Interviewer effects on patterns of non-response: Evaluating the impact on the reasons for contraceptive non-use in Indonesia

Mark Lyons-Amos, School of Health Sciences and Social Works, University of Portsmouth

Working Paper of Portsmouth-Brawijaya Centre for Global Health, Population, and Policy

Number 1

Abstract

Much social science research is reliant on generating data by questionnaires and interview. Understanding the processes by which these data are generated is therefore vital for ensuring validity of scientific results. Interviewers, as a primary means of collecting responses, are one mode through which the generation of data can be affected. This paper uses the reason for contraceptive non-use module of the Indonesian DHS to examine the effect of differential effects of interviewers on response patterns. We find that the probability of providing a response declines across the module, an effect which is robust to the introduction of controls. Using a cross classified multil-level model, we are able to partition the effect of this decline into respondent and interviewer effects. We find that although significant, the substantive effect of interviewers on the response profile is small and the majority of variation is accounted for by interviewee level variation. Therefore, while data collection via interviewers seems to be a reliable mechanism within the DHS, care should be taken to minimise respondent burden to ensure valid responses.

Introduction

Unmet need for contraception has fallen in Indonesia within the two decades. From a value of 17% in the 1991 DHS, the most recent DHS estimates that 11% of Indonesian women have unmet need, with demand for contraception, the percentage of demand met and the percentage of demand met by a modern method all contributing to this fall (Statistics Indonesia 2012). While this decline is impressive, in order to continue to reduce the number of women at risk of unwanted pregnancy- and in particular in the Indonesian context unmet need for limiting contraception- a thorough understanding of the reasons for contraceptive non-use is required. While this has been extensively studied in general, the usefulness of existing literature going forward is limited by the fact that reason for non-use at a population level shift over time (Sedgh et al. 2016).

To fully understand processes of contraceptive non-use, high quality data are required. Data quality in general has changed internationally (De Heer 1999). While DHS data quality has largely remained of a high quality (Stones and Lyons-Amos 2017), much evaluation has concentrated on demographic data such as age (Johnson et al.,2009, Pullum 2008, Robles and Goldman 1999) or basic health information (Channon et al. 2011, Pullum 2011) with scant attention paid to more complicated data collection modules. Indeed, what analysis does deal with complex modules tends to find some major quality concerns (e.g. Strickler et al. 1997).

A number of respondents characteristics are known to systematically influence the quality of data collected including the age of the respondent (Johnson et al. 2009) and well as method of recall (Channon et al. 2011). These influences manifest in a number of ways, including refusal to participate in the survey process entirely (Durrant and Steele 2009) or item non-response (Singer et al. 1983). Additionally, interaction between the respondent and the questionnaire can substantially affect the nature of the responses gathered, with excessively long questionnaires tending to decrease the quality of responses toward the end of the interview through respondent fatigue (Groves et al. 2002). Whilst respondent burden can be reduced through the introduction of skip and filters, it should be noted that this can introduce systematic biases into the responses elicited (Mathews et al. 2012, Beaujouan 2013).

In addition to respondent effects, interviewers can have a major impact on the quality of data collected. Interviewers systematically affect the rate at which survey respondents are both contacted and agree to participate in surveys (Durrant and Steele 2009), with systematic differences in interviewer success according to age, sex, interviewers experience, pay grade and years of experience and attitudes regarding the persuasion of reluctant respondents (Blom et al., 2010; Durrant et al., 2010, Hox and De Leeuw, 2002, Hansen, 2006, Haunberger, 2010). Interviewer characteristics tend to interact with those of their respondents when generating responses, with Durrant et al. (2010) finding that similarity between respondent and interviewer tends to improve survey response. Johnson et al. (2009) find this within the context of DHS data, with the sex of the interviewer and the presence of a translator having marked impact on the quality of data collected. Importantly, interviewers are subject to the same pressures as respondents, with the length of the interview assignment and the expectations of the interviewer playing a significant role in the quality of responses (Singer et al. 1983).

Whilst there has been a considerable literature devoted to evaluating DHS data quality (e.g. Johnson et al.,2009, Channon et al. 2011, Pullum 2006), little work has been dedicated to describing the mechanism by which data quality is generated. Where interviewer characteristics have been taken into account (Johnson et al 2009) by including characteristics of the interviewer on the nature of the response, this has largely dealt with only with issues in basic demographic data- such as age heaping which- while an important demographic factor in their own right are concentrated at the start of the questionnaire and as such are less likely to suffer from data quality issues related to respondent or interviewer fatigue (Teclaw et al. 2012) and moreover are relatively simple to conceptualise and operationalise in comparison to later modules, such as the reason for non-use (Morgan and Hagewen 2005). This paper therefore build on existing evaluations of data quality in two major respects: firstly by establishing the sources of variation between interviewers and respondents and identifying which of these has the largely relative effect, and secondly by looking at a more complicated module sited relatively late within the DHS questionnaire- specifically the reason for contraceptive non-use. The analysis relies on testing the significance of question order within the module: if neither respondent nor interviewers are affecting the responses gained, there should be no statistically significant effect of question order on the response pattern (net of controls). The significant effect of question order is taken to indicate either respondent (Groves et al. 2002) or interviewer (Singer et al. 1983) fatigue. Three research hypotheses are tested in performing this analysis. These research hypotheses are formulated under the assumption that there will be influences on response patterns from both respondents and interviewers: from this perspective rejection of the research hypotheses indicates higher quality collection of data, and support for these hypotheses indicates lower quality data collection:

Research Hypothesis I: Question order within the non-response module will affect the propensity to provide a positive response.

Research Hypothesis II: Interviewers will affect the propensity of obtaining a positive response within the non-response module

Research Hypothesis III: Interviewer will affect the impact of question order on positive response differently: diligent interviewing teams will mediate the effect of question order whereas lackadaisical interviewing teams will accentuate it.

Data

Data for this analysis are drawn for the reason for contraceptive non-use module from the 2012 Indonesian DHS. The DHS is a nationally representative household sample survey, which uses a cluster randomised sampling design. Primary sampling units are selected based on national level data, with complete enumeration of households within the PSU to create a sampling frame to provide a list for secondary sampling. Within selected households all eligible women are interviewed.

The 2012 Indonesia DHS employed 119 interviewing teams to collect the data. Each team was comprised of eight interviewers: one male supervisor, one female field editor, four female interviewers, and two male interviewers, one for currently married men and one for never-married men. In Papua and West Papua, each team consisted of five interviewers: one male supervisor, one female field editor, two female interviewers, and one male interviewer for married men and never-married men.

More than one interviewing team operated within each PSU, and each interviewing team operated within more than one PSU. A major advantage of this design is that interviewing teams are not nested within primary sampling units: this allows separation of the effect of interviewing teams from the area in which they are operating and is relatively rare in this type of analysis (Vasallo et al. 2017, Campanelli and O'Muircheartaigh 1999; Schnell and Kreuter 2005). Geographic influences have been shown to affect both the effect of interviewers on the interview outcomes (Durrant and Steele 2009, Durrant et al. 2010, Vasallo et al. 2017, Campanelli and O'Muircheartaigh 1999; Schnell and Kreuter

2005), but can potentially influence the responses of individuals due to contextual effects such as the presence of a family planning clinic in the local area (which could inhibit access to contraception) as well as other local geographic effects such as contraceptive networks (Lyons-Amos et al. 2010, Behrman et al. 2002). The cross classified design removes potential confounding of this nature and allows estimation of the relative size of the effect of interviewer, geographic and individual effects on the data generated.

Within the Indonesia DHS questionnaire, the module for contraceptive non-use is relatively vulnerable to respondent and interviewer burden. The reason for non-use question come in a list format, which require the respondent to affirm whether any of twenty potential reasons for non-use are relevant. Moreover, this comes at the end of the contraceptive use section of the DHS which uses similar question structures and as such the respondent will be familiar with the list format. Respondent and interviewer fatigue then can potentially manifest in the form of a string of negative responses to questions proffered to speed the process of completing the module, or in item non-response (Groves et al. 2002).

The analytic sample for this paper comprises all women with reported unmet need who were not using a contraceptive at the time of survey. Only women with unmet need are considered for this analysis since by definition they will not be using a contraceptive, which allows identification of missing responses due to item non-response as opposed to non-response due to skip patterns within the questionnaire. In total the analytic sample comprises 2,956 women. Women within the selected sample are nested within both 1188 Primary Sampling Units (PSU) and 26 interviewing teams, although as noted already each PSU will be served by more than one interviewing team, and each interviewing team is active in more than one primary sampling unit. This leads to a cross classified nesting structure with an average of 37.4 observations within each PSU and 1709.1 observations per interviewing team.

Method and model

This paper makes use of cross classified multi-level models to separate interviewer effects from the effect of local geography on response patterns, mirroring the approach of Durrant and Steele (2009), O'Muircheartaigh et al (1998) Durrant et al (2010) and Vassallo et al (2017). The cross classified multilevel model can be written in the form of equation 1 using the notation of Browne et al (2001).

$$logit(pr(positive\ response_i)) = \beta_0 + \beta_1 question\ order + \ u^{(2)}_{Interviewer,i} + u^{(3)}_{PSU,i}$$

$$u^{(2)}_{Interviewer} \sim N\left(0, \sigma^2_{u(2)}\right), u^{(3)}_{PSU} \sim N(0, \sigma^2_{u(3)})$$

Equation 1

In equation 1, the response variable takes the form of an indicator variable which takes the value 1 if the respondent proffers a positive response to the reasons for contraceptive non-use and zero if not. This operationalises the concept of declining data quality: respondent or interviewer fatigue can introduce a declining probability of responding positively to the reason for non-use which would introduce a significant and negative effect of question order (after controlling for non-questionnaire related factors influencing contraceptive utilisation, e.g. age, parity etc.). The effect of question order is captured by the coefficient $\beta_1$which is a linear fixed effect. Additional fixed effect control variable can be included in a similar manner. The effect of the interviewer and local geography is captured by the two random effect coefficients $u^{(2)}_{Interviewer}$ and $u^{(3)}_{PSUi}$ which partition variation in the response between interviewer and PSU clusters. Both of these coefficients are assumed to be Normally distributed with variance estimated at $\sigma^2_{u(2)}$ and $\sigma^2_{u(3)}$ respectively. This model is a variance partition model, and allows us to test the relative importance of interviewer effects in relation to all other sources of variation. This model can be extended in the form of equation 2 to account for variation in the effect of question order by interviewing team.

$$\begin{aligned} logit(pr(positive\ response_i)) &= \beta_0 + \beta_1 question\ order + u^{(2)}_{Interviewer,i} + u^{(3)}_{PSU,i} \\ &+ u^{(4)}_{Interviewer,i} question\ order \end{aligned}$$

$$\begin{bmatrix} u^{(2)} \\ u^{(4)} \end{bmatrix} \sim N(0, \Omega_u), \Omega_u \sim \begin{bmatrix} \sigma^2_{u(2)} & \\ \sigma_{u(3),u(4)} & \sigma^2_{u(4)} \end{bmatrix}, u^{(3)}_{PSU} \sim N(0, \sigma^2_{u(3)})$$

Equation 2

This model is defined similarly to equation 1, save for the addition of the random coefficient $u^{(4)}$ which allows the effect of the question order to vary depending on the interviewing team. The distribution of the interviewer random effects is now multivariate normal according to the variance-covariance matrix Ω_u, with the random intercept captured by the variance $\sigma^2_{u(2)}$ and the random slope by $\sigma^2_{u(4)}$ which have covariance $\sigma_{u(3),u(4)}$.

Modelling strategy

The modelling strategy builds successive models to test the research aims and test research hypotheses. Model I is a simple logit model which simply test whether the effect of question order on the probability of a positive response is statistically significant. This allows us to test the validity of research hypothesis I.

Random effects are then introduced into the model to test whether there is a significant impact of area and interviewer levels effects. Ideally, fixed effect controls would be introduced before the random effects, however, this leads to non-convergent solutions and this analysis therefore follows the recommendations of Browne (2017) and establishing a working random component of the model first, before adding fixed effect controls. Model II therefore comprises a random intercept model with the random part reflecting variation at the interviewer level. Model III extends this by introducing a cross classification with random intercepts for both interviewer and primary sampling unit, which should remove any confounding between area level characteristics and interviewer characteristics (*as per.* Vassalo *et al.* 2017). This model allows us to test research hypothesis II.

Model IV introduces fixed part controls for age, education, marital status and wealth. This establishes whether both the effects of question order, and the clustering by both PSU and interview team are robust to woman level determinants of contraceptive non use. Finally, Model V introduces a random slope for question order by interviewing team. This allows us to test research hypothesis III by allowing the effect of question order to depend on the interviewing team collecting data.

Model I is estimated using regression function in Stata 13.0 for Windows. All multilevel modelling is conducted in MLwiN 2.36 for Windows (Charlton *et al*. 2017) via the runmlwin function in Stata (Leckie and Charlton 2013). For models II to model IV, models are estimated via MCMC using 10000 samples with a 2000 sample burn in with initial values taken from models estimated using 2nd order Penalised Quasi-Likelihood. This follows the recommendation of Browne (2017) which recommends the reestimation of binary response models using MCMC since iterative (such as IGLS or RIGLS) estimation is likely to downward bias variance estimates. Model V was also estimated using MCMC with 10000 samples and a 2000 sample burn in, but initial values were taken from a bespoke input matrix since 2nd order PQL gave non-positive definite starting values.

Results

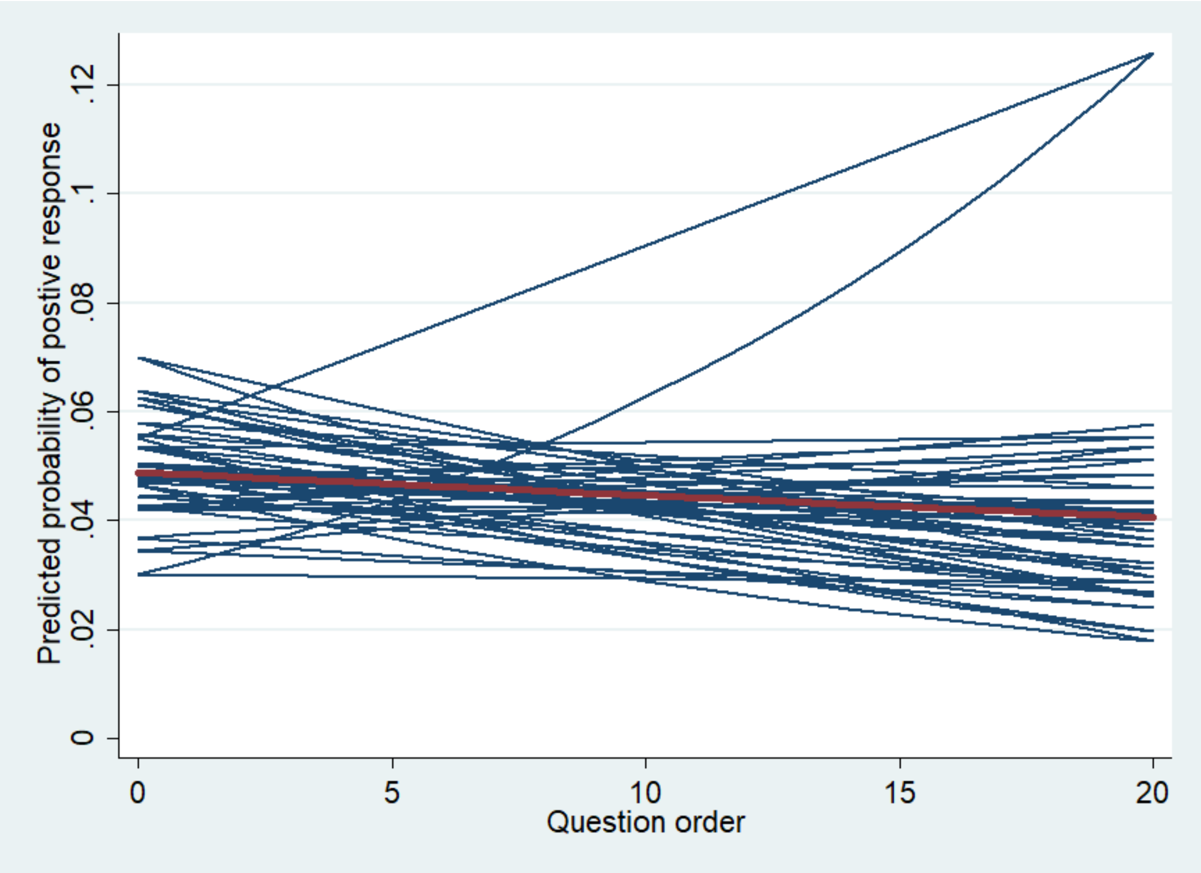
Results of the modelling procedure are presented in table 1. In initial data exploration it was found that a linear trend for the effect of question order provided the best fitting model, with terms for question order squared and cubed providing no improvement in model fit based on likelihood ratio tests.

Model I finds a significant effect of question order on the probability of providing a positive response to a reason for non-use ($p<0.01$). This effect is negative, indicating the probability of providing a positive response declines with increasing order. This is consistent with research hypothesis I: there is an effect of question order on the responses and the data generated.

Model II includes a random intercept at the interviewer level, while model III extend this model to include a random intercept at the PSU level. In both cases, the fixed effect for question order is statistically significant and relatively unaffected in magnitude compared to model I. Interviewer level effects are significant and robust to the introduction of PSU level random effects, indicating an interviewer influence on the responses generated. That said, the substantive size of these effects is trivial. Using a latent variable approximation with the woman level variance set to $\frac{\pi^2}{3}$, the estimated variance partition coefficient for the interviewer level effects is less than 1%. Similar results are obtained with the introduction of woman level fixed effect controls: the fixed effect for question order remain significant and negative, and the random effects for the interviewer effect remain significant but small.

Model V includes a random slope to account for differential effects of interviewers on the effect of question order. This random coefficient, again, is significant but relatively small in magnitude. This is reflected in the predicted probabilities of a positive response presented in Figure 1. Each blue line denotes the slope for one interviewer team, with the population average presented in red. Overall, there is little deviation from the population line, with most interview teams clustering around the overall downward trend. There are only two major outliers, both of whom show relatively rapid increases in the probability of positive responses with increasing question order.

Figure 1: Median predicted probability of positive response by question order for each interviewer team



Conclusions and discussion

This paper evaluates the effect of interviewers on the responses obtained from the reasons for contraceptive non-use module of the Indonesia DHS. Three major research hypotheses were tested, firstly that the probability of obtaining a positive response declines with question order, due either to respondent burden or interviewer induced question skipping. Secondly, the extent to which interviewers compared to respondents influence the pattern of positive responses. Thirdly, we tested whether there was a significant effect of interviewers on the effect of question order.

By and large, the analysis refuted the second and third research hypotheses. Although we obtain a consistent estimate that there is some variability due to interviewers, the estimated value is so small as to be practically trivial: less than 1% of variability in the response patterns can be attributed to interviewer level effects. This is in general a positive finding: DHS data quality seems to be robust and the generation of data in uninfluenced by the teams collecting it. Indeed, the only two major outlying interview team demonstrated increases in the probability of obtaining a positive response with increasing question order: contrary to the expected effect were interview teams responsible for declines in data quality. This should give confidence in the quality of data available from DHS surveys.

Table A: Estimated models from multilevel cross classified regression for probability of positive response

Parameter	Model I	Model II	Model III	Model IV	Model V
Fixed part parameters					
Question order	-0.15 ** (-0.02 - -0.006)	-0.014**(-0.023 - -0.007)	-0.015 ** (-0.023 - -0.007)	-0.014** (-0.023 - -0.007)	-0.009 (-0.057-0.049)
Respondent over 30				-0.054 (-0.163 - 0.536)	-0.059 (-0.163-0.0429)
Secondary or higher education				0.036 (-0.068 - 0.149)	0.040 (-0.063-0.147)
Currently married				-0.139 (-0.388 – 0.106)	-0.126 (-0.378-0.186)
Rich or richest wealth quintile				-0.085 (-0.196 – 0.024)	-0.084 (-0.184-0.023)
Constant	-3.08	-3.09	-3.09	-2.87	-2.96
Random part parameters					
Variance (PSU)			0.002 (0.001-0.006)	0.001 (0.000-0.003)	0.002 (0.000-0.005)
Variance (Interviewer)		0.0025 (0.000-0.010)	0.002 (0.000-0.007)	0.002 (0.000-0.009)	0.130 (0.052-0.284)
Variance (Question order)					0.028 (0.016-0.049)
Covariance (Interviewer, question order)					-0.009 (-0.042 -0.143)

Model II-V based on 10000 MCMC sample with 2000 sample burnin. Starting values from 2nd order penalised quasi-likelihood estimates for models II-IV, and from bespoke input matrix for model V.

Figure in (parentheses) indicate 95% credible intervals.

** denotes p<0.01, * denotes p<0.05

That said, a note of caution should be sounded. The declining probability of obtaining a positive response across question order supports the first research hypothesis and is robust to individual level controls, interviewer and area level effects. The persistent significance of this effect- even net of other determinants of contraceptive non-use- and the small effect of both area and interviewer level effects suggests a degree of respondent burden. This is consistent with other authors, who have noted the effect of DHS questionnaire design on response schedule- Mathews et al (2012) in particular finding that the DHS is vulnerable to framing effects and that question order can have systematic effects on responses collected. As noted previously, the module evaluated in this paper comes toward the end of a relatively length module with a number of question in a list format which increases the potential for respondent fatigue. Care should be taken to ensure that the requirements of the DHS interview on the respondent are not too onerous, and that a desire for data quantity does not compromise data quality.

References

Beaujouan, Eva. 2013. “COUNTING HOW MANY CHILDREN PEOPLE WANT: THE INFLUENCE OF QUESTION FILTERS AND PRE-CODES.” Demográfia English Edition 56 (5): 35–61.

Behrman, Jere R., Hans-Peter Kohler, and Susan Cotts Watkins. 2002. “Social Networks and Changes in Contraceptive Use over Time: Evidence from a Longitudinal Study in Rural Kenya.” Demography 39 (4): 713–38. doi:10.1353/dem.2002.0033.

Blom, A. G., De Leeuw, E. D. and Hox, J. J. (2010) Interviewer effects on nonresponse in the European Social Survey. Working Paper 2010-25. Institute for Social and Economic Research, University of Essex, Colchester

Browne, W.J. (2017) MCMC Estimation in MLwiN v3.00. Centre for Multilevel Modelling, University of Bristol.

Browne, W.J., Goldstein, H. & Rasbash, J. (2001a). Multiple Membership Multiple Classification (MMMC) models. Statistical Modelling, 1:103–124.
Charlton, C., Rasbash, J., Browne, W.J., Healy, M. and Cameron, B. (2017) MLwiN Version 3.00. Centre for Multilevel Modelling, University of Bristol.

Campanelli, P. and O’Muircheartaigh, C. (1999) Interviewers, interviewer continuity, and panel survey nonresponse. Qual. Quant., 33, 59–76.

Channon, Andrew A. R., Sabu S. Padmadas, and John W. McDonald. 2011. “Measuring Birth Weight in Developing Countries: Does the Method of Reporting in Retrospective Surveys Matter?” Maternal and Child Health Journal 15 (1): 12–18. doi:10.1007/s10995-009-0553-3.

De Heer, W. (1999) International response trends: results of an international survey. J. Off. Statist., 15, 129–142

Durrant, Gabriele B., and Fiona Steele. 2009. “Multilevel Modelling of Refusal and Non-Contact in Household Surveys: Evidence from Six UK Government Surveys.” Journal of the Royal Statistical Society: Series A (Statistics in Society) 172 (2): 361–81. doi:10.1111/j.1467-985X.2008.00565.x.

Durrant, G. B., Groves, R. M., Staetsky, L. and Steele, F. (2010) Effects of interviewer attitudes and behaviors on refusal in household surveys. Public Opinion Quarterly, 74, 1–36

Groves, R. M., Dillman, D. A., Eltinge, J. L. and Little, R. J. A. (eds) (2002) Survey Nonresponse. New York: Wiley

Hansen, K. M. (2006) The effects of incentives, interview length, and interviewer characteristics on response rates in a CATI study. International Journal of Public Opinion Research, 19, 112–121

Haunberger, S. (2010) The effects of interviewer, respondent and area characteristics on cooperation in panel surveys: a multilevel approach. Qual. Quant., 44, 957–969.

Hox, J. and De Leeuw, E. (2002) The influence of interviewers’ attitude and behavior on household survey nonresponse: an international comparison. In Survey Nonresponse (eds R. M. Groves, D. A. Dillman, J. L. Eltinge and R. J. A. Little), pp. 103–119. New York: Wiley

IRD. 1990. DHS Methodological Reports 1: An assessment of DHS-1 data quality. Washington, DC: Institute for Resource Development Inc.

Johnson, Kiersten, Monica Grant, Shane Khan, Zhuzhi Moore, Avril Armstrong, and Zhihong Sa. 2009. Fieldwork-Related Factors and Data Quality in the Demographic and Health Surveys Program

Leckie, G. and Charlton, C. (2013). runmlwin - A Program to Run the MLwiN Multilevel Modelling Software from within Stata. Journal of Statistical Software, 52 (11),1-40

Lyons-Amos, Mark J., Gabriele B. Durrant, and Sabu S. Padmadas. 2011. "IS TRADITIONAL CONTRACEPTIVE USE IN MOLDOVA ASSOCIATED WITH POVERTY AND ISOLATION?" Journal of Biosocial Science 43 (03): 305–327. doi:10.1017/S0021932010000775.

Mathews, P., Sear, R., Coast, E. and Iacovou, M. (2012): Do preceding questions influence the reporting of childbearing intentions in social surveys? In Population Association of America Annual Meeting. San Francisco, 3–5 May

Morgan, S. P. and K. J. Hagewen (2005). Fertility. Handbook of Population. D. L. Poston and M. Micklin. New York, Kluwer Academic/Plenum.

O 'muircheartaigh, Colm, Pamela Campanelli, and Colm O 'muircheartaight. 1998. "The Relative Impact of Interviewer Effects and Sample Design Effects on Survey Precision." Source Journal of the Royal Statistical Society. Series A (Statistics in Society) J. R. Statist. Soc. A 161 (1): 63–77.

Pullum, Thomas W. 2006. An Assessment of Age and Date Reporting in the DHS Surveys, 1985-2003. Methodological Reports No. 5. Calverton, Maryland: Macro International Inc.

Pullum, Thomas W. 2008. An Assessment of the Quality of Data on Health and Nutrition in the DHS Surveys, 1993-2003. Methodological Reports No. 6. Calverton, Maryland: Macro International Inc.

Robles, A., and Goldman, N. (1999). Can accurate data on birth weight be obtained from health interview surveys? International Journal of Epidemiology, 28, 925–931

Schnell, R. and Kreuter, F. (2005) Separating interviewer and sampling-point effects. Journal of Official Statististics, 21, 389–410

Sedgh, Gilda, Lori S. Ashford and Rubina Hussain "Unmet Need for Contraception in Developing Countries: Examining Women's Reasons for Not Using a Method." 2016. Guttmacher Institute. June 21. https://www.guttmacher.org/report/unmet-need-for-contraception-in-developing-countries.

Singer, Eleanor, Martin R Frankel, and Marc B Glassman. 1983. "The Effect of Interviewer Characteristics and Expectations on Response." Public Opinion Quarterly 47 (1): 68–83. doi:10.1086/268767.

Statistics Indonesia (Badan Pusat Statistik—BPS), National Population and Family Planning Board (BKKBN), and Kementerian Kesehatan (Kemenkes—MOH), and ICF International. 2013. Indonesia Demographic and Health Survey 2012. Jakarta, Indonesia: BPS, BKKBN, Kemenkes, and ICF International.

Stones, T. and M. J. Lyons-Amos 2017 "Trends in DHS data quality in Sub-Saharan Africa: An analysis of age heaping over time in 34 countries in Sub-Saharan Africa" Portsmouth-Brawijaya Centre for Global Health, Population, and Policy Number 4

Strickler, Jennifer A., Robert J. Magnani, H. Gilman McCann, Lisanne F. Brown, and Janet C. Rice. 1997. "The Reliability of Reporting of Contraceptive Behavior in DHS Calendar Data: Evidence from Morocco." Studies in Family Planning 28 (1): 44–53. doi:10.2307/2137970.

Teclaw, Robert, Mark C. Price, and Katerine Osatuke. 2012. “Demographic Question Placement: Effect on Item Response Rates and Means of a Veterans Health Administration Survey.” Journal of Business and Psychology 27 (3): 281–90. doi:10.1007/s10869-011-9249-y.

Vassallo, Rebecca, Gabriele Durrant, and Peter Smith. 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 (2): 531–50. doi:10.1111/rssa.12206.