chance of receiving each intervention and the investigators could not predict which intervention was next. If the investigators do not describe the allocation process and only use the words 'random' or 'randomly', the study is described as a controlled clinical trial.
See below for more details.

Was the study described as randomized?
Score YES, if the authors used words such as random allocation, randomly assigned, and random assignment. Score NO, if no mention of randomization is made.

Was the method of randomization described?
Score YES, if the authors describe any method used to generate a random allocation sequence. Score NO, if the authors do not describe the allocation method or describe methods of allocation such as alternation, case record numbers, dates of birth, day of the week, and any allocation procedure that is entirely transparent before assignment, such as an open list of random numbers of assignments. If NO is scored, then the study is a controlled clinical trial.

Was the method appropriate?
Score YES, if the randomization sequence allowed each study participant to have the same chance of receiving each intervention and the investigators could not predict which intervention was next. Examples of appropriate approaches include assignment of subjects by a central office unaware of subject characteristics, or sequentially numbered, sealed, opaque envelopes. Score NO, if the randomization sequence is open to the individuals responsible for recruiting and allocating participants or providing the intervention, since those individuals can influence the allocation process, either knowingly or unknowingly. If NO is scored, then the study is a controlled clinical trial.

Controlled Clinical Trial (CCT)
An experimental study design where the method of allocating study subjects to intervention or control groups is open to individuals responsible for recruiting subjects or providing the intervention. The method of allocation is transparent before assignment, e.g. an open list of random numbers or allocation by date of birth, etc.

Cohort analytic (two group pre and post)
An observational study design where groups are assembled according to whether or not exposure to the intervention has occurred. Exposure to the intervention is not under the control of the investigators. Study groups might be non-equivalent or not comparable on some feature that affects outcome.

Case control study
A retrospective study design where the investigators gather 'cases' of people who already have the outcome of interest and 'controls' who do not. Both groups are then questioned or their records examined about whether they received the intervention exposure of interest.

Cohort (one group pre + post (before and after))
The same group is pretested, given an intervention, and tested immediately after the intervention. The intervention group, by means of the pretest, act as their own control group.

Interrupted time series

A time series consists of multiple observations over time. Observations can be on the same units (e.g. individuals over time) or on different but similar units (e.g. student achievement scores for particular grade and school). Interrupted time series analysis requires knowing the specific point in the series when an intervention occurred.

C) CONFOUNDERS
By definition, a confounder is a variable that is associated with the intervention or exposure and causally related to the outcome of interest. Even in a robust study design, groups may not be balanced with respect to important variables prior to the intervention. The authors should indicate if confounders were controlled in the design (by stratification or matching) or in the analysis. If the allocation to intervention and control groups is randomized, the authors must report that the groups were balanced at baseline with respect to confounders (either in the text or a table).

D) BLINDING
(Q1) Assessors should be described as blinded to which participants were in the control and intervention groups. The purpose of blinding the outcome assessors (who might also be the care providers) is to protect against detection bias.
(Q2) Study participants should not be aware of (i.e. blinded to) the research question. The purpose of blinding the participants is to protect against reporting bias.

E) DATA COLLECTION METHODS
Tools for primary outcome measures must be described as reliable and valid. If 'face' validity or 'content' validity has been demonstrated, this is acceptable. Some sources from which data may be collected are described below:
Self reported data includes data that is collected from participants in the study (e.g. completing a questionnaire, survey, answering questions during an interview, etc.).
Assessment/Screening includes objective data that is retrieved by the researchers (e.g. observations by investigators).
Medical Records/Vital Statistics refers to the types of formal records used for the extraction of the data.
Reliability and validity can be reported in the study or in a separate study. For example, some standard assessment tools have known reliability and validity.

F) WITHDRAWALS AND DROP-OUTS
Score YES if the authors describe BOTH the numbers and reasons for withdrawals and drop-outs. Score NO if either the numbers or reasons for withdrawals and drop-outs are not reported. The percentage of participants completing the study refers to the percentage of subjects remaining in the study at the final data collection period in all groups (i.e. control and intervention groups).

G) INTERVENTION INTEGRITY
The number of participants receiving the intended intervention should be noted (consider both frequency and intensity). For example, the authors may have reported that at least 80 percent of the participants received the complete intervention. The authors should describe a method of measuring if the intervention was provided to all participants the same way. As well, the authors should indicate if subjects received an unintended intervention that may have influenced the outcomes. For example, co-intervention occurs when the study group receives an additional intervention (other than that intended). In this case, it is possible that the effect of the intervention may be over-estimated. Contamination refers to situations where the control group accidentally receives the study intervention. This could result in an under-estimation of the impact of the intervention.

H) ANALYSIS APPROPRIATE TO QUESTION
Was the quantitative analysis appropriate to the research question being asked?

An intention-to-treat analysis is one in which all the participants in a trial are analyzed according to the intervention to which they were allocated, whether they received it or not. Intention-to-treat analyses are favored in assessments of effectiveness as they mirror the noncompliance and treatment changes that are likely to occur when the intervention is used in practice, and because of the risk of attrition bias when participants are excluded from the analysis.

Component Ratings of Study:
For each of the six components A – F, use the following descriptions as a roadmap.
A) SELECTION BIAS
Strong: The selected individuals are very likely to be representative of the target population (Q1 is 1) and there is greater than 80% participation (Q2 is 1).
Moderate: The selected individuals are at least somewhat likely to be representative of the target population (Q1 is 1 or 2); and there is 60 - 79% participation (Q2 is 2). 'Moderate' may also be assigned if Q1 is 1 or 2 and Q2 is 5 (can't tell).
Weak: The selected individuals are not likely to be representative of the target population (Q1 is 3); or there is less than 60% participation (Q2 is 3) or selection is not described (Q1 is 4); and the level of participation is not described (Q2 is 5).

B) DESIGN Strong: will be assigned to those articles that described RCTs and CCTs.
Moderate: will be assigned to those that described a cohort analytic study, a case control study, a cohort design, or an interrupted time series.
Weak: will be assigned to those that used any other method or did not state the method used.

C) CONFOUNDERS Strong: will be assigned to those articles that controlled for at least 80% of relevant confounders (Q1 is 2); or (Q2 is 1). Moderate: will be given to those studies that controlled for 60 – 79% of relevant confounders (Q1 is 1) and (Q2 is 2).
Weak: will be assigned when less than 60% of relevant confounders were controlled (Q1 is 1) and (Q2 is 3) or control of confounders was not described (Q1 is 3) and (Q2 is 4).

D) BLINDING
Strong: The outcome assessor is not aware of the intervention status of participants (Q1 is 2); and the study participants are not aware of the research question (Q2 is 2).
Moderate: The outcome assessor is not aware of the intervention status of participants (Q1 is 2); or the study participants are not aware of the research question (Q2 is 2); or blinding is not described (Q1 is 3 and Q2 is 3).
Weak: The outcome assessor is aware of the intervention status of participants (Q1 is 1); and the study participants are aware of the research question (Q2 is 1).

E) DATA COLLECTION METHODS
Strong: The data collection tools have been shown to be valid (Q1 is 1); and the data collection tools have been shown to be reliable (Q2 is 1).
Moderate: The data collection tools have been shown to be valid (Q1 is 1); and the data collection tools have not been shown to be reliable (Q2 is 2) or reliability is not described (Q2 is 3).
Weak: The data collection tools have not been shown to be valid (Q1 is 2) or both reliability and validity are not described (Q1 is 3 and Q2 is 3).

F) WITHDRAWALS AND DROP-OUTS - a rating of: Strong: will be assigned when the follow-up rate is 80% or greater (Q2 is 1). Moderate: will be assigned when the follow-up rate is 60 – 79% (Q2 is 2) OR Q2 is 5 (N/A). Weak: will be assigned when a follow-up rate is less than 60% (Q2 is 3) or if the withdrawals and drop-outs were not described (Q2 is 4).

APPENDIX 3

QUALITY ASSESSMENT OF CROSS SECTIONAL STUDIES
NOKC Checklist. Translated by R. Berg and E. Denison, Oct 2009

Ref.number:_____ First author, year:_____

Checklist for cross sectional studies* Typically only used for answering questions regarding prevalence.		**Yes**	**Unclear**	**No**
1.	**Was the population from which the sample was drawn clearly defined?**			
Comments:				
2.	**Was the sample representative of the population?**			
Comments:				
3.	**Is it explained whether (and how) the participants who agreed to participate are different from those who refused to participate?**			
Comments:				
4.	**Is the response rate adequate?**			
Comments:				
5.	**Were standardized data collection methods used?**			
Comments:				
6.	**Were measures shown to be reliable and valid?**			
Comments:				
7.	**Were the statistical methods appropriate?**			
Comments:				

Quality rating: High ☐ Moderate ☐ Low ☐

* ”Hvordan vurdere en prevalensstudie”, Avdeling for kunnskapsstøtte, Shdir 2003 (Basert på EBM Notebook, Guidelines for evaluating prevalence studies. May 1998, No 2 p 37-9).

Appendix 4

Data extraction form for Review question 1: C2 SW 2009-12

Reviewer:_____ Date:_____

STUDY DATA	DESCRIPTION		
Study identifier: (first author, publ. year) Other references to this study			
Publication type for main reference			
Study location (country, region/city)			
METHODS	DESCRIPTION		
Study design			
Study quality (from QA)			
Study duration (base-fu)			
Targeted stakeholder group(s)			
Ethnic group(s)			
Religious affiliation(s)			
Setting/community			
	Total:	Intervention:	Comparison:

Number of participants			
% female	Intervention:	Comparison:	
Age	Intervention:	Comparison:	
Other participant data			
Intervention(s)			
Behaviour change techniques used (from coding manual)			
Duration of intervention(s), follow-up details			
Control/comparison			
Outcome(s)			

RESULTS	DESCRIPTION	
	Intervention	Comparison
Outcome:	Before: N; fx/md,IQR/m,sd After: N; fx/md,IQR/m,sd/eff est	Before: N; fx/md,IQR/m,sd After: N; fx/md,IQR/m,sd
Outcome:	Before: N; fx/md,IQR/m,sd	Before: N; fx/md,IQR/m,sd

	After: N; fx/md,IQR/m,sd/ eff est	After: N; fx/md,IQR/m,sd
Outcome:	Before: N; fx/md,IQR/m,sd After: N; fx/md,IQR/m,sd eff est	Before: N; fx/md,IQR/m,sd After: N; fx/md,IQR/m,sd
Outcome:	Before: N; fx/md,IQR/m,sd After: N; fx/md,IQR/m,sd/ eff est	Before: N; fx/md,IQR/m,sd After: N; fx/md,IQR/m,sd
Outcome:	Before: N; fx/md,IQR/m,sd After: N; fx/md,IQR/m,sd/ eff est	Before: N; fx/md,IQR/m,sd After: N; fx/md,IQR/m,sd
Outcome:	Before: N; fx/md,IQR/m,sd After: N; fx/md,IQR/m,sd/ eff est	Before: N; fx/md,IQR/m,sd After: N; fx/md,IQR/m,sd

Outcome:	Before: N; fx/md,IQR/m,sd After: N; fx/md,IQR/m,sd/ eff est	Before: N; fx/md,IQR/m,sd After: N; fx/md,IQR/m,sd
Outcome:	Before: N; fx/md,IQR/m,sd After: N; fx/md,IQR/m,sd/ eff est	Before: N; fx/md,IQR/m,sd After: N; fx/md,IQR/m,sd

Comments:_____

_____

_____

_____

Appendix 5

Data extraction form for Review question 2: C2 SW 2009-12

Reviewer:_____ Date:_____

STUDY

Study identifier: (first author, pub.year) _____

Other references to this study: _____

Publication type for main reference: ☐journal article ☐report ☐other: _____

Location where study was conducted: (country, region/city) _____

TYPE OF STUDY DESIGN ☐Qualitative ☐Quantitative

Details about study design: _____

Study quality (from QA): _____

PARTICIPANTS /Stakeholder group(s)

☐Women from FGM/C community _____

☐Men from FGM/C community _____

☐Health workers _____

☐Youth/students _____

☐Other: _____

N= _____ **Age:** mean (SD) _____ Other age data: _____

Cutting status if women from FGM/C community: _____

Ethnic group(s): _____

Education: _____

Religious affiliation: _____

Other participant data: _____

OUTCOMES: _____ →

RESULTS

QUANTITATIVE - Outcome data

Outcome:	N; fx/md,IQR/m,sd
Outcome:	N; fx/md,IQR/m,sd
Outcome:	N; fx/md,IQR/m,sd
Outcome:	N; fx/md,IQR/m,sd
Outcome:	N; fx/md,IQR/m,sd

QUALITATIVE - Factors related to continuance of FGM/C:

QUALITATIVE - Factors related to discontinuance of FGM/C:

Comments:

APPENDIX 6

Figure1: Realist synthesis approach