Field interviewer effects on the quality of malaria diagnosis in Malawi

Ngianga II Kandala, Mark Amos, Saseendran Pallikadavath

School of Health Sciences and Social Works, University of Portsmouth

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

Number 8

Abstract

Malaria is a life threatening disease affecting 216 million people in 91 countries and remains endemic in 106 countries. Sub-Saharan Africa has a disproportionate high share of the global malaria burden. Yet though parasite-based diagnosis is increasing, most suspected cases of malaria are still not properly diagnosed. Of particular concern in variability in the quality of diagnosis associated with the team performing diagnosis. This analysis examines variability in the quality of malaria diagnoses using Demographic and Health survey data from the sub Saharan Africa region. We examine the incidence of false and false negative using multilevel random intercept logistic regressions. With each regression we evaluate variability in the accuracy in test administration by clustering individual respondents within the interviewing team. Variation partition coefficients are then calculated to examine the proportion of variation in the rate of false diagnoses to examine the proportion of variation attributable to interviewer teams. Around 15% of variation in the rate of false positive diagnosis is attributable to interviewing teams, with 43% attributable to interviewers for false negatives. Such large variation in the consistency of the diagnostic test procedure attributable to interviewing teams is concerning and should be addressed during training for field test procedures.

Introduction

Malaria is a life threatening disease affecting 216 million people in 91 countries and remains endemic in 106 countries. Sub-Saharan Africa has a disproportionate high share of the global malaria burden (WHO 2017). Yet though parasite-based diagnosis is increasing, most suspected cases of malaria are either not properly diagnosed. The increased usage of Rapid Diagnostic Techniques (RDTs) has increased the potential for improvements in treatment of malaria, since much care seeking is still triggered by the onset of symptoms (Romay-Barja et al. 2016; Kassile et al. 2014). Earlier diagnosis by the use of RDTs offers the possibility for more managed intervention, and potentially avoiding life threatening cases. However, care seeking is currently widely delayed, with an overwhelming majority (more than 79%) delaying seeking care until more than one day after the onset of fever symptoms (Birhanu et al. 2016; Sonkong, Suggaravetsiri, and Chaiklieng 2015)

Given the high prevalence of malaria in many setting, accuracy of testing is of vital importance to ensure that false positive cases do not lead to overutilization of care in resource poor context and that false negatives do not lead to delays in seeking treatment. Malaria can be easily missed even in highly controlled setting and using laboratory and clinical diagnosis (Allan and Tahir 2006). Other studies have analysed the rate of false diagnosis for RDTs, with estimated sensitivities and specificities in the region of 90% in controlled conditions (Ishengoma et al. 2011; Chong et al. 2014). However, these values fall considerably when used in a community survey context- sensitivity has been reported as low as 63.4% (Ishengoma et al. 2011).

Demographic and Health Surveys are widely used cross sectional sample surveys, which have recently started to include more biomarker information such as malarial diagnosis. This is a major advantage, due to the consistent nature of DHS in terms of questions collected as well as the consistency of questions allowing for cross national comparison. Variation in the accuracy of responses from cross sectional surveys has been noted in general, and within the specific context of the DHS (Pullum et al. 2018). Often, these take the form of evaluating the accuracy of quantities such as age (Johnson et al. 2009; Pullum 2006; Lyons-Amos and Stones 2017) required for skip patterns or denominators in the survey, or basic health information (Channon, Padmadas, and McDonald 2011; Pullum 2008) required to generate numerators for disease rates.

The role of interviewers in the process of data generation in an increasingly studies phenomenon. These influences manifest in a number of ways, including refusal to participate in the survey process entirely (Durrant and Steele 2009; Durrant et al. 2010) or item non-response (Singer, Frankel, and Glassman 1983). Interviewers can affect survey responses in two major ways; role restricted (where interviewers influence outcomes through their behaviour(Durrant et al. 2010)) or role independent (where interviewer characteristics influence the response) (Pullum et al. 2018)- for instance, measurement errors may be influenced by the age of the interviewer (Cleary, Mechanic, and Weiss 1981; Ford and Norris 1997). The role of interviewers in diagnosis is vital in the context of malaria: the discordance between relatively high theoretical sensitivity and specificity values compared to those used in the field stymies much of the advantage of RDTs, and the misapplication of testing kits by interviewers is one potential mechanism that could explain this discrepancy.

There is considerable evidence that DHS produce high quality research in an international context (Lyons-Amos and Stones 2017) and that the effect of interviewers can be extremely limited (Amos 2018). That said, the diagnosis of malaria is more complicated than previously examined items, and more critically can be verified by laboratory testing- indeed where verification of DHS data has been possible data has often proved to be deficient (Strickler et al. 1997). We therefore focus on both the specificity (false positive) and sensitivity (false negative rate) of diagnosis in the Malawian context as

two measures of diagnostic accuracy. Using multilevel modelling techniques we are able to extract interviewer effects net of all other cluster effects due to an interpetrated design within the Malawi DHS- the allocation of one interview team to a cluster would otherwise preclude this measurement (Amos 2018; Durrant and Steele 2009; Vassallo, Durrant, and Smith 2017; O'Muircheartaigh and Campanelli 1999; Pullum et al. 2018). This is a major advantage over previous community based surveys, where the effect of interviewers was confounded by other sampling design based factors e.g (Ishengoma et al. 2011). By removing this confounding, we are able to identify the proportion of misdiagnosis attributable to interviewing teams, which may go some way to explaining disappointing accuracy of RDTs when used in the field.

Data

Data for this analysis are drawn from the 2015 Malawi Demographic and Health Survey (NSO 2017). The MDHS is a nationally representative cross sectional sample survey collecting information on contraceptive use, reproductive health and other health indicators. The DHS uses a complex cluster randomised design. Strata were created between rural and urban areas in each census area. Primary sampling units were defined based on census standard enumeration areas, and were selected within each stratum with probability proportional to size. Where SEAs were large, they were separated into segments, so each cluster within the dataset represents each a SEA or segment on an SEA. Within each PSU a household list is used to take a second stage of sampling of household. All eligible respondents within a household are interviewed. A critical aspect of this analysis is that more than one interview team operates in each cluster, and each interview team operates in more than one cluster. This interpenetration of sampling cluster and interview team allows identification of interviewer effect net of local geographic effects in the manner of (Amos 2018; Vassallo, Durrant, and Smith 2017; O'Muircheartaigh and Campanelli 1999; Durrant and Steele 2009)

The relevant sub module for this analysis is the malaria indicator, which is included in the HIV/AIDS sub module. Within this module, two tests for malarial status were recorded. The field test for malaria was administered as part of the anthropometry module, and consisted of a finger prick test for RDT. Five blood spots from the finger prick were collected on a filter paper card, and linked to the respondent via barcode. Laboratory confirmation of the field test result was linked via this barcode.

Method

Two logistic regressions are used to identify variation in false diagnosis. The first model examines the probability of receiving a false positive diagnosis (specificity), which is defined as a field diagnosis of malaria which was later shown to be malaria free by lab testing. We assume in all cases that the lab tests are the gold standard: RDT diagnoses are verified by comparing their result to the lab result. The response variable is a 0/1 indicator variable which takes the value 1 in the case of a false positive diagnosis, and zero in all other cases. The second model examines the probability of false negative diagnosis (sensitivity), which is defined as a field diagnosis of no malaria, where the lab diagnosis was positive. Again, this is modelled as a 0/1 indicator variable with 1 indicating a false negative and 0 a false positive.

In both cases, interviewer effects on the probability of a false diagnosis are captured by a cross classified multilevel model in the form of equation 1.

$$logit\big(pr(incorrect\ diagnosis_{ij})\big) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + u^{(2)}_{Interviewer,j} + u^{(3)}_{Cluster,j}$$

$$u^{(2)}_{Interviewer} \sim N\big(0, \sigma^2_{u(2)}\big), u^{(3)}_{Cluster} \sim N(0, \sigma^2_{u(3)})$$

In this equation, i indexes individual respondents and j indexes higher order aggregations, namely interviewers and clusters. Two separate random effects are specified due to the interpenetration of interviewers with clusters, allowing a cross classified multilevel model to be specified in the manner of Amos (2018), (Durrant and Steele 2009), (O'Muircheartaigh and Campanelli 1999), (Durrant et al. 2010) and (Vassallo, Durrant, and Smith 2017). The vector $\boldsymbol{x}$ contains a vector of control variables pertaining to the individual, namely age, region, residence, education, marital status, and occupation with their estimated effect captured by the vector $\boldsymbol{\beta}$.

Model interpretation is conducted by evaluating the proportion of variation in false diagnosis attributable to interviewers. This is taking $u^{(2)}_{Interviewer}$ as a proportion of all residual variation, with variation at the individual level approximated as $\frac{\pi^2}{3}$. 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). 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 advises the reestimation of binary response models using MCMC since iterative (such as IGLS or RIGLS) estimation is likely to create downward bias variance estimates.

Results

Estimated models are presented in Table 1. Even in the presence of control variables, the second level residuals for both interviewer and PSU are relatively large, in contrast to other studies (Amos 2018). In the case of the false positive, the estimated variance attributable to the interviewers amounts to 15.6% of all unexplained variation within the model. In the case of false negatives, this variation is much larger, accounting for 43% of unexplained variability.

Table 1: Estimated models for probability of false positive and false negative

	False positive		*False negative*	
	Parameter estimate	*95% credible interval*	*Parameter estimate*	*95% credible interval*
Fixed effects				
Constant	-1.22		0.36	
Age group ref=15-19				
20-24	0.19	(-0.64, 0.46)	-0.09	(-0.31, 0.14)
25-29	0.32	(0.05, 0.61)	-0.11	(-0.35, 0.13)
30-34	0.18	(-0.11, 0.49)	0.01	(-0.24, 0.28)
35-39	0.06	(-0.25, 0.38)	0.09	(-0.18, 0.37)
40-44	0.40	(0.04, 0.78)	-0.17	(-0.48, 0.14)
45-49	0.23	(-0.16, 0.64)	-0.07	(-0.42, 0.29)
Province ref=Kinshasa				
Bandushu	-1.18	(-2.45, -0.24)	0.73	(-0.37, 1.53)
Bas-Congo	0.30	(-0.86, 1.30)	-0.90	(-2.00, 0.04)
Equateur	-0.68	(-1.83, 0.15)	0.25	(-0.77, 1.02)
Kasai-Occidental	0.41	(-0.76, 1.34)	-0.92	(-2.08,0.04)
Kasai-Oriental	0.45	(-0.82, 1.35)	-1.08	(-2.21, -0.10)
Katanga	0.42	(-0.75, 1.39)	-1.05	(-2.28, -0.19)
Maniema	0.19	(-1.03, 1.23)	-1.01	(-2.14, -0.16)
Kivu	-2.53	(-3.89, -1.41)	1.64	(0.47, 2.71)
Residence ref=Urban				
Rural	-0.29	(-0.69, 1.44)	-0.03	(-0.37, 0.36)
Highest educational level ref=No education				
Primary	-0.07	(-0.27, 0.13)	0.19	(-0.00, 0.38)
Secondary	-0.43	(-0.73, -0.20)	0.53	(0.29, 0.77)
Higher	-0.99	(-1.89,-0.19)	0.58	(-0.06, 1.25)
Wealth index ref=Poorest				
Poor	0.09	(-0.11, 0.30)	0.02	(-0.15, 0.20)
Medium	-0.20	(-0.43, 0.20)	0.40	(0.20, 0.60)
Rich	-0.26	(-0.53, 0.02)	0.58	(0.33, 0.84)
Richest	-1.20	(-1.73, -0.69)	1.18	(0.76, 1.65)
Marital status ref=Never in union				
Married	-0.21	(-0.49, 0.05)	-0.15	(-0.38, 0.08)
Living with partner	-0.18	(-0.49, 0.11)	-0.04	(-0.30, 0.22)
Formerly partnered	-0.15	(-0.48, 0.17)	-0.11	(-0.41, 0.17)

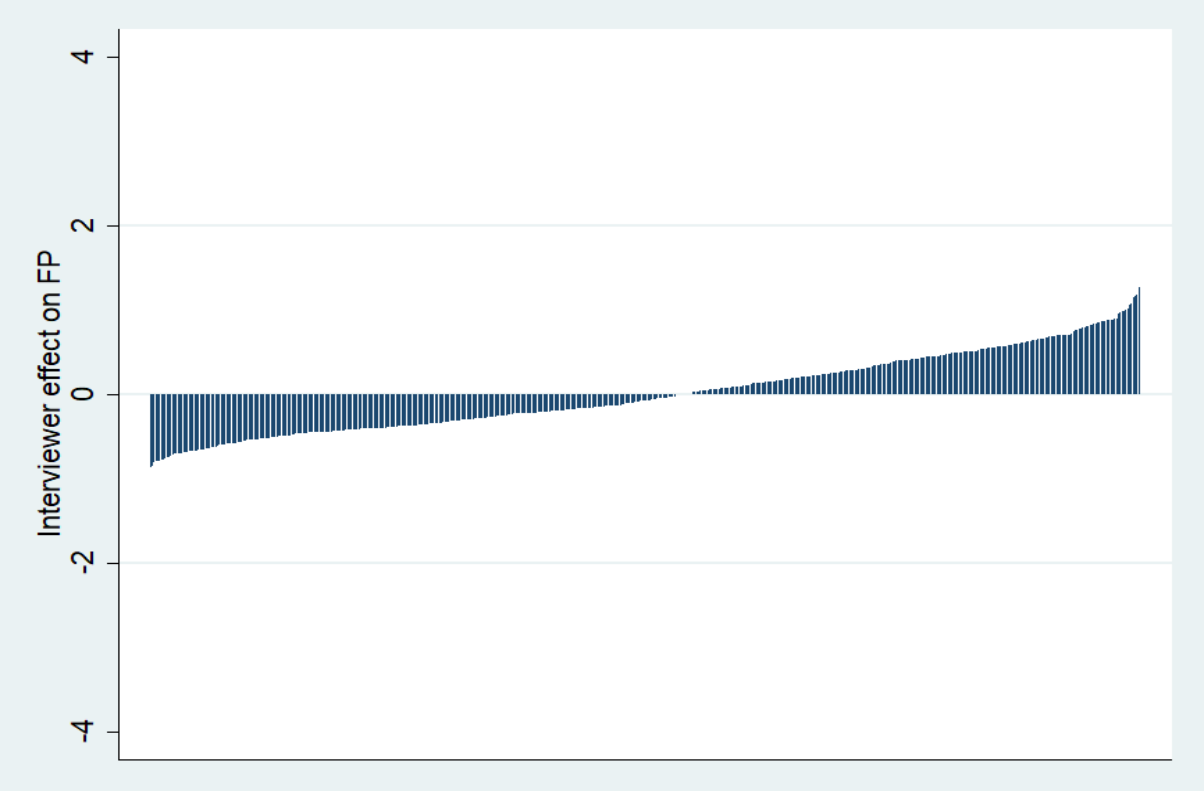
Occupational status ref=Not working				
Professional	-0.04	(-0.28, 0.18)	0.06	(-0.14, 0.27)
Agriculture	0.17	(-0.06, 0.40)	0.07	(-0.13, 0.28)
Non-agricultural manual	0.04	(-0.34, 0.39)	0.17	(-0.14, 0.48)
Random effect				
Interviewer effect ($u^{(2)}$)	0.86	(0.02, 1.98)	2.57	(2.11, 3.10)
PSU effect ($u^{(3)}$)	1.36	(0.39, 2.43)	0.04	(0.00, 0.18)

Notes: Models estimated via MCMC with 10000 samples and 2000 sample burn in. Starting values for MCMC chain are taken from 2nd order PQL.

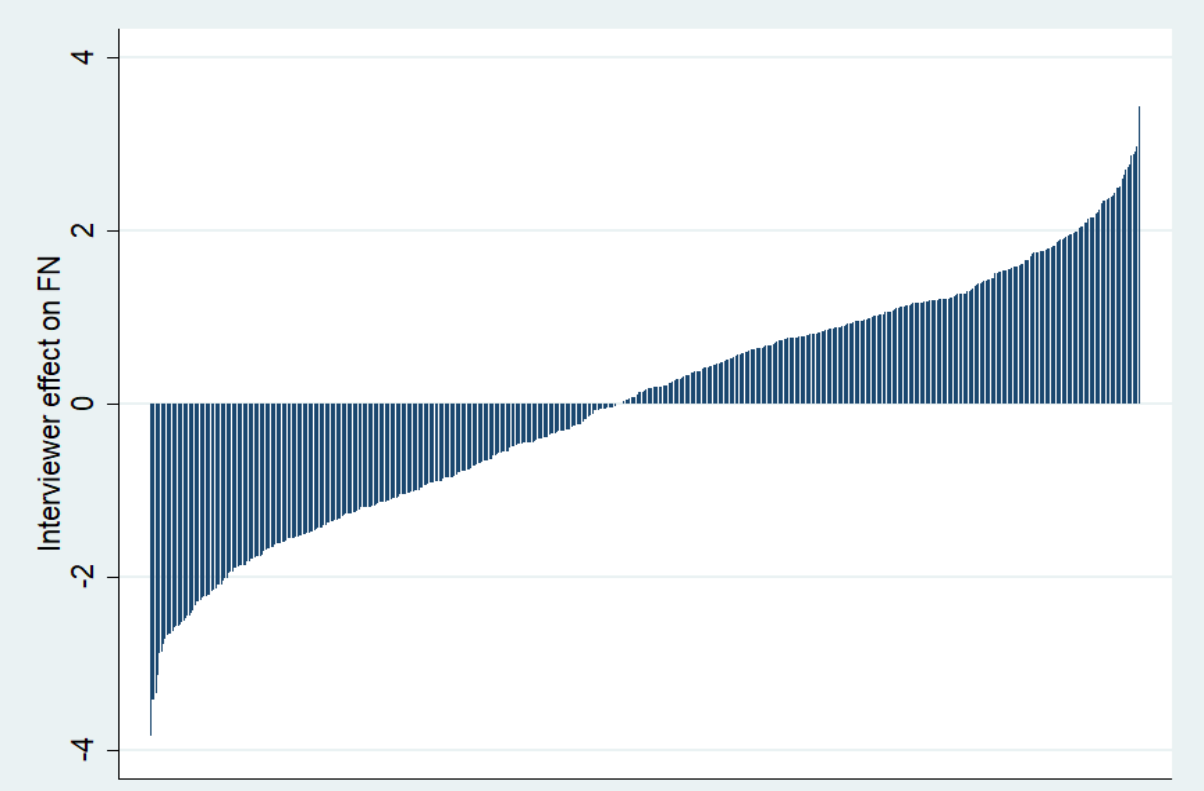
The difference in effects on the probability of a false report is illustrated in Figure 1. This figure presents the Empirical Bayes residuals for the interviewing team ranked on the residual size. The relative size of the two interviewer variance terms observed in table 1 is also reflected in the figure, the level of variation around the median interviewing team is much smaller (although still substantial) for the false positive diagnosis compared to the false negative diagnosis. This reflects the relative size of the estimated interviewer level variance in the regression: the interviewer attributable variance for the false positive model of 0.86 is considerably smaller than in the false negative model (2.57)

Figure 1: Estimated residuals ranked by size of residual indexed by interviewing team for a) false positive diagnosis and b) false negative diagnosis

a) False positive



b) False negative



Conclusions and discussion

This paper examines the effect of interviewing team on variation in inaccurate reporting of malarial diagnosis from RDTs. We examine both the specificity (false positive) and sensitivity (false negative) of the malarial diagnosis performed by field teams based on the gold standard of laboratory diagnosis. Our work build on previous analyses of the accuracy of other community based surveys (Ishengoma et al. 2011; Bisoffi et al. 2010) by taking advantage of the interpenetrated design of the DHS we are able to separate the effect of interviewing teams from confounded effects, and to establish the degree of variability in the accuracy of diagnosis attributable to the interviewer alone.

Overall we found considerable variation in the probability of receiving a false diagnosis, and this is for a large part attributable to membership of the interviewing team. This is consistent with existing analysis of the validity of RDTs in that the accuracy of testing in the field is considerably lower (Ishengoma et al. 2011; Bisoffi et al. 2010) than the high accuracy levels found in more controlled environments e.g (Chong et al. 2014). Even for the false positive rate, the proportion of variation attributable to interviewing teams is still large in substantive terms. Of greater concern is the fact that the variation attributable to interviewing teams for false negative diagnoses which accounts for over 2/5ths of variation in malarial accuracy. This is particularly troubling in light of the false negative diagnosis potentially delaying care seeking treatment (Kassile et al. 2014; Romay-Barja et al. 2016).

Limitations

There are some limitations to this analysis and scope for further work. While we have included controls at the respondent level, we were unable to include interviewer characteristics since these were not readily available in our dataset. This could potentially inform and target training programmes directed at improving the delivery of the RDT: it is highly likely that a role restricted influence such as RDT administration would vary by interviewer characteristic (Durrant et al. 2010), and the integration of such paradata is possible for certain DHS datasets (Pullum et al. 2018). A further extension might be to reevaluate the validity of diagnostic tools for a different survey round or by time at which the survey was administered: the overall accuracy of the RDT depends heavily on the underlying malarial case load, with lower rates being associated with better accuracy (Ishengoma et al. 2011; Bisoffi et al. 2010). Verifying our findings accounting for the season of interview would be enlightening.

Policy recommendations

While the high level of variability of diagnostic accuracy is concerning, the fact that the variation is attributable to interviewing team does at least offer an opportunity to improve diagnostics through interviewer training. Care should be taken to improve the consistency of diagnosis between interviewers in performing field diagnoses and improving the overall accuracy of field testing in general. Due to the fact that this variation is most likely due to some application of the testing rather than an intrinsic quality of the interviewer, the role restricted (Pullum et al. 2018) nature of the application of the test affords the potential for greater application of training to improve diagnostic accuracy.

References

Allan, Philip J, and Hasan I S Tahir. 2006. "How Easily Malaria Can Be Missed." *Journal of the Royal Society of Medicine* 99 (4): 201–2.

Amos, Mark. 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 (14): 415–30. https://doi.org/10.4054/DemRes.2018.39.14.

Birhanu, Zewdie, Lakew Abebe, Morankar Sudhakar, Gunawardena Dissanayake, Yemane Ye-ebiyo Yihdego, Guda Alemayehu, and Delenasaw Yewhalaw. 2016. "Malaria Related Perceptions, Care Seeking after Onset of Fever and Anti-Malarial Drug Use in Malaria Endemic Settings of Southwest Ethiopia." *PLOS ONE* 11 (8): e0160234. https://doi.org/10.1371/journal.pone.0160234.

Bisoffi, Zeno, Sodiomon B. Sirima, Joris Menten, Cristian Pattaro, Andrea Angheben, Federico Gobbi, Halidou Tinto, et al. 2010. "Accuracy of a Rapid Diagnostic Test on the Diagnosis of Malaria Infection and of Malaria - Attributable Fever during Low and High Transmission Season in Burkina Faso." *Malaria Journal* 9 (1): 192. https://doi.org/10.1186/1475-2875-9-192.

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. https://doi.org/10.1007/s10995-009-0553-3.

Chong, Chom-Kyu, Pyo Yun Cho, Byoung-Kuk Na, Seong Kyu Ahn, Jin Su Kim, Jin-Soo Lee, Sung-Keun Lee, et al. 2014. "Evaluation of the Accuracy of the EasyTest™ Malaria Pf/Pan Ag, a Rapid Diagnostic Test, in Uganda." *The Korean Journal of Parasitology* 52 (5): 501–5. https://doi.org/10.3347/kjp.2014.52.5.501.

Cleary, P. D., D. Mechanic, and N. Weiss. 1981. "The Effect of Interviewer Characteristics on Responses to a Mental Health Interview." *Journal of Health and Social Behavior* 22 (2): 183–93.

Durrant, Gabriele B., Robert M. Groves, Laura Staetsky, and Fiona Steele. 2010. "Effects of Interviewer Attitudes and Behaviors on Refusal in Household Surveys." *Public Opinion Quarterly* 74 (1): 1–36. https://doi.org/10.1093/poq/nfp098.

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. https://doi.org/10.1111/j.1467-985X.2008.00565.x.

Ford, Kathleen, and Anne E. Norris. 1997. "Effects of Interviewer Age on Reporting of Sexual and Reproductive Behavior of Hispanic and African American Youth." *Hispanic Journal of Behavioral Sciences* 19 (3): 369–76. https://doi.org/10.1177/07399863970193008.

Ishengoma, Deus S., Filbert Francis, Bruno P. Mmbando, John PA Lusingu, Pamela Magistrado, Michael Alifrangis, Thor G. Theander, Ib C. Bygbjerg, and Martha M. Lemnge. 2011. "Accuracy of Malaria Rapid Diagnostic Tests in Community Studies and Their Impact on Treatment of Malaria in an Area with Declining Malaria Burden in North-Eastern Tanzania." *Malaria Journal* 10 (1): 176. https://doi.org/10.1186/1475-2875-10-176.

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." *DHS Analytical Studies No. 19*, no. September. https://doi.org/DHS Analytical Studies No. 19.

Kassile, Telemu, Razack Lokina, Phares Mujinja, and Bruno P. Mmbando. 2014. "Determinants of Delay in Care Seeking among Children under Five with Fever in Dodoma Region, Central Tanzania: A Cross-Sectional Study." *Malaria Journal* 13 (1): 348. https://doi.org/10.1186/1475-2875-13-348.

Leckie, George, and Chris Charlton. 2013. “Runmlwin: A Program to Run the MLwiN Multilevel Modeling Software from within Stata.” *Journal of Statistical Software* 52 (1): 1–40. https://doi.org/10.18637/jss.v052.i11.

Lyons-Amos, Mark, and Tara Stones. 2017. “Trends in Demographic and Health Survey Data Quality: An Analysis of Age Heaping over Time in 34 Countries in Sub Saharan Africa between 1987 and 2015.” *BMC Research Notes* 10 (December): 760. https://doi.org/10.1186/s13104-017-3091-x.

O’Muircheartaigh, Colm, and Pamela Campanelli. 1999. “A Multilevel Exploration of the Role of Interviewers in Survey Non-Response.” *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 162 (3): 437–446. https://doi.org/10.1111/1467-985X.00147.

Pullum, Thomas W. 2006. “An Assessment of Age and Date Reporting in the DHS Surveys, 1985-2003.” *DHS Methodological Reports No. 5* 4. http://dhsprogram.com/pubs/pdf/MR5/MR5.pdf.

Pullum, Thomas W. 2008. “An Assessment of the Quality of Data on Health and Nutrition in the DHS Surveys, 1993-2003.” https://dhsprogram.com/publications/publication-mr6-methodological-reports.cfm.

Pullum, Thomas W., Christina Juan, Nizam Khan, and Sarah Staveteig. 2018. “The Effect of Interviewer Characteristics on Data Quality in DHS Surveys.” https://dhsprogram.com/publications/publication-MR24-Methodological-Reports.cfm.

Romay-Barja, Maria, Jorge Cano, Policarpo Ncogo, Gloria Nseng, Maria A. Santana-Morales, Basilio Valladares, Matilde Riloha, and Agustin Benito. 2016. “Determinants of Delay in Malaria Care-Seeking Behaviour for Children 15 Years and under in Bata District, Equatorial Guinea.” *Malaria Journal* 15 (March). https://doi.org/10.1186/s12936-016-1239-0.

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. https://doi.org/10.1086/268767.

Sonkong, Krit, Pornnapa Suggaravetsiri, and Sunisa Chaiklieng. 2015. “Treatment Seeking Behaviour and Prevalence of Treatment Delay among Malaria Patients along Thailand-Myanmar Border in Tak Province.” *Asian Pacific Journal of Tropical Disease* 5 (4): 279–83. https://doi.org/10.1016/S2222-1808(14)60783-0.

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. https://doi.org/10.2307/2137970.

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. https://doi.org/10.1111/rssa.12206.

WHO. 2017. “WHO | Key Points: World Malaria Report 2017.” WHO. Accessed January 5, 2019. http://www.who.int/malaria/media/world-malaria-report-2017/en/.