How Accurate is Our Misinformation?

A Randomized Trial to Assess the Cost-Effectiveness of Administering Alternative Survey Modes to Youth at Risk

DOMINICAN REPUBLIC CASE STUDY

SOCIAL PROTECTION AND HEALTH DIVISION

PAULA LÓPEZ-PEÑA
SANDRA MCCOY
RODRIGO MUÑOZ
SIGRID VIVO

IDB
Inter-American Development Bank

with the support of
THE REPUBLIC OF KOREA
FINLAND

AUGUST, 2011
Publication code: IDB-MG-117

Inter-American Development Bank

How Accurate is Our Misinformation?

A Randomized Trial to Assess the Cost-Effectiveness of Administering Alternative Survey Modes to Youth at Risk

Dominican Republic Case Study

SCL/SPH

Paula López-Peña

Sandra McCoy

Rodrigo Muñoz

Sigrid Vivo

IDB-MG-117

Preface

This research and publication came about in response to the need for a growing body of knowledge to support the healthy and integral development of young people in the region, so that interventions targeting this segment of the population can be more effective. The Social Sector of the Inter-American Development Bank (IDB), through its Division of Social Protection and Health, has been leading this project with the support of the Finnish Technical Assistance Fund.

In order to learn more about tools to measure high risk behaviors and confirm their accuracy, it was agreed to work with the Youth Employment Program (PJE) being implemented by the Ministry of Labor of the Dominican Republic. This research will complement the knowledge gleaned by that government through a rigorous evaluation of the impact of its program. The IDB and the World Bank are providing technical leadership for the impact assessment.

It is important to note that the PJE won the 2009 award for Best Practices in Youth Policies and Programs in Latin America and the Caribbean, conferred by the IDB and UNESCO and supported by the IDB-managed Korean Poverty Reduction Fund.

This experience is an example of the kind of partnership espoused in the Paris Declaration on Development Effectiveness, according to which the principles of ownership, alignment, harmonization, results, and mutual accountability should lead government ministries, multilateral development banks, and bilateral donors to work in tandem to deliver development aid as effectively as possible.

Acknowledgements

The project team wishes to thank the Ministry of Labor of the Dominican Republic, particularly the Office of the Minister of Labor and the Program Coordination Unit, for promoting and generating rigorous data collection in the region. Special recognition goes to Minister of Labor Francisco Dominguez Brito and the Director of the Coordination Unit, José Luis Polanco de León, for their support throughout the process. We also thank Brígida García Romero and Douglas Hasbún for their guidance and technical input.

Thanks also goes to the World Bank, particularly Cornelia Tesliuc, Paloma Azevedo, and Carlos Asenjo, for their support and contributions to the study. We thank Paul Gertler, Sebastián Martínez, Caridad Araujo, and Alexandra Minnis for their guidance, assistance, and technical input.

We express our deep appreciation to the Finnish Technical Assistance Fund and the Korean Poverty Reduction Fund—both managed by the IDB—for making this project possible and for their continuous support in generating knowledge about youth, so that the Paris Declaration can bear fruit.

We also thank the IDB Office of External Relations and the Development Communication Unit, as well as the IDB Youth Program. We thank Elena Suarez for her leadership and shared effort, and particularly Isabel Álvarez-Rodriguez for her support, guidance, and input throughout the process. We thank Dolores Subiza for graphic design and production support.

Recognition is accorded to Suzanne Duryea for the idea of generating knowledge about youth; to Rosario Londoño for spreading youth development, and to Jorge Torres for his unconditional support in the field. We thank Pablo Celhay for his technical input, Drina Saric and Juliana Arbelaez for their valuable contributions and support throughout this publication process, and Laura Ripani for her commitment and constant support. This study was led by Fernando Regalía, Chief of the Division of Social Protection and Health.

Finally, we wish to thank the young participants in the Youth and Jobs Program in the Dominican Republic for allowing us to learn from them.

September 2011

Table of Contents

Acronyms and Abbreviations

SAI	Self-Administered (Pen-and-Paper) Interview
ACASI	Computer-Assisted Audio Self Interview
FTFI	Face-to-Face Interview
CAFE	Computer-Assisted Field Edits
CAPI	Computer-Assisted Personal Interview
CATI	Computer-Assisted Telephone Interview
CE	Cost Error
STD	Sexually Transmitted Disease
RCI	Response Consistency Index
PJE	Programa Juventud y Empleo (*Youth and Employment Program*)
TTM	Technical Training Module
LSM	Life Skills Module

Summary

The study reports on a randomized trial of 1,200 young adults enrolled in an employment training program executed by the Ministry of Labor in the Dominican Republic "Youth and Employment Program" (PJE, for its Spanish acronym), to determine the most cost-effective and appropriate interview mode for measuring youth risk behaviors. Four different survey administration modes –two interviewer-assisted (FTFI and CATI) and two self-administered modes (SAI and ACASI)–were randomly assigned to young adults between the ages of 18 and 30. The findings contribute to knowledge of the Latin American and the Caribbean region, where similar experiments are scarce. The authors have centered the study on the question of cost-effectiveness, which integrates both actual implementation costs and estimates of measurement bias into the decision about an appropriate choice of interview mode for a given research study. The research also includes randomization of interviewer gender in order to assess the interaction between gender and data quality. The research shows that the target population is likely to underreport sensitive questions in self-administered surveys, and thus the degree to which a mode improves self-reporting of a particular risk behavior or set of behaviors is likely to be context specific. More research is needed in the region to support these findings. To validate results, it is suggested that biomarkers be integrated into the study.

JEL Classification: C93, I15, J13 y O54

1 Introduction

Accurately measuring the effects of public policies on the healthy development of youth is of utmost importance. However, the accurate measurement of youth behavior is challenging and likely a function of the mode of survey administration and perceived privacy of the responses. There is evidence that measurement methodology can strongly influence respondents' answers and that applying one administration mode or another can bias the results of a survey. Although progress has been made in understanding the effect of the survey mode on the responses given by youth, the mechanism underlying response variation remains unknown (Eaton et al. 2010). Moreover, several of those studies have had inconsistent results over different geographic areas.

Epidemiologists typically consider three main sources of bias in measuring self-reported risk behaviors: social desirability, recall, and comprehension. Another relevant bias, in both self-administered modes (ACASI and SAI), is introduced through the interview design itself (e.g. errors in measurement). The primary focus of comparison in this study is to present evidence for a reduction in social desirability bias and errors related to actual completion of the survey (e.g., missing items, incorrect skips).

Using surveys to measure youth risk indicators, such as drug use or sexual behavior, is particularly difficult, since surveys may contain questions considered sensitive or taboo. For example, many youth may be fearful of disciplinary actions if they admit to having used drugs, or married respondents may be afraid that their spouses learn of other sexual partners. These types of questions are vulnerable to socially desirable responses, which is what occurs when respondents give false or vague answers because they fear that their social acceptance may be compromised by their responses.

Likewise, survey respondents exhibit a social desirability bias when they over-report socially approved behaviors (like voting) and underreport socially disapproved behaviors (like using illicit drugs). Because social measurement research usually assumes that higher-prevalence estimates are more valid than lower estimates, methodological factors shown to increase prevalence estimates, such as setting and mode, should be considered when planning surveys (Brener et al., 2006). However, is important to highlight that what is regarded as sensitive or normative in one population or region may not be elsewhere. Likewise, the broad consensus is that a bias might result due to the socially desirability associated with a

certain response (Gregson et al. 2002; Pienaar 2009); moreover, such bias may be different for men and women (Mensch et al. 2003).

One way to reduce the number of socially desirable responses is to increase the **confidentiality** and **privacy** of the mode (Sedlak 2010, Lothen-Kline et al., 2003). To that end, interview settings specially designed to offer privacy and to assure the respondent that no relative, close acquaintance or even the interviewer will know the participants' responses are particularly attractive. For example, previous studies have repeatedly shown that risk behavior surveys administered at school produce prevalence rates of risk indicators higher than those administered at the household level (Eaton et al. 2010). The former setting allows respondents to participate anonymously, and there is no risk that parents may see the responses.

In addition to increasing the privacy of the interview setting (e.g., school vs. home), one recommendation mentioned in several studies is to increase the privacy of the survey mode itself. For example, the use of self-administered questionnaires to be filled out by the youths themselves, without the assistance of an interviewer, offers a higher degree of privacy than when an interviewer is present[1] (Tourangeau and Smith, 1998; Langhaug, 2010; Brenner et al., 2003, or the impact of CATI or T-ACASI interviewing, as in Gribble et al., 2000). On the other hand, the greatest weakness of self-administered questionnaires is that, without the interviewer's assistance, the quality of responses and subsequent measurement error are highly dependent on the difficulty level of the questionnaire the respondent's cognitive level and motivation.

Most studies on the effect of the survey method focus on comparing some of the following modes: ACASI, Face-to-Face (FTFI), paper Self-Administered Interview (SAI), CATI, and Informal confidential voting interview (Africa, low-cost alternative to ACASI). Moreover, most interview mode comparison studies focus solely on the accuracy of self reports, often by comparing the prevalence of reported risk behaviors between two interview modes.

Along these lines, Brener et al. 2006 examined the effect of various survey modes over 55 risk behaviors. He found that, only seven risk behaviors showed significant differences after controlling for the setting and student characteristics. For those seven indicators (injury, alcohol, drug use, physical activity, tobacco use, sexual behavior and weight control), the young adults assigned to ACASI tended to report more risk behaviors than those assigned to the self-administered questionnaire. These results are

[1] This offers greater privacy than the mode where the interviewer reads the questions aloud, and this is precisely the strongest argument raised in favor of using self-administered questionnaires to measure sensitive indicators.

consistent with studies that demonstrate that mode effects are stronger for more sensitive behaviors (Turner et al. 1998; Wright, Aquilino, and Supple 1998). Those findings were supported by Vereecken and Maes (2006), Beebe et al. (1998) and Hallfors et al. (2000) also compared these two methods and validated the aforementioned findings. These studies on setting and mode demonstrate that, when holding the mode of administration constant, prevalence of risk behaviors is equal or higher when questionnaires are administered in schools compared with when they are administered in students' homes. The effect of mode, however, appears to vary by setting. To date, no study has systematically varied both setting and mode of administration to understand the effects of each.

ACASI has been extensively used to obtain "sensitive" information, although findings have not been consistent across studies. Some studies show optimal results (Langhaug et al. 2010; Ghanem et al. 2008; Rogers et al., 2005; Ghanem et al., 2005; Rathod S. et al., 2011[2]). . Other studies have yielded inconsistent results regarding their effectiveness (i.e., Mensch et al., 2003; Jaya et al., 2008; Johnson et al., 2001, Jennings et al., 2002) and even negative ones (Testa et al. 2005; Hallfors et al., 2000; Mensch, 2008). Thus, the degree to which a mode improves self-reports of a particular behavior or set of risk behaviors is likely to be context specific.

Likewise, the limited evidence from developing countries also suggests mixed results with respect to the use of ACASI. A study that employed ACASI to elicit sensitive information about sexual behavior from male and female adolescents in Kenya found that among some subgroups, boys reported a higher incidence of behaviors such as forcing a partner to have sex or having ever had a sexually transmitted infection when ACASI was used than when the face-to-face interviewing was used (Mensch et al. 2003; Hewett et al., 2004a and 2004b). Among some subgroups in the same study, however, respondents' fear of computers appeared largely to negate the advantages of privacy and confidentiality associated with ACASI (Hewett et al., 2004a). A study of Zimbabwean women found that the efficacy of ACASI varied significantly by respondents' educational level: those having a middle-school or higher education performed with greater ease on the computer than did less-educated women (van de Wijgert et al., 2000). Two other studies, conducted in Mexico and Zimbabwe concluded that other, less expensive and less technologically sophisticated methods yielded higher levels of reporting of sexual and reproductive health behaviors than did ACASI (Lara et al., 2001; Gregson et al., 2002). In contrast, a study of college students in Thailand found that Audio-CASI improved the reporting

[2] For some risks, perceived therapeutic benefit in reporting an outcome (e.g., experience of interpartner violence) may yield higher reports in FTFI interviews relative to ACASI, where reporting to an individual who may be able to facilitate a linkage to care is perceived by respondents as having value (see Rathod S. et al., 2011).

of sexual behaviors, particularly among female students (Rumakom et al., 1999). The effectiveness of ACASI in eliciting information about such behaviors in India and in South Asia more generally remains unknown, and more evidence is needed in Latin America and the Caribbean, where survey method effect has been only minimally explored. There is an ACASI feasibility study covering multiple countries from several regions, including Peru (NIMH, 2007), a comparative study of four interview modes conducted in Mexico measuring abortion prevalence (Lara, 2004) and another in Brazil comparing FTFI and ACASI that integrated STI results into the analysis of mode effects (Hewett, 2008).

Minnis et al. (2009) and Mensch and Abott et al. (2011) conclude that questionnaire delivery modes do affect self-reported sexual behaviors and that the use of ACASI can significantly reduce reporting bias, They also suggest that triangulation of self-reported data using biomarkers is recommended. In Zimbabwe (Minnis et al., AJE, 2009) found that, when comparing self-reported behavioral data, ACASI improved self-reports over FTFI, but that when the biomarker data was integrated, the underlying bias was so great that the mode effects were modest. However, study results in South Africa, Mensch and Abott et al. (2011) lean more in favor of ACASI.

Finally, it is important to consider that because populations are purposefully selected for presumed “high-risk" behavior, interviewer attitudes may impact the socially desirable responses of the interviewee, especially since socially stigmatized behaviors such as male-to-male sex or anal sex are especial sensitive in specific contexts (van der Elst et al., 2009) or because the interviewer has explicit knowledge of the high-risk conditions of the interviewee (see studies related to HIV-infected prisoners, Bautista-Arredondo et all 2011).

In general, no studies incorporate the cost element (with the exception of Aitken et al., 2004 and Brown et al., 2008), even though cost is an underlying variable in deciding on the type of administration mode.

2 Objective, context and limitations of the research

The purpose of this study is to assess the cost-effectiveness of various survey administration modes and determine which one is the most accurate in measuring risk behaviors among youth. Four different administration modes–two interviewer-assisted and two self-administered modes–were randomly assigned to measure risk behaviors among young adults between the ages of 18 and 30. The experiment was conducted in the Dominican Republic in November and December 2010 and covered a sample of 1,200 young adults enrolled in the PJE (executed by the Ministry of Labor).

PJE is an employment training intervention for youth between the ages of 16 and 29 who have not completed secondary school, have dropped out of school, or who are not employed. The objective of the program is to increase employment opportunities for lower-income youth by improving the supply of training and the demand for labor force from the private sector.

There were two specific questions examined in the experiment. The first one was to compare the **quality** of data obtained through each administration mode. Potential *socially desirable response bias*[3] was detected by comparing the risk indicators reported for each administration mode, while other quality problems–such as those arising out of cognitive or motivational difficulties–were identified by creating a *Response Consistency Index (RCI*) and analyzing the reasons behind non-responses. The second specific question was to determine the **effect of interviewer sex** on data quality. Some studies related to the interviewer gender, interviewer choice, and self-reported sexual behavior are Chun et al. (2011) and Catania et al. (1996).

We hope the results of this study will help designers of youth-at-risk surveys to choose the administration mode that best suits their needs. Even though the experiment provides a multitude of useful lessons, two factors put a limitation to its external validity. First, the experiment was conducted on a specific population of young adults in a concurrent employment training program; they differ from the general youth population both in the Dominican Republic and abroad. These participants were sampled from an ongoing impact evaluation and were accustomed to working with researchers and filling out surveys. Thus, they may be a highly motivated sample that is different from a general-population sample. Second, the cost and effectiveness of any given administration mode are not constant (administration mode effectiveness should be understood as being equivalent to its accuracy, or to the inverse of the measurement errors it generates). Even if instruments and samples remain constant, the effectiveness of an administration mode can vary significantly depending on the budget involved and the efforts put into the design, preparation and execution of the survey.

There are two major assumptions in the study that need to be taken into consideration. First, the experiment considers that the mode that detects the highest prevalence is the most accurate. The underlying hypothesis is that

[3] "Social desirability bias" refers to the tendency to present oneself in a favorable light. Survey respondents exhibit this bias when they over-report socially approved behaviors (like voting) and underreport socially disapproved behaviors (like using illicit drugs)". Methods of Data Collection. (2009). In R. Groves, F. Fowler, M. Couper et al. (Eds.), *Survey Methodology - 2nd Ed.* Hoboken, New Jersey: Wiley & Sons, Inc.

self-reported interviews allow more honest answers. This is a reasonable assumption to make, but it is true that in some populations, youth may exaggerate or underreport risk behaviors depending on the prevailing social norms.

A second assumption is that there are no pre-designed conditions in the research design to verify the validity of what the participants report, especially regarding sexual behavior. Thus, without any kind of external verification (e.g., medical record review for the case of STDs), we assume that self-report is accurate.

In other words, it is not possible to provide an answer to the questions "What is the cost of an administration mode?" or "What is its effectiveness?" because designers may determine the survey costs from a wide range or margin, and effectiveness is the direct result of their decisions. A more appropriate question would be "How does the effectiveness of an administration mode vary in relation to the money invested?" or "For any given budget, what is the most effective administration mode?" These two factors –context and investment– are determinants of the effectiveness of an administration mode. Therefore, the results of one particular experiment should not be extrapolated to any other situation. Still, extra care was taken to maintain control over as many parameters as possible, trying to keep the same level of quality in the design, preparation and execution of each administration mode.

3 Methodology

3.1 Administration Modes

The four administration modes used in this experiment—Self-Administered Interview (SAI), Face-to-Face Interview (FTFI), Computer-Assisted Telephone Interview (CATI) and Audio Computer-Assisted Self Interview (ACASI)—are actually subtypes of four larger modes of survey administration, each of which allows for several different ways to implement a survey. The following paragraphs review the types of implementation used in this particular experiment and describe their main characteristics. Chapter X [of the Technical Note] provides further details about these administration modes.

3.1.1. Face-to-Face Interview (FTFI)

The FTF administration mode relies on a team of enumerators to conduct face-to-face interviews and record respondents' answers on paper. Implementation can vary substantially depending on the location chosen for the interview (home, school, etc.), as location is a primary determinant of cost and effectiveness in the FTF mode. In this experiment, interviews were

conducted at the respondent's home and were entered in a central office, making it harder to resolve errors spotted during data entry.

3.1.2. Computer-Assisted Telephone Interview (CATI)

The CATI mode relies on a team of enumerators who use a computer during the telephone interview. The CATI mode may be centralized, with all interviewers working in the same location, or decentralized, with each enumerator working from his/her own home. For this experiment, the centralized mode was selected in order to better monitor the quality of the enumerator team's work. The interview is carried out in "real time," making use of a software platform that controls and corrects for errors, making the correct skip patterns. Contact numbers were provided by the program and refer to cell phone and landline phone numbers.

3.1.3. Self-administered Interview (SAI)

The SAI mode utilizes a printed form for respondents to complete by themselves. SA has been typically implemented in different ways: (i) by delivering forms in person or by mail to a service-provision point (e.g., schools, health clinics or communities),and (ii) by delivering forms in person or by mail to the home.

For the purposes of this experiment, enumerators used the home-based delivery mode because respondents had already finished their participation in the PJE and could not be gathered in one single place, and because the mail delivery mode usually has low response rates. Also, comparing the SAI and FTFI modes is more interesting when both are administered in the home, since any differences between them can only be attributed to the interviewer being present or not.

3.1.4. Audio Computer-Assisted Self Interview (ACASI)

The ACASI is another type of self-completion interview in which a computer presents the questions in text and/or audio format for the respondent to enter their responses directly. Like the other modes, ACASI may be applied in different ways, thus affecting its cost and effectiveness. For the purposes of this experiment, ACASI interviews were conducted at the respondents' homes using portable computers with an electronic version of the risk behavior questionnaire. The software presented one question at a time (as text and audio), adapting interview skips to the respondent's responses. Each question was presented with an accompanying list of response alternatives or a box to enter numbers or text. Respondents were given the chance to choose between a male or a female voice, or to work in silent mode. Data were encrypted with a password defined by the respondent so that nobody could access their responses except for survey analysts. No

inconsistency checks were used, except to pinpoint blank or out-of-range responses.

3.2 Considerations to choose an administration mode

The four administration modes reviewed can be classified in the following categories: (i) presence (or absence) of an interviewer, (ii) presence (or absence) of computer-assisted checks, (iii) location of the interview[4], and (iv) mode of the interview (see Table 1). These are all essential characteristics that determine the cost-effectiveness and advantages and disadvantages of each mode.

Location (where data are entered) is a factor that heavily influences the cost-effectiveness of the mode. Centralized data entry rarely offers the possibility of using a computer to prevent typographical, sequencing and consistency errors. In contrast, field entry allows using a computer to identify errors and correct them during a visit to the home.[5]

Another dimension to consider is location defined as where the interview is performed with the presence of the interviewer. One of the most recognized advantages of CATI is its substantially lower cost compared to FTFI, since there is no need for enumerator transportation (Tourangeau and Smith, 1988). This method makes it possible to cover larger samples at a fraction of the cost. However an important disadvantage to consider is the fact that fewer and fewer homes now have landlines, and thus one must rely on participants giving out their cell phone numbers in advance. Another consideration to take into account when using CATI is to find a motivation mechanism that limits the interviewee's propensity to not respond.

Presence of computer-assisted checks is also a key factor for quality control, cost and effectiveness. For example, one of the main advantages of CATI is that, unlike the SAI method, the computer used to enter responses helps the enumerator prevent transcription, skipping and consistency errors. The main advantage of the SAI mode over assisted modes (FTFI and CATI) is its capacity to provide greater interviewer privacy. Its main drawback is that respondents may find it difficult to fill out forms that are relatively complex. This can lead to a large number of errors that may be correlated to the respondents' level of understanding. Furthermore, assisted modes reduce

[4] Although all of the interviews can be done in the home, the difference is that the interviewer is not present in the home with the telephone or CATI surveys.

[5] The CAFE method (Computer-Assisted Field Edits) should not be confused with CAPI (Computer-Assisted Personal Interview). The CAFE mode utilizes paper questionnaires that are entered in portable computers before the team leaves the field (thus permitting any errors spotted to be corrected by paying a second visit to the respondent's home). The CAPI mode has replaced the paper questionnaire with a portable computer, allowing the enumerator to conduct the interview and enter the data at the same time.

the number of errors in complex questionnaires, avoid missing data, and smooth out differences between individuals with disparate comprehension levels.

Each individual was administered a questionnaire about tobacco, alcohol and drug use; violence and crime; sexuality; reproductive health and family planning, and education. An additional 5-question test was applied to determine the respondent's cognitive level, together with an evaluation for the interviewer to assess conditions during the interview.

Finally, and aligned to the above arguments, the ACASI mode offers important comparative advantages, such as: (i) it is considered "private" and considered by many to be the best way to capture very sensitive information, since there is no interviewer in the room **(confidentiality);** (ii) questions are asked in the same way across interviews **(standardization);** (iii): ACASI can be used with literate or illiterate populations and in different languages and dialects, or color-coding or pictures can be adopted to indicate responses which work for very short questionnaires **(language flexibility)**; (iv) data from interviews are automatically stored in a database **(security); and (v)** the computer can recognize skipped questions and detect questionnaire errors automatically **(quality controls),** so many of the quality factors traditionally attributed to enumerators and data entry operators are now the responsibility of the ACASI designer.

Table 1. ***Essential Characteristics of the Different Administration Modes***

	Administration Mode			
Characteristics	FTFI	CATI	SAI	ACASI
Assisted Interview	Yes	Yes	No	No
Automatic Checks by a Computer	No	Yes	No	Yes
Interview Location for the Experiment	Home	Telephone	Home	Home
Interview Mode	Verbal Face to face	Verbal Telephone	Paper Text	Computer Audio+Text

Source: Authors.

3.3 Instruments

The questionnaire contained questions related to tobacco, alcohol and drug use, violence and crime, sexuality, sexual identity, reproductive health and family planning, and education. Table 2 shows the risk indicators chosen for the experiment together with their definitions, reference population and estimated direction of bias (disaggregated by gender). These indicators

were chosen to represent a wide variety of risk indicators and reporting sensitivities; estimated direction of bias was hypothesized by previous analysis of the PJE population and qualitative evaluations carried out in a context similar to that of this study.

Table 2. Risk Behavior Indicators Considered Plus the Hypothesized Direction of Bias

#	Indicator	Definition	Reference Population	Estimated rection of Bias
1	Ever regular smoker	Have you ever smoked at least one cigarette a day for more than 30 days?	All	+
2	Current smoker	Have you smoked at least one cigarette in the last 30 days?	All	+
3	Binge drinking, last 30 days	Have you had more than X consecutive drinks or glasses of wine in the last 30 days?[6]	All	+
4	Drug use	Have you ever used any drugs, including recreational illicit drugs or prescription drugs without a prescription?[7]	All	-
5	Marijuana use	Have you ever consumed marijuana	All	+ (men) - (women)
6	Gang affiliation	Have you ever belonged to or been affiliated with a street gang?	All	-
7	Fights	Have you been involved in a fight with other young adults in the last 12 months?	All	-
8	Intimate partner violence	Has your partner beat you or physically hurt you in the last 12 months?	All	-
9	Sexual identity	Heterosexual, homosexual, bisexual, other	All	- (for non heterosexual)
10	STD diagnosis, ever	Have you ever been diagnosed with a sexually transmitted disease?[8]	All	-
11	Pregnancy intentions	Plans to get pregnant in the next 6 months	Non-pregnant women	-
12	Ever had sex	Have you ever had sexual intercourse, vaginal or anal?	All	+ (men) - (women)
13	Age of sexual debut	Age at first sex	Ever had sex	-
14	Lifetime sexual partners	How many sexual partners have you had in your entire life?	Ever had sex	+ (men) - (women)
15	No. of sexual	Number of sexual partners in the last 12 months	Ever had sex	+ (men)

[6] For men, X=5. For women, X=4.

[7] Drugs in the questionnaire: Marijuana or hashish; cocaine (powder, crack, paste or injection); glue or rubber cement sniffing, aerosol or spray inhalation; heroin; methamphetamines; ecstasy; hallucinogens (LSD, acids, PCP, mushrooms, etc.); steroid pills or injections without a medical prescription; medicines (such as painkillers, stimulants, antidepressants, etc.) used as narcotics.

[8] Diseases in the questionnaire: genital herpes; gonorrhea; chlamydia, syphilis; trichomoniasis; hepatitis B; lice infestation; condyloma (warts, HPV); pelvic inflammatory disease; HIV or AIDS.

Table 2. Risk Behavior Indicators Considered Plus the Hypothesized Direction of Bias

#	Indicator	Definition	Reference Population	Estimated rection of Bias
	partners in the last 12 months			- (women)
16	Concurrent sex	Six months prior to the survey, were you having sex with two or more partners at the same time?[9]	Ever had sex	+ (men) - (women)
17	Same gender sex	Were any of your last three sexual partners of your same sex, or have you reported having had sexual relations with someone of your same sex ever in your life?	Ever had sex	-
18	Safe casual sex	Did you wear a male or female condom during your last sexual relation with a casual sexual partner(s)?[10]	Youth with casual partner(s)	+
19	Commercial sex, ever	Have you ever paid for sex?	Ever had sex	-
20	Transactional sex	Have you ever been paid or received gifts or drugs for having sex?	Ever had sex	-
21	Sex under the influence	Used alcohol or drugs during the last sexual encounter	Ever had sex	-

Source: Authors.

The existence of differences between males and females in their responses to surveys of sexual behavior has been documented in the literature[11]. Some studies have shown men reporting two to four times as many opposite-sex partners as women (Johnson et al., 1992; Brown, 1999; Smith, 1992). Nevertheless, gender differences tend to be smaller or even inexistent when men and women are asked about the frequency and duration of sexual activity, engagement in oral and anal sex, and the number of sexual partners they had in the past year (Brown 1999; Johnson et al., 1992; Laumann et al., 1994; Morris, 1993; Smith, 1992). These results suggest differences in the recall process between men and women as a major explanatory cause for the discrepancy in their responses. However, most of these studies were conducted among adult populations. We hypothesize that non-recall

[9] UNAIDS Reference Group on Estimates, Modeling, and Projections: Working Group on Measuring Concurrent Sexual Partnerships, 2010.

[10] The survey requests information on the last three sex partners only.

[11] Brown, N.R., and Sinclair, N.C. (1999): Estimating Number of Lifetime Sexual Partners: Men and Women Do It Differently, The Journal of Sex Research, Vol. 36, No. 3;
Johnson, A. M., Wadsworth, J., Wellings, K., Bradshaw, S., & Field, J. (1992). Sexual lifestyles and HIV risk. Nature, 360, 410-412.
Smith, T. W. (1992). Discrepancies between men and women in reporting number of sexual partners: A summary from four countries. Social Biology, 39, 203-211.

processes account for a larger part of the discrepancy between young men and women when asked about sexual behavior and partners.

All questionnaires administered were by design identical in content, paper questionnaires for the FTFI and SAI modes were almost identical, except for some instructions specifically designed for the enumerator or the respondent (see Figure 1). In the CATI mode, the software showed a data entry screen identical to the paper questionnaires. In the ACASI mode, the software showed the questions one at a time: that is, there was never more than one question per screen. As illustrated in Figure 2, multiple choice questions were separated into a series of Yes/No questions.

Figure 1. Instruments used in the different administration modes

FTFI

1 **a. Have you ever smoked a tobacco cigarette, even if you did not finish it and had just one puff?**

☐ Yes

☐ No → *Skip to question 2 on the next page*

b. How old were you the first time you smoked a cigarette?

Years of Age

☐

c. Have you ever smoked cigarettes on a regular basis, that is, at least one cigarette a day for more than 30 days?

☐ Yes

☐ No

d. In the last 30 days, have you smoked cigarettes at least once?

SAI

1 **a. Have you ever smoked a tobacco cigarette, even if you did not finish it and had just one puff?**

☐ Yes

☐ No → *Skip to question 2 on the next page*

b. How old were you the first time you smoked a cigarette?

Years of Age

☐

c. Have you ever smoked cigarettes on a regular basis, that is, at least one cigarette a day for more than 30 days?

☐ Yes

☐ No

d. In the last 30 days, have you smoked cigarettes at least once?

CATI

1 **a. Have you ever smoked a tobacco cigarette, even if you did not finish it and had just one puff?**

☐ Yes

☐ No → *Skip to question 2 on the next page*

b. How old were you the first time you smoked a cigarette?

Years of Age

☐

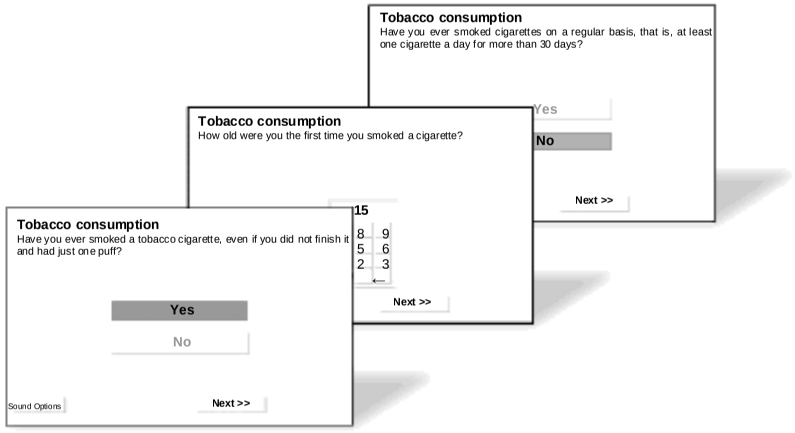
c. Have you ever smoked cigarettes on a regular basis, that is, at least one cigarette a day for more than 30 days?

☐ Yes

☐ No

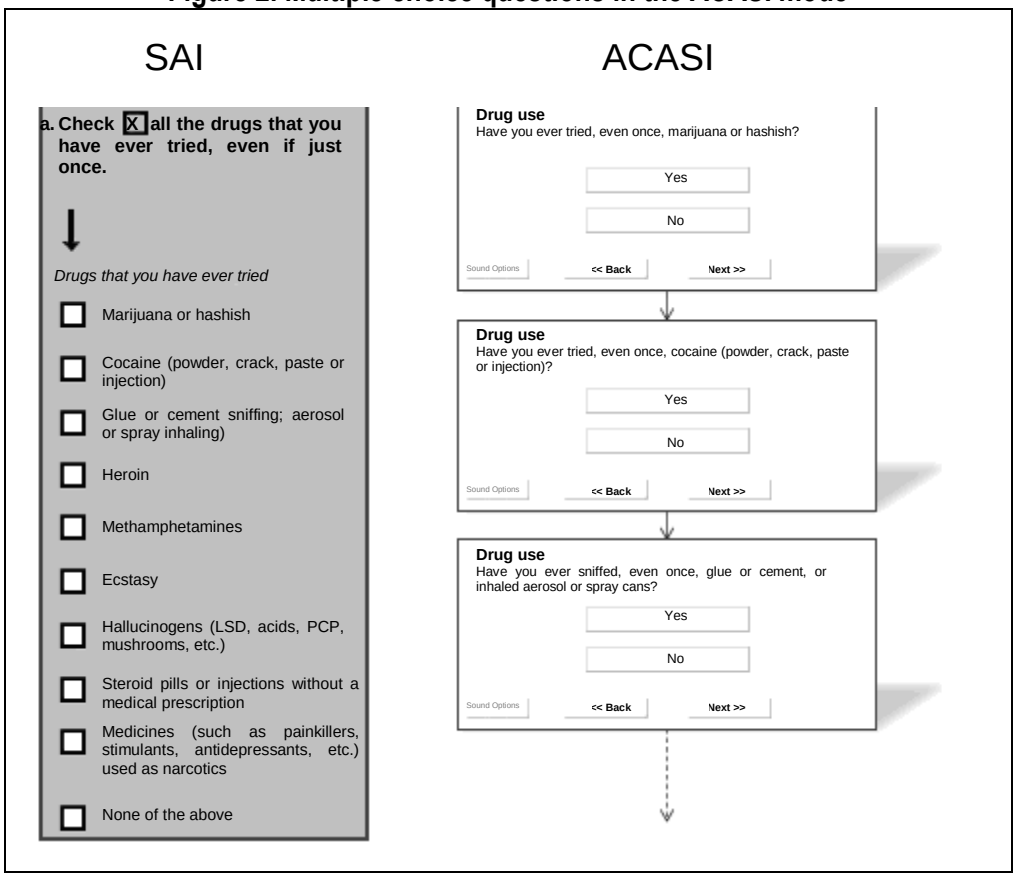
d. In the last 30 days, have you smoked cigarettes at least once?

ACASI



Source: Authors.

Figure 2. Multiple choice questions in the ACASI mode



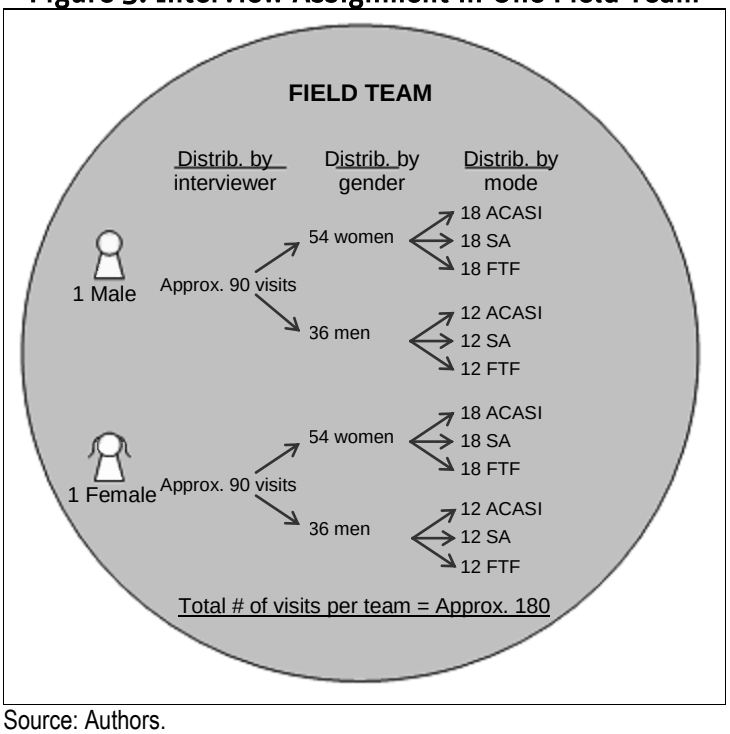
Source: Authors.

3.4 Data collection

Data were collected by a national specialized survey firm. This company recruited 10 interviewers specializing in face-to-face surveys (5 male and 5 female) and assigned them to cover the sample under the three home-applied modes. Enumerators were divided into 5 teams, with one male and one female interviewer in each team (see Figure 3). Each team was assigned one supervisor. Teams were assigned by geographical area based on company budget criteria (see Figure 4).

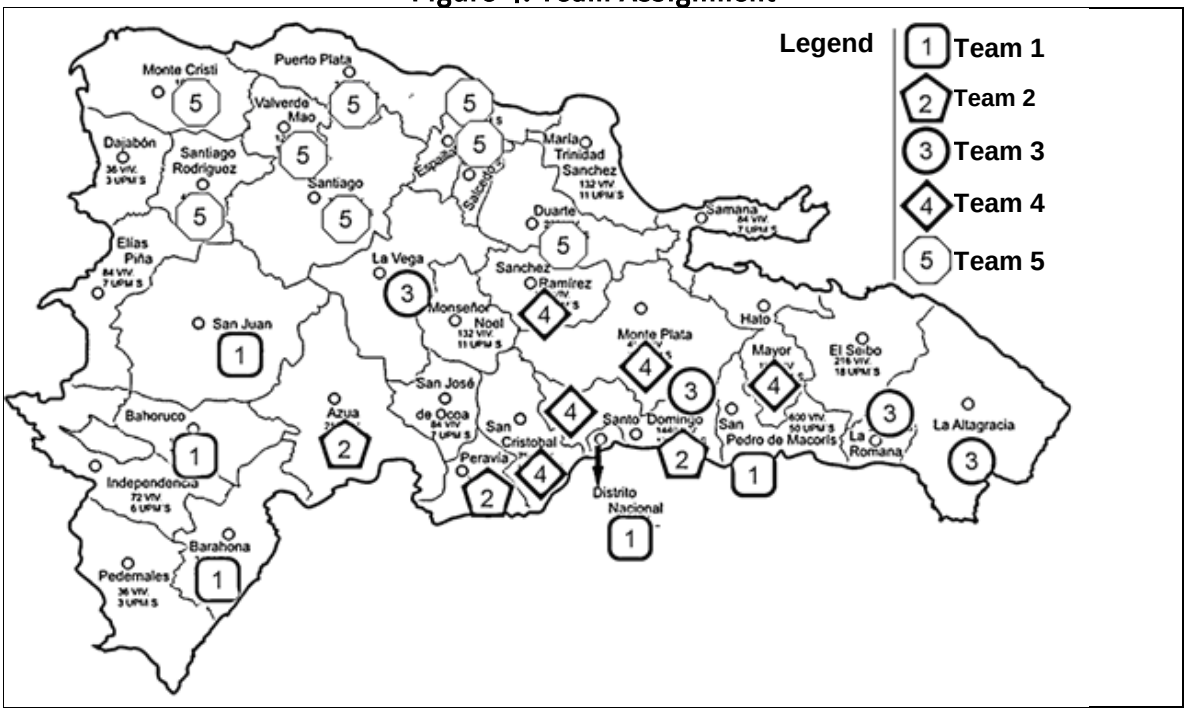
The company also enlisted 4 enumerators specializing in telephone surveys (2 male and 2 female) and assigned them to the CATI sample. Enumerators and supervisors received their training from the company and were later evaluated and supervised by the authors.

Figure 3. Interview Assignment in One Field Team



Source: Authors.

Figure 4. Team Assignment



Source: Authors.

3.5 The Sample

The experiment included a sample of 1,200 youth between the ages of 18 and 30 enrolled in the Youth and Employment Program (PJE). The sample was randomly assigned to four groups of 300 youth each, one for each of the 4 administration modes used. All participants received a cell phone calling card for RD$ 150 (approximately US$4.00) as an incentive to take part of the survey.

The sample was stratified by PJE assigned treatment, geographical area and sex of the respondent[12]. The PJE treatment group was divided into three levels, corresponding to the three experimental groups defined for the original PJE impact evaluation design[13]: (1) Technical Training Module (TTM) + Life Skills Module (LSM) treatment + internships, (2) LSM treatment + internships, and (3) Control Group. Stratifying by PJE treatment group prior to randomly assigning the administration mode ensured that the distribution of modes within each PJE treatment group would be the same. This meant that any effect modes could not bias the differences in the PJE treatment groups.

The enumerator's characteristics may have had a strong impact on the answers of respondents, especially in the case of interviewer-assisted modes (e.g., FTFI, CATI). In order to obtain information in the three home-applied modes, 10 field enumerators were organized into five teams, each made up of one male and one female interviewer. Teams were assigned by geographical area according to budget criteria, trying to give the same amount of interviews to each team ($n \approx 180$).14 Stratifying by geographical area prior to assigning the administration modes randomly ensured the same mode distribution for all teams. In other words, no team-related effect could bias the differences between these home-applied modes.

[12] The study assumes there are no regional differences in terms of poverty, literacy, or exposure to technology that might affect how participants responded to the survey modes..

[13] The PJE program had the eligible applicant youth organized into 35 persons per course, in 520 courses in total. The treatment subsamples of were organized as follows: 341 courses with life-skills and a complete sample of 520 courses with technical/vocational training.

[14] Geographical areas coincide with the country's provinces, except Santo Domingo, which was divided into three areas: east, north and west, plus the National District.

Table 3. Minimum detectable effect size, for different levels of prevalence

Base prevalence	Minimum detectable effect size
1.0%	+4.7
5.0%	+7.4
10.0%	+9.3
20.0%	+11.5
30.0%	+12.6
40.0%	+13.1
50.0%	+13.0
60.0%	+12.4
70.0%	+11.3
80.0%	+9.5
90.0%	+6.6
95.0%	+4.4

Source: Authors.

In order to obtain data in the CATI mode, two additional teams were included. This means that the difference between the CATI mode and any other home-applied mode may be biased by differential effects between the five home-applied teams and the two CATI teams. This is why certain comparisons exclude the CATI mode.

Within each stratum resulting from the 150 possible combinations between PJE status, sex, and geographical area, each individual was randomly assigned one of the 4 administration modes.

The entire sample was once again stratified, this time by administration mode, respondent sex, and team. Within each stratum resulting from the 56 possible combinations, each individual was randomly assigned, with equal probability, a male or female enumerator. This ensured a balanced sample for coincidence analysis between the enumerator's and the respondent's sex, separately for each mode.

Table 3 shows the minimum detectable effect size between modes for different levels of a prevalence indicator (level of significance $\alpha = 0.05$; statistical power $1-\beta = 0.9$). The sample allows us to detect biases from approximately 5 points (for very low or very high prevalence) to approximately 13 points (for medium prevalence rates).

3.6 Data Entry

Paper questionnaires produced in the FTFI and SAI modes were entered by company personnel using data entry software programmed to automatically check for errors at the question level (see definition of RCI later on). In the CATI mode, enumerators used software specifically designed for telephone interviews and entered their responses directly into the computer. The software featured the same automatic checks as for the FTFI and SAI modes. In the ACASI mode, respondents entered their responses directly into the computer. The ACASI software skipped between questions automatically, thus

preventing skip errors. It also featured automatic checks to detect blank or out-of-range responses, though checks for inconsistencies between two or more questions were excluded.

The company received the FTFI and SAI data entry software and the CATI and ACASI software from the authors.

3.7 Cost-Effectiveness Analysis

The following cost-effectiveness indicator was adopted, equal to the product between the cost and the measuring error attributable to the administration mode:

$$CE = Cost \cdot Error \qquad \text{Eq. 1}$$

Assuming that the effectiveness of an administration mode is the same as its accuracy or the inverse value of the measurement errors it generates, the above definition comes very close to the definition of cost-effectiveness ratio used in intervention cost-effectiveness analyses.

3.7.1. Cost Estimation

A bottom-up cost analysis was conducted for each administration mode; this considered staff costs, per diems, transportation, materials, office space and technical assistance. Then the fixed and variable components were calculated in relation to the sample size, and the authors' assumed input parameters were sensitized.

For the FTFI and SAI modes we set the estimate cost of software development at US$10,000. These two modes require developing a standard data entry program that includes error checking routines. For the ACASI and CATI modes we increased the cost of software development by 50% to account for the additional functionalities required by the data entry program. In the ACASI mode, the program must carry out the interview and be designed to be used by an untrained respondent, instead of a trained data entry operator. Also, recordings are needed for the audio. In the CATI mode, the program must allow the interviewer to carry out the interview on the computer screen, as well as manage the phone number database and monitor and record the calls.

3.7.2. Error Estimation

Errors attributable to each particular administration mode were estimated by means of several methods. **First**, data were reviewed to detect errors visible *at the observation unit level*, such as blank responses, incorrect question skipping, out-of-range values, and inconsistencies between two or more questions. **Second**, errors visible *at the aggregate level*[15]-such as socially desirable response bias-were identified. **Third**, *theoretical models* were utilized to estimate the effect of sample errors.

[15] Some important examples of inconsistencies between 2 or more questions are: Times consumed drugs in last 12 months is consistent with frequency of consumption /Number of lifetime sexual partners is greater than or equal to number of sexual partners in last 12 months /Number of lifetime sexual partners in question is equal to the number of sexual partners reported in the sexual history

3.7.3. Errors visible at the individual level

Errors visible at the **individual level** were manifold: First, non-response at the individual level because the respondent could not be located or because s/he declined to respond to any of the questions; and second, question-level errors, which were measured only among individuals who could be located and responded to at least one of the questions in the questionnaire. **Question-level errors** may take the following forms:

(1) **Non-response errors:** Option DON'T KNOW/NO RESPONSE has been selected (only for questions that offer one of these options explicitly).

(2) **Blank errors:** A data box that should not be empty is actually empty.

(3) **Skip errors:** A data box that should be empty is not empty.

(4) **Out-of-range errors:** The data contains impossible values (e.g. someone aged 3 years in this group, or a single-option question with multiple choices provided).

(5) **Consistency errors between two or more questions**: Responses to two or more questions have impossible values (e.g. the date of the last sexual relation with a partner comes earlier than the date of the first sexual relation).

The **Response Consistency Index (RCI)** is defined as the *percentage of data questions for each person without detectable errors*. To calculate the RCI, a rectangular matrix is constructed where each row represents an interview and each column represents one question from the questionnaire. The total amount of data (or cells) in the matrix is the product of the number of interviews times the number of questions. In each cell a number 1 is entered if the question is answered without errors and a 0 if it is not. The RCI is the sum of all 1s in the matrix divided by the total number of cells.

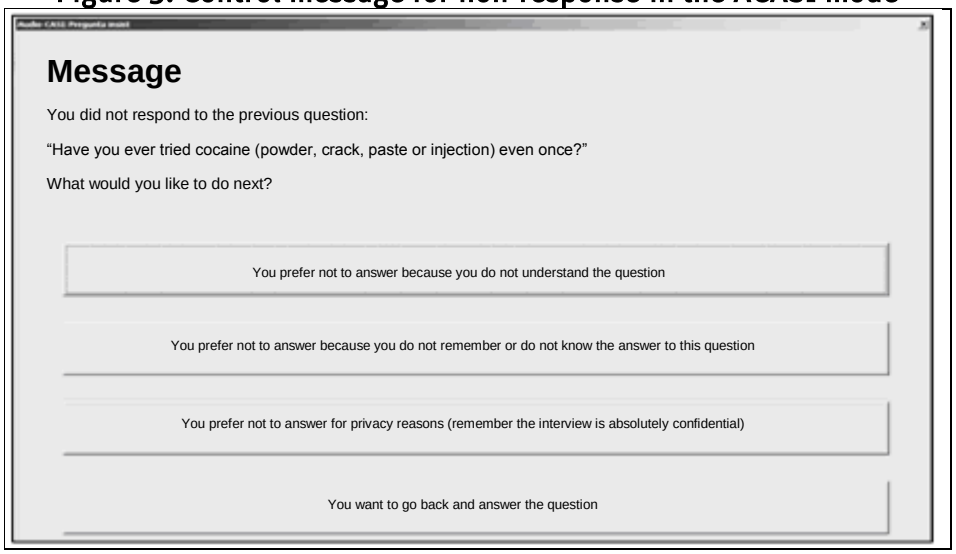
Non-response[16] was measured by using enumerator, data entry operator and ACASI software records. In the assisted modes, if a respondent declined to respond to a certain question because they felt uneasy, the enumerator entered the words DECLINES

section /If person did not respond to the question "Do you want to get pregnant?" because she is currently pregnant, then the response to the question "Are you currently pregnant?" should be "yes"/If a person responded affirmatively to any of the questions on reproductive health, then the question "Have you ever had sex?" should be "yes:/If the number of lifetime sexual partners is greater than or equal to 2, then the question "Have you had sex with someone else before your last sexual partner?" should be "yes"/If the number of lifetime sexual partners is greater than or equal to 3, then the question "Have you had sex with someone else before your next-to-last sexual partner?" should be "yes"/For each sexual partner in the sexual history section, the date of first sex should be prior to the date of last sex./For each sexual partner in the sexual history section, if the date of last sex is within 12 months of the interview date, then the frequency of sex during the last 12 months must be at least 1 time. Conversely, if the date of last sex is *not* within 12 months of the interview date, than the frequency of sex during the last 12 months must be zero/If any of the sexual partners in the sexual history section is of the same sex as the respondent, then question "Have you ever had sex with a *man* (men's questionnaire)/*woman* (women's questionnaire)?" should be "yes"/If the education level is primary, education grades takes values between 1 and 8/If the education level is secondary, education grades takes values between 1 and 4.

[16] Non-responses offered are cases where the respondent selected the option "*don't know/no response"* (available in some questions in the self-administered questionnaires). The SAI mode shows how offering these options may induce respondents to use them more than they would if an interviewer were coding the response without offering the answer options.

TO RESPOND. If the respondent did not respond because they did not know or did not remember, the enumerator entered DOES NOT KNOW. In the ACASI mode, if the respondent failed to respond to a question, the software displayed the message shown in Figure 5. In the SAI mode, by design no distinction was made between DECLINES TO RESPOND and DOES NOT KNOW.[17]

Figure 5. Control message for non-response in the ACASI mode



Source: Authors.

3.7.4. Errors visible at aggregate level

In order to detect errors visible at the aggregate level, we compared the mean values given by each administration mode, for the twenty one risk indicators. Since assignment to treatment groups was conducted randomly, any statistically significant difference between modes is the effect of mode of administration.

3.7.5. Sample error

The sample error depends on the size of the sample and the variance of the indicator to be estimated. It may be quantified by using the following formula to calculate the margin of error with 95% reliability (for a simple random sample without correction for finite population):

$$\text{Error margin} = \pm 1,96\sqrt{\frac{\sigma^2}{n}} \qquad \textit{Eq. 2}$$

[17] In the SAI mode, the cause of an error could not be determined unless options for DOES NOT KNOW or DECLINES TO RESPOND were explicitly provided. Including these responses explicitly is not recommended, as it is well known that they induce non-response. This mode shows over twice as many blank and non-responses as the other modes.

where: σ^2 is the sample variance

n is the sample size

If the indicator is biased, the total error is calculated as the sum of the bias and the margin of error:

$$\text{Total error} = \beta \pm 1,96\sqrt{\frac{\sigma^2}{n}} \qquad \textit{Eq. 3}$$

Where: β is the bias

Equation 3 indicates that bias is not dependent on sample size. This is true in theory, but in practice bias can indeed depend on the size of the sample. Typically, bias increases progressively with the sample size, at least for the following two reasons: (i) if increasing the sample means that the field staff should also be increased, monitoring the quality of collected data becomes more difficult; and (ii) if increasing the sample implies that the duration of data collection should also be extended, it is possible that field staff enter the downward part of the learning curve.

This study assumes that bias does not change with sample size. In order to increase the sample size, the data collection period is extended, leaving everything else unchanged. Then, it is assumed that the decreasing effects of the learning curve are negligible.

4 Results

4.1 Costs

Table A.1 in the Annex presents the details of the bottom-up cost results. Some entry parameters were obtained by empirical means (marked with a dagger), while others correspond to assumptions by the authors (marked with an asterisk). The total cost of surveying 300 youths can range between US$60,000 and US$80,000, depending on the administration mode applied: Conducting a survey in the CATI mode costs roughly US$60,000; in the SAI mode approximately US$70,000; and in the FTFI and ACASI modes as much as US$ 80,000. The difference between the CATI mode and the others is basically attributable to savings in transportation and localization time costs. The primary difference between the SAI mode and the other home-based modes is the lower level of expertise and training required from interviewers and supervisors.

An optimistic scenario was defined, in which all entry parameters assumed by the authors were divided by a factor of 1.5, and a pessimistic scenario, in which these parameters were multiplied by a factor of 1.5. Relative differences between the administration modes were not very sensitive to this variation.

The fixed component of the total cost is independent on the sample size, and is equal to the minimum investment required. The fixed cost is basically the same in the FTFI, ACASI and CATI modes (see Table 4). In the SAI mode, the fixed cost is lower than in other modes because it requires a shorter training period.

The variable component of the cost is a function of sample size and is expressed as the marginal cost of each additional unit in the sample. The marginal cost was estimated by varying the size of the sample while keeping all other entry parameters constant, including the amount and productivity of interviewers, supervisors, etc. The number of days in the field was left as a parameter that can vary freely with the size of the sample. In other words, each additional unit is produced by extending the duration of the survey. Assuming that production conditions remain constant in time, we have the following: (i) the marginal cost is constant for any sample size, (ii) investment costs are independent of the sample size and equal to fixed costs, and (iii) recurrent costs are independent of the sample size.

The marginal cost is substantially lower in the CATI mode than in the other modes due to the savings in transportation and location time costs. As illustrated in Table 4, as the sample size grows, the CATI mode quickly becomes the most economical option.

Table 4. Fixed and variable components of cost as a function of the sample size (US$)

	Administration Mode			
Cost Component	FTFI	CATI	SAI	ACASI
Fixed cost[18]	52,569.30	52,369.09	45,922.87	55,861.34
Marginal cost[19]	90.77	20.49	77.82	79.85
Total cost for different sample sizes:				
n = 100	61,645.88	54,418.38	53,704.45	63,846.25
n = 1,000	143,335.10	72,862.02	123,738.64	135,710.45
n = 10,000	960,227.38	257,298.48	824,080.62	854,352.42
Average cost per unit for different sample sizes:				
n = 100	616.46	544.18	537.04	638.46
n = 1,000	143.34	72.86	123.74	135.71
n = 10,000	96.02	25.73	82.41	85.44

Source: Authors.

[18] Fixed costs are costs that are independent of output and include technical assistance (design software, training materials, methodology), machinery, project team leader, and data team leader.

[19] For each additional unit in the sample.

4.2 Errors

4.2.1. Errors visible at the individual level

The non-response rate at the individual level (refusal rate) was 15%, and no statistically significant differences were observed between modes. Roughly 9 out of every 10 non-response cases correspond to situations where the individual could not be located, while the other cases were declines (the individual was located, but did not answer any questions). No differences between the three home-based modes were expected, since they use exactly the same location method. However, the CATI mode uses a different location method, achieving the same level of efficiency as the home-applied mode.

The non-response rate at the question level and the RCI were measured from among those who answered at least one question in the questionnaire (85% of the sample, distributed evenly among modes). Table 5 shows the percentage of data with errors, according to the type of error and administration mode. The SAI mode shows the poorest performance out of all four modes, with an RCI equal to 83%. This mode showed over twice as many blank and non-responses as the other modes. It also produced a large number of skip errors, which are rarely observed in the other modes.

The FTFI and CATI modes control non-response, blank responses, and skip errors because they rely on skilled interviewers. However, FTFI mode interviewers produce a larger number of complex inconsistencies (inconsistencies involving two or more questions) than CATI mode interviewers. This is because in the CATI mode, computer-assisted checks help interviewers eliminate complex inconsistencies.

The ACASI mode manages to control non-responses, blank responses and skip errors as efficiently as the FTFI and CATI modes. This is done by replacing the interviewer and instead using software that controls blank responses and out-of-range values and adapts the flow of the interview to prevent incorrect question skipping. The ACASI mode, by software design, fails to effectively control complex consistencies.

Table 5. Percentage of data with visible errors at the question level, according to type of error and administration mode

	Administration Mode				
Type of error	FTFI	CATI	SAI	ACASI	Total
Non-responses offered	0.5	0.8	1.6	0.9	0.9
Blank	1.2	1.1	3.3	1.3	1.7
Skip errors	0.2	0.0	6.5	0.0	1.7
Out-of-range errors	0.0	0.0	0.2	0.0	0.1
Consistency errors between two or more questions	3.2	0.1	5.5	4.8	3.4
Total errors	5.1	2.0	17.0	7.0	7.8
RCI	**94.9**	**98.0**	**83.0**	**93.0**	**92.2**

Source: Authors.

Non-response to individual risk indicators shows a similar situation (Tables A.2 and A.3).[20] Self-completion modes generated higher non-response rates than assisted modes. Among the self-completion modes, the ACASI mode shows a lower level of non-response than the SAI mode because the former relies on automatic checks to detect blank responses and range errors and prompts the participant for a response.

At least four causes for non-response can be identified in the SAI mode. *First*, in the case of indicators developed based on multiple-choice questions –such as drugs and STDs– non-responses mostly resulted from the way the question is formatted in paper. Many respondents would leave those questions blank even though the "none of the above" option was offered, so it was difficult to distinguish a non-response from a participant who had not experienced any of the listed risk behaviors.

A second cause of non-response is incorrectly following a skip pattern. For example, in the case of indicators developed based on questions that (by chance) were listed at the bottom of a page–such as gangs, fights, sexual identity, and transactional sex–non-responses mostly occurred due to skip instructions in other questions at the top of the page. Skip instructions were given at the top of the page for all those indicators where respondents were instructed to skip to the question at the bottom of the page. Most of the youths who left the bottom question blank chose the skip option in the question at the top of the page and moved on to answer the first question on the following page. This means that they either did not read or did not understand the skip instructions correctly and skipped to the following page rather than to the question at the bottom of the page.

[20] The number of observations used to calculate risk indicators may be equal to or higher than the number of consistent observations according to the RCI. Some of the errors considered in the RCI may be ignored or corrected for the purposes of calculating risk indicators. This means that the actual data percentage for indicator analysis may be higher than the RCI.

Third, non-response to indicators based on questions involving high cognitive difficulty, such as concurrent sexual partnerships, results from youth not being able to recall detailed information (for example, recalling the dates of the first and last sexual intercourse with their partners). This situation applies to the FTFI, ACASI and CATI modes as well.

Obviously, a *fourth* and very important cause of non-response is refusal. In contrast, non response in the other indicators with high non-response levels in the SAI mode–such as *age of sexual debut*, *number of lifetime sexual partners, and sex under the influence of alcohol or drugs*–originated in other causes that cannot be determined (e.g., privacy, fatigue, lack of interest–all of which are types of refusal).

In the ACASI mode, non-response also has several potential causes. First, in ACASI, multiple choice indicators are separated into a series of Yes/No questions (Figure 2); leaving any single question blank results in a non-response for the indicator. For example, in 11 of the 14 non-response cases in the *drug* indicator ("Have you ever used any of the following drugs"? followed by 9 different drugs choices categories), the respondent had left only one of the 9 questions blank. For the STD indicator, which consisted of 11 Yes/No questions, non-response is lower than for drugs, probably because the STD question comes later in the survey, and respondents had learned how to avoid the control message for non-response by the time the question was asked (Figure 5).

The control message for non-response in the ACASI mode asks the participant to indicate the reason for the non-response (does not understand, does not recall/know, or declines to respond on privacy grounds). For example, 5 non-response cases for *sexual identity* were "does not understand" and 2 cases were "declines to respond for privacy reasons." The remainder 11 cases corresponded to "other" (offered as a fourth option in addition to "heterosexual," "homosexual" and "bisexual"). In the "*ever had sex"* indicator, 2 cases corresponded to "does not understand," 1 case to "does not recall/know," and 2 cases to "do not answer for privacy reasons." In the "*lifetime sexual partners"* indicator, one case answered "does not recall," and 2 cases gave extreme values that were disregarded (Figure 5).

4.2.2. Errors visible at the aggregate level[21]

Some risk indicators show statistically significant differences between modes (Table A.4). There is some evidence that the ACASI mode shows a prevalence of "*ever had sex"* approximately 30 points lower than the FTFI and SAI modes. To confirm that the ACASI mode is biased, we used a telephone survey conducted by the World Bank between 2009 and 2010 among the same youth cohort, which recorded data on the participants' number of children. We compared the number of respondents who reported that they had never had sex but who had children (Table 6). In the FTFI and

[21] The findings of this section exclude the CATI mode from the analysis because the group of interviewers used in the CATI mode differed from the group of interviewers used in the home-applied modes.

CATI modes, the number of such inconsistencies is almost zero. The SAI mode shows some inconsistencies, while in the ACASI mode, the number of inconsistencies is abnormally high, particularly among women. Such a difference cannot be explained by a socially desirable response bias and may correspond to fatigue or lack of interest or seriousness. We believe that due to the length of the questionnaire and the learning process for the software, where the interviewer got to know how to advance faster in the questionnaire, a skipping pattern was developed that was independent from reality or circumstances. The fact that prevalences for specific indicators were lower with ACASI (usually those located at the end of the survey) means that they might use "No" as a systemic response pattern.

In the particular case of this experiment, no data post-processing issues were experienced. The frequencies directly obtained from ACASI laptops were equal to those obtained after post-processing and to those in the report. Likewise, the software used in this mode properly recorded which key was pressed (Yes or No) for a certain question.

The explanation that youths have been systematically pressing "No" before reading a question may be because the question follows a long series of questions on STDs and symptoms. After 14 or 15 questions like: "Has any health practitioner, such as a doctor or a nurse, ever told you that you had a sexually transmitted disease called…?" it is highly likely that many youths have realized that all of those questions dealt with medical diagnoses and knew that the answer to all of them was "No," and thus they just pressed "No" repeatedly. When suddenly they reached the question "Have you ever had sex?" it may well be ascertained that some just pressed the "No" key accidentally (thus continuing the pattern they had been following of pressing the "No" key).

This is a good example of how the structure of a questionnaire and the order of questions have an effect on the potential responses.

Table 6. Number of respondents who reported that they had never had sex but who have children

Mode	Men, ever sex = 0		Women, ever sex = 0	
	Total	With children	Total	With children
FTFI	6	0	16	0
CATI	5	0	10	1
SAI	13	1	17	5
ACASI	40	8	64	44

Source: Authors, based on data from PJE longitudinal evaluation telephone survey (World Bank)

This problem hinders the measurement of all indicators of sexual behaviors in the ACASI mode. The "*ever had sex"* question is a filter for all sex-related questions in the questionnaire (i.e., the respondents who choose *ever had sex* = 0 automatically skip all sex-related questions). Erroneously responding *ever had sex* = 0, not only results in a downward bias in the prevalence of this indicator, but also in a non-response for all the other indicators related to sexual behaviors (this type of non-response is not included in the non-response calculated in Tables A.2 and A.3). A similar phenomenon is observed in the *binge drinking* indicator: The ACASI system shows a prevalence that is 10 to 15 points lower than that from the FTFI and SAI.

The ACASI mode shows a higher prevalence of drug and marijuana consumption than the FTFI and SAI modes. This higher prevalence of drug consumption could be explained by the fact that multiple-choice questions are broken down into a series of independent questions, and this results in the participants spending more time on a single question which might increase recall. Nevertheless, the foregoing does not explain the higher prevalence of marijuana consumption observed in women, which may be explained by the attitude of women to marijuana and hence desirable response bias. Higher levels of reported use may be explained by increased privacy conditions ensured by ACASI. Harrison et al. (1997) shows that measuring levels and patterns of illicit drug use, their correlates, and related behaviors requires the use of self-report methods. However, the validity of self-reported data on sensitive and highly stigmatized behaviors such as drug use has been questioned, and the authors suggest the need to use biomarkers to validate results (urinalysis or hair analysis). Also, the literature suggests that familiarity with the interviewer, as measured by number of prior interviewing contacts, depresses drug-use reporting. We speculate that interviewer familiarity increases the salience of normative standards and that participants respond not only in terms of their past familiarity but also in terms of their subjective expectations regarding the probability of a future encounter with the interviewer.

Among women, there is weak evidence that the FTFI mode shows a younger *age of sexual debut* than the other modes. This difference is less than one year. This could be explained by a socially desirable response bias, where the response considered socially desirable by young women is an *age of sexual debut* younger than the real one. In fact, when the "*privacy during the interview"* variable is introduced as an explanatory variable in models, a privacy effect equal to +0.9 years is observed. In other words, whenever privacy during a FTFI interview is insufficient, women tend to report ages almost a year younger than when privacy exists. This is consistent with the hypothesis that, among women, it is socially desirable to boast a younger *age of sexual debut* than the real one[22].

[22] The levels of statistical significance of these effects range from 0.05 to 0.10. In addition, the privacy variable is not controlled in this experiment and is vulnerable to endogeneity bias.

Findings about women's sexual debut and binge drinking highlight how the direction bias can be population specific. For socially accepted behaviors, suggesting that less risky behaviors might be the more accurate response. Unlike what Brener et al. suggest, higher prevalence estimates may be less valid than lower estimates. Data from a qualitative evaluation[23] using focus groups among Dominican Republic sample of youth ages 15-18 reinforces this hypothesis; the results shows that sexuality is understood by youth as a right and tolerated by adults. Alcohol consumption is considered to be common entertainment among youth and actually not harmful to health[24].

There is evidence that, among women, the SAI mode shows a higher prevalence of partnership concurrency (e.g., partnerships that overlap in time) than the other modes. In the SAI mode, concurrency is equal to 0.17, while in the other modes it ranges from 0.02 to 0.05. It seems that this difference cannot be explained solely by the non-response rate of the SAI mode (use of Eq. 2 can show that the highest downward bias is equal to -0.04). Concurrency may well be a sensitive subject for this group of women, who opt to not reveal overlapping sexual partnerships when interviewed thru the FTFI or ACASI modes. Also, it may be related to the underlying trust of computers, the youth's relationship to technology and their willingness to report sensitive information in the ACASI mode.

There is weak evidence that, among men, the FTFI mode shows a higher prevalence of *safe casual sex* than the other modes. This difference may originate in a socially desirable response bias, where the response the youth consider socially desirable is a greater use of condoms than the real one.

Also among men, there is weak evidence that the FTFI mode shows a higher prevalence of *transactional sex* than the other modes. This could also be explained by a socially desirable response bias.

The sex of the interviewer seems to generate socially desirable response biases in the *sexual identity* indicator. In the FTFI mode, the prevalence of men interviewed by women and who reported same-gender *sex* is higher than that among men interviewed by men (Table A.5). This could be explained by the assumption that men consider is undesirable to admit to other men that they have had homosexual sex.

4.2.3. CATI Mode

This section is devoted to comparing the CATI mode and the other modes. It should be noted, however, that differences may be contaminated due to differential effects between the two groups of interviewers.

[23] Bautista-Arredondo et al., 2011. Qualitative evaluation under the "Solidaridad" Program context.

[24] A statistically significant effect in *binge drinking* is observed, explained by the underlying correlated behavior: for example, a relationship between alcohol consumption and sexual activities in the Dominican context.

The CATI mode generated findings close to those of the FTFI mode. Statistically significant differences were observed in four indicators; however, only two of them might be caused by socially desirable responses.

- *Same-gender sex*, in men and women: In the CATI mode, the prevalence of *same-gender sex* was lower than in the FTFI mode, which may originate in a socially desirable response bias, if telephone calls inspire less privacy or confidence than face-to-face interviews.
- *Binge drinking*, in women: The prevalence of *binge drinking* in the CATI mode was lower than in the FTFI mode. This may also originate in a socially desirable response bias.
- *Age of sexual debut* and *pregnancy intentions*, in women: The prevalence of these indicators was higher in the CATI mode than in the FTFI mode. These differences do not seem to result from socially desirable responses, since that would not be consistent with the differences observed between the FTFI mode and self-completion modes.

4.3 Cost-Effectiveness

The cost-effectiveness ratio of a survey depends on: (i) the administration mode, which determines the size of the bias, the fixed cost, and the marginal cost; and (ii) the size of the sample, which determines the sample error and the variable cost. The total error corresponds to the addition of the bias and the sampling error. Accordingly, in the absence of bias, the sampling error becomes the only relevant error, but the larger the bias, the less important the sampling error.

4.3.1. Unbiased case

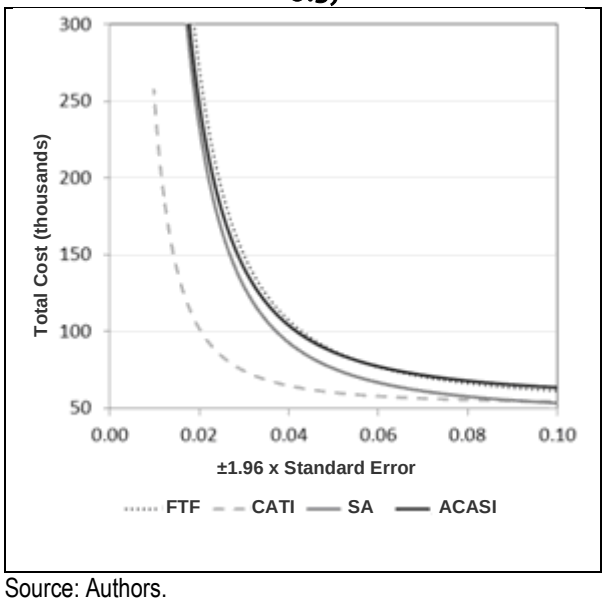
Figure 6 shows the cost involved in obtaining a determined level of sample error for a prevalence indicator equal to 0.50. The Y axis shows the cost calculated with the cost function described under section 3.1, scanning the size of the sample between 10 and 10,000. The cost tends to infinity when the sampling error approaches zero (or equivalently, when the size of the sample tends to infinity) and to a minimum–corresponding to the fixed cost of each mode–as the error grows bigger (or equivalently, when the size of the sample approaches zero). It is noted that the CATI mode provides the lowest sampling error. For any level of investment, the sampling error in the CATI mode is almost half of that in the FTF and ACASI modes. The SAI mode is fairly similar to the FTF and ACASI modes for high levels of investment, although its accuracy improves for lower levels of investment.

The cost-effectiveness of the sample varies depending on the level of investment (Figure 7). The cost-error has an infinite value when the level of investment is equal to the fixed cost. This is the minimum level of investment and corresponds to a sample of zero. As the level of investment increases, the cost-error drops swiftly because the size of the sample increases and the sampling error decreases. The cost-error reaches a minimum point, close to US$100,000, and then starts increasing again. This is because

the sampling error declines with the inverse of the square root of the sample size: that is, in order to cut the error in half, the size of the sample needs to grow fourfold. This implies that to cut the error in half, the variable cost needs to increase fourfold. For samples below the minimum point, the fixed cost takes predominance over variable costs, and increasing the size of the sample becomes cost-effective. However, for samples above the minimum point, the variable cost predominates; therefore, the cost-effectiveness of further increasing the size of the sample is limited.

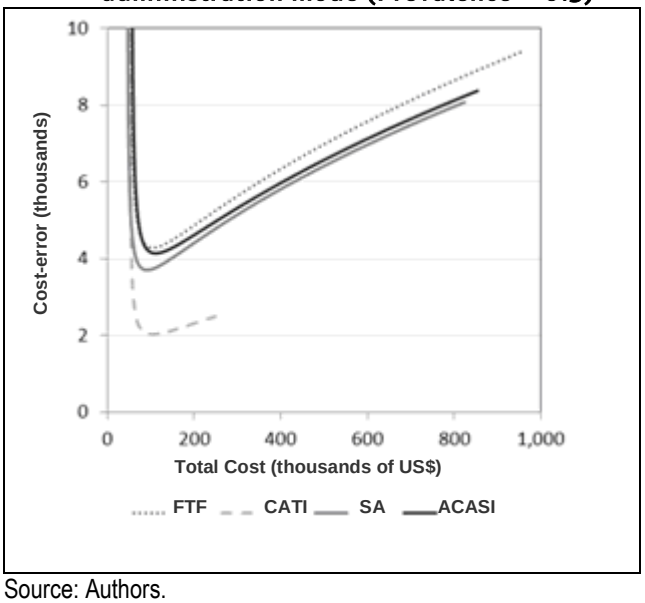
Table 7 shows the optimal level of investment for every mode, as well as the relevant sample size, the generated sampling error and the resulting error cost. The optimal level of investment will range from US$90,000 to US$110,000, depending on the mode. In this unbiased scenario, the CATI mode is approximately twice as cost-effective as the other modes because its cost is similar to those of the other modes but it generates half the sampling error.

Figure 6. Total cost versus sampling error, based on the administration mode (Prevalence = 0.5)



Source: Authors.

Figure 7. Cost-error versus total cost, based on the administration mode (Prevalence = 0.5)



Source: Authors.

Table 7. Optimal levels of investment for the unbiased case

	Administration Mode			
Statistics	FTFI	CATI	SAI	ACASI
Total cost ($)	105,213.46	104,831.01	91,834.18	111,755.72
Sampling size	580	2,560	590	700
Margin of error*	0.041	0.019	0.040	0.037
Cost-error*	4,281.30	2,030.43	3,705.07	4,139.41

* Calculated for prevalence = 0.50
Source: Authors.

4.3.2. Biased Case

When a bias exists, the optimal level of investment and sample size decrease. Table A.6 shows the optimal levels of investment for the biased case. Unlike the unbiased case, the optimal level of investment is sensitive to the prevalence level of the indicator. Accordingly, calculations are shown for prevalence equal to 0.05, 0.20, and 0.50. It may be observed that the optimal level of investment drops quickly as the bias increases. For large biases, optimal samples are very small, which reflects how sampling error has little importance when biases are large.

Table 8 shows two examples of how Table A.6 can be used to compare the cost-effectiveness of the various modes under different bias assumptions. In both examples, an indicator with prevalence equal to 0.05 is assumed. In the first case, the bias in the assisted modes is equal to +0.01, and a non-response bias in the SA mode is equal to +0.02. In the second case, the bias in the assisted modes equals +0.02, and the bias in

the SA mode is equal to +0.01. Table A.6 is intended to help calculate the cost-error values for each mode. Accordingly, the mode showing the lowest cost-error corresponds to the most cost-effective mode. This means that there is no mode or alternative sample size capable of producing a lower error at the same cost. It also means that any more economical alternative (regarding mode or sample size) necessarily generates a larger error.

It can be noted that the result is highly sensitive to small bias differences (the cost difference between both examples is almost US$40,000). Just a 1 point prevalence variation in the bias is enough to trigger a change of preferred mode and generate significant variations in the optimal levels of investment. This experiment lacks the statistical power to detect such small biases; however, this means that if biases have been detected, they are sufficiently large to generate large changes in the cost-effectiveness of the modes.

Table 8. Examples of how to use Table A.6

Examples	Case 1	Case 2
Entry parameters		
Prevalence	0.05	0.05
Bias in assisted modes	+0.01	+0.02
Non-response bias in SAI mode	+0.02	+0.01
Error cost according to Table A.6		
FTFI	2,765.03	3,537.22
CATI	1,712.73	2,411.05
SAI	3,073.17	2,399.46
ACASI	1,804.33	1,804.33
Result		
Most cost-effective mode	CATI	ACASI
Total cost ($)	73,271.88	111,755.72
Sample size	1,020	700
Bias + 1.96 x standard error	0.023	0.016

Source: Authors.

5 Conclusions

5.1 Costs

The costing exercise proves that the contact mode is the characteristic that has the largest impact on cost: telephone or home. Transportation costs absorb a large portion of the survey budget. This explains why the marginal cost of the CATI mode is four times lower than the marginal cost of the other modes. The fixed cost does not vary much from one administration mode to the other. Therefore, the CATI mode is the least costly of all, starting from relatively small sample sizes.

The characteristic having the second largest impact on cost is the level of experience and training required for interviewers and supervisors. The FTFI mode requires more experienced and skilled field personnel than the SAI mode, since they must play the interviewer's role, rather than a mere visitor's role. The ACASI mode requires personnel with an intermediate level of experience and training, since although they are visitors who are not required to conduct a face-to-face interview, they do need to be prepared to operate the ACASI software.

Another variable cost that may be relevant for the FTFI and SAI modes, particularly in the case of large samples, is the cost of printed copies, photocopies, pencils and any other material necessary to administer a paper form.

5.2 Errors

No differences have been identified in terms of location efficiency between home-based modes and the CATI mode. The list of addresses and telephone numbers shows the same quality and the decline rate (refusal rate?) at the individual level and is relatively low in all modes. But remember that this is an unusual sample–youth that are already in another study—and so may be highly motivated, willing, and enthusiastic about completing the survey. This limits generalizability/external validity.

The SAI mode generated the highest non-response rate at the question level, and the lowest RCI. This is caused by various cognitive flaws arising as a direct consequence of youth responding without any assistance or supervision whatsoever, with the exception of the written instructions on the paper form.

The ACASI mode introduced a downward bias in the "*ever had sex"* indicator, as a direct consequence of unsupervised responses by the youth. Many youth may have responded "No" to complete the questionnaire faster. This motivation may be higher in ACASI than in SAI, since the ACASI mode did not offer any hints as to how much of the questionnaire was left to complete, whereas in the SAI mode, youths knew approximately how many pages and questions were yet to be filled out. Although the ACASI mode solves several of the quality problems observed in the SAI mode, it shows quality problems of its own, which are related to lack of supervision or motivation.

There is evidence of socially desirable response biases in some risk indicators: (1) the FTFI mode introduced an upward bias in the *protected casual sex* in men; (2) using male interviewers introduced a downward bias in the *homosexual sex* in men; (3) the FTF mode introduced an upward bias in the *"sale of sexual services/transactional sex"* indicator in men; (4) the FTF mode introduced a downward bias in the "*partnership concurrency"* indicator in women; and (5) the FTFI mode introduced a downward bias in the "*age of sexual debut"* indicator in women. (It seems like this type of information should come much earlier in the results section.)
The length of the questionnaire is a determining factor when choosing an administration mode. This study used a questionnaire with some 100 questions to measure a comprehensive set of risk indicators. The experiment shows that self-completion modes are not suitable for a questionnaire of this size. The SAI mode is vulnerable to high non-response rates, partly because of cognitive difficulties. The longer the questionnaire, the harder it will be to avoid using complex questions and instructions. The ACASI mode, in turn, is vulnerable to high rates of inconsistent responses that may originate in the fact that the ACASI tool transforms multiple-choice questions into a series of Yes/No questions (see Figure 1). This forces the respondent to go through some 200 screens in the ACASI mode: that is, almost twice as many questions as the other modes. It is likely that some youths will get tired and will systematically opt for responding "No" just to finish sooner.

The CATI mode is not negatively affected by the length of the questionnaire. The experiment shows that a long questionnaire may be applied both in the home and by telephone among this population. Administering the survey by telephone does not generate completeness issues or higher decline rates in the interview. Moreover, the CATI mode offers better data quality than the FTF mode.

5.3 Cost-Effectiveness

In the absence of bias or whenever all modes are subject to the same bias, the CATI mode is always the most cost-effective one. When the bias is zero, the optimal investment in the CATI mode is approximately US$ 105,000 for a sample consisting of 2,560 youths. The optimal investment drops significantly in the presence of bias. With a bias equal to +0.01 (in a prevalence equal to 0.5), the optimal investment decreases to US$ 82,000, and the size of the sample diminishes to 1,440. With a bias equal to +0.1, the optimal investment drops to US$ 62,000, and the size of the sample decreases to 470. This proves that investing in large samples when there is a bias is futile.

Whenever biases differ among modes, the most cost-effective mode is usually the one subject to the lowest bias. Bias differences of roughly 0.01 may suffice to justify a change of mode. The biases detectable thru this experiment go from 0.04 and up; therefore, they practically ensure a change in favor of the mode subject to the lowest bias.

6 Recommendations

6.1 Selection of the most cost-effective administration mode

The decision on which administration mode should be chosen to conduct a risk behavior survey and how much should be invested depends on the number and type of indicators the experiment seeks to measure. The number of indicators determines the length of the questionnaire and the self-completion modes are not suitable for long questionnaires (over 50 questions) due to data quality issues and the resulting bias.

Should the SAI mode be selected, special attention should be paid to the logical and graphic design of the tool. The following lessons are drawn from this experiment:

- A large percentage of youth do not follow skip instructions correctly, unless they are relatively simple. For example "skip to the next page" usually works, but "skip to question x" may generate errors.
- A large percentage of youth do not respond well to questions requiring a large cognitive effort, such as the variables necessary to calculate partnership concurrency (i.e., they need to report dates of the partnership).
- Multiple-choice questions should be avoided, since they result in high level of missing data. In addition to questions being left blank, even for those who checked one response option the analyst doesn't know for sure that it is a complete answer unless the questionnaire is designed so that each item requires a response.
- Key filter questions (like "ever had sex") should be placed earlier in the questionnaire, but not at the beginning, unless the ACASI is being used only to asses risk and some less sensitive questions precede the ACASI assessment.

If the mode of choice is ACASI, special attention should be paid to avoid an excessive number of screens and to show respondents a progress indicator for the interview. If a large number of screens are involved, key filter questions (like "*ever had sex"*) should be placed earlier in the questionnaire.

If the FTFI mode is chosen, efforts should be focused on reducing complex inconsistencies. One way to obtain data quality similar to that of the CATI mode is by integrating new technologies that allow interviewers to administer questionnaires on computers and record the data as they would in a CATI setting. For example, CAFE mode and new technologies (e.g., computerized tablets, PDAs) make this much easier for real-time data collection in the field.

The type of indicator also has an effect. On the one hand, the self-completion modes are not suitable for indicators involving high cognitive difficulty, such as concurrency dates and multiple-choice questions. On the other hand, assisted modes are not advisable to measure sensitive indicators (e.g., same-gender sex). If enumerators wish to combine difficult and sensitive questions in the same survey, one alternative is to use a hybrid model with two modes, an assisted mode for the difficult questions and a

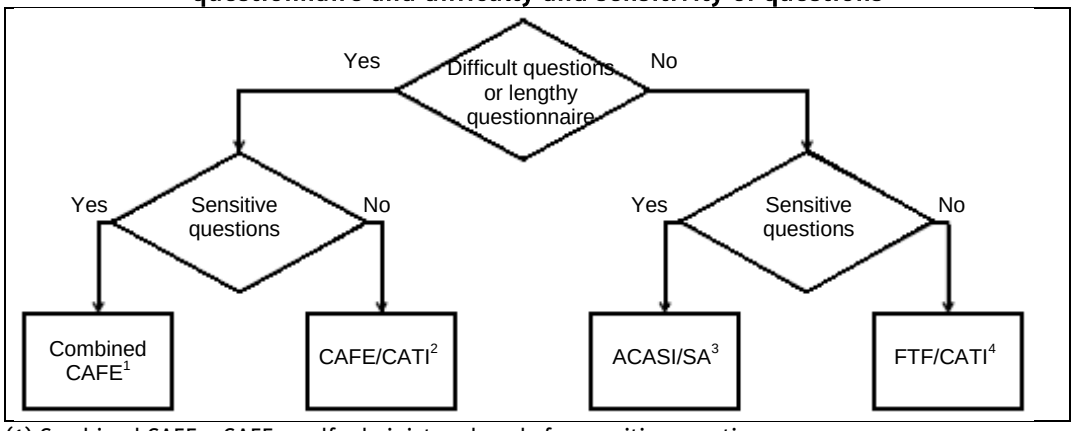
different mode for sensitive questions. For example, Langhaug (2002) uses the FTFI mode with confidential voting boxes; which are equivalent to the SAI mode.

There is a temptation to choose self-administered modes to deal with socially desirable response bias and to cut costs (as well as to avoid using the face-to-face mode, which requires trained interviewers). However, it turns out that properly applying a self-administered mode is extremely complex, and there is a risk that the final result will be worse than that of a face-to-face mode. Small details in the format of a paper questionnaire (like skipping to question X rather than to the following page) may result in an equally big or even greater bias than the one originating in the socially desirable response intended to be eliminated in the first place. Likewise, decisions that may seem relatively innocuous when designing an ACASI questionnaire (like separating a multiple-choice question into a series of Yes/No questions), may lead to the same negative results. In conclusion, developing a self-administered instrument entails its own difficulties, which are inherent to each mode and which may only be properly solved by conducting more extensive field testing than that required for an assisted mode.

Questionnaire design is particularly important. Responses may be susceptible to interview length (tedium and fatigue), graphic layout, placement of skip patterns on the page (SAI), and question placement within the interview. Certainly, greater creativity can be brought into survey design for ACASI administration. Several HIV prevention trials have integrated pictures into ACASI to bring greater clarity to the question being asked (e.g., include a picture of the contraceptive method asked about in the question) and to address low literacy.

Figure 8 summarizes the foregoing recommendations in a decision tree that shows the recommended modes based on the number and type of indicators intended to be measured. The biases generated by the length of a questionnaire and the difficulty and sensitivity of the questions are assumed to be sufficiently large to justify a change of mode.

Figure 8. Recommended administration modes, according to length of the questionnaire and difficulty and sensitivity of questions



(1) Combined CAFE = CAFE + self-administered mode for sensitive questions.
(2) CAFE or CATI depending on the relative accuracy of address or telephone number records.
(3) ACASI is preferred. SAI only for very simple questionnaires.
(4) FTF or CATI depending on the relative accuracy of address or telephone number records.
Source: Authors.

6.2 Selection of the administration mode under budgetary constraints

Up to this point, selection of the optimal administration mode has not yet considered budgetary constraints. Both Table A.6 and Figure 8 provide formulas to determine the most cost-effective administration mode; however, the level of investment is not defined a priori, as it is rather an outcome of the exercise. The level of investment cannot be optimized in a scenario of budgetary constraints. The only possible avenue is choosing the most effective administration mode (i.e., the one with the smallest error) for that investment level. This procedure is shown in Table 9.

Table 9. Procedure for selection of the most cost-effective administration mode under budgetary constraints

#	Procedure	Example	
1	Choose the available budget, B	$B = \$70.000$	
2	Choose two administration modes to be compared.	Mode 1: CATI	Mode 2: ACASI
3	Using Table 4, determine the fixed costs associated with each mode (CF_1 and CF_2, respectively), and the marginal costs (CM_1 and CM_2, respectively).	$CF_1 = \$52.369{,}09$ $CM_1 = \$20{,}49$	$CF_2 = \$55.861{,}34$ $CM_2 = \$79{,}85$
4	Using the formula below, calculate the sample sizes supported by this budget (n_1 and n_2, respectively): $n_i = {(B - CF_i)}/{CM_i}$	$n_1 = {(70.000 - 52.369{,}09)}/{20{,}49}$ $n_1 = 860$	$n_2 = {(70.000 - 55.861{,}34)}/{79{,}85}$ $n_2 = 177$
5	Define the level of prevalence of the indicator intended to be measured, p	$p = 0{,}20$	
6	Define the bias in each mode (β_1 and β_2, respectively).	$\beta_1 = 0{,}05$	$\beta_2 = 0{,}01$
7	Using the formula below, calculate the error generated by each mode: $e_i = \beta_i + 1{,}96\sqrt{\frac{p(1-p)}{n_i}}$	$e_1 = 0{,}05 + 1{,}96\sqrt{\frac{0{,}2(1-0{,}2)}{860}}$ $e_1 = 0{,}077$	$e_2 = 0{,}01 + 1{,}96\sqrt{\frac{0{,}2(1-0{,}2)}{177}}$ $e_2 = 0{,}069$
8	Choose the mode with the smallest error.	ACASI Mode	

Source: Authors.

References

Aitken JF, Youl PH, Janda M, Elwood M, Ring IT and Lowe JB. 2004. Comparability of Skin Screening Histories Obtained by Telephone Interviews and Mailed Questionnaires: A Randomized Crossover Study. American Journal of Epidemiology, Vol. 160, No. 6

Ashenfelter, K.. 2010. Eye-tracking Study Report: Examining User Patterns for Demographic Items on the 2007 and 2008 ACS Mail Forms, Statistical Research Division, US Census Bureau.

Bautista-Arredondo et al, 2011. Qualitative evaluation under the "Solidaridad" Program context.

Bautista-Arredondo S, Conde C, González A, López-Ridaura R, Sierra-Madero J, Volkow P. Public Health in the Penitentiary System. A study on prevalence of HIV and other STI, and health-related risk behaviors in Mexico City. National Institute of Public Health, 2011. Unpublished manuscript.

Beebe, T., P. Harrison, J. McRae Jr., J. Evans. 2006.The Effects of Data Collection Mode and Disclosure on Adolescent Reporting of Health Behavior, Social Science Computer Review Winter 2006 vol. 24 no. 4 476-488.

Borgers, N. and J. Hox. 2001. Item Non-response in Questionnaire Research with Children, Journal of Official Statistics, Vol. 17, No. 2, 321-335.

Brener, N.D., D.K. Eaton, L. Kann, J.A. Grunbaum, L.A. Gross, T.M. Kyle and J.G. Ross. 2006. The Association of Survey Setting and Mode with Self-Reported Health Risk Behaviors Among High School Students, Public Opinion Quarterly, Vol. 70, No. 3.

Brener, N.D., J.O.G. Billy and W.R. Grady. 2003. Assessment of Factors Affecting the Validity of Self-Reported Health-Risk Behavior Among Adolescents: Evidence From the Scientific Literature, Journal of Adolescent Health;33:436 –457.

Brown, N.R. and Sinclair, N.C. 1999. Estimating Number of Lifetime Sexual Partners: Men and Women Do It Differently. The Journal of Sex Research, Vol. 36, No. 3.

Brown JL, Vanable PA and Eriksen MD. 2008. "Computer-assisted self-interviews: A cost effectiveness analysis", Behavior Research Methods, 40 (1), 1-7

Catania, J., Binson, D., Hauck, W., and Canchola, J. 1996. Effects of Interviewer Gender, Interviewer Choice, and Item Context on Responses to Questions Concerning Sexual Behavior. Public Opinion Quarterly, Vol. 60, No. 3: 345-375.

Caeyers, B., N. Chalmers and J. De Weerdt. 2009. A Comparison of CAPI and PAPI through a Randomized Field Experiment, Draft Working Paper.

Centers for Disease Control and Prevention. 2004. Methodology of the Youth Risk Behavior Surveillance System, Department of Health and Human Services.

Centers for Disease Control and Prevention. 2009. A Guide to Conducting Your Own Youth Risk Behavior Survey, Department of Health and Human Services.

Chun H, Tavarez MI, Dann GE, Anastario MP. 2010. Interviewer gender and self-reported sexual behavior and mental health among male military personnel. Int J Public Health. 2011 Apr;56(2):225-9. Epub 2010 Aug 31; 8.

Cochran, W., 1999. Sampling Techniques - Third Edition, John Wiley and Sons.

Couper, M. 1999. Completion of Self-Administered Questionnaires in a Sex Survey - Statistical Data Included, Journal of Sex Research, http://findarticles.com/p/articles/mi_m2372/is_4_36/ai_58459532/

de Leeuw, E., J. Hox and G Snijkers. 1995. The Effect of Computer-Assisted Interviewing on Data Quality - A Review, Journal of the Market Research Society, 1995, Vol. 37, No. 4.

Denniston, M. , N. Brener, L. Kann, D. Eaton, T. McManus, T. Kyle, A. Roberts, K. Flint and J. Ross. 2010. Comparison of paper-and-pencil versus Web administration of the Youth Risk Behavior Survey (YRBS): Participation, data quality, and perceived privacy and anonymity, Computers in Human Behavior Volume 26, Issue 5, Pages 1054-1060.

Dillman, D., C. Jenkins, B. Martin and T. DeMaio. 1996. Cognitive and Motivational Properties of Three Proposed Decennial Census Forms, Center for Survey Methods Research, U.S. Census Bureau.

Eaton DK, Lowry R, Brener ND, Grunbaum JA, Kann L. 2004. Passive versus active parental permission in school-based survey research: does type of permission affect prevalence estimates of self-reported risk behaviors?. Evaluation Review. (In press).

Eaton DK, Brener ND, Kann L, Denniston MM, Hallfors TD, Khatapoush S, Kadushin C, Watson K and Saxe L. 2010. A comparison of paper vs. computer-assisted self interview for school alcohol, tobacco, and other drug surveys. Evaluation and Program Planning 23:149-55

Fernandez, L., E. Gerber, M. Clifton, G. Higbie and M. Meyers. 2009. Cognitive Pretesting of 2010 Alternative Questionnaire Experiment (AQE) Race and Hispanic Origin Treatment Panels, US Bureau of the Census.

Gauthier, J. 2002. Measuring America: the Decennial Censuses From 1790 to 2000, U.S. Census Bureau.

Ghanem KG, Hutton HE, Zenilman JM, Zimba R, Erbelding EJ. 2005. Audio computer assisted self interview and face to face interview modes in assessing response bias among STD clinic patients. Johns Hopkins University School of Medicine, Bayview Medical Center, Division of Infectious Diseases, Baltimore, MD 21224, USA. kghanem@jhmi.edu. http://www.ncbi.nlm.nih.gov/pubmed/16199744)

Gregson SA, Zhuwau T, Ndlovu J, et al. 2002. Methods to reduce social desirability bias in sex surveys in low-development settings: experience in Zimbabwe. Sex Transm Dis 2002;29:568e75.

Gribble, J., H. Miller, et al. 2000. The impact of T-ACASI interviewing on reported drug use among men who have sex with men. Subst Use Misuse 35(6-8): 869-890.

Grosh, M. and J. Muñoz. 1996. A Manual for Planning and Implementing the Living Standards Measurement Study Survey, LSMS Working Paper Number 126, The World Bank.

Groves, R.. 1989. Survey Error and Survey Costs, John Wiley and Sons.

Groves, R., R. Cialdini and M. Couper. 1992. Understanding the Decision to Participate in a Survey, Public Opinion Quarterly, Vol. 56:475-495.

Hallfors, D., S. Khatapoush, C. Kadushin, K Watson and L. Saxe. 2000. A comparison of paper vs computer-assisted self interview for school alcohol, tobacco, and other drug surveys, Evaluation and Program Planning Volume 23, Issue 2, 1 May 2000, Pages 149-155.

Hanscom, B., J. Lurie, K. Homa and J. Weinstein. 2002. Computerized Questionnaires and the Quality of Survey Data, Spine Volume 27, No. 16, 1797-1801.

Hewett PC, Mensch BS, Erulkar AS. 2004. Consistency in the reporting of sexual behaviour by adolescent girls in Kenya: a comparison of interviewing methods. Sex Transm Infect. 2004 ;80(suppl 2):ii43–ii48.

Hewett PC, Mensch BS, Ribeiro MC, Jones HE, Lippman SA, Montgomery MR, van de Wijgert JH. 2008. Using sexually transmitted infection biomarkers to validate reporting of sexual behavior within a randomized, experimental evaluation of interviewing methods. Population Council, New York, NY 10017, USA. phewett@popcouncil.org. http://www.ncbi.nlm.nih.gov/pubmed/18525081

Hunter, J., E. Gerber and D. Norris, 2009. Be Counted Form: Respondent Problems Encountered in Cognitive Testing, U.S. Census Bureau.

Jaya, Hindin, MJ and Ahmed S. 2008. Differences in Young People's Reports of Sexual Behaviors According to Interview Methodology: A Randomized Trial in India. American Journal of Public Health, Vol 98, No. 1, p.169-174

Jenkins, C. 1997. Improving the Navigational Qualities of the Decennial Census Short Form Requires Paying Attention to the Entire Mailing Package, US Bureau of the Census.

Jenkins, C. and D. Dillman. 1993a. Combining Cognitive and Motivational Research Perspectives for the Design of Respondent-Friendly Self-Administered Questionnaires, U.S. Census Bureau.

Jenkins, C. and D. Dillman. 1993b. The Language of Self-Administered Questionnaires As Seen Through the Eyes of Respondents, US Bureau of the Census.

Jennings T. E., Lucenko B. A., Malow R. M., Dévieux J. G. 2002. Audio-CASI vs. interview method of administration of an HIV/STD risk of exposure screening instrument for teenagers. International Journal of STD & AIDS, 13, 781-784.

Johnson, A. M., Wadsworth, J., Wellings, K., Bradshaw, S., & Field, J. 1992. Sexual lifestyles and HIV risk. Nature, 360, 410-412.

Kreuter, F., S. Presser, R. Tourangeau. 2008. Social Desirability Bias in CATI, IVR, and web surveys, Public Opinion Quarterly, Vol. 72, No. 5 2008, pp. 847–865.

Langhaug, L.F., Cheung, Y.B., Pascoe, S.J.S., et al. 2010. How You Ask Really Matters: Randomized Comparison of Four Sexual Behaviour Questionnaire Delivery Modes in Zimbabwean Youth, Sex Transm Infect.

Lara D, Strickler J, Olavarrieta Diaz C, et al. 2004. Measuring induced abortion in Mexico: a comparison of four methodologies. Sociol Methods Res 2004;21:529e58.

Laumann, Edward O., John H. Gagnon, Robert T. Michael, and Stuart Michaels. 1994. *The Social Organization of Sexuality: Sexual Practices in the United States*. Chicago: University of Chicago Press.

Lothen-Kline, Christine , Donna E Howard, Ellen K Hamburger, Kevin D Worrell, Bradley O Boekeloo. 2003. Truth and consequences: ethics, confidentiality, and disclosure in adolescent longitudinal prevention research. Journal of Adolescent Health. Volume 33, Issue 5 , Pages 385-394, November.Top of Form.

Martin, E., J.H. Childs, T. DeMaio, J. Hill, C. Reiser, E. Gerber, K. Styles and D. Dillman. 2007. Guidelines for Designing Questionnaires for Administration in Different Modes, U.S. Census Bureau.

Mensch BS, Hewett PC, Erulkar AS. 2003. The reporting of sensitive behavior by adolescents: a methodological experiment in Kenya. Population Council, New York, USA. www.ncbi.nlm.nih.gov/pubmed/12846131. Consultado en Julio, 2011.

Mensch BS, Hewett PC, Jones HE, Luppi CG, Lippman SA, Pinho AA, Diaz J. 2008. Consistency in women's reports of sensitive behavior in an interview mode experiment, São Paulo, Brazil. Population Council, New York, USA. bmensch@popcouncil.org. http://www.ncbi.nlm.nih.gov/pubmed/19201677

Mensch BS, Hewett PC, Abbott S, Rankin J, Littlefield S, Ahmed K, Cassim N, Patel S, Ramjee G, Palanee T, Mierzwa S, Skoler-Karpoff S. 2011. Assessing the reporting of adherence and sexual activity in a simulated microbicide trial in South Africa: an interview mode experiment using a placebo gel. AIDS Behav. Feb;15(2):407-21.

Minnis, Alexandra, M. Steiner, M. Gallo, L. Warner, M. Hobbs, A. van der Straten, T. Chipato, M. Macaluso, and N. Padian. 2009. Biomarker Validation of Reports of Recent Sexual Activity: Results of a Randomized Controlled Study in Zimbabwe. American Journal of Epidemiology. Vol. 170, No. 7. Downloaded from http://aje.oxfordjournals.org at University of California, Berkeley on June 18, 2010.

Morris, M. (1993) Telling tails explain the discrepancy in sexual partner reports. Nature, 365, 437-440.

NIMH Collaborative HIV/STD Prevention Trial Group. 2007. The feasibility of audio computer-assisted self-interviewing in international settings. AIDS; 21 Suppl 2:S49-58

Pienaar, JC. 2009. Efficacy of the informal confidential voting interview in enhancing self-disclosure and reducing social desirability bias: a comparative analysis with the SAQ and FTFI. Thesis (M.Sc.) - University of KwaZulu-Natal, Pietermaritzburg.

Potdar, Rukmini y Koenig, Michael. 2005. Does Audio-CASI improve reports of risky behaviour? Evidence from randomized field trial among young urban men in India. Studies in Family Planning. Hidhbeam Business journal. Consultado en Julio 2011.

Rathod, Sujit, Alexandra M. Minnis, Kalyani Subbiah, Suneeta Krishnan. 2011. ACASI and Face-to-Face Interviews Yield Inconsistent Estimates of Domestic Violence Among Women in India: The Samata Health Study 2005-2009. Journal of Interpersonal Violence. 2011. 26: 2437 originally published online 30 January 2011.

Redline, C., R. Smiley, M. Lee, T. DeMaio and D. Dillman. 1998. Beyond Concurrent Interviews: an Evaluation of Cognitive Interviewing Techniques for Self-Administered Questionnaires, U.S. Census Bureau.

Reininger, B., Evans, A.E., Griffin, S.F., Valois, R.F., Vincent, M.L., Parra-Medina, D., Taylor, D.J. and Zullig, K.J. 2003. Development of a Youth Survey to Measure Risk Behaviors, Attitudes and Assets: Examining Multiple Influences, Health Education Research Theory & Practice Vol.18 no.4.

Richardson, S.C., K. Politikou, M. Terzidou, Z. Maka and A. Kokkevi. 2006. The Quality of Data Obtained from Self-Completed Questionnaires in a Survey of High School Students, Quality and Quantity, 40:121-127.

Rogers SM, Willis G, Al-Tayyib A, Villarroel MA, Turner CF, Ganapathi L, Zenilman J, Jadack R. 2005. Audio computer assisted interviewing to measure HIV risk behaviours in a clinic population. Program in Health and Behavior Measurement, RTI International, Washington DC, 20036, USA. smr@rti.org http://www.ncbi.nlm.nih.gov/pubmed?term=measurement%20AND%20adolescent%20AND%20risk%20AND%20ACASI

Schonlau, M., R. Fricker and M. Elliott. 2002. Conducting Research Surveys Via E-mail and the Web - Chapter 2 Background on the Survey Process, RAND.

SEE - Secretaría de Estado de Educación, Fundación Familia Sana, Universidad Autónoma de Santo Domingo (UASD), Washington University en San Louis. 2009. Estudio sobre el Comportamiento y Salud en los Jóvenes Escolares de la República Dominicana Resultados de 2009.

Sedlak, Andrea J. 2010. Introduction to the Survey of Youth in Residential Placement, U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention.

Smith, T. W. 1992. Discrepancies between men and women in reporting number of sexual partners: A summary from four countries. Social Biology, 39, 203-211.

Strange, V., S. Forest, A. Oakley and The Ripple Study Team. 2003. Using Research Questionnaires with Young People in Schools: the Influence of the Social Context, Int. Journal of Social Research Methodology, Vol. 6, No. 4, 337-346.

Testa M, Livingston JA, VanZile-Tamsen C . 2005. The Impact of Questionnaire Administration Mode on Response Rate and Reporting of Consensual and Nonconsensual Sexual Behavior. Psychology of Women Quarterly, 29(4): 345-352.

Tourangeau, R. and T.W. Smith. 1998. Collecting Sensitive Information with Different Modes of Data Collection. In Computer Assisted Survey Information Collection, John Wiley and Sons, Inc., pp. 431–454.

Turner CF, Ku L, Rogers SM, et al. 1998. Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. Science 280:867e73.

UNAIDS. 2010. Reference Group on Estimates, Modelling, and Projections: Working Group on Measuring Concurrent Sexual Partnerships, HIV: Consensus Indicators Are Needed for Concurrency, The Lancet.

Van der Elst, E., H. Okuku, P. Nakamya, A. Muhaari, A. Davies, S. McClelland, M. Price, A. Smith, S. Graham & E. Sanders. 2009. Is Audio Computer Assited Self-Interview (ACASI) Useful in Risk Behabiour Assessment of Female and Male Sex Workers, Mombasa, Kenya?. Article published online. Plos One.

Van de Wijgert J, Padian NS, Shiboski S, et al. 2000. Is audio computer assited interviewing a feasible method of surveying in Zimbabwe? Int J Epidemiol;29:885e90.

Vereecken, C.A. and L. Maes. 2006. Comparison of a Computer-Administered and Paper-and-Pencil Administered Questionnaire on Health and Lifestyle Behaviors, Journal of Adolescent Health 38:426–432.

Wright, D., W.S. Aquilino, A.J. Supple. 1998. A Comparison of Computer-Assisted and Paper-and-Pencil Self-Administered Questionnaires in a Survey on Smoking, Alcohol, and Drug Use, Public Opinion Quarterly, Vol. 62:334-353.

ANNEX

Table A.1 Administration Mode Costs (US$)

	Mode			
ITEM	FTFI	SAI	ACASI	CATI
Sample†	300	300	300	300
Data entry operator †	1	1	0	0
Enumerators†	10	10	10	4
Supervisors†	5	5	5	1
Months of preparation*	3	3	3	3
Field days	10	10	10	10
Training days*	6	3	4	6
# Survey/day/ data entry operator	30	30		
Survey/day/enumerator†	3	3	3	7.5
Survey/day/supervisor	6	6	6	30
Field Personel Transport Costs				
Lease/day/car ($)*	50.00	50.00	50.00	0.00
Fuel/day/car ($)*	10.00	10.00	10.00	0.00
Total Cost ($)	3,900.00	3,450.00	3,600.00	0.00
Wage/day/driver($)*	30.00	30.00	30.00	0.00
Per diem/Driver ($)*	20.00	20.00	20.00	0.00
Total Cost Drivers($)	3,250.00	2,875.00	3,000.00	0.00
Total Cost Transportation($)	7,150.00	6,325.00	6,600.00	0.00
Wages and per diem personnel				
Wage/day/Data entry operator ($)*	30.00	30.00		
Total Cost Data entry operator ($)	300.00	300.00	0.00	0.00
Wage/day/ enumerator ($)*	50.00	30.00	40.00	50.00
Per diem/ enumerator ($)*	50.00	50.00	50.00	0.00
Total Cost enumerator ($)	13,000.00	8,900.00	10,600.00	3,200.00
Wage/day/Supervisor ($)*	100.00	60.00	80.00	200.00
Per diem/supervisor ($)*	50.00	50.00	50.00	0.00
Total Cost supervisors ($)	10,500.00	6,400.00	8,100.00	3,200.00
Wage/month/Project leader($)*	1,000.00	1,000.00	1,000.00	1,000.00
Per diem/ Project leader ($)*	100.00	100.00	100.00	0.00

Table A.1 Administration Mode Costs (US$)

	Mode			
ITEM	FTFI	SAI	ACASI	CATI
Total Cost Project leader ($)	4,713.28	4,298.29	4,436.62	3,613.28
Wage/month/field chief ($)*	700.00	700.00	700.00	
Per diem/ field chief ($)*	100.00	100.00	100.00	
Total Cost field chief ($)	4,129.30	3,748.80	3,875.63	0.00
Wage/month/data operator chief($)*	700.00	1,000.00	1,000.00	1,000.00
Total Cost data operator chief ($)*	2,529.30	3,498.29	3,536.62	3,613.28
Wage/month/Asistant($)*	250.00	250.00	250.00	250.00
Total cost Asistant ($)*	903.32	874.57	884.15	903.32
Total cost personnel ($)	36,075.19	28,019.95	31,433.03	14,529.88
Technical Assistance				
Wage/trainer ($)*	500.00	500.00	500.00	500.00
Cost/training ($)	3,000.00	1,500.00	2,000.00	3,000.00
Instruments/guidesdevelopment ($)*	10,000.00	10,000.00	10,000.00	10,000.00
Software development ($)*	10,000.00	10,000.00	15,000.00	15,000.00
Analysis ($)*	10,000.00	10,000.00	10,000.00	10,000.00
Total cost technical assistance ($)	33,000.00	31,500.00	37,000.00	38,000.00
Supplies				
Computers†	2	2	6	5
Price/computers ($)*	600.00	600.00	600.00	800.00
Total cost computers ($)	1,200.00	1,200.00	3,600.00	4,000.00
Surveys †	330	330	330	330
Survey print cost ($)*	3.00	3.00	0.00	0.00
Total Cost ($)	990.00	990.00	0.00	0.00
Telephone operator	0	0	0	4
Supervisors	5	5	5	1
Other †	3	3	3	3
Mobile minutes/day/operator †	0	0	0	480
Mobile minutes/day/supervisor*	60	60	60	60
Mobile minutes/day/management*	30	30	30	30
Monthly plan/phone ($)*				350.00
Monthly contract plan operators				1
Mobile minute cost ($)*	0.08	0.08	0.08	0.03

Table A.1 Administration Mode Costs (US$)

	Mode			
ITEM	FTFI	SAI	ACASI	CATI
Montly plan cost / operator ($)	0.00	0.00	0.00	1,400.00
Montly plan / operators ($)	0.00	0.00	0.00	748.80
Montly plan / supervisors ($)	312.00	276.00	288.00	23.40
Montly plan / management ($)	791.85	766.65	775.05	296.94
Total cost comunications ($)	1,103.85	1,042.65	1,063.05	2,469.14
Total cost supplies ($)	3,293.85	3,232.65	4,663.05	6,469.14
Work place				
Cost/day/data entry room($)*	10.00	10.00	0.00	0.00
Cost/day/calling room($)*	0.00	0.00	0.00	20.00
Cost/day/training room($)*	30.00	30.00	30.00	20.00
Total cost work places($)	280.00	190.00	120.00	320.00
Total Cost($)	79,799.04	69,267.60	79,816.08	59,319.02
Optimistic scenario ($)	48,263.09	42,728.19	49,193.37	36,554.64
Pessimist scenario ($)	136,845.86	115,995.04	133,744.55	99,243.68

† Empirical values; * Estimate values by authors; values without * or † are calculated.

Source: Authors.

Table A.2 Risk Indicators by Mode, Males

Indicator	Administration Mode																			
	FTFI				SAI				ACASI				CATI				Total			
	Obs.	NR	Mean	IC 95%	Obs.	NR	Mean	IC 95%	Obs.	NR	Mean	IC 95%	Obs.	NR	Mean	IC 95%	Obs.	NR	Mean	IC 95%	Prob.
ever_regular_smoker	101	0.0%	0.04	(0.00; 0.08)	91	2.2%	0.05	(0.01; 0.10)	94	4.1%	0.03	(0.00; 0.07)	100	0.0%	0.03	(0.00; 0.06)	386	1.5%	0.04	(0.02; 0.06)	0.812
current_smoker	101	0.0%	0.03	(0.00; 0.06)	91	2.2%	0.02	(-0.01; 0.05)	95	3.1%	0.05	(0.01; 0.10)	100	0.0%	0.05	(0.01; 0.09)	387	1.3%	0.04	(0.02; 0.06)	0.631
binge_drinking	101	0.0%	0.34	(0.24; 0.43)	86	7.5%	0.35	(0.25; 0.45)	95	3.1%	0.20	(0.12; 0.28)	100	0.0%	0.25	(0.16; 0.34)	382	2.6%	0.28	(0.24; 0.33)	0.071
Drug consumption	100	1.0%	0.10	(0.03; 0.17)	65	30.1%	0.08	(0.01; 0.14)	91	7.1%	0.18	(0.10; 0.25)	100	0.0%	0.07	(0.02; 0.12)	356	9.2%	0.11	(0.07; 0.14)	0.085
marihuana	100	1.0%	0.07	(0.02; 0.12)	65	30.1%	0.06	(0.00; 0.12)	94	4.1%	0.07	(0.02; 0.13)	100	0.0%	0.06	(0.01; 0.11)	359	8.4%	0.07	(0.04; 0.09)	0.977
Gang affiliation	101	0.0%	0.14	(0.07; 0.20)	70	24.7%	0.11	(0.04; 0.19)	96	2.0%	0.09	(0.03; 0.16)	100	0.0%	0.17	(0.10; 0.24)	367	6.4%	0.13	(0.09; 0.17)	0.437
Fight	101	0.0%	0.12	(0.04; 0.19)	76	18.3%	0.05	(0.00; 0.10)	98	0.0%	0.07	(0.02; 0.12)	100	0.0%	0.15	(0.07; 0.23)	375	4.3%	0.10	(0.07; 0.13)	0.118
Victim	101	0.0%	0.07	(0.02; 0.12)	91	2.2%	0.05	(0.01; 0.10)	96	2.0%	0.06	(0.01; 0.11)	99	1.0%	0.05	(0.01; 0.10)	387	1.3%	0.06	(0.04; 0.08)	0.948
sexual_identity	101	0.0%	0.03	(0.00; 0.06)	79	15.1%	0.00		91	7.1%	0.02	(-0.01; 0.05)	98	2.0%	0.00		369	5.9%	0.01	(0.00; 0.03)	0.181
STD	100	1.0%	0.02	(-0.01; 0.05)	83	10.8%	0.04	(0.00; 0.08)	96	2.0%	0.05	(0.01; 0.10)	99	1.0%	0.06	(0.01; 0.11)	378	3.6%	0.04	(0.02; 0.06)	0.506
plans_pregnancy	.	.	.		.	.	.		.	.	.		.	.	.		.	.	.		.
ever_sex	101	0.0%	0.94	(0.89; 0.99)	92	1.1%	0.86	(0.78; 0.94)	97	1.0%	0.59	(0.50; 0.68)	99	1.0%	0.95	(0.91; 0.99)	389	0.8%	0.84	(0.80; 0.88)	0.000
age_first_sex	93	2.1%	15.74	(15.3; 16.1)	71	10.1%	15.69	(15.1; 16.3)	56	1.8%	16.52	(15.8; 17.2)	93	1.1%	16.00	(15.6; 16.5)	313	3.7%	15.95	(15.7; 16.2)	0.177
partners_lifetime	92	3.2%	7.90	(5.85; 9.95)	68	13.9%	7.96	(5.77; 10.14)	54	5.3%	6.57	(4.14; 9.00)	93	1.1%	8.14	(6.33; 9.95)	307	5.5%	7.75	(6.69; 8.82)	0.745
partners_12months	95	0.0%	3.02	(1.97; 4.07)	67	15.2%	2.40	(1.67; 3.13)	55	3.5%	3.29	(1.29; 5.29)	93	1.1%	2.20	(1.73; 2.68)	310	4.6%	2.69	(2.16; 3.22)	0.425
Concurrency	83	12.6%	0.08	(0.02; 0.15)	54	31.6%	0.02	(-0.02; 0.06)	51	10.5%	0.08	(0.00; 0.16)	89	5.3%	0.08	(0.02; 0.14)	277	14.8%	0.07	(0.04; 0.10)	0.451
same_gender_sex	95	0.0%	0.09	(0.03; 0.16)	78	1.3%	0.04	(-0.01; 0.08)	57	0.0%	0.05	(-0.01; 0.11)	94	0.0%	0.03	(0.00; 0.07)	324	0.3%	0.06	(0.03; 0.08)	0.240
condom_casual	69	0.0%	0.81	(0.71; 0.91)	41	0.0%	0.63	(0.49; 0.78)	31	0.0%	0.52	(0.32; 0.71)	66	0.0%	0.77	(0.67; 0.88)	207	0.0%	0.72	(0.65; 0.79)	0.008
paid_sex	95	0.0%	0.22	(0.13; 0.31)	78	1.3%	0.19	(0.11; 0.28)	56	1.8%	0.29	(0.16; 0.41)	94	0.0%	0.22	(0.14; 0.30)	323	0.6%	0.23	(0.18; 0.27)	0.648
transaction_sex	94	1.1%	0.16	(0.08; 0.24)	70	11.4%	0.07	(0.01; 0.13)	56	1.8%	0.09	(0.01; 0.17)	94	0.0%	0.15	(0.07; 0.23)	314	3.4%	0.12	(0.08; 0.16)	0.260
sex_alcohol_drugs	94	1.1%	0.29	(0.19; 0.38)	71	10.1%	0.32	(0.22; 0.43)	56	1.8%	0.34	(0.22; 0.45)	93	1.1%	0.28	(0.19; 0.36)	314	3.4%	0.30	(0.26; 0.35)	0.840

Note: NR = item level non response rate in reference population. Obs. = number of effective observations used to calculate mean (ie. after dropping non response).
Prob. = p-value of null difference between means across modes.

Table A.3 Risk Indicators by Mode, Females

Indicator	Administration Mode																				
	FTFI				SAI				ACASI				CATI				Total				
	Obs.	NR	Mean	IC 95%	Obs.	NR	Mean	IC 95%	Obs.	NR	Mean	IC 95%	Obs.	NR	Mean	IC 95%	Obs.	NR	Mean	IC 95%	Prob.
ever_regular_smoker	160	0.0%	0.01	(-0.01; 0.02)	153	3.8%	0.01	(-0.01; 0.02)	157	0.6%	0.01	(0.00; 0.03)	155	0.0%	0.01	(-0.01; 0.03)	625	1.1%	0.01	(0.00; 0.02)	0.879
current_smoker	160	0.0%	0.01	(-0.01; 0.02)	153	3.8%	0.01	(-0.01; 0.02)	157	0.6%	0.03	(0.00; 0.06)	155	0.0%	0.01	(-0.01; 0.02)	625	1.1%	0.01	(0.00; 0.02)	0.111
binge_drinking	159	0.6%	0.26	(0.20; 0.33)	138	13.2%	0.25	(0.18; 0.33)	156	1.3%	0.12	(0.06; 0.17)	155	0.0%	0.15	(0.09; 0.22)	608	3.8%	0.20	(0.16; 0.23)	0.001
Drug consumption	154	3.8%	0.03	(0.00; 0.05)	114	28.3%	0.01	(-0.01; 0.03)	151	4.4%	0.07	(0.03; 0.11)	155	0.0%	0.01	(-0.01; 0.02)	574	9.2%	0.03	(0.01; 0.04)	0.006
marihuana	154	3.8%	0.00		114	28.3%	0.00		158	0.0%	0.03	(0.00; 0.06)	155	0.0%	0.01	(-0.01; 0.02)	581	8.1%	0.01	(0.00; 0.02)	0.018
Gang affiliation	159	0.6%	0.04	(0.01; 0.08)	120	24.5%	0.04	(0.01; 0.08)	157	0.6%	0.04	(0.01; 0.07)	155	0.0%	0.07	(0.03; 0.11)	591	6.5%	0.05	(0.03; 0.07)	0.530
Fight	159	0.6%	0.08	(0.04; 0.12)	131	17.6%	0.05	(0.01; 0.09)	156	1.3%	0.04	(0.01; 0.08)	155	0.0%	0.05	(0.02; 0.09)	601	4.9%	0.06	(0.04; 0.08)	0.516
Victim	160	0.0%	0.06	(0.03; 0.10)	150	5.7%	0.08	(0.03; 0.13)	158	0.0%	0.05	(0.02; 0.08)	155	0.0%	0.10	(0.05; 0.14)	623	1.4%	0.07	(0.05; 0.09)	0.416
sexual_identity	159	0.6%	0.00		119	25.2%	0.00		147	7.0%	0.02	(0.00; 0.04)	152	1.9%	0.02	(0.00; 0.04)	577	8.7%	0.01	(0.00; 0.02)	0.131
STD	159	0.6%	0.18	(0.12; 0.23)	138	13.2%	0.14	(0.08; 0.20)	154	2.5%	0.14	(0.08; 0.19)	155	0.0%	0.23	(0.16; 0.31)	606	4.1%	0.17	(0.14; 0.20)	0.090
plans_pregnancy	153	1.9%	0.07	(0.03; 0.10)	137	8.7%	0.11	(0.06; 0.16)	146	1.4%	0.11	(0.06; 0.16)	147	0.0%	0.20	(0.13; 0.26)	583	3.0%	0.12	(0.09; 0.15)	0.005
ever_sex	160	0.0%	0.90	(0.85; 0.95)	157	1.3%	0.89	(0.85; 0.94)	154	2.5%	0.58	(0.50; 0.67)	155	0.0%	0.94	(0.89; 0.98)	626	0.9%	0.83	(0.80; 0.86)	0.000
age_first_sex	143	0.7%	16.20	(15.8; 16.6)	128	8.6%	16.95	(16.5; 17.4)	89	1.1%	16.82	(16.2; 17.4)	144	0.7%	16.80	(16.4; 17.2)	504	2.9%	16.67	(16.5; 16.9)	0.054
partners_lifetime	143	0.7%	2.66	(2.28; 3.03)	125	10.7%	3.16	(2.59; 3.73)	90	0.0%	3.14	(2.49; 3.80)	145	0.0%	2.41	(2.08; 2.75)	503	3.1%	2.80	(2.55; 3.05)	0.063
partners_12months	143	0.7%	1.13	(1.04; 1.23)	122	12.9%	1.25	(1.13; 1.36)	90	0.0%	1.22	(1.10; 1.35)	145	0.0%	1.06	(0.97; 1.16)	500	3.7%	1.16	(1.10; 1.21)	0.053
Concurrency	134	6.9%	0.04	(0.01; 0.07)	104	25.7%	0.17	(0.10; 0.24)	88	2.2%	0.05	(0.00; 0.09)	137	5.5%	0.02	(0.00; 0.05)	463	10.8%	0.06	(0.04; 0.09)	0.000
same_gender_sex	144	0.0%	0.04	(0.01; 0.07)	137	2.1%	0.07	(0.03; 0.11)	89	1.1%	0.02	(-0.01; 0.05)	142	2.1%	0.01	(-0.01; 0.02)	512	1.3%	0.04	(0.02; 0.05)	0.053
condom_casual	21	0.0%	0.67	(0.45; 0.89)	22	8.3%	0.55	(0.31; 0.78)	23	0.0%	0.48	(0.26; 0.70)	27	0.0%	0.67	(0.48; 0.85)	93	2.1%	0.59	(0.48; 0.70)	0.480
paid_sex	144	0.0%	0.00		133	5.0%	0.00		89	1.1%	0.00		142	2.1%	0.00		508	2.1%	0.00		.
transaction_sex	141	2.1%	0.05	(0.01; 0.09)	126	10.0%	0.07	(0.03; 0.12)	90	0.0%	0.08	(0.02; 0.13)	142	2.1%	0.07	(0.03; 0.11)	499	3.9%	0.07	(0.04; 0.09)	0.822
sex_alcohol_drugs	140	2.8%	0.15	(0.09; 0.21)	119	15.0%	0.21	(0.13; 0.29)	na	na	na	(na; na)	142	2.1%	0.11	(0.06; 0.17)	na	na	na	(na; na)	na

Note: NR = item level non response rate in reference population. Obs. = number of effective observations used to calculate mean (ie. after dropping non response).
Prob. = p-value of null difference between means across modes. na = data not available because of a programming error in Audio-CASI instrument.

Table A.4 Differences in risk indicators by mode

	Males									Females								
	Diferencia simple					Diferencia ajustada				Diferencia simple					Diferencia ajustada			
Indicator	SAI-FTF		AC-SAI		FTF-AC	SAI-FTF		AC-SAI		FTF-AC	SAI-FTF		AC-SAI		FTF-AC	SAI-FTF		AC-SAI		FTF-AC	
ever_regular_smoker	0.015		-0.023		0.008	-0.017		0.009		0.007	0.000		0.006		-0.006	0.003		0.005		-0.008	
current_smoker	-0.008		0.031		-0.023	-0.007		0.032		-0.025	0.000		0.025		-0.026	0.007		0.026		-0.032	
binge_drinking	0.012		-0.149	**	0.137 **	0.062		-0.164	**	0.102	-0.011		-0.138	***	0.149 ***	0.025		-0.157	***	0.132	***
Drug consumption	-0.023		0.099	*	-0.076	-0.065		0.116	**	-0.051	-0.017		0.057	***	-0.040 *	0.004		0.050	**	-0.054	**
marihuana	-0.008		0.013		-0.004	-0.043		0.041		0.003	0.000		0.032	**	-0.032 **	0.001		0.037	**	-0.038	***
Gang affiliation	-0.024		-0.021		0.045	-0.041		-0.005		0.046	-0.002		-0.003		0.006	0.014		-0.008		-0.006	
Fight	-0.066		0.019		0.047	-0.079		0.021		0.058	-0.028		-0.009		0.037	-0.032		0.010		0.023	
Victim	-0.014		0.008		0.007	0.007		0.006		-0.012	0.018		-0.029		0.012	0.032		-0.059	*	0.027	
sexual_identity	-0.030	*	0.022		0.008	-0.023		0.008		0.015	0.000		0.020	*	-0.020 *	0.003		0.016		-0.019	*
STD	0.016		0.016		-0.032	0.004		0.004		-0.008	-0.038		-0.001		0.040	-0.073		0.023		0.050	
plans_pregnancy	.		.		.	.		.		.	0.044		0.000		-0.044	0.035		0.016		-0.051	
ever_sex	-0.082	*	-0.271	***	0.353 ***	-0.049		-0.292	***	0.341 ***	-0.008		-0.307	***	0.316 ***	-0.037		-0.288	***	0.324	***
age_first_sex	-0.052		0.828	*	-0.776 *	-0.155		0.667		-0.512	0.750	**	-0.125		-0.624 *	0.677	**	-0.153		-0.524	
partners_lifetime	0.054		-1.382		1.328	-0.195		-1.140		1.334	0.503		-0.016		-0.487	0.618		0.041		-0.659	*
partners_12months	-0.618		0.888		-0.270	-0.919		1.071		-0.152	0.113		-0.024		-0.089	0.152	*	-0.027		-0.125	
Concurrency	-0.066	*	0.060		0.006	-0.048		0.082	*	-0.034	0.136	***	-0.128	***	-0.008	0.114	***	-0.108	**	-0.005	
same_gender_sex	-0.056		0.014		0.042	-0.045		0.014		0.031	0.024		-0.043	*	0.019	0.018		-0.043		0.025	
condom_casual	-0.177	**	-0.118		0.295 ***	-0.117		-0.158		0.275 **	-0.121		-0.067		0.188	-0.064		-0.111		0.175	
paid_sex	-0.029		0.093		-0.065	-0.062		0.120		-0.058	0.000		0.000		0.000	.		.		.	
transaction_sex	-0.088	*	0.018		0.070	-0.115	**	0.031		0.084	0.022		0.006		-0.028	0.026		0.005		-0.031	
sex_alcohol_drugs	0.037		0.015		-0.052	0.077		-0.015		-0.062	0.060		nd		nd	0.023		nd		nd	

1: Unadjusted difference is between simple means. Adjusted difference adjusts for covariates.

2: AA-Pr = Difference between AA Papel and Presencial. AC-AA = Difference between ACASI and AA Papel. Pr-AC = Difference between Presencial and ACASI.

3: level of significance p≤0.1 (*), p≤0.05 (**), and p≤0.01 (***).

4: na = data not available because of a programming error in Audio-CASI instrument.

Table A.5 Effect on risk indicators when interviewer gender changes from same sex as respondent to opposite sex

	Males				Females			
Variable	Total 3 modes	FTFI	SAI	ACASI	Total 3 modes	FTFI	SAI	ACASI
ever_regular_smoker	0.017	-0.025	0.073	0.015	0.001	-0.012	0.016	0.001
current_smoker	0.008	-0.051	0.053	0.035	-0.010	-0.014	-0.011	-0.006
binge_drinking	0.077	0.023	0.355 ***	-0.089	-0.022	-0.036	-0.033	0.000
Drug consumption	0.014	-0.068	0.084	0.055	0.033	0.036	0.038	0.027
marihuana	-0.016	-0.005	0.061	-0.073	0.001	0.006	0.011	-0.011
Gang affiliation	-0.032	0.004	0.057	-0.128 *	0.003	0.014	-0.050	0.030
Fight	-0.046	-0.136 *	0.016	0.003	0.007	0.060	-0.039	-0.009
Victim	0.008	0.012	0.125 **	-0.095	0.013	0.044	-0.033	0.025
sexual_identity	0.024	0.019	0.007	0.044	0.005	0.001	0.001	0.012
STD	-0.011	-0.016	-0.034	0.011	-0.014	-0.036	0.029	-0.029
plans_pregnancy	.	.	.	.	0.053	-0.005	0.096	0.072
ever_sex	0.041	0.009	0.215 **	-0.071	-0.006	-0.038	-0.061	0.077
age_first_sex	-0.201	0.965 *	-1.327 *	-0.816	0.181	0.034	0.121	0.471
partners_lifetime	-0.401	-3.934	1.855	2.849	-0.209	-0.187	0.365	-0.953
partners_12months	-0.995	-2.098	1.466	-1.934	0.045	-0.045	0.173	0.022
Concurrency	-0.058 *	-0.039	0.012	-0.161 **	0.013	-0.019	0.079	-0.011
same_gender_sex	0.115 ***	0.154 **	0.121 *	0.044	0.010	0.033	0.048	-0.072 *
condom_casual	-0.037	-0.076	-0.210	0.218	-0.051	0.281	-0.129	-0.203
paid_sex	0.053	0.097	-0.011	0.062	.	.	.	.
transaction_sex	0.011	0.050	0.009	-0.051	0.001	-0.011	0.010	0.007
sex_alcohol_drugs	-0.024	-0.116	0.063	0.024	-0.027	0.001	-0.060	nd

1: effects from fixed-effects model.
2: level of significance p≤0.1 (*), p≤0.05 (**), and p≤0.01 (***).

Table A.6 Optimun Investment Levels

	Total Cost ($)				Sample size				Bias + 1,96 x estándar error				Cost·error			
Bias	FTFI	CATI	SAI	ACASI	FTFI	CATI	SAI	ACASI	FTFI	CATI	SAI	ACASI	FTFI	CATI	SAI	ACASI
Prevalencia = 0,05																
0,00	105.213,46	104.831,01	91.834,18	111.755,72	580	2.560	590	700	0,018	0,008	0,018	0,016	1.866,17	885,04	1.615,00	1.804,33
0,01	81.614,35	73.271,88	70.823,92	85.405,51	320	1.020	320	370	0,034	0,023	0,034	0,032	2.765,03	1.712,73	2.399,46	2.750,67
0,02	74.353,09	67.328,93	64.598,66	77.420,60	240	730	240	270	0,048	0,036	0,048	0,046	3.537,22	2.411,05	3.073,17	3.561,07
0,05	66.184,17	61.181,05	57.595,24	69.435,69	150	430	150	170	0,085	0,071	0,085	0,083	5.617,57	4.319,36	4.888,56	5.746,64
0,10	61.645,88	58.312,04	53.704,45	65.443,24	100	290	100	120	0,143	0,125	0,143	0,139	8.797,88	7.293,90	7.664,51	9.096,26
0,20	58.922,90	56.262,74	51.369,97	62.249,27	70	190	70	80	0,251	0,231	0,251	0,248	14.792,94	12.996,12	12.896,73	15.422,78
0,30	57.107,59	55.238,10	49.813,66	60.652,29	50	140	50	60	0,360	0,336	0,360	0,355	20.582,15	18.565,63	17.953,35	21.540,46
Prevalencia = 0,20																
0,00	105.213,46	104.831,01	91.834,18	111.755,72	580	2.560	590	700	0,033	0,015	0,032	0,030	3.425,04	1.624,34	2.964,06	3.311,53
0,01	87.967,96	79.624,70	76.271,02	91.793,44	390	1.330	390	450	0,050	0,031	0,050	0,047	4.371,89	2.507,96	3.790,57	4.310,38
0,02	80.706,70	72.452,17	70.045,76	84.607,02	310	980	310	360	0,065	0,045	0,065	0,061	5.207,79	3.263,50	4.519,87	5.188,07
0,05	70.722,46	64.869,78	62.264,18	75.025,13	200	610	210	240	0,105	0,082	0,104	0,101	7.456,70	5.302,63	6.481,71	7.547,98
0,10	65.276,51	60.771,19	57.595,24	68.637,20	140	410	150	160	0,166	0,139	0,164	0,162	10.852,80	8.430,08	9.446,32	11.117,82
0,20	61.645,88	57.902,18	53.704,45	64.644,74	100	270	100	110	0,278	0,248	0,278	0,275	17.162,12	14.343,06	14.951,24	17.761,15
0,30	59.830,56	56.672,60	52.148,13	63.047,76	80	210	80	90	0,388	0,354	0,388	0,383	23.193,45	20.067,78	20.215,34	24.124,55
Prevalencia = 0,50																
0,00	105.213,46	104.831,01	91.834,18	111.755,72	580	2.560	590	700	0,041	0,019	0,040	0,037	4.281,30	2.030,43	3.705,07	4.139,41
0,01	89.783,28	81.878,92	78.605,50	94.188,91	410	1.440	420	480	0,058	0,036	0,058	0,055	5.243,15	2.933,30	4.544,83	5.154,95
0,02	82.522,01	74.706,39	72.380,23	87.002,50	330	1.090	340	390	0,074	0,050	0,073	0,070	6.102,19	3.711,62	5.294,40	6.057,40
0,05	73.445,43	66.509,21	63.820,50	76.622,11	230	690	230	260	0,115	0,087	0,115	0,111	8.418,18	5.806,74	7.314,98	8.487,88
0,10	67.091,82	62.000,77	58.373,39	71.032,67	160	470	160	190	0,177	0,145	0,177	0,171	11.907,08	9.002,71	10.359,78	12.153,36
0,20	62.553,53	58.721,90	54.482,60	66.241,73	110	310	110	130	0,293	0,256	0,293	0,286	18.355,56	15.012,80	15.987,25	18.941,83
0,30	60.738,22	57.287,39	52.926,29	63.846,25	90	240	90	100	0,403	0,363	0,403	0,398	24.495,67	20.810,08	21.345,13	25.410,69