university of groningen / faculty of economics and business

14017-EEF

Sensitive Survey Questions: Measuring Attitudes Regarding Female Circumcision through a List Experiment

Elisabetta De Cao
Clemens Lutz

SOM is the research institute of the Faculty of Economics & Business at the University of Groningen. SOM has six programmes:
- Economics, Econometrics and Finance
- Global Economics & Management
- Human Resource Management & Organizational Behaviour
- Innovation & Organization
- Marketing
- Operations Management & Operations Research

Research Institute SOM
Faculty of Economics & Business
University of Groningen

Visiting address:
Nettelbosje 2
9747 AE Groningen
The Netherlands

Postal address:
P.O. Box 800
9700 AV Groningen
The Netherlands

T +31 50 363 7068/3815

www.rug.nl/feb/research

university of groningen / faculty of economics and business

Sensitive Survey Questions: Measuring Attitudes Regarding Female Circumcision through a List Experiment

Elisabetta De Cao
Faculty of Economics and Business, University of Groningen

Clemens Lutz
Faculty of Economics and Business, University of Groningen

Sensitive survey questions: Measuring attitudes regarding female circumcision through a list experiment[1]

Elisabetta De Cao
University of Groningen, Department of Economics, Econometrics and Finance, University of Groningen, The Netherlands
Corresponding author: e.de.cao@rug.nl

Clemens Lutz
University of Groningen, Department of Innovation Management & Strategy, University of Groningen, The Netherlands

Abstract

A list experiment is designed to measure the attitudes among women toward Female Genital Mutilation (FGM) by using new data collected in Ethiopia. The results of multivariate regression methods recently developed for the list experiments show that educated women are less in favor of FGM compared to the uneducated ones (6% versus 47%). Using the results of a direct question about FGM support, we show that the social desirability bias is the greatest among uneducated women. In particular, uneducated women that are targeted by a NGO intervention have a stronger incentive to reveal a biased answer.

Key words: FGM; female circumcision; sexual and reproductive health; list experiment; sensitive questions; Ethiopia.

JEL-Classification: I15; O10; C13; C83.

[1]We would like to thank IFPRI and in particular Getaw Tadesse and Samson Jemaneh for supervising the data collection. We would like to thank Aljar Meester, Robert Lensink, Rob Alessie, Viola Angelini, Andreas Rauch, Petros Milionis and Mariko Klasing as well as seminar participants at the 2013 IFP conference in Addis Ababa, at the PEG seminar series at the University of Groningen, the University of Wageningen for useful comments. All errors are our own.

1 Introduction

Female genital mutilation (FGM) or female circumcision includes all procedures that alter or cause injure to the female genital organs. They are mainly carried out on young girls. FGM can lead to reproductive health complications, among which: obstructed menstruation, difficulty in conceiving, and neonatal death. FGM is recognized as an extreme form of discrimination and violence against women. Worldwide about 140 million girls and women are living with the consequences of FGM. The WHO estimates that in Africa more than 3 million girls are at risk for FGM annually (WHO, 2012).

Conventional theories suggest that in a context of extreme resource inequality, FGM is a mean to secure a better marriage by signaling fidelity, or a prerequisite for marriage (Mackie, 1996). Rai and Sengupta (2013) show in a game theoretic framework that pre-marital confinement of women affect their prospects in the marriage market. The authors consider seclusion, footbinding, and FGM as examples of confinement. A recent study by Shell-Duncan et al. (2011) shows that being circumcised serves as a signal to other circumcised women that a girl or woman has been trained to respect the authority of her circumcised elders and is worthy of inclusion in their social network. Therefore, Shell-Duncan et al. (2011) suggest that interventions aimed at eliminating FGM should target women's social networks.

However, a recent study by Bellemare and Steinmetz (2013) shows that factors at village level only account for 15% of the relationship between begin circumcised and support for the practice, while 85% is attributed to individual and household level factors. They also show that a woman who has undergone FGM is 40 percentage points more likely to be in favor of circumcision. Wagner (2011) demonstrates that ethnic and religious identity foster this health-destructing practice and that being cut increases marriage prospects by almost 40%. Wagner (2011) suggests that the concerned women may conceive the FGM-induced health impairments as less important than the socio-cultural gains.

Remarkably, all the quantitative research regarding FGM attitudes is based on direct survey questions. A large body of literature shows that answers to direct questions about sensitive issues may be biased due to incentives to lie (see for example, Jones and Forrest, 2010; Gilens et al., 1998; Presser and Stinson, 1998; Janus, 2010; Comşa and Postelnicu, 2012). In our study, social pressure (social desirability effects) is expected to play an important role as circumcision is formally prohibited but, concomitantly, a widespread advocated custom in the local culture. Therefore, further research on the attitudes that sustain the practice is needed (Obermeyer, 2001).

This paper aims at identifying the true perceptions about FGM using new data collected in Ethiopia. According to the Demographic and Health Survey (DHS), Ethiopia is one of the countries with the highest FGM prevalence, about 74-80% (Yoder and Khan, 2008). Rahlenbeck et al. (2010) explore factors influencing attitudes towards the practice of FGM in Ethiopia. Religion is often used as a justification, even if there is no doctrinal basis for this practice in Islam, Christianity or Judaism. FGM is traditionally believed to ensure hygiene and preserve a girl's chastity and fertility. Hence, the practice is considered beneficial to girls, but also as a prerequisite for a honorable marriage.

In 2004 the Ethiopian government introduced the Criminal Code Proclamation No. 414/2004, that criminalizes harmful traditional practices among which FGM. The Proclamation became law in 2005. In December 2012, the United Nations General Assembly unanimously passed Resolution 67/146, condemning FGM and related harmful practices and urging member states to take measures to accelerate its elimination.[2]

Even though FGM is formally banned in Ethiopia, the practice still exists. Generating knowledge about the causes and consequences of FGM is one of the strategies to eliminate FGM. Survey data are necessary to empirically study these issues. However, eliciting truthful answers in surveys is challeng-

[2]http://www.un.org/sg/statements/index.asp?nid=6529

ing, especially when studying sensitive issues such as attitudes toward FGM. If asked directly, individuals may lie or refuse to answer, leading to biased results. To account for this problem, this paper considers a survey technique called the list experiment. The idea behind a list experiment, also called item count or unmatched count technique, is that if a sensitive question is asked indirectly, the respondent may reveal a truthful response.

The method presents respondents with a list of items and asks to indicate the total number of items with which they agree. The respondents are randomly divided in a control and treatment group. The control group respondents receive a list of non-sensitive items. The treatment group respondents receive the same list of non-sensitive item plus one sensitive item. The difference in the total number of items between control and treatment group identifies the proportion of people in the population that agree with the sensitive item. The list experiment technique has been mainly used in political science to understand voters' attitudes and racial attitudes (for example see Kuklinski et al., 1997; Redlawsk et al., 2010). It has also been used to study sexual risk behavior (LaBrie and Earleywine, 2000). More recently it is also applied in development economics to study sensitive issues. In micro-finance, for example, Karlan and Zinman (2012) used a list experiment to understand how people spend their loan proceeds, showing that direct elicitation under-reports the non-enterprise uses of loan proceeds. In reproductive health, list experiments have been developed to get truthful answers on topics such as condom use, number of sexual partners, unfaithfulness, and attitude changes with respect to the social acceptability of these behaviors (Jamison et al., 2013; Chong et al., 2013).

Even though the difference-in-means estimator is commonly used in the literature to analyze the list experiment (see for example Karlan and Zinman, 2012), it does not allow to estimate the relationship between preferences over the sensitive item and the respondent's characteristics. Moreover, the effect of social pressure on the answers given to direct sensitive questions may differ among groups in the population. A multivariate statistical analysis is indeed needed to study both the list experiment and the difference between direct and indirect questioning (social desirability bias) (Corstange, 2009; Holbrook and Krosnick, 2010; Imai, 2011; Blair and Imai, 2012).

In this paper we present four important contributions. First, we design a list experiment to measure attitudes regarding FGM. Second, we use new data collected in one of the areas where FGM prevalence is among the highest. Third, we use a statistical multivariate regression model to analyze the list experiment. Fourth, we provide a further analysis of the social desirability bias.

The results show that educated women support female circumcision less than uneducated women, the social desirability bias is the greatest among uneducated women, and finally uneducated women targeted by a NGO intervention have a stronger incentive to lie. This confirms the relevance of potential bias in responses to direct sensitive questions and is highly relevant for impact studies measuring the effect of treatments concerning sensitive issues. Our analysis shows that the seemingly positive effect of the NGO intervention on reducing support for the FGM practice seems to disappear if the sensitive question is asked indirectly through the list experiment.

The paper is structured as follows. In Section 2, we present the new data collected in Ethiopia. In Section 3, we describe the list experiment technique and its use in our survey. Section 4 describes the list experiment results. Section 5 presents the social desirability bias analysis. In Section 6, we present some robustness checks. Finally, Section 7 concludes.

2 Data

In this paper we focus on the Afar region, one of the most remote and poorest regions in Ethiopia. The Afar region (pop. 1.5m) is a pastoralist area characterized by conflict, food insecurity and drought. Pastoralists mainly depend on the services of traditional health providers who are not formally trained and are not linked to the formal health referral system. The overall health status of the Afar population is poor, with women and children particularly vulnerable: under-five child mortality is 123/1,000, 20% of the women are pregnant during adolescence, only 0.6% of the children (12-23 months) are fully vaccinated (Macro International Inc., 2008). According to the Afar Regional Health bureau, in 2000, Afar counted 2 hospitals, 14 Health Centers and 112 Health Posts serving a population of 1.5m people (http://www.moh.gov.et).

Figure 1: Map of Afar, Ethiopia.

Note: This map shows the Afar region in Ethiopia (UN OCHA, http://www.unocha.org).

Since 2011 a NGO program is working in some Afar areas to provide comprehensive sexuality education programs and health services. In October 2012, we collected data in Afar to evaluate the impact of the NGO interventions. We used a multi-stage stratified sampling method in which strata were defined by zones representing different target groups and villages. In particular, we selected some of the beneficiaries from areas where the project was implemented (zones 3 and 5), and some non-beneficiaries without access to any of the NGO activities from a different area (zone 1). Figure 1 shows the map of Afar with the different zones highlighted. Since the NGO program mainly targets young people and women of reproductive age, our survey consists of women/mothers aged between 15 and 49 (n=631), and unmarried girls (n=217) aged between 15 and 24 mainly from the same household, for a total of 848 respondents.

The information covered in the questionnaire concerns: the socio-economic back-ground of the respondent, access to sexual and reproductive health services, knowledge about sexual and reproductive health services, attitudes towards sexual and reproductive health practices, use of sexual and reproductive health services, intentions to use sexual and reproductive health services, household water supply and sanitation.

Table 1 reports the descriptive statistics of some variables. As described above, the survey contains individuals that were exposed to the NGO's program (67%). Most of the respondents are Muslim (95%) and of Afar ethnicity (78%). Not many of the respondents have ever participated in any sexual and reproductive health education or training program in the last two years (24%), while the average number of health service providers available in the area (e.g., traditional health services, community health promoters, health extension worker, health centre) is 2.5 (maximum 4), and the average number of health services (e.g., pregnancy test, counseling on pregnancy/child care/contraceptives, medical treatment, condoms, contraceptives) easily accessible is 2.6 (maximum 5). About 72% of the respondents are mothers,

Table 1: Descriptive statistics

Variable	**N**	**Mean**	**Std. Dev.**
Age	845	28.226	9.512
Religion (1=Christian; 0=Muslim)	839	0.049	
Ethnic group (proportions)			
Afar	848	0.78	
Other ethnic minorities	848	0.22	
Areas in Afar (proportions)			
Zone 1	848	0.33	
Zone 3	848	0.34	
Zone 5	848	0.33	
Health education/training (1=yes; 0=no)	835	0.243	
Health providers available (0-4)	848	2.514	1.003
Health services accessible (0-5)	848	2.637	2.041
Having children (1=yes; 0=no)	846	0.722	
Ever being married (1=yes; 0=no)	843	0.770	
Educated† (1=yes; 0=no)	844	0.213	
Sex and HIV knowledge (0-6)*	847	4.046	1.279
NGO program target (1=yes; 0=no)	848	0.667	

Note. * This variable is the percentage of correct answers of a battery of 6 questions related to sexual knowledge and HIV. † includes people that have at least completed elementary school.

and 77% have been married (this includes widows and divorced women). The level of education is very low, with 62% of the sample being illiterate, 5% with adult education, 11% with few years of elementary school (these three classes are considered together for a total of 79%), and 21% with different levels of education (elementary (13%), secondary or tertiary education (5%)).

3 Methods

3.1 Standard list experiment design

In order to measure the true perception about female circumcision we added to the survey a list experiment. The list experiment or unmatched count technique works by aggregating the sensitive item with a list of other non-sensitive items (Miller, 1984). The survey sample is composed by N respondents, that are randomly divided in two groups: treatment and control groups. $T_i = 1$ ($T_i = 0$) implies that the respondent i belongs to the treatment (control) group. The control group respondents receive a list of J non-sensitive, yes/no items and they have to tell the interviewer how many of the listed items they agree on, but not which items. The treatment group respondents instead receive the same list of non-sensitive, yes/no items plus a sensitive, yes/no item ($J+1$ in total), where $j = J+1$ is the sensitive one. The sensitive item measures the sensitive topic. As for the control group respondents, the treatment group respondents have to tell the interviewer the number of items they agree on.

To formalize, we use the same notation as in Imai (2011) and Blair and Imai (2012). Let us define $Z_{ij}(t)$ a dummy variable that indicates the respondent i's preference for the jth control item ($j = 1, \ldots, J$) under the treatment status $t = 0, 1$. The respondent i's answer to the sensitive item for the treatment group is indicated as $Z_{i,J+1}(1)$. Z^*_{ij} corresponds to the respondent i's truthful answer to the jth item where $j = 1, \ldots, J+1$. The potential answer respondent i would give under the control or treatment group, is respectively: $Y_i(0) = \sum_{j=1}^{J} Z_{ij}(0)$ or $Y_i(1) = \sum_{j=1}^{J+1} Z_{ij}(1)$. Finally, $Y_i = Y_i(T_i)$ represents the observed response, and X_i the vector of observed covariates for respondent i, where $X_i \in \chi$ and χ is the support of the covariate distribution.

This design relies on three important assumptions (Imai, 2011). The first assumption is the *randomization of the treatment* and it implies that potential and truthful responses are jointly independent of the treatment variable, hence, for any respondent $i = 1, \ldots, N$, the following needs to hold: $\{\{Z_{ij}(0), Z_{ij}(1)\}_{j=1}^{J}, Z_{i,J+1}(1)\} \perp T_i$. The second assumption called *no design effect* implies that the addition of the sensitive item does not change the sum of affirmative answers to the control items, hence, for each $i = 1, \ldots, N$, we have $\sum_{j=1}^{J} Z_{ij}(0) = \sum_{j=1}^{J} Z_{ij}(1)$. The third assumption is called *no liars* and it implies that the respondents give truthful answers for the sensitive item, for each $i = 1, \ldots, N$, we have $Z_{i,J+1}(1) = Z^*_{i,J+1}$.

If these assumptions hold, then the unbiased estimate of the population proportion of those that agree on the sensitive item can be computed using a difference-in-means estimator:

$$\hat{\tau} = \frac{1}{N_1} \sum_{i=1}^{N} T_i Y_i - \frac{1}{N_0} \sum_{i=1}^{N} (1 - T_i) Y_i,$$

where $N_1 = \sum_{i=1}^{N} T_i$ is the size of the treatment group and $N_0 = N - N_1$ is the size of the control group. The joint distribution of $(Y_i(0), Z^*_{i,J+1})$ can be identified and it characterizes each respondent's type ($2 \times (J + 1)$ types in total).

Imai (2011) proposes new multivariate regression estimators, that also rely on the assumptions of no design effect and no liars, to analyze the relationship between preferences over the sensitive item and the respondent's characteristics. One of the estimators reduces to a linear regression with interaction terms (see also Holbrook and Krosnick, 2010):

$$Y_i = X_i^T \gamma + T_i X_i^T \delta + \epsilon_i, \tag{1}$$

where $E(\epsilon_i | X_i, T_i) = 0$, and (γ, δ) are unknown parameters.[3] To estimate

[3] The estimator in this case is a nonlinear least squares estimator: $Y_i = f(X_i, \gamma) + T_i g(X_i, \delta) + \epsilon_i$, where $f(x, \gamma)$ and $g(x, \delta)$ represent the regression models for the conditional expectations of the control and sensitive items given the covariates. In the case of X_i that contains only an intercept, the difference-in-means estimator is obtained. If linearity is assumed for the two sub-models $f(x, \gamma) = x^T \gamma$ and $g(x, \delta) = x^T \delta$ then the estimator reduces to a linear regression with interaction terms. For further details about the different

(γ, δ) heteroskedasticity-consistent standard errors are computed to account for the difference in the variance of error term between the treatment and control groups.

In this paper we analyze the list experiment using the difference-in-means estimator to estimate the overall proportion of respondents that agree on the sensitive item. We then apply the linear regression estimator to study the different preferences over the sensitive item and the main respondent's characteristics. This estimator more efficiently estimates the relationships between the sensitive item and respondent's characteristics compared to a subgroup analysis. This technique is easy to interpret, but rarely used in the empirical research of the list experiment. Moreover, to the best of our knowledge, it is the first time that this approach is used in the context of reproductive health, hence we believe this is an important contribution.

3.2 Our list experiment

In our survey, the control group was presented with the following question:

I want you to give me a secretive answer for the following statements. I will give you 3(4) stones and you have to hold them in your right hand. Keep your hands (both) on your back side. If you agree on the statement I will soon be reading to you, you transfer one stone to your left hand behind you (I will not see it, you shouldn't also tell me), but if you don't agree, do not transfer any stone. At the end, I would like to know the total number of statements you agreed on. Now, I read the statements:

1. *HIV can be transmitted through witchcraft or other supernatural means*
2. *It is acceptable to use contraceptives to avoid pregnancy*
3. *In a marriage both partners should decide on how many children they should have*

estimators, please see Imai (2011).

For the treatment group, we asked an identical question, but with an extra item, a sensitive item, concerning female circumcision:

4. A girl should be circumcised

Given that the Ethiopian law prohibits FGM, people are expected to be less prone in revealing their true belief about the fourth item that is indeed considered as a sensitive issue. The choice of the non-sensitive items needs to be such that the so called ceiling and floor effects are avoided (Kuklinski et al., 1997). Ceiling effects occur when a respondent would honestly respond "yes" to all nonsensitive items, and in the treatment group the respondent no longer has the protection to honestly report her/his response to the sensitive item. Floor effects instead occur when the respondent in the treatment group whose truthful answer is affirmative only for the sensitive item replies negatively to all the items to cover his/her identity.

Table 2 reports the observed data from the list experiment. As we said the list experiment has three non-sensitive items and one sensitive item. The sample size is 848, of which 443 are in the treatment groups for the circumcision item. We observe that the responses are well distributed and there are few responses in the extreme cases (0 and 3 for the control group, and 0 and 4 for the treatment group). Having many responses in the extreme cases can indicate the presence of ceiling effects, or floor effects.

4 Results of the list experiment

Table 3 reports the results from the list experiment, using the difference in-means estimator, commonly used to analyze the list experiment. The results indicate that 39% is the estimated proportion of women who agree with the sensitive item "a girl should be circumcised".

In addition to knowing the overall proportion of women that agree with FGM, it is interesting to know what type of respondent is more in favor

Table 2: Observed data from the experiment result.

	Control group		Treatment group	
Response value	**Frequency**	**Proportion (%)**	**Frequency**	**Proportion (%)**
0	6	1.5	3	0.68
1	87	21.75	40	9.13
2	246	61.5	231	52.74
3	61	15.25	152	34.7
4			12	2.74
Total	400	100	438	100

Note. The table displays the number of respondents for each value of the observed outcome variable (total number of items the respondent agree on) and its proportions, separately for the control and the treatment group where the sensitive item is “a girl should be circumcised”.

Table 3: List experiment difference-in-means result.

	Control group	**Treatmen group**	**Diff-in-means estimate**
Mean	1.905	2.297	0.392***
S.E.	0.032	0.033	0.047
N	400	438	

Note. The table displays the average response to the observed outcome variable (total number of items the respondent agree on), separately for the control and the treatment group where the sensitive item is “a girl should be circumcised”, and the difference-in-means estimation.
Robust S.E. Signif. codes: (*) if p < .05, (**) if p < .01, (***) if p < .001.

of FGM. One can do the difference-in-means estimator separately in each subgroup, as commonly done (some examples are Kuklinski et al., 1997; McKenzie and Siegel, 2013), but this leads to a small number of respondents at the subgroup level and to an increase in the standard errors.

To address this problem, we apply a linear regression model developed to analyze our list experiment (Holbrook and Krosnick, 2010; Imai, 2011). Table 4 presents the model results.[4] Several variables are highly correlated and we selected some of them for the analysis. In particular, we consider only the ever married variable, and not the mother variable because 94% of the ever married women have kids. We do not include religion, given that 95% of the sample is composed by Muslims. Instead of adding the different geographical zones, we add a dummy variable that indicates if the woman is a beneficiary of the NGO program or not. We then control also for the age of the woman and her education level.

The results show that the coefficient for the education variable in the model for the sensitive item (treatment status=1) is negative, and it is statistically significantly different from zero with a p-value below 1%. This implies that on average educated women are 41% less likely to be in favor of circumcision even after controlling for different individuals' characteristics.

We present in Figure 2 a comparison of the difference-in-means and linear model results considering education as the main variable. Figure 2 is based on the fitted model presented in Table 4 and on the model without covariates (diff-in-means). Figure 2 presents the estimated proportions of uneducated (circle) and educated people (triangle) who agree that "a girl should be circumcised". The difference between those proportions is also shown (diamond). To obtain the estimated proportion for each subgroup in the models with covariates, we computed the predicted probability by setting

[4]The link between mothers and daughters is not available in the data, therefore we cannot use clustered standard error. However, the analysis considering only the mothers do not differ much from the current analysis on the entire sample.

Table 4: Results of the linear regression model for the list experiment.

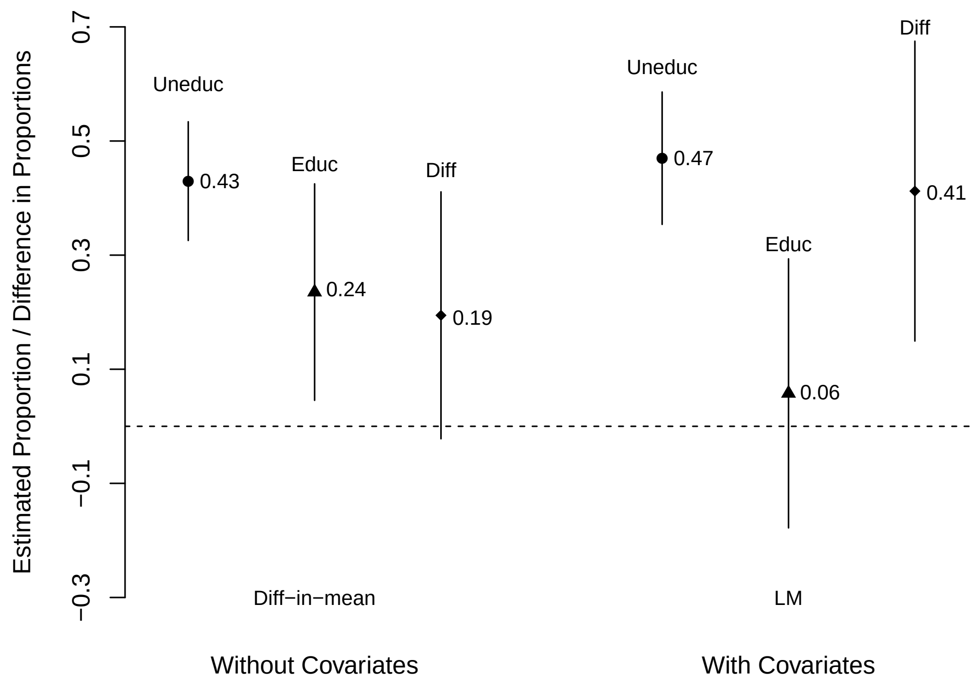
Variables	**Est**	**SE**
Sensitive item		
T	0.813***	0.209
Age×T	-0.007	0.007
Ever married×T	-0.179	0.167
Educated×T	-0.412**	0.147
Other ethnic minorities×T	0.009	0.119
NGO program target×T	-0.022	0.100
Control items		
Intercept	1.721***	0.141
Age	-0.000	0.005
Ever married	0.097	0.117
Educated	0.138	0.097
Other ethnic minorities	0.152*	0.076
NGO program target	0.084	0.068

Note. Estimated coefficients from the item count technique linear regression model 1 where the sensitive item is whether or not "a girl should be circumcised".
T corresponds to the treatment status dummy (1 treated; 0 control).
The sensitive item estimated parameters correspond to δ in equation 1,
The control item estimated parameters correspond to γ in equation 1.
Robust S.E. Signif. codes: (†) if p < .1, (*) if p < .05, (**) if p < .01, (***) if p < .001.

all the other covariates to their observed values.[5] The solid lines correspond to the 95% asymptotic confidence intervals. The model without covariates (diff-in-means) does not present significantly different effects between uneducated and educated women, while the linear regression model confirms the significant difference in attitudes between educated and not educated women. In particular, 47% of the uneducated women agree with FGM, while 6% is the proportion of educated women in the multivariate linear model.

[5] In the case of the list experiment, by keeping a particular X constant and all the other covariates to their observed values, we estimate the difference in the Y predictions between the control and the treatment group.

Figure 2: Estimated proportion of women who are in favor of FGM based on the linear regression model for the list experiment design.



Note. Predictions are based on the difference-in-means estimator when no covariates are used, and on the linear regression model when covariates are considered. The solid lines correspond to 95% confidence interval for the estimated proportions. In the linear regression model the results are averaged over the sample distribution of covariates.

5 Social desirability bias

To assess the impact of sensitivity on responses, we compare the attitudes toward FGM measured when the question is asked directly and when it is asked indirectly via the list experiment. The direct question is not necessary for the purpose of analyzing the list experiment, but it can provide a good comparison between direct and indirect means of eliciting respondent attitudes (Corstange, 2009)

As in Blair and Imai (2012), we define $Z_{i,J+1}(0)$ as the respondent i's potential answer to the sensitive item when asked directly. Since the social desirability bias can also vary across respondents as a function of their characteristics, it is defined as:

$$S(x) = Pr(Z_{i,J+1}(0) = 1|X_i = x) - Pr(Z^*_{i,J+1} = 1|X_i = x), \quad \text{for any} \quad x \in \chi.$$

The first term can be estimated by regressing (using for example a logistic regression or a linear probability model) the observed value of $Z_{i,J+1}(0)$ on X_i. The second term can be estimated using the linear regression estimator described in the methodological section.

In the survey, we asked: "Do you agree on the following statements? A girl should be circumcised."[6] The possible answers were totally agree (200 answers), somehow agree (52), neither agree nor disagree (50), somehow disagree (35) and totally disagree (511). We created the direct question variable by dichotomizing the survey question, assuming that totally agree and somehow agree correspond to 1, and the remaining ones to 0. If asked directly, about 30% of the women agree upon the fact that a girl should be circumcised[7]. The proportion obtained using the difference-in-means estimator is

[6] We asked the direct question both to the control and the treatment group. The questionnaire was extensive and it included many different questions about sexual and reproductive health, and we believe that it did not affect the response to the list experiment.

[7] Note that if we dichotomize the direct question considering as 1 also the respondents who replied neither agree nor disagree, we obtain 35.6% as proportion of women that are

39.2%, hence the difference is 9.2%, slightly statistically different from zero at 10% level (see the no covariates results in Table 6).

Since also the answer to the direct sensitive question might vary as a function of respondent's characteristics, we apply a linear probability model to analyze the responses to the direct question. Table 5 reports the results of the regression where the dependent variable is a dummy variable where 1 corresponds to agreeing that a girl should be circumcised, and 0 the opposite. In particular, Table 5 shows that education and age are slightly statistically significant, being educated reduces the probability of being in favor of FGM by 7.4%, while being one year older increases the probability by 0.4%. Moreover, this fitted model says that, holding all other variables fixed, the probability that members of the other ethnic minority groups are in favor of the sensitive question is 12.3% lower than for Afar people. Being exposed to the NGO program is also significant and negatively affects the outcome variable. In particular, the probability for targeted people of being in favor of FGM is 11.5% lower than the probability for not-targeted people.

However, in the list experiment multivariate analysis (Table 4), we find that being targeted by the NGO program is not significant. The positive effect of the NGO program on reducing the support for the practice seems to disappear when the sensitive information is asked indirectly. This indicates that respondents targeted by the NGO program may have a stronger incentive to reveal a biased answer, i.e. the treatment makes them less willing to share publicly the real attitudes concerning this sensitive issue.

Table 6 shows the differences in estimated proportions of respondents answering the sensitive question if the direct or indirect question is used. In particular, we use the linear model to predict answers to the list experiment, and the linear probability model to predict answers to the direct question. Table 6 includes also the results for the model without covariates and the model with covariates (age, ethnic group, marital status, education and being

in favor of the sensitive issue.

Table 5: Results of the linear probability model applied to responses to the direct question.

Variables	**“A girl should be circumcised”**
Age	0.004†
	(0.002)
Ever married	0.060
	(0.051)
Educated	-0.074†
	(0.044)
Other ethnic minorities	-0.123**
	(0.036)
NGO program target	-0.115**
	(0.034)
Intercept	0.251***
	(0.064)
Observations	836

Note. The dependent variable is a dummy variable whether or not a girl should be circumcised.

Standard errors are in parentheses.

Signif. codes: (†) if $p < .1$, (*) if $p < .05$, (**) if $p < .01$, (***) if $p < .001$.

targeted by the NGO). We also report the estimated proportions for different groups by controlling for all the other covariates. The differences between the direct and indirect questions are always positive (except for the educated people) and slightly statistically significant in the no covariates and covariates models, as well as for the unmarried group, the other ethnic minorities group and the NGO targeted group. The difference is instead highly statistically significant and positive for the uneducated group. Therefore, it seems that the group that lies the most is the group of uneducated people where the direct question produces a 31% of the women in favor of circumcision, compared to 47% obtained through the list experiment. Overall, the social desirability bias is present, even if not statistically significant for some of the groups, and it is as expected, the proportions are almost always higher if the indirect question is asked.

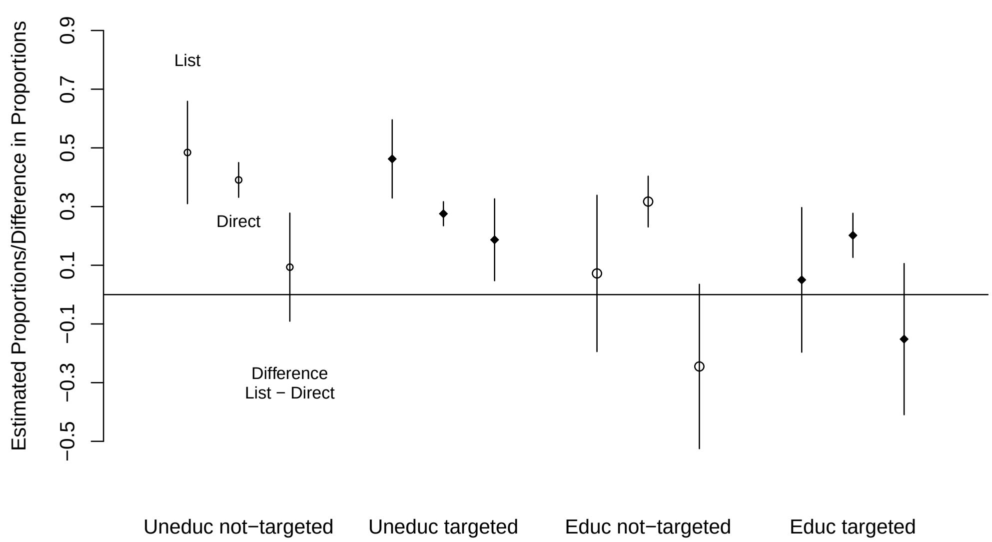
Since the two most relevant variables seem to be the respondent's education and her NGO target status, we analyze the social desirability bias across education levels and NGO target status. Figure 3 shows the difference between responses to the direct and indirect questions in the four subgroups: uneducated non-targeted, uneducated targeted, educated non-targeted and educated targeted. Even after adjusting for age, ethnic group and marital status, it is clear that the social desirability bias is estimated to be the greatest among uneducated people that are targeted by the NGO intervention.

Table 6: Estimated proportion of women answering the sensitive item in the affirmative way by socio-demographic characteristics, and differences between direct and indirect questioning.

	Direct question		List experiment		Differences	
	Est	SE	Est	SE	Est	SE
No covariates	0.300	0.016	0.392	0.047	0.092†	0.049
Covariates	0.298	0.015	0.384	0.047	0.086†	0.049
Uneducated	0.313	0.019	0.470	0.059	0.156**	0.062
Educated	0.240	0.037	0.057	0.120	-0.182	0.126
Never married	0.251	0.042	0.522	0.142	0.271†	0.148
Ever married	0.312	0.020	0.343	0.057	0.031	0.061
Ethnic group Afar	0.325	0.018	0.382	0.053	0.057	0.056
Other ethnic minorities	0.202	0.031	0.391	0.106	0.189†	0.110
NGO not-targeted	0.375	0.028	0.398	0.081	0.023	0.086
NGO targeted	0.260	0.018	0.376	0.058	0.116†	0.061

Note. The sensitive item corresponds to "a girl should be circumcised".
Predictions are based on the linear probability model for the direct question, and on the linear model for the indirect question.
The results are averaged over the sample distribution of covariates.
Signif. codes: (†) if p < .1, (*) if p < .05, (**) if p < .01, (***) if p < .001.

Figure 3: Estimated proportion of women answering the sensitive item in the affirmative way by education and NGO targeting status, and their differences between direct and indirect questioning.



Note. Predictions are based on the linear model for the indirect question, and on the linear probability model for the direct question. The solid lines correspond to the 95% confidence interval for the estimated proportions. The results are averaged over the sample distribution of covariates.

6 Robustness

In this section we carefully test for potential violations of the three key assumptions of the list experiments.

The first assumption is the *randomization of the treatment.* Table 7 provides sample means for the main variables in the treatment group and the control group. Comparing the means allows us to see that the randomization of the list experiment (control group and treatment group) was successful given that the observable characteristics of the respondents do not significantly differ between the two groups.

The second assumption is called design effects and it happens when the inclusion of a sensitive item affects some respondents' answer to control items. The population proportion of each respondent type is defined as $\pi_{yz} = Pr(Y_i(0) = y, Z^*_{i,J+1} = z)$ for $y = 0, \ldots, J$ and $z = 0, 1$. The π_{yz} is identified for all $y = 0, \ldots, J$ as:

$$\pi_{y1} = Pr(Y_i \leq y | T_i = 0) - Pr(Y_i \leq y | T_i = 1),$$

$$\pi_{y0} = Pr(Y_i \leq y | T_i = 1) - Pr(Y_i \leq y - 1 | T_i = 0).$$

There are design effects if at least one of these proportions is negative (Glynn, 2013). Table 8 reports the test to verify the presence of design effects. It reports the estimated proportion of each respondent type. They are all positive, hence, the assumption *no design effects* holds.

The third possible problem is the violation of the assumption *no liars*. As we can see from Table 2, the responses are well distributed and there are few responses in the extreme cases (0 and 3 for the control group, and 0 and 4 for the treatment group). Blair and Imai (2012) have developed a formal test to verify the existence of these effects and to compute the population proportion of liars. There can be two types of liars: liars that give the answer $Y_i = J$ if assigned to the treatment condition even if the truthful answer would be $Y_i = J + 1$, affirmative for both sensitive and control items

Table 7: Tests of randomization for the list experiment.

Respondent's characteristics	**Control mean**	**Treatment mean**	**T test/ chi-squared p-value**
Age	28.225	28.227	0.998
Religion (1=Christian; 0=Muslim)	0.052	0.046	0.675
Ethnic (1=Afar; 0=Other ethnic minorities)	0.778	0.785	0.784
Areas in Afar (proportions)			0.885
Zone 1	0.489	0.511	
Zone 3	0.474	0.526	
Zone 5	0.470	0.530	
Health education/training (1=yes; 0=no)	0.239	0.247	0.773
Health providers available (0-4)	2.472	2.553	0.238
Health services accessible (0-5)	2.654	2.621	0.811
Having children (1=yes; 0=no)	0.732	0.713	0.544
Ever being married (1=yes; 0=no)	0.787	0.754	0.253
Educated (1=yes; 0=no)	0.196	0.229	0.242
Sex and HIV knowledge (0-6)	4.064	4.029	0.693
Agree circumcision (1=yes; 0=no)	0.301	0.294	0.804
NGO targeted (1=yes; 0=no)	0.659	0.675	0.628
N	405	443	

Note. A good randomization of the list experiment is a crucial assumption. The most important characteristics do not vary between the two groups.

Table 8: Design effects. Estimated respondent types for the list experiment.

y **value**	π_{y0}	**se**	π_{y1}	**se**
0	0.68%	0.004	0.82%	0.007
1	8.32%	0.016	13.43%	0.026
2	39.31%	0.031	22.19%	0.029
3	12.51%	0.020	2.74%	0.008
Total	60.82%		39.18%	

Note. The table shows the estimated proportion (and standard error) of respondent types, $\hat{\pi}_{yz}$, characterized by the total number of affirmative answers to the control questions, y, and the truthful answer for the sensitive item.

(ceiling effects); and liars that give the answer $Y_i = 0$ if assigned to the treatment condition even if the truthful answer is affirmative only for the sensitive item (floor effects). Since both types of lies lower the observed mean response of the treatment, the presence of ceiling and/or floor effects lead to the underestimation of the population proportion of those who agree with the sensitive item.

Using Blair and Imai (2012) test, we estimate the population proportions of liars to be 4.35% (s.e. = 0.001) for the ceiling liars, and 0.30% (s.e. = 1e-04) for the floor liars.[8] Since both proportions are small, the assumption of no liars holds.

7 Conclusions

The list experiment is a relatively simple technique that could be added to many surveys when asking about sensitive issues. We believe this is the first attempt to use the list experiment technique to overcome major identification

[8] For details about the test see Blair and Imai (2012). We estimate the proportions of liars using the intercept-only model.

problems related to individual perceptions regarding sexual and reproductive health. Ethiopia is one of the countries with the highest FGM prevalence, and this paper is relevant in shedding light on the real attitudes and beliefs about FGM to understand how to intervene and who to target most.

Our results indicate that when asking a direct question about circumcision 30% of the women are in favor of the practice. If, instead, we take into consideration the question's sensitivity by asking indirectly, we find that the proportion of women in favor of FGM is much higher, 39.2%. In general, a lot needs to be done to change these perceptions and to lower these high percentages. Rahlenbeck et al. (2010) use the Ethiopia Demographic and Health Survey to show that in 2005 in Oromia, 29.7% favored the continuation of FGM. Although, these results are not directly comparable to ours, given that they focus on a different region of Ethiopia, the percentage is consistent with the one that we obtained asking the direct question (30%). In the DHS the question about perceptions toward circumcision is a direct question: "Do you think that female circumcision should be continued, or should it be stopped?". However, we expect that the proportion of people that are actually in favor of continuation is higher. The analysis shows that some respondent's characteristics seem relevant in explaining the consent towards female circumcision. In particular, the women's education turned out to be the most critical variable in explaining differences in attitudes. Moreover, uneducated respondents seem to be most hesitant to share their real attitudes regarding this sensitive issue.

Interestingly, the positive effect of the NGO program on reducing the support for the practice seems to disappear when the sensitive information is asked indirectly. It is well possible that the interventions of the NGO increase the social pressure on this sensitive issue and result in a stronger incentive to reveal a biased answer. In particular, we show that the social desirability bias is large among uneducated women and it is the greatest among uneducated women targeted by the NGO intervention. The intervention focuses on the dissemination of sexual and reproductive health knowledge that contradicts the local FGM customs. Being targeted by the NGO program makes these women less willing to share publicly the real attitudes concerning FGM support. This is an interesting result for studies that try to measure impact of policy interventions. Measuring impact on sensitive attitudes requires a proper survey design.

Overall we conclude that the list experiment technique is a relatively simple approach to overcome major identification problems related to sensitive issues. Multivariate models are helpful to analyze the association between respondent's characteristics and their responses to sensitive items in the list experiment.

References

Bellemare, M.F., Steinmetz, T.L., 2013. All in the family: Explaining the persistence of female genital cutting in the Gambia. Munich Personal RePEc Archive, Paper No. 47628 .

Blair, G., Imai, K., 2012. Statistical analysis of list experiments. Political Analysis 20, 47–77.

Chong, A., Gonzales-Navarro, M., Karlan, D., Valdivia, M., 2013. Effectiveness and spillovers of online sex education: Evidence from a randomized evaluation in Colombian public schools. NBER Working Paper No. 18776 .

Comşa, M., Postelnicu, C., 2012. Measuring social desirability effects on self-reported turnout using the item count technique. International Journal of Public Opinion Research doi:10.1093/ijpor/eds019.

Corstange, D., 2009. Sensitive questions, truthful answers? Modeling the list experiment with LISTIT. Political Analysis 17, 45–63.

Gilens, M., Sniderman, P.M., Kuklinski, J.H., 1998. Affirmative action and the politics of realignment. British Journal of Political Science 28, 159–183.

Glynn, A.N., 2013. What can we learn with statistical truth serum? Design and analysis of the list experiment. Public Opinion Quarterly 77, 159–172.

Holbrook, A.L., Krosnick, J.A., 2010. Social desirability bias in voter turnout reports: Tests using the item count technique. Public Opinion Quarterly 74, 37–67.

Imai, K., 2011. Multivariate regression analysis for the item count technique. Journal of the American Statistical Association 106, 407–416.

Jamison, J., Karlan, D., Raffler, P., 2013. Mixed method evaluation of a passive mHealth sexual information texting service in Uganda. Economic Growth Center Yale University, Discussion Paper No. 1025 .

Janus, A.L., 2010. The influence of social desirability pressures on expressed immigration attitudes. Social Science Quarterly 91, 928–946.

Jones, E.F., Forrest, J.D., 2010. Underreporting of abortion in surveys of U.S. women: 1976-1988. Demography 29, 113–126.

Karlan, D.S., Zinman, J., 2012. List randomization for sensitive behavior: An application for measuring use of loan proceeds. Journal of Development Economics 98, 71–75.

Kuklinski, J.H., Cobb, M.D., Gilens, M., 1997. Racial attitudes and the "New South". The Journal of Politics 59, 323–349.

LaBrie, J.W., Earleywine, M., 2000. Sexual risk behaviors and alcohol: higher base rates revealed using the unmatched-count technique. Journal of Sex Research 37, 321–326.

Mackie, G., 1996. Ending footbinding and infibulation: A convention account. American Sociological Review 61, 999–1007.

Macro International Inc., 2008. Ethiopia Atlas of Key Demographic and Health Indicators, 2005. Technical Report. Calverton, Maryland, USA: Macro International Inc.

McKenzie, D., Siegel, M., 2013. Eliciting illegal migration rates through list randomization. Policy research working paper 6426, World Bank .

Miller, J.D., 1984. A new survey technique for studying deviant behavior. Ph.D. thesis. The George Washington University.

Obermeyer, C.M., 2001. Complexities of a Controversial Practice. Female "Circumcision" in Africa: Culture, Controversy, and Change by Bettina Shell-Duncan; Ylva Hernlund; The Female Circumcision Controversy: An Anthropological Perspective by Ellen Gruenbaum. Science, New Series. Science, New Series 5520, 1305–1306.

Presser, S., Stinson, L., 1998. Data collection mode and social desirability bias in self-reported religious attendance. American Sociological Review 63, 137–145.

Rahlenbeck, S., Mekonnen, W., Melkamu, Y., 2010. Female genital cutting starts to decline among women in Oromia, Ethiopia. Reproductive BioMedicine Online 20, 867–872.

Rai, B., Sengupta, K., 2013. Pre-marital confinement of women: A signaling and matching approach. Journal of Development Economics 105, 48–63.

Redlawsk, D.P., Tolbert, C.J., Franko, W., 2010. Voters, emotions, and race in 2008: Obama as the first black president. Political Research Quarterly 63, 875–889.

Shell-Duncan, B., Wandera, K., Hernlundc, Y., Moreau, Y., 2011. Dynamics of change in the practice of female genital cutting in Senegambia: Testing predictions of social convention theory. Social Science & Medicine 73, 1275–1283.

Wagner, N., 2011. Why female genital cutting persists. Working Paper - International Economics Section. The Graduate Institute of International and Development Studies at Erasmus University of Rotterdam .

WHO, 2012. Female Genital Mutilation, Fact Sheet 241. Technical Report 241. World Health Organization.

Yoder, P.S., Khan, S., 2008. Numbers of women circumcised in Africa: The production of a total. DHS Working Paper 39.

List of research reports

12001-HRM&OB: Veltrop, D.B., C.L.M. Hermes, T.J.B.M. Postma and J. de Haan, A Tale of Two Factions: Exploring the Relationship between Factional Faultlines and Conflict Management in Pension Fund Boards

12002-EEF: Angelini, V. and J.O. Mierau, Social and Economic Aspects of Childhood Health: Evidence from Western-Europe

12003-Other: Valkenhoef, G.H.M. van, T. Tervonen, E.O. de Brock and H. Hillege, Clinical trials information in drug development and regulation: existing systems and standards

12004-EEF: Toolsema, L.A. and M.A. Allers, Welfare financing: Grant allocation and efficiency

12005-EEF: Boonman, T.M., J.P.A.M. Jacobs and G.H. Kuper, The Global Financial Crisis and currency crises in Latin America

12006-EEF: Kuper, G.H. and E. Sterken, Participation and Performance at the London 2012 Olympics

12007-Other: Zhao, J., G.H.M. van Valkenhoef, E.O. de Brock and H. Hillege, ADDIS: an automated way to do network meta-analysis

12008-GEM: Hoorn, A.A.J. van, Individualism and the cultural roots of management practices

12009-EEF: Dungey, M., J.P.A.M. Jacobs, J. Tian and S. van Norden, On trend-cycle decomposition and data revision

12010-EEF: Jong-A-Pin, R., J-E. Sturm and J. de Haan, Using real-time data to test for political budget cycles

12011-EEF: Samarina, A., Monetary targeting and financial system characteristics: An empirical analysis

12012-EEF: Alessie, R., V. Angelini and P. van Santen, Pension wealth and household savings in Europe: Evidence from SHARELIFE

13001-EEF: Kuper, G.H. and M. Mulder, Cross-border infrastructure constraints, regulatory measures and economic integration of the Dutch – German gas market

13002-EEF: Klein Goldewijk, G.M. and J.P.A.M. Jacobs, The relation between stature and long bone length in the Roman Empire

13003-EEF: Mulder, M. and L. Schoonbeek, Decomposing changes in competition in the Dutch electricity market through the Residual Supply Index

13004-EEF: Kuper, G.H. and M. Mulder, Cross-border constraints, institutional changes and integration of the Dutch – German gas market

13005-EEF: Wiese, R., Do political or economic factors drive healthcare financing privatisations? Empirical evidence from OECD countries

13006-EEF: Elhorst, J.P., P. Heijnen, A. Samarina and J.P.A.M. Jacobs, State transfers at different moments in time: A spatial probit approach

13007-EEF: Mierau, J.O., The activity and lethality of militant groups: Ideology, capacity, and environment

13008-EEF: Dijkstra, P.T., M.A. Haan and M. Mulder, The effect of industry structure and yardstick design on strategic behavior with yardstick competition: an experimental study

13009-GEM: Hoorn, A.A.J. van, Values of financial services professionals and the global financial crisis as a crisis of ethics

13010-EEF: Boonman, T.M., Sovereign defaults, business cycles and economic growth in Latin America, 1870-2012

13011-EEF: He, X., J.P.A.M Jacobs, G.H. Kuper and J.E. Ligthart, On the impact of the global financial crisis on the euro area

13012-GEM: Hoorn, A.A.J. van, Generational shifts in managerial values and the coming of a global business culture

13013-EEF: Samarina, A. and J.E. Sturm, Factors leading to inflation targeting – The impact of adoption

13014-EEF: Allers, M.A. and E. Merkus, Soft budget constraint but no moral hazard? The Dutch local government bailout puzzle

13015-GEM: Hoorn, A.A.J. van, Trust and management: Explaining cross-national differences in work autonomy

13016-EEF: Boonman, T.M., J.P.A.M. Jacobs and G.H. Kuper, Sovereign debt crises in Latin America: A market pressure approach

13017-GEM: Oosterhaven, J., M.C. Bouwmeester and M. Nozaki, The impact of production and infrastructure shocks: A non-linear input-output programming approach, tested on an hypothetical economy

13018-EEF: Cavapozzi, D., W. Han and R. Miniaci, Alternative weighting structures for multidimensional poverty assessment

14001-OPERA: Germs, R. and N.D. van Foreest, Optimal control of production-inventory systems with constant and compound poisson demand

14002-EEF: Bao, T. and J. Duffy, Adaptive vs. eductive learning: Theory and evidence

14003-OPERA: Syntetos, A.A. and R.H. Teunter, On the calculation of safety stocks

14004-EEF: Bouwmeester, M.C., J. Oosterhaven and J.M. Rueda-Cantuche, Measuring the EU value added embodied in EU foreign exports by consolidating 27 national supply and use tables for 2000-2007

14005-OPERA: Prak, D.R.J., R.H. Teunter and J. Riezebos, Periodic review and continuous ordering

14006-EEF: Reijnders, L.S.M., The college gender gap reversal: Insights from a life-cycle perspective

14007-EEF: Reijnders, L.S.M., Child care subsidies with endogenous education and fertility

14008-EEF: Otter, P.W., J.P.A.M. Jacobs and A.H.J. den Reijer, A criterion for the number of factors in a data-rich environment

14009-EEF: Mierau, J.O. and E. Suari Andreu, Fiscal rules and government size in the European Union

14010-EEF: Dijkstra, P.T., M.A. Haan and M. Mulder, Industry structure and collusion with uniform yardstick competition: theory and experiments

14011-EEF: Huizingh, E. and M. Mulder, Effectiveness of regulatory interventions on firm behavior: a randomized field experiment with e-commerce firms

14012-GEM: Bressand, A., Proving the old spell wrong: New African hydrocarbon producers and the 'resource curse'

14013-EEF: Dijkstra P.T., Price leadership and unequal market sharing: Collusion in experimental markets

14014-EEF: Angelini, V., M. Bertoni, and L. Corazzini, Unpacking the determinants of life satisfaction: A survey experiment

14015-EEF: Heijdra, B.J., J.O. Mierau, and T. Trimborn, Stimulating annuity markets

14016-GEM: Bezemer, D., M. Grydaki, and L. Zhang, Is Financial Development Bad for Growth?

14017-EEF: De Cao, E., and C. Lutz, Sensitive survey questions: Measuring attitudes regarding female circumcision through a list experiment www.rug.nl/feb