

Depends Who's Asking: Interviewer Effects in Demographic and Health Surveys Abortion Data

Tiziana Leone, Laura Sochas, and Ernestina Coast

ABSTRACT Responses to survey questions about abortion are affected by a wide range of factors, including stigma, fear, and cultural norms. However, we know little about how interviewers might affect responses to survey questions on abortion. The aim of this study is to assess how interviewers affect the probability of women reporting abortions in nationally representative household surveys: Demographic and Health Surveys (DHS). We use cross-classified random intercepts at the level of the interviewer and the sampling cluster in a Bayesian framework to analyze the impact of interviewers on the probability of reporting abortions in 22 DHS conducted worldwide. Household surveys are the only available data we can use to study the determinants and pathways of abortion in detail and in a representative manner. Our analyses are motivated by improving our understanding of the reliability of these data. Results show an interviewer effect accounting for between 0.2% and 50% of the variance in the odds of a woman reporting ever having had an abortion, after women's demographic characteristics are controlled for. In contrast, sampling cluster effects are much lower in magnitude. Our findings suggest the need for additional effort in assessing the causes of abortion underreporting in household surveys, including interviewers' skills and characteristics. This study also has important implications for improving the collection of other sensitive demographic data (e.g., gender-based violence and sexual health). Data quality of responses to sensitive questions could be improved with more attention to interviewers—their recruitment, training, and characteristics. Future analyses will need to account for the role of interviewer to more fully understand possible data biases.

KEYWORDS Abortion • Interviewer • Cross-classified • DHS

Background

Abortion data, particularly in low- and middle-income countries (LMICs), are of variable quality and coverage (Chae et al. 2017; Sedgh and Keogh 2019; Sedgh et al. 2012) and often contain little information about the characteristics of women who have induced abortions. Induced abortion is consistently underreported in nationally representative surveys, including in contexts where abortion is legally available (Jones and Forrest 1992; Jones and Kost 2007). Nationally representative household surveys have been

ELECTRONIC SUPPLEMENTARY MATERIAL The online version of this article (<https://doi.org/10.1215/00703370-8937468>) contains supplementary material.

used to describe the characteristics of women who have abortions, but abortion self-reports are unreliable. This results in the use of alternative, indirect estimation methods (abortion incidence complications method, best friend approach, confidant method, list experiment) (Sedgh and Keogh 2019), but these are not always based on nationally representative data or able to provide direct and reliable information on the individual- and household-level characteristics of women having abortions (Jones and Kost 2007; Sedgh et al. 2012). Although self-reported abortion data are known to be unreliable, they are at present our main source of nationally representative data on the characteristics of individuals who have sought abortion in Africa, Latin America, and Eastern Europe.

Survey item response on any topic is affected by a range of factors, including mode of interview (e.g., face-to-face interview vs. online interface), question wording and order, language used, presence of others during the interview, and interviewer's attitudes. Survey questions might elicit either a nonresponse or a response that may or may not be valid or reliable. Survey questions on sensitive topics are likely to elicit higher nonresponse rates or larger measurement error than nonsensitive questions (Tourangeau and Yan 2007). There is no standardized definition of a sensitive question, and Tourangeau et al. (2000) suggested that the concept of a sensitive question has three distinct meanings in the survey literature: intrusiveness, the threat of disclosure, and social undesirability. Typical sensitive topics, depending on context, might include drug use, sexual behaviors, voting, and income. Survey questions on abortion—particularly those relating to personal experience or behaviors (as opposed to opinions)—are considered sensitive in nearly every context (Tourangeau et al. 2000). Reluctance to respond may be due to either the stigma of reporting an abortion or the fear of repercussions for the respondent or someone close to them (Tourangeau and Yan 2007), particularly in settings where abortion is highly restricted or stigmatized. The *interviewer effect* refers to the influence that an interviewer's skills, beliefs, or personal characteristics have on respondents' responses to survey questions. Sensitive topics, including but not limited to questions on abortion, are likely to be particularly prone to interviewer effects.

Although the literature on stigma and abortion is growing (Levandowski et al. 2012; Lindberg and Scott 2018), we know very little about interviewer effects in the context of nationally representative household survey data in LMICs and nothing about the impact of interviewer effects on abortion survey data in LMICs. Current global evidence around abortion questions concentrates on interview mode, finding that it has a considerable impact on response quality. Face-to-face interviews (FTF) are more prone to underreporting of sensitive issues (Mensch et al. 2008). Evidence on abortion-related data collection from both high-income countries (HICs) and LMICs shows drawbacks and benefits for a wide range of methods (Rossier 2003). Self-administered survey modes, such as assisted self-interviewing (ACASI), are usually more reliable in reporting confidential information (Jones and Forrest 1992; Lindberg and Scott 2018; Rossier 2003; Sedgh and Keogh 2019; Sedgh et al. 2012). However, they are less suitable when respondents have no education or a low level of education. Although ACASI might make people feel more at ease in answering questions on abortion, thereby lowering the chance of underreporting (Hewitt 2002; Mensch et al. 2008; Sedgh and Keogh 2019), there is limited and context-specific evidence on the impact of ACASI with respect to abortion. The limited evidence shows that the interviewer effect remains whether the questionnaire is filled out by the interviewer or by the interviewee and that it is the fear of disclosure that matters (Tourangeau and

Yan 2007). Research from Malawi showed mixed results when comparing FTF with ACASI (Mensch et al. 2008). Particularly in low-resource settings with low levels of education, FTF surveys remain a crucial survey mode, meaning that the potential impact of the interviewer on responses needs to be better understood.

Qualitative data have shown that interviewers can influence responses in many ways, including making decisions about who meets inclusion criteria for the survey, their power of persuasion in eliciting a response, and their relative socioeconomic position compared with the respondent (Biruk 2018; Randall et al. 2013). Quantitative evidence on interviewer effects relates to survey questions other than abortion. Weinreb and Sana (2009) used the 1998 Kenya DHS to analyze the effect of the interviewer's translation of the questionnaire, included random effects for the interviewer and district, and showed a clear interviewer effect in relation to questions on HIV and pregnancy. However, a recent study showed very little interviewer impact on nonresponse related to contraceptive use questions in the Philippines and Indonesia (Amos 2018). Quantitative studies have described interviewer effects in both HICs and LMICs (Becker et al. 1995; Bignami-Van Assche et al. 2003; Couper and Groves 1992; Flores-Macias and Lawson 2008). A study from the United States showed that survey respondents interviewed by more experienced interviewers were more likely to agree or strongly agree with attitude questions, regardless of the question (Olson and Bilgen 2011). In a study in Kenya and Malawi, Bignami-Van Assche et al. (2003:60) concluded that questionnaire translation is less important than "the selection, training, and supervision of interviewers," underlining the importance of interviewers for collecting high-quality household survey data.

Interviewer effects are often deemed to be due to interviewers' different propensity for reproducing the influence of community-level stigma and cultural norms within the interview interaction (Couper and Groves 1992; Hox and De Leeuw 2002; Randall et al. 2013). Community effects are substantively but also methodologically important. Because interviewers operate within a given geographical area, including interviewer effects on their own could be picking up community effects if these were not separately included. Community factors must be accounted for to ensure that the interviewer effect is not simply a confounding factor for community effects. In addition, community effects may vary because of geographic differences in both underreporting and the true incidence of abortion.

There is little evidence on how community effects impact responses to sensitive questions. Studies on access to maternal and child healthcare that included community random effects or community-level variables showed that community effects are important (Gabrysch and Campbell 2009; Mahmud Khan et al. 2005). Most studies have used survey sampling clusters as a proxy for community effects and often use the terms "cluster" and "community" interchangeably (Koenig et al. 2003). In this article, we use "sampling cluster" to refer to our methods and describe results, and we use "community" to situate our findings against the literature and interpret results substantively.

Community effects partly reflect the clustering of norms around service use or around who makes decisions about healthcare access in the household. In addition, the community level is important to understand the availability, accessibility, and quality of services and health workers (Gabrysch and Campbell 2009; Mahmud Khan et al. 2005; Stephenson et al. 2006). In the case of abortion, the availability of services

can differ widely within a given country, and access to abortion depends partially on health workers' attitudes toward abortion (Haaland et al. 2020; Nandagiri 2019). Understanding the relative impact of community effects and interviewer effects on abortion reporting allows us to identify the extent to which quality issues with survey reporting of abortion experiences are explained by factors beyond the survey's control, such as community-level stigma, and by factors that can be addressed through improvements in methodological quality.

Separation of community effects from interviewer effects can also help to explain whether reporting of abortion is related to issues with survey implementation or to actual differences in abortion rates, given that we would expect there to be a high level of variance in abortion rates between communities. Accounting for both community and interviewer effects can inform decisions about the value of including questions about abortion in household surveys as well as the measures needed to improve the quality of these data.

The aim of our study is to test the impact of interviewer effects on the probability of reporting having ever had an abortion in a DHS. This study also contributes to research aiming to improve data collection on sensitive issues (e.g., domestic violence, sexual practices) within a nationally representative survey. To our knowledge, no other study to date has included both the interviewer and community effects in the analysis of responses to a sensitive question in a nationally representative survey.

The DHS is a key nationally representative and internationally comparable household survey for LMICs that includes abortion questions in several countries. In the DHS, individual interviewers work as part of a team, both for training and fieldwork. They report to a team leader and usually attend the same training at national or regional level that follows standard guidelines (ICF 2020; ICF Macro 2009).

With an estimated 25 million unsafe abortions taking place annually, causing 4.7% to 13.2% of maternal deaths, accurate abortion estimates are vital (Ganatra et al. 2017). Abortion rates are often estimated indirectly. The widely validated abortion incidence complications method (AICM) relies on collecting data from women receiving post-abortion care, including health facility data, to indirectly estimate the incidence of abortion in countries where abortion is restricted or highly stigmatized (Sedgh and Keogh 2019). These are the most reliable estimates, and they bear little resemblance to rates calculated from DHS data. For example, Malawi had an indirectly estimated abortion ratio of 22 per 100 live births in 2015, compared with 0.6 per 100 live births calculated from DHS data collected in the same year. Pakistan's AICM study in 2012 estimated 17.5–21.6 abortions per 100 live births, compared with a DHS estimate of 1.7 per 100 live births. In Brazil, the AICM estimated 26.7–44.4 abortions per 100 live births in 1991 versus 2.5 per 100 from DHS data in the same year (Singh and Wulf 1994; Singh et al. 2017). Although the AICM is an effective method for estimating abortion rates, it is able to estimate only national or regional abortion prevalence and is silent on individual-, household-, or community-level determinants of seeking an abortion. Other approaches that use nationally representative community-based surveys (e.g., best friend approach, confidante method) rely on respondents' reporting of others, not themselves (Sedgh and Keogh 2019). Indirect data can be used to collect demographic- and abortion-related characteristics information, but it also presents an inherent risk of bias as well as a more limited scope

to collect a wide range of characteristics. Therefore, understanding how to improve the collection of abortion survey data that includes individual- and household-level characteristics is particularly important.

DHS evidence on abortion has often been disregarded because of its low quality, leaving a persistent gap in abortion information from LMICs (MacQuarrie et al. 2018; Polis et al. 2017; Sedgh and Keogh 2019) and resulting in a wealth of unused data, with only a few studies attempting to analyze them (Bradley et al. 2019; Chae et al. 2017) or using them to benchmark a minimum abortion rate (Sedgh et al. 2016). The low quality has been attributed to abortion being legally restricted in many settings, combined with high levels of abortion-related stigma, leading to low levels of disclosure about ever having had an abortion. The literature has suggested modifications of the abortion questions, such as using the confidant method (asking the woman to report about a friend); asking about miscarriages and stillbirths separately as well as about total terminations in order to indirectly estimate induced abortions; or using the list experiment approach, with abortions included within a wider list of events that a woman might have experienced (Rossier 2003; Sedgh and Keogh 2019; Sedgh et al. 2012).

The DHS program has little metadata available on the level and quality of interviewer training. A review of available evidence from the DHS along with personal communications with experts at the DHS and the Guttmacher Institute demonstrates a general lack of information about the content and quality of training given to interviewers on abortion data collection beyond what is reported in the general DHS trainers' manuals. The DHS manuals include in-depth detail only on the mechanics of data collection, mainly related to calendar data, and do not focus on the social aspects of interviewing. We also know that that one day of two to three weeks of training is used to train interviewers on collecting birth histories, which include pregnancy terminations (ICF 2020; ICF Macro 2009).

Starting from the premise that DHS data on abortion are of poor quality, we hypothesize that the interviewer influences the quality of responses. The ultimate objective is to establish whether in addition to the efforts being made to modify sensitive questions, more effort should be invested into understanding and mitigating interviewer effects.

Context

Our study sample comprises all DHS surveys that included a question on either ever having had an induced abortion (online appendix, Table A1) or the number of lifetime induced abortions. See Table 1 for the reproductive health indicators for these 22 countries included in this study. Selected countries have a range of legal frameworks in place at the time of the DHS survey: legal only to save a woman's life (Haiti, Malawi, Congo, Côte d'Ivoire, and Madagascar); legal to save a woman's life and in cases of rape (Mali); and legal to save a woman's life, in cases of rape, to preserve the woman's physical and/or mental health (Cameroon), or in cases of fetal impairment (Colombia, Ghana, and Gabon). Less restrictive settings also include provisions for socioeconomic issues (India), whereas other countries have legalized abortion on request (Albania, Azerbaijan, Armenia, Cambodia, Moldova, Kyrgyz Republic, Kazakhstan, Tajikistan, Turkey, Ukraine, and Vietnam) (Centre for Reproductive Rights 2020; U.N. Population Division 2014).

Table 1 Reproductive health indicators for the countries included in this study

Country	Year of DHS Survey	MMR ^a (survey year)	% Unmet Need for Contraception ^b	Grounds on Which Abortion Permitted ^c
SSA				
Cameroon	2011	782 (2011)	18	3, r
Congo	2011–2012	426 (2012)	17.9	1
Côte d'Ivoire	2011–2012	614 (2012)	30.5	1
Gabon	2012	316 (2012)	27.9	1
Ghana	2014	450 (2007)	26.3	3, r, f
Madagascar	1997	498 (1997)	19	1
Mali	2012–2013	465 (2013)	17.2	1, r
Malawi	2015–2016	570 (2014)	19.4	1
Europe				
Albania	2008–2009	20.9 (2009)	13	5, r, f
Armenia	2015	19 (2012)	13	5, r, f
Azerbaijan	2006	28.9 (2006)	15	5, r, f
Kazakhstan	1999	50 (2001)	12	5, r, f
Kyrgyz Republic	2012	49 (2012)	18	5, r, f
Moldova	2005	45 (2010)	10	5, r, f
Tajikistan	2012	33 (2012)	23	5, r, f
Turkey	2003	29 (2005)	6	5, r, f
Ukraine	2007	13.1 (2004)	5	5, r, f
Asia				
Cambodia	2014	170 (2014)	13	5, r, f
India	1998–1999	549 (1998)	13	4, r, f
Vietnam	2002	170 (2001)	6	5, r, f
Latin America				
Colombia	2015	64 (2015)	7	3, r, f
Haiti	2016–2017	359 (2017)	38	1

^a MMR national estimates are from the World Bank Database (<https://data.worldbank.org/indicator/SH.STA.MMRT.NE>).

^b Data are from the DHS.

^c Data are from the U.N. Population Division Policy Database (https://esa.un.org/PopPolicy/about_data-base.aspx) and are coded as follows: 1 = to save woman's life only; 3 = to save woman's life or to preserve her mental/physical health; 4 = on socioeconomic grounds; 5 = on request; r = because of rape; or f = because of fetal impairment.

Data and Methods

To select the countries for analysis, we first review all DHS data sets that have ever asked any questions on induced abortion ($n=101$).¹ Of these, we analyze the most recent survey for those that included a question about ever having had an abortion and/or the lifetime number of abortions, had data available, and had no comparability issues with the data set (e.g., data for Liberia were collected among only young women, and India's latest NFHS asked about abortions in the last five years only, and thus the data were not comparable). We also exclude countries that asked about abor-

¹ We started from the DHS user forum discussing abortion questions, where a complete list had been posted (<https://userforum.dhsprogram.com/index.php?t=msg&goto=17456&S=Google>).

tions only in the last five years because this is not comparable with questions collecting abortion data over the lifetime and was asked by only two countries (Bangladesh and Indonesia). We include 22 countries: 8 from sub-Saharan Africa (SSA), 9 from Europe, 3 from Asia, and 2 from Latin America (LA).²

Our review reveals a high level of heterogeneity in abortion-related question wording and sequencing within DHS surveys, highlighting the challenge of cross-country comparative analysis. In contrast to the homogeneity of typical DHS data (e.g., V201 is widely recognized as the code for the parity variable), we also find substantial diversity in DHS abortion question codes.

The surveys included in this study used four key sets of questions with various phrasing (Table 2 and the online appendix, Table A1). A group of countries asked a question about any termination (i.e., miscarriage, stillbirth, or induced abortion) followed by a question about number of induced abortions (Cameroon, Congo, and Malawi). A second group asked about any termination, and then asked about any induced abortion (Côte d'Ivoire, Gabon, Mali, and Vietnam). A third group asked about any induced abortion, and then asked about the number of induced abortions (Armenia, Azerbaijan, Cambodia, Kazakhstan, Kyrgyz Republic, Moldova, Tajikistan, Turkey, and Ukraine). A fourth group asked about any termination followed by any induced abortion, and then asked about the number of induced abortions (Cambodia, Haiti, and India). Ghana is the only country that asked about the number of induced abortions directly, and then recorded the timing of abortions using the reproductive calendar. Finally, one country (Madagascar) asked only about ever having an abortion. In each of the 22 surveys, we used multilevel multivariable logistic (or ordinary least squares [OLS]) regressions in a Bayesian framework to test the impact of the interviewer on women's responses to three questions: ever having had a termination; ever having an induced abortion; and, conditional on ever having had an induced abortion, the lifetime number of induced abortions. We model each country separately because a pooled analysis would have been computationally intractable.

We test the interviewer effect across different questions within the same survey to understand whether the interviewer effect tends to be stronger for more sensitive questions. For example, one would expect that the question about any termination would be less sensitive than the induced abortion question. However, when a woman answered the question about any induced abortion or provided a nonzero response to the number of abortions over her lifetime, it is unclear whether the number of abortions is an reliable measure. Given the diverse formulation of questions, we cannot produce a definitive comparison between different question sequences (e.g., by asking directly about abortion vs. by asking about all terminations first). We do, however, suggest possible reasons for the results we obtain.

We test the interviewer effect using (primarily) logistic models, estimated according to Eq. (1). Here, y is the outcome of interest for the i th respondent, interviewed by the j th interviewer in the k th sampling cluster. θ is a vector of individual-level demographic characteristics, and ζ represents the crossed-classified interviewer and cluster-level random intercepts. The random intercepts are assumed to be normally distributed around mean β_0 according to, respectively, variance σ_j and σ_k (conditional on covariates) and are assumed to be independent from each other and across interviewers and clusters.

² For a complete list, see Table A1 in the online appendix.

$$\text{logit}\left(\Pr(y_{ijk} = 1 | \theta_i, \zeta_j, \zeta_k)\right) = \beta_0 + \lambda\theta_i + \zeta_j + \zeta_k. \quad (1)$$

An equivalent equation is estimated using OLS for the continuous outcome of number of abortions. OLS models are preferred over Poisson models for the continuous outcome because it is not possible to estimate the intraclass correlation coefficient (ICC) for Poisson models in a comparable way with OLS models (Rabe-Hesketh and Skrondal 2008). Regardless, point estimates of the variance of interviewer and cluster-level random effects in a sensitivity analysis employing Poisson models are within credible intervals of the OLS models³ (see the online appendix).

Because of the binomial logistic crossed-classified random-effects specification, frequentist estimation of the variance components using Laplacian approximation with one integration point is highly biased (Rabe-Hesketh and Skrondal 2008). We therefore use Markov chain Monte Carlo estimation in a Bayesian framework—specifically, a Gibbs sampler (JAGS, implemented using the R package *rjags*)—to generate our results. We use noninformative priors, 5,000 iteration burn-in and 5,000 saved posterior samples for the logistic models. The OLS models are run for 80,000 samples. Chains with different starting points give similar results, and the Raftery-Lewis diagnostic indicates an appropriate number of burn-in and saved samples to obtain results with a 0.05 margin of error and 95% accuracy (Lunn et al. 2012).

In total, we analyze data from 344,623 women aged 15–49. The data sets report information on sampling clusters/primary sampling units (used as a proxy for community) and interviewers, but they do not usually report on interviewers' characteristics. The inclusion of interviewers' characteristics is an innovation included only toward the end of the most recent DHS phase (Phase 7, 2015–2018).⁴ A description of the sampling clusters and number of interviewers is reported in Table 2.

The multilevel approach has cross-classified random intercepts at the level of the sampling cluster and the level of the interviewer. In this way, we can simultaneously consider the amount of variance in the outcome (e.g., ever had an abortion) associated with different interviewers and the amount of variance associated with different communities, while controlling for the respondent's demographic characteristics. Cross-classified random effects are used because interviewers and clusters are not nested within one another. Assuming that interviewers were equally likely to be assigned to respondents with a low or high probability of reporting an abortion, the variance in the average odds of reporting an abortion by interviewer should be close to zero if all respondents reported a valid response or if all interviewers had a similar capacity to elicit valid responses. Under the same assumption, a large variance indicates that interviewers are differentially likely to elicit valid responses. The ICC calculates the share of the total variance made up by the different components—for example, the share of the variance accounted for by the interviewer variance relative to the sum of the respondent-level variance and the cluster-level variance, conditional on covariates. The ICC allows us to compare the magnitude of the interviewers' random intercept variance across questions

³ *Credible interval* is a term used in Bayesian inference to define an interval within which a parameter falls with a certain probability, conditional on the data and model (http://mc-stan.org/rstanarm/reference/posterior_interval.stanreg.html). Credible intervals are reported in Tables 2 and 3, and in Table A2 of the online appendix. The equivalent in frequentist statistics is a confidence interval, but although they are both used for statistical inference, the interpretation differs.

⁴ This information was obtained from an informal conversation with DHS staff in December 2018.

and settings. In the Results section, we refer to the ICC for the interviewer random effect as simply the “interviewer effect.” In logistic models, the respondent-level variance is fixed at 3.29, whereas it is directly estimated in the OLS models.

The inclusion of a cross-classified random intercept at the cluster level allows us to control for the fact that some interviewers might be assigned to communities with a different probability of reporting an abortion (Vassallo et al. 2017). Community effects could capture factors such as culture, religion, and stigma. We also control for region in the main effects such that the cluster-level variance is estimated conditional on regional factors. Previous research in Bangladesh has shown the importance of accounting for location when looking at sensitive questions in survey data (Koenig et al. 2003). The sampling cluster is also an indication of where abortion-related services might be available and of the extent to which the community knows about them.

To implement our method, we must be able to accurately distinguish between cluster-level variance and interviewer-level variance, which requires sufficient interpenetration of the two levels. Vassallo et al. (2017) suggested that having three areas per interviewer provides sufficient interpenetration. A description of the sampling clusters is reported in Table 2. In all countries, the majority of interviewers worked in three clusters or more, providing a robust level of interpenetration.

In addition to testing for cluster and interviewer effects, we control for respondents’ characteristics to control for the possibility that some interviewers may have been systematically assigned to women more (or less) likely to report an abortion. We would expect younger, nonmarried, poorer, and less-educated women to be more reticent in answering abortion questions because abortion stigma is more evident across groups of women of low socioeconomic status (Jones and Forrest 1992; Lindberg and Scott 2018; Tourangeau and Yan 2007). However, previous studies have tested only the likelihood of responding given women’s characteristics; less clear is how women’s characteristics have an impact on the outcome once interviewer effects are accounted for. Amos (2018) modeled nonresponse to a question about reasons for not using contraception by including interviewer random effects and respondent-level covariates. Although less-educated and poorer women were less likely to respond, the study did not find a strong interviewer effect (Amos 2018).

The respondent characteristics included in our study are age, marital status (Turkey, India, and Vietnam interviewed only married women), rural-urban residence, geographic region, education, and wealth. Wealth was calculated using principal component analysis of asset indicators at the household level weighted separately by rural and urban areas (Filmer and Pritchett 2001). All variables are included in the model to ensure that the models are comparable across countries.

Results

Individual Interviewers or Communities?

We estimate the models with interviewer and sampling cluster random effects with respect to three different questions (Table 3). In the first set of models, the outcome is “ever having had a termination” (i.e., any abortion, miscarriage, or stillbirth). In these models, the interviewer effect is stronger than the sampling cluster effect, with two exceptions: (1) Armenia and Vietnam, where the sampling cluster effect is stronger, and

Table 2 Interpenetration checks with DHS cluster structure

	Sample Size	Question Type ^a	% Reporting Ever Having Had an Abortion in the DHS (% missing) ^b	Interviewers per Cluster	Clusters per Interviewers	% of Interviewers Covering Three or More Clusters
SSA						
Cameroon 2011	15,002	1, 3	6.34 (0.32)	2–7	4–38	100
Congo 2011	10,626	1,3	20.01 (0)	2–5	1–35	98
Côte d'Ivoire 2012	9,278	1,2	7.00 (0.14)	2–6	1–23	94
Gabon 2012	8,218	1,3	19.04 (0.04)	1–4	1–24	88
Ghana 2017	25,062	1,3	14.77 (0)	2–4	28–36	100
Madagascar 1997	6,978	2	3.12 (0)	1–7	1–41	93
Malawi 2015	24,562	1,3	6.98 (0.84)	2–5	2–28	99
Mali 2012–2013	10,424	1,2	1.38 (0)	1–4	1–22	89
Europe						
Albania 2008–2009	7,398	1,2	3.77 (0)	1–3	1–43	99.53
Armenia 2015	6,116	1,2,3	25.87 (0)	2–7	1–31	94.95
Azerbaijan 2006	8,444	1,2,3	37.78 (0)	3–7	1–35	68.83
Kazakhstan 1999	4,800	1,3	40 (0)	2–7	1–61	82.82
Kyrgyz Republic 2012	8,208	1,2,3	17.31 (0)	3–5	1–38	77.8
Moldova 2005	7,318	1,3	37.43 (0)	2–5	26–30	100
Tajikistan 2012	9,496	1,3	9.65 (0)	1–5	10–29	100
Turkey 2003	8,075	1,2,3	21.68 (0)	1–10	1–95	94.25
Ukraine 2007	6,841	1,3	31.79 (0)	1–4	1–31	76.71
Asia						
Cambodia 2014	17,578	1,2,3	6.19 (0)	2–7	1–61	81.82
India 1998–1999	90,303	1,2,3	1.07 (0.01)	3–55	1–195	90
Vietnam 2002	5,665	1,2	6.76 (0)	1–6	2–49	97.69
Latin America						
Colombia 2015	38,718	1, 2	1.62 (0)	1–8	1–334	96.55
Haiti 2016–2017	15,513	1,2,3	11.3 (7.36)	2–4	2–30	98.18

^a 1 = any terminations, 2 = any abortions, and 3 = number of abortions.

^b Data are unweighted.

(2) Cameroon, where the variance is the same (3%). Mali has the strongest interviewer effect (26%), and Turkey has the lowest (0.8%). Countries in SSA have stronger and more similar levels of interviewer effects between countries (average 19.75%) compared with countries in Eastern Europe and Central Asia (average 9.1%), which are former communist countries with a history of less restrictive abortion laws.

In the second set of models, for the question “ever having had an abortion,”⁵ the interviewer effect is stronger than in the “ever having had a termination” models in the majority of countries. This suggests that the interviewer effect is stronger for

⁵ In some countries, this question is directly asked as number of abortions, but here the variable is recoded as 0 or 1 if the woman had one or more abortions (Congo, Cameroon, Gabon, Malawi, Cambodia, Kazakhstan, Moldova, Tajikistan, and Ukraine).

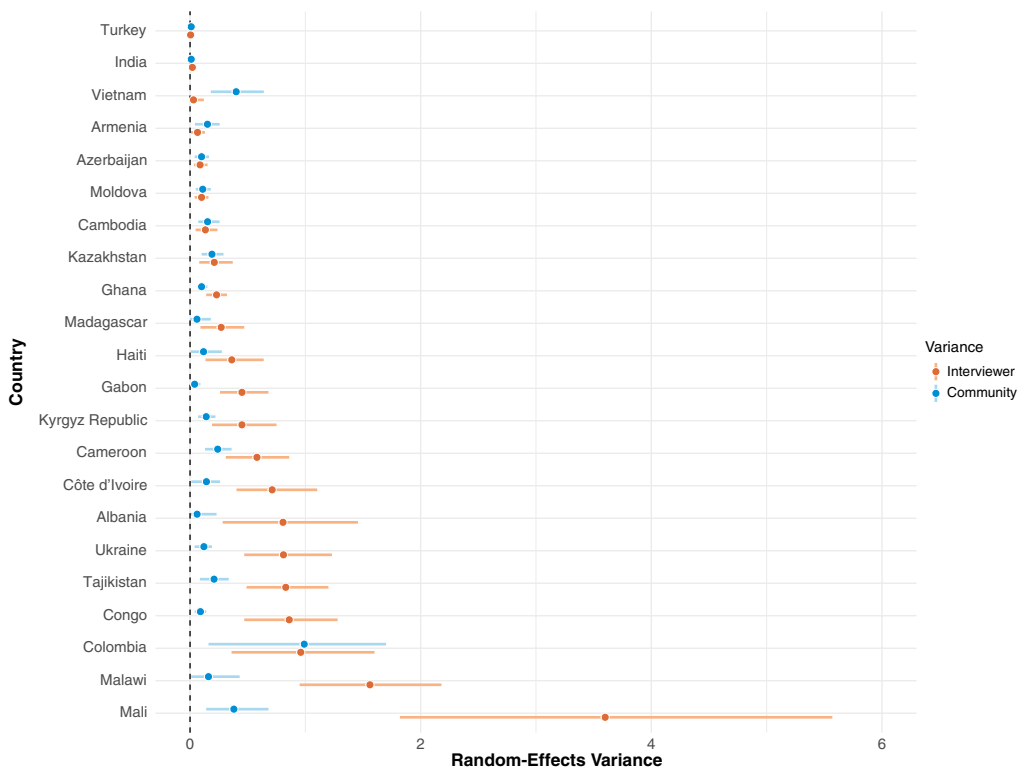


Fig. 1 Multilevel logistic model predicting “ever having had an abortion” for all countries, 1997–2015: Logit coefficients, with 95% credible intervals

more sensitive questions. Again, the interviewer effect is generally stronger than the sampling cluster effect; exceptions are for Armenia, Azerbaijan, Moldova, Turkey, Cambodia, Vietnam, and Colombia. The interviewer effect for “ever having had an abortion” ranges from 50% in Mali to 0.15% in Turkey (see Figure 1 and Table 3). On average, the interviewer effect is strongest in SSA (19%). However, the 95% credible intervals for the variance of the interviewer random effects are quite wide and are wider the larger the variance (Figure 1).

The last set of models analyzes the question on the number of abortions over a woman’s lifetime. For this question, which we recode as a number greater than 0 (and missing if equal to 0), the interviewer effect is much smaller. This aligns with our expectations given that reporting ever having had an abortion (the first question) is more sensitive and therefore more prone to interviewer effects (i.e., initial disclosure) compared with subsequent reporting of the number of abortions (although we would also expect this to be underreported). Across all countries, both the interviewer and sampling cluster effects for this question are very small, suggesting that there may be less variability across interviewers in the number of abortions a woman reports, conditional on her reporting an abortion. In this set of models, more countries reported a higher level of variance for the interviewer effect than for the sampling cluster effect. This could be due to the smaller size of the interviewer effect and/or to communities’

Table 3 Variance level for models at interviewer and sampling cluster level for both questions on ever had an abortion and number of abortions^a

Country	Model 1: Ever Having Had a Termination With All Control Variables						Model 2: Ever Having Had an Abortion With All Control Variables						Model 3: Number of Abortions With All Control Variables					
	Variance (cluster)		Variance (residual)		ICC (cluster) (%)		Variance (cluster)		Variance (residual)		ICC (cluster) (%)		Variance (cluster)		Variance (residual)		ICC (cluster) (%)	
	inter-viewer	viewer	inter-viewer	viewer	inter-viewer	viewer	inter-viewer	viewer	inter-viewer	viewer	inter-viewer	viewer	inter-viewer	viewer	inter-viewer	viewer	inter-viewer	viewer
SSA																		
Cameroon 2011	0.096	0.098	3.29	3.29	3	3	0.576	0.24	3.29	14	7	0	0	0	1.3	0	0	0
Congo 2011	0.451	0.058	3.29	3.29	12	2	0.862	0.087	3.29	20	2	0	0	1	0	0	0	0
Côte d'Ivoire 2012	0.441	0.125	3.29	3.29	11	4	0.712	0.142	3.29	17	4	0	0	0	1.5	0	0	0
Gabon 2012	0.453	0.025	3.29	3.29	12	1	0.449	0.039	3.29	12	1	0	0	0	0.8	0	0	0
Ghana 2017							0.225	0.101	3.29	6	3	0	0	0				
Madagascar 1997							0.274	0.06	3.29	8	2	0	0	0				
Malawi 2015	0.228	0.073	3.29	3.29	6	2	1.556	0.16	3.29	31	4	0	0	0.1	1.1	0	0	8
Mali 2012–2013	1.206	0.196	3.29	3.29	26	5	3.598	0.38	3.29	50	9							
Europe																		
Albania 2008–2009																		
DHS 2015	1.035	0.266	3.29	3.29	22.54	7.03	0.806	0.061	3.29	19.39	1.72							
Armenia 2015 DHS	0.038	0.081	3.29	3.29	1.11	2.39	0.065	0.151	3.29	1.85	4.36	0.1	0.012	2.171	4.38	0.54		
Azerbaijan 2006 DHS	0.117	0.099	3.29	3.29	3.34	2.89	0.087	0.1	3.29	2.50	2.93	0.141	0.042	2.597	5.07	1.56		
Kazakhstan 1999 DHS	0.093	0.074	3.29	3.29	2.69	2.18	0.211	0.191	3.29	5.72	5.40	0.091	0.101	4.738	1.85	2.08		

Table 3 (continued)

Country	Model 1: Ever Having Had a Termination With All Control Variables			Model 2: Ever Having Had an Abortion With All Control Variables			Model 3: Number of Abortions With All Control Variables								
	Vari- ance (inter- viewer) (cluster)	Vari- ance (resid- viewer) (cluster)	ICC (cluster) (%)	Vari- ance (inter- viewer) (cluster)	Vari- ance (resid- viewer) (cluster)	ICC (cluster) (%)	Vari- ance (inter- viewer) (cluster)	Vari- ance (resid- viewer) (cluster)	ICC (cluster) (%)						
Kyrgyz Republic	0.484	0.119	3.29	12.43	3.37	0.447	0.143	3.29	11.52	4.03	0.01	0.014	0.552	1.74	2.40
2012 DHS															
Moldova						0.096	0.11	3.29	2.75	3.21	0.011	0.027	2.132	0.51	1.25
2005 DHS															
Tajikistan	0.827	0.073	3.29	19.74	2.05	0.827	0.208	3.29	19.12	5.64	0.012	0.007	0.831	1.41	0.82
2012 DHS															
Turkey 2003	0.027	0.009	3.29	0.81	0.27	0.005	0.01	3.29	0.15	0.30	0.018	0.01	1.378	1.28	0.71
DHS															
Ukraine 2007						0.812	0.115	3.29	19.26	3.20	0.094	0.099	1.416	5.84	6.29
DHS															
Asia															
Cambodia	0.219	0.124	3.29	6.03	3.57	0.133	0.152	3.29	3.72	4.37	0.048	0.038	5.812	0.81	0.65
2014 DHS															
India	0.062	0.007	3.29	1.85	0.21	0.02	0.008	3.29	0.60	0.24	0.002	0.004	0.334	0.59	1.16
1998–1999															
DHS															
Vietnam 2002	0.065	0.133	3.29	1.86	3.86	0.032	0.402	3.29	0.86	10.86					
DHS															
Latin America															
Colombia	0.104	0.08	3.29	2.99	2.35	0.962	0.99	3.29	18.35	22.18					
2015 DHS															
Haiti	0.245	0.021	3.29	6.89	0.62	0.362	0.117	3.29	9.60	3.34	0.001	0	0.334	0.30	0.00
2016–2017															
DHS															

^a We control for women's age, residence, region, marital status, education, and wealth.

differential availability of abortion-related services, which is more likely to affect the number of abortions accessed.

Patterns According to Question Sequencing and Prevalence and Legal Status of Abortion

We consider possible patterns according to the sequencing of the questions (e.g., any terminations followed by any abortion versus number of abortions). We conclude that the interviewer effect tends to be stronger than the sampling cluster effect regardless of the sequencing of the questions. This finding holds even for countries such as Ghana, where a question about the number of abortions was asked directly after asking about any terminations.

In Ghana, where the data come from a special DHS maternal health survey with a more complex set of questions, the interviewer effect is lowest in Africa, and the percentage of women reporting having had an abortion is so high (14.8% in [Table 2](#)) that it would appear that underreporting could be lower. A recent study reported an AICM estimate of the abortion rate of 26.8 per 1,000 women aged 15–49 in comparison ([Polis et al. 2020](#)), perhaps because including abortion questions within a maternal health survey reduces stigma by signaling that abortion services are part of reproductive health-care. Alternatively, higher-quality interviewer training might have reduced the bias.

We find no discernible difference in the interviewer and sampling cluster effects across different levels of legal status of abortion at the time of the survey ([Table 1](#)). For example, across all the European/central Asian countries with the most liberal laws on abortion, there is a wide range of variance at both the interviewer (from 0.2% in Turkey to 19.4% in Albania) and the sampling cluster level (from 0.3% in Turkey to 5.64% in Tajikistan) within the models on the question, “Have you ever had an abortion?” This finding is in line with evidence that abortion stigma is present irrespective of legality ([Coast et al. 2018](#)).

Respondent Characteristics

We report the coefficients on respondents’ characteristics for ever having had an abortion in the appendix ([Table A2](#), online appendix), although we include these control variables for all three questions. We find a positive impact of age on the probability of reporting having an abortion, except in Colombia, Cambodia, Moldova, and Albania, where the relationship is negative. This finding is to be expected given that older women are more likely to feel confident in reporting an abortion and would also have been more exposed to the need for an abortion over their lifetime ([Jones and Forrest 1992](#); [Jones and Kost 2007](#)). For each country, where the coefficient for urban is positive and the Bayesian credible intervals around the coefficient do not include 0, our model estimates greater than a 97.5% probability that women in rural areas are less likely to report ever having had an abortion. Wealth shows a positive gradient, with wealthier women in 13 of our 22 study countries (Armenia, Kyrgyz Republic, Tajikistan, Ukraine, Cameroon, Madagascar, Ghana, India, Malawi, Congo, Kazakhstan, Moldova, and Haiti) being the most likely to report ever having had an abortion.

For the remaining countries, the credible intervals of the estimate include 0 for most of the coefficients. This pattern is to be expected given the substantial administrative, cognitive, and financial barriers to procuring an abortion, especially—but not only—in places where it is legally restricted (Sedgh et al. 2012).

In all SSA countries except for Malawi, having at least a primary education is associated with a higher probability of reporting ever having had an abortion. However, this relationship is less clear in other world regions. Marital status also shows mixed results, although ever having been in a union and currently being in a union are typically associated with a higher probability of reporting. Across all models (where available), having never been in a union shows a lower reporting of ever having had an abortion.

Overall, we find a higher probability of reporting ever having had an abortion among urban, wealthier, older, and more-educated women.

Discussion

To inform strategies, policies, and program to reduce unsafe abortion, it is critical to improve the availability of representative data on abortion incidence by subgroup, including abortion methods, sources, safety, and experiences. Similar data about contraceptive use have been crucial for programs aiming to reduce unmet need for contraception. However, household survey data on abortion have been lacking or underused because of concerns about quality and underreporting, and insufficient efforts have gone into assessing the quality of abortion survey data. The results of this study show a clear interviewer effect on responses to abortion questions. The effect is strongest for the question on ever having had an abortion relative to the question on the number of abortions. Interviewer effects are typically stronger than community effects, which proxy for potentially different community levels of stigma and context as well as service availability. The fact that community effects are weaker could also be attributed to the fact that conditional on regions and place of residence (rural vs. urban), which are controlled for in the main effects, communities are more homogeneous.

However, this study cannot estimate the magnitude of the interviewer effect in absolute terms, nor can it estimate how much better the data would be had the interviewer effect not existed. This study cannot determine whether a low interviewer ICC is evidence of the abortion question having high validity for those surveys because it is possible that all interviewers homogeneously yet negatively affected the validity of the response. However, we believe that this is unlikely, given evidence from our study and others showing that less sensitive questions have lower interviewer variance, and we interpret low variance across interviewers as a sign of high validity.

The findings further show that the magnitude of the interviewer effect is sensitive to the question asked, with questions on any termination and number of abortions showing a smaller interviewer effect. In a further analysis, we choose a random subsample of countries ($n=5$) from our sample of 22 and run the same model on less sensitive questions, such as current use of contraception and number of children. The results show a considerably lower variance at interviewer level and a more prominent effect at community level (results not shown here).

Previous studies looking at interviewer effects highlighted that a sense of being judged is a key barrier to answering sensitive questions (Durrant et al. 2010; Randall

et al. 2013). Gender, social status, and age are also key characteristics that are sensitive in context-specific ways and could lead to a bias in responses (Becker et al. 1995; Singer et al. 1983)

Running the models with a cross-classified random effect at the community level gives us a greater sense of the significance of the interviewer effect. To further test this, we first run separate models with random effects at the sampling cluster level only for the “ever having had an abortion” outcome to gauge the impact that geographically specific culture, stigma, and/or abortion services availability might have on the willingness to report an abortion. The community effect is much larger when interviewer effects are not included, which implies that including only sampling cluster-level random effects incorrectly picks up interviewer effects. The share of variance accounted for by the community varied from 32.4% (Mali) to 0.5% (Turkey) (results not shown). We cannot ascertain any relationship between the level of liberalization of the abortion law and the interviewer effect or between different question sequencing patterns (e.g., simply asking the number of abortions vs. three separate questions). This would need to be further analyzed in future research to attempt to dampen the influence of legal status of abortion on the likelihood of a valid response to abortion questions. However, we would expect that even in fairly liberal contexts, stigma and lack of knowledge on legislation would still be a barrier to a valid answer (Rossier 2003).

In accordance with previous literature that did not account for interviewer effects (Chae et al. 2017), our study shows that women from a poorer background, with a lower level of education, and from rural areas are less likely to report ever having had an abortion. This, in addition to the positive correlation with age, could be a combination of both being less likely to have accessed an abortion and being more afraid or ashamed to report one.

Notwithstanding the robustness of the analysis, this study has limitations. First, question wording and sequencing varied across surveys. We cannot test the impact that the phrasing, the order, and type (e.g., ever having had an induced abortion vs. asking directly the number) of the questions might have on the probability of responding to the question. We descriptively address this issue by looking at overall patterns of variance given a set of questions. We cannot assess whether larger interviewer effects are due to the context, the phrasing of questions, how sensitive the questions are in that context, or to the quality of interviewer training and supervision across countries, which is unobserved.

In addition, by considering the three different set of questions, we include all possible variations around the sequencing of the questioning, and we exclude those countries or questions that would not yield a comparable estimate of abortion incidence.

Conclusions

This study highlights the importance of providing better training and supervision to interviewers when collecting abortion data and more generally sensitive data in household surveys to improve data quality. It also demonstrates that despite an attempt at using standardized training tools, the impact of individual interviewers is credibly greater than that of communities after region and rural/urban residence are controlled for. Because abortion stigma and shame are locally (re-)produced, one of the key issues might yet be the standardization of questionnaires and training. How-

ever, we are not able to show differences in the scale of interviewer effects across different levels of legalization of abortion, nor can we tease out which wording or typology of questions might yield more validity. Ghana's abortion module, which is the most detailed of the group of countries analyzed, shows some improvement in the quality of the data, but the module is lengthy and costly to administer. The Ghana results might have been due to the questions being asked within a wider maternal health survey, which could have made the reporting of induced abortion less stigmatizing as well as possibly reflecting greater emphasis on training and selecting interviewers. A recent estimation of the abortion rate in Ghana made using AICM sets the national level at 26.8 abortions per 1,000 women versus 14.8 in the DHS (Polis et al. 2020). This estimate is possibly one of the DHS estimates that is closest to the real value, although it is still likely to be underestimated. Our design cannot assess the counterfactual: what the reporting of abortions would be without an interviewer effect.

Given widespread lack of trust in DHS abortion data, should abortion questions be excluded from interviewer-led household sample surveys, such as the DHS? Or should greater efforts be made to recruit and train interviewers? Our analyses cannot answer all these questions, but they point to the need for a more careful understanding of the value of asking sensitive questions in general and abortion more specifically, as well as the impact that improved questions and/or interviewer training might have on the quality of abortion data from household sample surveys. The DHS is always under pressure to include additional questions, both for existing and new topics (Kishor 2015); these pressures must be balanced against survey length and costs. Questions that yield low-quality data and/or are not fully exploited might become vulnerable to future exclusion. DHS data provide crucial evidence about abortion, and better understanding of the interaction between demographic and behavioral determinants will improve our understanding of abortion. Better understanding of the social interaction between interviewers and respondents could be key to improving abortion reporting as well as reporting on other sensitive questions.

The DHS remains a valuable source of information on reproductive histories. It is the most complete household survey capable of showing linkages between reproductive histories, socioeconomic status, and abortion. Removing abortion questions would present a partial understanding of sexual and reproductive health realities and could signal that abortion is not worthy of measurement and understanding. Countries have invested considerable resources running DHS surveys, and we suggest that additional attention would leverage these investments to generate better abortion data.

Interviewer effects have been neglected in the quest to improve abortion data, with most attention given to question wording and survey mode. Even where they have been acknowledged to have an impact (MacQuarrie et al. 2018), they have not been analyzed in depth as this study does. Studies of interviewer effects should be further extended by including interviewer characteristics in the analysis, across other types of surveys asking sensitive questions in both high- and low-and-middle-income countries. This would enable us to test whether interviewers' attitudes, demographics, and social status impact the validity of responses and would allow differentiation between the interviewer's demographic characteristics hypothesis and the interviewer's skills hypothesis. The DHS has recently included this information in its latest rounds, making this the right time to better understand interviewer effects.

We also need more qualitative information on how to improve survey responses and reduce potential interviewer effects. Given the high cost of collecting household survey data such as the DHS, there is an urgent need to further investigate the suitability of current modules. Cognitive interviewing could elicit better understanding of how women interpret and answer terms (e.g., abortion, miscarriage, stillbirth) and the suitability of existing question wording and sequencing. Much research has explored indirect techniques to gather abortion information from surveys, such as the Guttmacher Institute's work on ACASI (Lindberg and Scott 2018). Although no consensus has been reached on the best mode of interviewing, national surveys could provide an excellent testing ground. More localized efforts to test abortion questions led by initiatives such as the Performance Monitoring and Accountability 2020 project (PMA2020) could also inform the way forward for larger data collection exercises like the DHS (Bell et al. 2019). Interviewer effects should also be tested in the context of these alternative data collection exercises.

Interviewer effects on abortion survey data have not been previously identified and need to be included in quantitative studies as a further quality check. Although we do not identify patterns in relation to macro factors such as abortion legislation and type of question, our study indicates the need for more methodological work to identify such possible influences. Our findings suggest a substantial interviewer impact on the probability of reporting an abortion, highlighting the need for greater awareness of the impact of interviewers on data outcomes, in particular—but not only—when questions involve sensitive or stigmatized topics. If the interviewer effect holds for other sensitive questions, there is an opportunity for broader improvements in data quality from interviewer-administered household survey data. ■

Acknowledgments We would like to acknowledge the funding of the LSE Department of International Development's School Research Incentive Fund. We thank the incredibly useful feedback of the LSE Global Health Initiative reading group, and in particular of Dr. Rishita Nandagiri.

References

- Amos, M. (2018). Interviewer effects on patterns of nonresponse: Evaluating the impact on the reasons for contraceptive nonuse in the Indonesia and the Philippines DHS. *Demographic Research*, 39, 415–430. <https://doi.org/10.4054/DemRes.2018.39.14>
- Becker, S., Feyisetan, K., & Makinwa-Adebusoye, P. (1995). The effect of the sex of interviewers on the quality of data in a Nigerian family planning questionnaire. *Studies in Family Planning*, 26, 233–240.
- Bell, S., Shankar, M., Omoluabi, A. F., Khanna, A., Emmanuel, M. D. D., OlaOlorun, F., Ahmad, D., . . . Moreau, C. (2019, April). *Methodological advances in survey-based abortion incidence estimation: Promising findings from Nigeria, India and Côte d'Ivoire*. Presented at the annual meeting of the Population Association of America, Austin, TX.
- Bignami-Van Assche, S., Reniers, G., & Weinreb, A. A. (2003). An assessment of the KDICP and MDICP data quality: Interviewer effects, question reliability and sample attrition. *Demographic Research*, 11, 31–76. <https://doi.org/10.4054/DemRes.2003.S1.2>
- Biruk, C. (2018). *Cooking data: Culture and politics in an African research world*. Durham, NC: Duke University Press.
- Bradley, S. E. K., Polis, C. B., Bankole, A., & Croft, T. (2019). Global contraceptive failure rates: Who is most at risk? *Studies in Family Planning*, 50, 3–24.

- Centre for Reproductive Rights. (2020). *The world's abortion laws* [Database]. Retrieved from <https://reproductiverights.org/worldabortionlaws>
- Chae, S., Desai, S., Crowell, M., Sedgh, G., & Singh, S. (2017). Characteristics of women obtaining induced abortions in selected low- and middle-income countries. *PLoS ONE*, *12*(3), e0172976. <https://doi.org/10.1371/journal.pone.0172976>
- Coast, E., Norris, A. H., Moore, A. M., & Freeman, E. (2018). Trajectories of women's abortion-related care: A conceptual framework. *Social Science & Medicine*, *200*, 199–210.
- Couper, M. P., & Groves, R. M. (1992). The role of the interviewer in survey participation. *Survey Methodology*, *18*, 263–278.
- Durrant, G. B., Groves, R. M., Staetsky, L., & Steele, F. (2010). Effects of interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly*, *74*, 1–36.
- Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. *Demography*, *38*, 115–132.
- Flores-Macias, F., & Lawson, C. (2008). Effects of interviewer gender on survey responses: Findings from a household survey in Mexico. *International Journal of Public Opinion Research*, *20*, 100–110.
- Gabrysch, S., & Campbell, O. M. R. (2009). Still too far to walk: Literature review of the determinants of delivery service use. *BMC Pregnancy and Childbirth*, *9*, 34. <https://doi.org/10.1186/1471-2393-9-34>
- Ganatra, B., Gerdt, C., Rossier, C., Johnson, B. R., Jr., Tunçalp, Ö., Assifi, A., . . . Alkema, L. (2017). Global, regional, and subregional classification of abortions by safety, 2010–14: Estimates from a Bayesian hierarchical model. *Lancet*, *390*, 2372–2381.
- Haaland, M. E. S., Haukanes, H., Zulu, J. M., Moland, K. M., & Blystad, A. (2020). Silent politics and unknown numbers: Rural health bureaucrats and Zambian abortion policy. *Social Science & Medicine*, *251*, 112909. <https://doi.org/10.1016/j.socscimed.2020.112909>
- Hewitt, M. (2002). Attitudes toward interview mode and comparability of reporting sexual behavior by personal interview and audio computer-assisted self-interviewing: Analyses of the 1995 National Survey of Family Growth. *Sociological Methods & Research*, *31*, 3–26.
- Hox, J., & De Leeuw, E. D. (2002). The influence of interviewers' attitude and behavior on household survey nonresponse: An international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 103–120). New York, NY: Wiley.
- ICF. (2020). *Demographic and Health Survey interviewer's manual*. Rockville, MD: ICF.
- ICF Macro. (2009). *Training field staff for DHS surveys*. Calverton, MD: ICF Macro.
- Jones, E. F., & Forrest, J. D. (1992). Underreporting of abortion in surveys of U.S. women: 1976 to 1988. *Demography*, *29*, 113–126.
- Jones, R. K., & Kost, K. (2007). Underreporting of induced and spontaneous abortion in the United States: An analysis of the 2002 National Survey of Family Growth. *Studies in Family Planning*, *38*, 187–197.
- Kishor, S. (2015). *Strengthening the demographic data base in the post-2015 era: Lessons from the DHS program* (DHS Questionnaires and Manuals). Retrieved from <https://www.un.org/en/development/desa/population/events/pdf/expert/23/Presentations/EGM-S3-Kishor%20presentation.pdf>
- Koenig, M. A., Ahmed, S., Hossain, M. B., & Mozumder, A. B. M. K. A. (2003). Women's status and domestic violence in rural Bangladesh: Individual- and community-level effects. *Demography*, *40*, 269–288.
- Levandowski, B. A., Kalilani-Phiri, L., Kachale, F., Awah, P., Kangaude, G., & Mhango, C. (2012). Investigating social consequences of unwanted pregnancy and unsafe abortion in Malawi: The role of stigma. *International Journal of Gynecology & Obstetrics*, *118*(S2), S167–S171.
- Lindberg, L., & Scott, R. H. (2018). Effect of ACASI on reporting of abortion and other pregnancy outcomes in the US National Survey of Family Growth. *Studies in Family Planning*, *49*, 259–278.
- Lunn, D., Jackson, C., Best, A. T., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: CRC Press.
- MacQuarrie, K. L. D., Winfrey, W., Meijer-Irons, J., & Morse, A. (2018). *Consistency of reporting of terminated pregnancies in DHS calendars* (DHS Methodological Reports No. 25). Rockville, MD: ICF.
- Mahmud Khan, M., Hotchkiss, D. R., Berruti, A. A., & Hutchinson, P. L. (2005). Geographic aspects of poverty and health in Tanzania: Does living in a poor area matter? *Health Policy and Planning*, *21*, 110–122.
- Mensch, B. S., Hewitt, P. C., Gregory, R., & Helleringer, S. (2008). Sexual behavior and STI/HIV status among adolescents in rural Malawi: An evaluation of the effect of interview mode on reporting. *Studies in Family Planning*, *39*, 321–334.

- Nandagiri, R. (2019). "Like a mother-daughter relationship": Community health intermediaries' knowledge of and attitudes to abortion in Karnataka, India. *Social Science & Medicine*, 239, 112525. <https://doi.org/10.1016/j.socscimed.2019.112525>
- Olson, K., & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly*, 75, 99–114.
- Polis, C. B., Castillo, P. W., Otipiri, E., Keogh, S. C., Hussain, R., Nakua, E. K., . . . Bell, S. O. (2020). Estimating the incidence of abortion: Using the abortion incidence complications methodology in Ghana, 2017. *BMJ Global Health*, 5, e002130. <https://doi.org/10.1136/bmjgh-2019-002130>
- Polis, C. B., Mhango, C., Philbin, J., Chimwaza, W., Chipeta, E., & Msusa, A. (2017). Incidence of induced abortion in Malawi, 2015. *PLoS ONE*, 12(4), e0173639. <https://doi.org/10.1371/journal.pone.0173639>
- Rabe-Hesketh, S., & Skrondal, A. (2008). Chapter 10. Dichotomous or binary responses. In *Multilevel and longitudinal modelling using Stata* (pp. 431–472). College Station, TX: Stata.
- Randall, S., Coast, E., Compaore, N., & Antoine, P. (2013). The power of the interviewer: A qualitative perspective on African survey data collection. *Demographic Research*, 28, 763–792. <https://doi.org/10.4054/DemRes.2013.28.27>
- Rossier, C. (2003). Estimating induced abortion rates: A review. *Studies in Family Planning*, 34, 87–102.
- Sedgh, G., Bearak, J., Singh, S., Bankole, A., Popinchalk, A., Ganatra, B., . . . Alkema, L. (2016). Abortion incidence between 1990 and 2014: Global, regional, and subregional levels and trends. *Lancet*, 388, 258–267.
- Sedgh, G., & Keogh, S. C. (2019). Novel approaches to estimating abortion incidence. *Reproductive Health*, 16, 44. <https://doi.org/10.1186/s12978-019-0702-0>
- Sedgh, G., Singh, S., Shah, I. H., Åhman, E., Henshaw, S. K., & Bankole, A. (2012). Induced abortion: Incidence and trends worldwide from 1995 to 2008. *Lancet*, 379, 625–632.
- Singer, E., Frankel, M. R., & Glassman, M. B. (1983). The effect of interviewer characteristics and expectations on response. *Public Opinion Quarterly*, 47, 68–83.
- Singh, S., Remez, L., Sedgh, G., Kwok, L., & Onda, T. (2017). *Abortion worldwide 2017: Uneven progress and unequal access* (Report). New York, NY: Guttmacher Institute. Retrieved from <https://www.guttmacher.org/report/abortion-worldwide-2017>
- Singh, S., & Wulf, D. (1994). Estimated levels of induced abortion in six Latin American countries. *International Family Planning Perspectives*, 20, 4–13.
- Stephenson, R., Baschieri, A., Clements, S., Hennink, M., & Madise, N. (2006). Contextual influences on the use of health facilities for childbirth in Africa. *American Journal of Public Health*, 96, 84–93.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133, 859–883.
- United Nations, Department of Economic and Social Affairs, Population Division. (2014). *Abortion policies and reproductive health around the world* (United Nations Publication, Sales No. E.14.XIII.11). New York, NY: United Nations.
- Vassallo, R., Durrant, G., & Smith, P. (2017). Separating interviewer and area effects by using a cross-classified multilevel logistic model: Simulation findings and implications for survey designs. *Journal of the Royal Statistical Society: Series A, Statistics in Society*, 180, 531–550.
- Weinreb, A. A., & Sana, M. (2009). The effects of questionnaire translation on demographic data and analysis. *Population Research and Policy Review*, 28, 429–454.

Tiziana Leone (corresponding author)
t.leone@lse.ac.uk

Leone • Department of International Development, LSE, London, UK

Sochas • Department of Social Policy, LSE, London, UK

Coast • Department of International Development, LSE, London, UK