

**The Malawi Diffusion and Ideational Change Project 2004-06:
Data Collection, Data Quality, and Analysis of Attrition**

Philip Anglewicz¹

jimi adams²

Francis Obare-Onyango³

Hans-Peter Kohler⁴

Susan Watkins⁵

Abstract

In this paper, we evaluate the quality of data for the Malawi Diffusion and Ideational Change Project, and investigate four potential sources of bias: sample representativeness, interviewer effects, response unreliability and sample attrition. We discuss the results of our analysis and implications of our findings on the collection of data particularly for HIV/AIDS research.

¹ (A.M.) Population Studies Center, University of Pennsylvania, 239 McNeil Building, 3718 Locust Walk, Philadelphia, PA 19104-6298 Phone: (215) 898-6441 Fax: (215) 898-2124 Email: philipaz@sas.upenn.edu

² (M.A.) Department of Sociology, Ohio State University. Email: adams.644@sociology.osu.edu

³ (A.M.) Population Studies Center, University of Pennsylvania. Email: fonyango@pop.upenn.edu

⁴ (PhD) Department of Sociology, University of Pennsylvania. Email: hpkohler@pop.upenn.edu

⁵ (PhD) California Center for Population Research, University of California Los Angeles. Email: swatkins@ccpr.ucla.edu

1. Introduction

Data analysis in demographic publications typically involves hypothesis testing, where a series of factors are evaluated for their effect on a demographic outcome. It is less common that research is published that looks at the data used to carry out such analysis. This paper examines the data quality of the Malawi Diffusion and Ideational Change Project (MDICP) by investigating several sources of potential bias in a longitudinal dataset: sample representativeness, interviewer effects, response unreliability, and sample attrition.

The analysis in this paper builds on a similar study, conducted by Bignami et al (2003). We continue and expand this previous research for several reasons. First, as MDICP is a longitudinal study that has now completed four waves of data collection (1998, 2001, 2004 and 2006), some sources of bias may increase with time: as attrition increases over MDICP waves, the selection bias may also increase. In addition, changes made to the MDICP sample in 2004 warrant a re-examination of the data. As described below, a sample of approximately 1500 adolescents was added to the MDICP sample in 2004. Due to changes in the MDICP sample both by addition (adding the adolescent sample) and subtraction (attrition) we also investigate the representativeness of the MDICP sample to rural Malawi by comparing with data from the Malawi Demographic and Health Surveys (DHS). These analyses are conducted in the same manner as the Bignami et al paper to allow for comparison of the results.

2. Data

The Malawi Diffusion and Ideational Change Project (MDICP) is a longitudinal research project that examines the role of social interactions on attitudes related to contraceptive use and family planning, and HIV/AIDS knowledge and risk behavior. MDICP data collection takes place in three sites in rural Malawi, each representing one of the three regions of the country: Balaka (southern region), Mchinji (central), and Rumphi (north).

The first wave of MDICP data collection took place in 1998, at which time MDICP completed interviews for 1,541 ever-married women between 14-49 years old, and 1,065 of their husbands. The first follow-up wave collected data for the same respondents in 2001.

In 2004, MDICP returned to the field to interview the same respondents as in 1998 and 2001, but added two new data collection components. First, a sample of approximately 400 married and never-married adolescents aged 14-28 were added for each data collection site¹. This sample added completed interviews for a total of 717 women (255 never-married) and 769 men (411 never-married) to the MDICP sample.

Secondly, MDICP offered tests for HIV, gonorrhea, Chlamydia and trichosomiasis in 2004 to all MDICP respondents². Because the administration of these tests required trained personnel, MDICP recruited nurses to conduct the biomarker tests and administer a questionnaire. As a result, the 2004 MDICP fieldwork teams were split into two: main survey and biomarker collection. In 2006, MDICP returned to the field for a fourth wave of data collection, during which time biomarkers were collected for HIV only.

3. Sample representativeness

The purpose of sample surveys is to derive measurements which can be used for making inferences about a subject matter of concern for a population of interest. As such, the validity and reliability of the inferences depend on the manner in which the sample is chosen which in turn determines how representative it is of the population covered as well as on the way in which the measurements are made (Levy and Lemeshow 1999). With respect to MDICP, it is the case that the initial sample that was drawn in 1998 was not designed to be representative of the rural population of Malawi (Watkins et al. 2003).

¹ Description of sampling strategy for the 2004 adolescents can be found at:
<http://www.malawi.pop.upenn.edu/Level%203/Malawi/docs/Sampling3.pdf>

² Protocol for 2004 MDICP biomarker collection by Bignami Van-Assche et al can be found at
http://www.malawi.pop.upenn.edu/Level%203/Papers/level3_papers_byauthor.htm

This might raise concerns about whether the findings and conclusions from the study can be generalized to the rural population of Malawi or other rural areas of sub-Saharan Africa. But does the fact the sample was not designed to be representative of the rural population imply that the MDICP sample is different from the national rural population of Malawi in terms of such basic socio-demographic characteristics as age, sex, and highest educational attainment?

In this section we compare the MDICP sample characteristics with those of the rural population of Malawi obtained from the Malawi Demographic and Health Surveys (MDHS) that covered nationally representative samples. In terms of age, sex, and highest educational attainment, the MDHS characteristics pertain to the rural population aged 15 years and above in the MDHS households and not just those who were included in the individual interviews. For comparisons based on marital status at the time of the survey, the MDHS characteristics pertain to those who were included in the individual interviews (ages 15-49 years for women and 15-54 years for men). This is because information on marital status was not collected for all household members. We obtained the average distribution of the rural populations in terms of basic socio-demographic characteristics in the 1996 and 2000 MDHS (hereafter referred to as MDHS 96-00) and used this as the standard for comparison with the 1998 MDICP sample. We also compared the 2004 and 2006 MDICP sample characteristics with those of the rural population in the 2004 MDHS. The results of these comparisons are shown in Figures **Xa** to **Xd**.

<Figures Xa-Xd about here>

The 1998 MDICP sample targeted ever-married women aged 15-49 years and their husbands (Watkins et al. 2003). But it is interesting to note that in terms of such socio-demographic characteristics as age, sex, highest educational attainment, and current marital status, the sample was not any different from that of the rural MDHS 96-00 (see Figures **Xa-Xd**). We reach similar conclusions if we compare the 2004-06 MDICP sample characteristics with those of the 2004 MDHS. Perhaps, the only notable difference is that the initial MDICP sample (1998) under-represented those in the youngest (15-19 years) and oldest (50 years and above) age groups (Figure **Xa**) but those

in the middle age-groups (20-49 years) were fairly represented. It is also interesting to note that with time, these under-represented age groups gained fair representation if we compare the 2004-06 MDICP sample characteristics with those of the 2004 MDHS. Two factors contributed to this: first, the inclusion of a new sample of adolescents in the 2004 survey was responsible for the increased representation of those in the youngest age group, and second, the natural ‘aging’ of the sample increased the proportion of sample members in the oldest age group.

We also made comparisons based on other household-level attributes such as mean household size, and ownership of radio and bicycle (two of the most common household items to be found in the rural areas), as well as on fertility-related characteristics, that is, reported ever use of any method of contraception (traditional or modern), mean number of children ever born, and mean number of living children. The results are given in **Table X**. Again, the MDICP figures closely resemble those of the MDHS in terms of these characteristics. We therefore conclude that the MDICP sample, though not initially designed to be representative of the rural population in Malawi, is not any different from nationally representative rural samples in the MDHS. It is also worth noting that Watkins et al. (2003) reached a similar conclusion based on the comparison of the 1998 MDICP sample characteristics with those of the 1996 MDHS.

4. Interviewer effects

Interviewer Recruitment and Training

As described in the 2003 Bignami et al paper, the MDICP recruits interviewers from each of our three fieldwork sites, all of whom are secondary school graduates and who have fluency in English and the local language(s) spoken at each site. All interviewers are given several days of training prior to the start of fieldwork. This training is given by graduate students, most of whom have experience with MDICP fieldwork; and a group of locally-hired supervisors who are graduates of Chancellor College in Malawi and have had considerable experience with the MDICP. Although all interviewers are from the

sample area, very few knew any of the respondents in the MDICP sample³. Recent research has revealed that “inside” interviewers tend to get more consistent responses from respondents (Weinreb 2006), but the MDICP has recruited under the assumption that sensitive information (e.g. marital infidelity, number of sexual partners) that is critical for research on HIV/AIDS is less likely to be revealed to an interviewer who knows the respondent or the respondent’s family.

Interviewer Effects

To estimate interviewer effects, we use the interclass correlation coefficient, ρ , as was the technique used in the Bignami et al paper. One important difference with this paper is that, in addition to assessing the interviewer effect of main survey interviewers, we also examine the effect of nurse interviewers on responses to questions in the VCT surveys in 2004 and 2006, and also on acceptance of HIV testing and receiving test results. Furthermore, we break down the characteristics of interviewers to examine whether gender (both of the interviewer and the interviewer-respondent) has any effect on survey or HIV testing responses.

5. Response reliability

A common method of identifying response validity is by testing the reliability of responses across data waves. Any significant changes in responses that are predictable across waves (e.g. age, level of education, number of children) represents lack of response validity, which can also shed doubt on other survey responses that are not predictable but are critical for research on HIV/AIDS.

In this section, we examine two different types of response reliability. First, we utilize background information gathered separately by the main survey and the HIV testing teams in 2004 and 2006. By comparing the information collected by these teams we can gauge reliability within data collection waves. Next, we compare responses across

³ Of men and women interviewed in 2004, interviewers reported less than 5% of respondents ‘very well’ or ‘quite well’, the two categories established by Weinreb as determining interviewers who were ‘insiders’.

MDICP waves 3-4 to examine the reliability over time. To do so, we focus on background characteristics that change predictably over time: age, level of education (for the respondent and the respondent's spouse), number of children, and child mortality. In addition, we utilize another method of establishing reliability in survey responses: in 2006, survey interviewers were asked to verify whether respondents had a pit latrine. We compare the responses of the interviewer with the respondents' answer to whether they have a pit latrine.

Cross-sectional response reliability

Consistency in responses was measured within MDICP waves 3 (for women only) and 4. Table x reveals results, which reveal that differences in reporting background characteristics are typically for less than 10% of the sample in both 2004 and 2006. For example, approximately 9% of women reported a different level of education (none, primary or secondary), and 8% reported an age difference of five years or greater to MDICP interviewers.

The cross-sectional results are comparable across waves: 9% of women report different levels of education both in 2004 and 2006, and similar reports are found for women's age. Results by gender in 2006 are also similar, with differences in response consistency found in spouse's education, where men are more consistent; and number of living children, where women are more consistent in their response.

Comparison of interviewer and respondent's reports of the presence of a pit latrine are also similar across gender, with more than 15% of men and women who claim to have a pit latrine that was not found by the interviewer.

Longitudinal response reliability

In general, greater differences in reporting are found for responses across MDICP waves than within waves, as shown in table X. For example, approximately 14% of both men and women report differences in their level of education between 2004 and 2006,

compared with 9% within waves. Similarly, larger reporting discrepancies across waves are found for age.

We also included tests of reliability across waves for two questions that may be considered sensitive to Malawian respondents: use of family planning and reporting child mortality. Consistency in reporting child mortality was evaluated by the percentage of respondents reporting a larger number of child deaths in 2004 than 2006, for those who report having had at least one death. Similar to reports of total numbers of children, men show greater likelihood of inaccuracy in reporting children's death, which probably reflects the greater involvement of rural Malawian women in childbearing and rearing.

6. Attrition

One of the frequent problems facing longitudinal data collection is sample attrition, which is the failure to recruit individuals who were interviewed in one round in subsequent interview rounds. Sample attrition can produce problems in analyses using data from subsequent rounds, particularly if attritors are substantially different from non-attritors in systematic ways. All longitudinal data collection faces some level of attrition; and in settings of rural sub-Saharan Africa rates of attrition have previously been observed to be particularly high (Alderman et al. 2001; Bignami-Van Assche, Reniers and Weinreb 2003; Maluccio 2000). Numerous events can lead to sample attrition, including mobility – whether for work family (Ford and Hosegood 2005; Reniers 2001, 2003), or other reasons – mortality (Doctor and Weinreb; Ford and Hosegood 2005; Grassly et al. 2004; Timaeus and Jasseh 2004), and any number of other sources of population fluctuations. Table x.1 presents recruitment status and reason for attrition for MDICP Wave 4, where column 1 represents the full sample, and columns 2-4 splitting the sample across the three different research locations. Panel A represents figures for the full MDICP sample in 2006, while Panel B represents only those individuals in the sample who were successfully interviewed in 2004.

[Insert Table x.1 about here]

Table x.1 demonstrates several trends that will be explored further below, namely that the vast majority of attritors are missing from the sample due to migration. Men are more likely to leave the sample than women, particularly in Balaka, the research location in the southern region of Malawi. This particular finding is likely linked to high marital instability (Reniers 2003), combined with the largely matrilocal residential patterns followed in this district. Refusal rates within this study remain remarkably low, due in part to substantial resources allocated to follow up in the MDICP (adams et al. 2006; Bignami-Van Assche et al. 2003; Weinreb, Madhavan and Stern 1998).

While researchers ideally would like to keep levels of attrition as low as possible, the more important question for later usage of these data concern whether those who migrate vary systematically from those who remain in the sample. In the sections that follow, we present data demonstrating the variations that do exist between those who were subsequently successfully recruited and those who were not. First, we present a series of tables that descriptively compare attritors to non-attritors. All presented variables in **Table x.2** come from 2004 (Wave 3) data, so are limited to those respondents who were successfully interviewed in 2004.⁴ Panel A presents the figures for women, while Panel B Represents men.

[Insert **Table x.2** about here]

For both men and women, attritors are younger, have fewer children, and are less likely to be from Rumphi than are respondents who were successfully re-recruited. Several other differences by recruitment status hold only for men or women. Specifically, women who attrit, in addition to the observations above, are also more likely to be from Mchinji, to be Roman Catholic, and to have achieved higher levels of education, while they also are less likely to be Mission Protestant, and to live in a household with a pit latrine than are women who were re-interviewed in 2006. Men who attrit are more likely to be from Balaka, to be Muslim, are less worried about contracting HIV and are less likely to be

⁴ We also calculated these differences for some of the variables using the entire sample, including respondents *not* interviewed in 2004. On those select variables, we found no substantial differences from the figures presented here.

members of African Independent Churches than are men who were re-interviewed in 2006.

One perhaps significant way in which the sample changed between 2004 and 2006 is among those who were found HIV-positive in 2004. Respondents who were HIV positive in 2004 were less likely to be successfully recruited in 2006. While initial interpretation might suggest this is primarily a matter of health, additional analyses, presented in Table x.3 lend another explanation. The effect is exacerbated based on whether or not the respondent received their HIV test results in 2004. So, while respondents who were HIV-positive in 2004 are less likely to be re-interviewed in 2006, those who received this result in 2004 are even at greater risk of attrition than those who did not receive their results. For women, this effect becomes non-significant, though is likely due more to a problem of small-N than to actually representing a different trend for women and men.

[Insert Table x.3 about here]

In the last set of analyses of attrition here, we present in Table x.4 series of OLS and logistic regression models predicting several outcomes of particular interest within the MDICP studies. We estimate a global-interaction for each of four outcomes by attrition status on each of the predictor variables, and present the coefficients and summary statistics for the models (Alderman et al. 2001; Beckett et al. 1988; Bignami-Van Assche et al. 2003). Model I predicts a respondent's level of "AIDS Worry" as an ordered logistic regression, with "not worried at all" as the omitted category, "worried a little" and "worried a lot" as the other categories. Models II and III are each OLS regressions respectively predicting the number of people (other than their spouse) with whom the respondent has discussed AIDS and the respondent's reported number of sexual partners. Model IV is a logistic regression predicting whether the respondent has ever used contraception.

[Table x.4 about here]

The bold-faced rows present the interaction of each of the predictor variables with attrition status. Only two such relationships appear significant. For examples respondents

who attrit and have some education (as compared to no education) are likely to have more persons with whom they discuss issues regarding AIDS than are those who remain in the sample. Similarly, though a lower association, attritors who have lived outside their home district for more than 6 months since the age of 16 have fewer sexual partners than those who remain in the sample. While these do present two variables upon which researchers may want to proceed with caution in interpretation of analyses, the overall trend in these models suggests that while attritors are significantly different on a handful of univariate characteristics from those who remain in the sample, parameter estimates in multivariate analyses are largely unaffected by changes in the sample due to attrition.

Acknowledgements

The Malawi Diffusion and Ideational Change Project is supported by grants from the Rockefeller Foundation; NICHD (R01-HD4173, R01 HD372-276); NIA (AG1236-S3); and the Center for AIDS Research and the Center on the Demography of Aging, both at the University of Pennsylvania.

Tables and Figures

Section 3 Tables: Sample Representativeness

Table X: Comparison of MDICP and MDHS with respect to other socio-demographic characteristics

Characteristics	MDHS 96-00	MDICP 1998	MDHS 2004	MDICP 2004
Mean household size ^a	4.4 [6,507]	n/a	4.5 [11,940]	5.7 [1,767]
Ownership of radio (percent) ^a	49.3 [6,498]	56.6 [1,533]	58.8 [11,939]	69.4 [1,767]
Ownership of bicycle (percent) ^a	43.4 [6,504]	50.9 [1,534]	43.2 [11,921]	51.5 [1,766]
Ever use of contraception (percent) ^{b,c}	42.8 [5,850]	50.4 [1,509]	50.0 [10,058]	46.5 [1,569]
Mean children ever born ^b	3.3 [5,850]	4.2 [1,511]	3.2 [10,058]	4.0 [1,573]
Mean number of living children ^b	2.5 [5,850]	3.3 [1,410]	2.6 [10,058]	3.7 [1,306]

Notes: ^aMDICP figures pertain to responses by women aged 15 years and above while MDHS figures (unweighted) are based on household-level data for rural areas; ^bBoth MDICP and MDHS figures are based on responses by women aged 15-49 years; ^cEver use of contraception refers to ever use of any method, traditional or modern; MDHS 96-00 figures refer to the average of the 1996 and 2000 MDHS distributions; Sample sizes are given in brackets; n/a- not applicable because the information was not collected in the 1998 MDICP survey.

Figure Xa: Comparison of MDICP and MDHS with respect to age

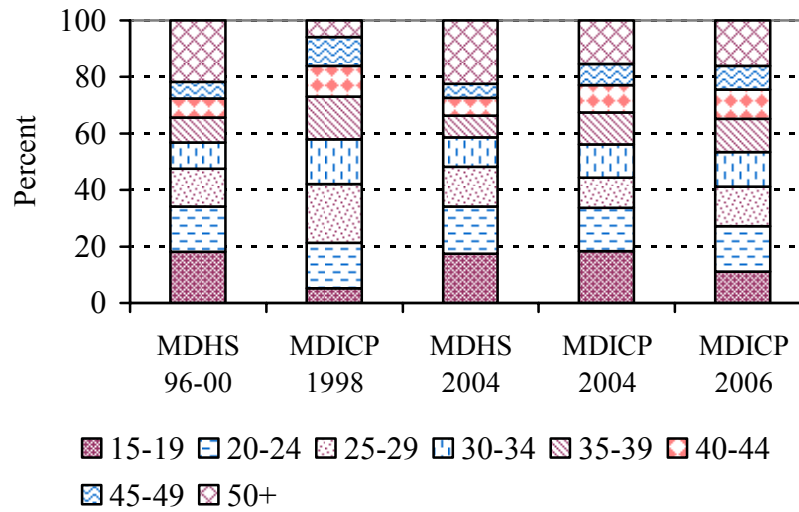


Figure Xb: Comparison of MDICP and MDHS with respect to sex

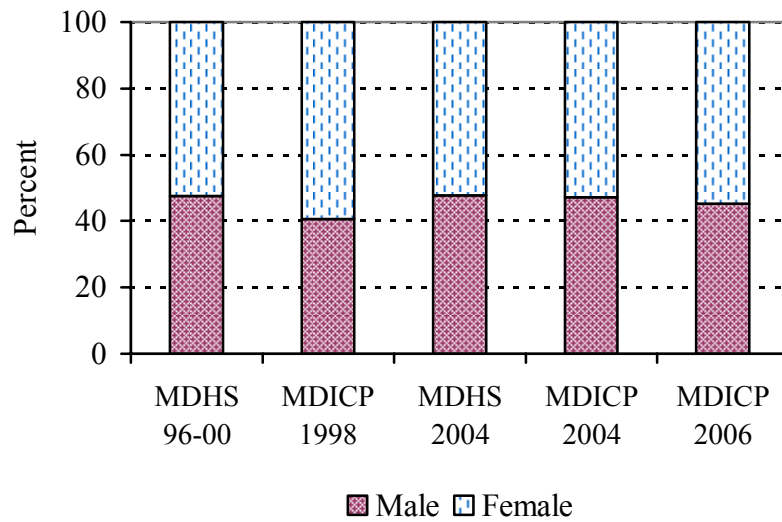


Figure Xc: Comparison of MDICP and MDHS
with respect to highest education level

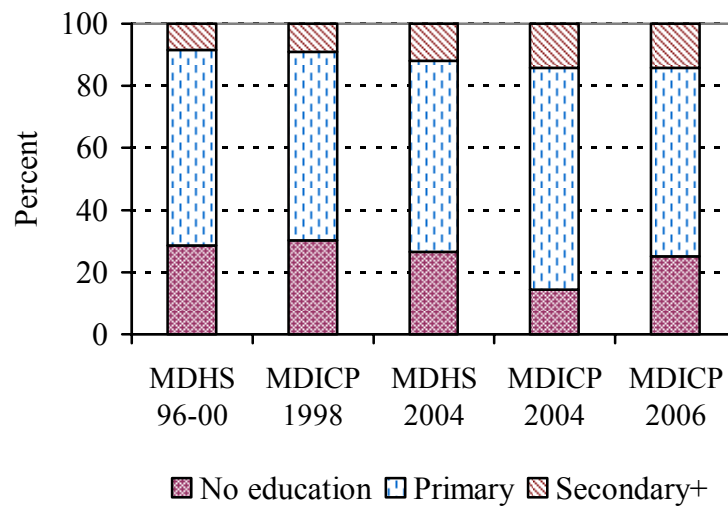
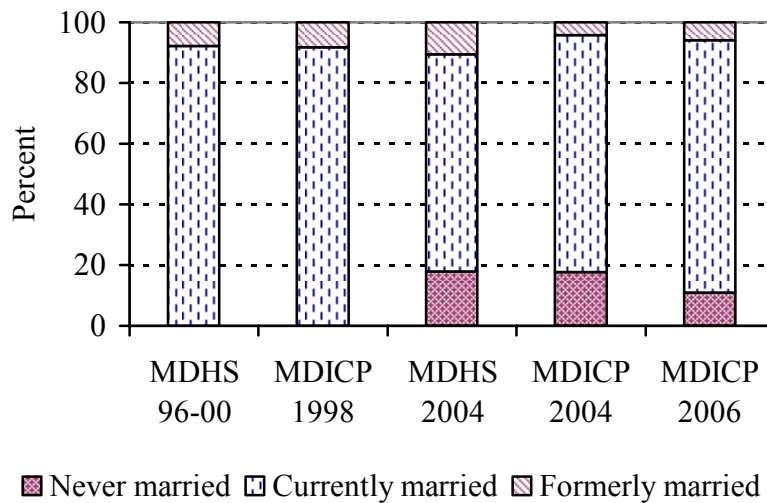


Figure Xd: Comparison of MDICP and MDHS
with respect to marital status



Section 5 Tables: Response Reliability

Within Wave Consistency				
	2004		2006	
	Women		Men	Women
Pit Latrine	n/a		17%	20%
			(752)	(1034)
Level of Education (3 cat)	9%		9%	9%
	(993)		(753)	(1031)
Spouse's Education (3 cat)	n/a		10%	23%
			(720)	(1031)
Age				
> 5 years	8%		7%	8%
	(1114)		(961)	(1109)
> 10 years	3%		2%	3%
	(1114)		(961)	(1109)
Number Living Children	14%		22%	10%
	(951)		(755)	(1032)

Across Wave Consistency			
		Men	Women
Level of Education (3 cat)		14%	13%
		(619)	(1290)
Age			
> 5 years		17%	12%
		(894)	(1103)
> 10 years		6%	5%
		(894)	(1103)
Number Living Children		16%	10%
		(820)	(1241)
Underreporting Child Mortality		35%	27%
		(389)	(614)
Ever Used Family Planning		13%	11%
		(760)	(1167)

Section 6: Attrition

Table x.1a 2006 Family Listing Outcome, for all Respondents in MDICP Sample

	Full-Sample				Balaka				Mchinji				Rumphi			
	<u>Total</u>	<u>% Men</u>	<u>% Women</u>	<u>%</u>	<u>Total</u>	<u>% Men</u>	<u>% Women</u>	<u>%</u>	<u>Total</u>	<u>% Men</u>	<u>% Women</u>	<u>%</u>	<u>Total</u>	<u>% Men</u>	<u>% Women</u>	<u>%</u>
Complete	3702	70.1	1733	68.8	1969	71.3	1192	65.4	546	60.4	646	70.2	1204	73.2	559	74.4
Refusal	63	1.2	38	1.5	25	0.9	31	1.7	17	1.9	14	1.5	22	1.3	16	2.1
Hospitalized	10	0.2	2	0.1	8	0.3	5	0.3	1	0.1	4	0.4	4	0.2	1	0.1
Dead	145	2.7	74	2.9	71	2.6	50	2.7	28	3.1	22	2.4	64	3.9	29	3.9
Not Found	304	5.8	157	6.2	147	5.3	160	8.8	86	9.5	74	8.0	25	1.5	4	0.5
Temp Abs	93	1.8	60	2.4	33	1.2	45	2.5	35	3.9	10	1.1	2	0.1	1	0.1
Moved	862	16.3	408	16.2	454	16.4	307	16.8	174	19.2	133	14.5	303	18.4	132	17.6
Other	100	1.9	47	1.9	53	1.9	34	1.9	17	1.9	17	1.8	21	1.3	9	1.2
N	5279		2519		2760		1824		904		920		1645		751	

Table x.1b 2006 Family Listing Outcome, for MDICP Respondents Successfully Interviewed in 2004

	Full-Sample				Balaka				Mchinji				Rumphi			
	<u>Total</u>	<u>% Men</u>	<u>% Women</u>	<u>%</u>	<u>Total</u>	<u>% Men</u>	<u>% Women</u>	<u>%</u>	<u>Total</u>	<u>% Men</u>	<u>% Women</u>	<u>%</u>	<u>Total</u>	<u>% Men</u>	<u>% Women</u>	<u>%</u>
Complete	2239	84.4	951	83.2	1288	85.2	730	80.4	278	75.1	452	84.0	728	82.5	312	82.8
Refusal	39	1.5	20	1.7	19	1.3	22	2.4	9	2.4	13	2.4	10	1.1	8	2.1
Hospitalized	5	0.2	1	0.1	4	0.3	2	0.2	0	0.0	2	0.4	3	0.3	1	0.3
Dead	39	1.5	20	1.7	19	1.3	10	1.1	6	1.6	4	0.7	21	2.4	9	2.4
Not Found	43	1.6	11	1.0	32	2.1	23	2.5	8	2.2	15	2.8	14	1.6	2	0.5
Temp Abs	38	1.4	29	2.5	9	0.6	20	2.2	16	4.3	4	0.7	1	0.1	0	0.0
Moved	227	8.6	100	8.7	127	8.4	93	10.2	48	13.0	45	8.4	99	11.2	42	11.1
Other	24	0.9	11	1.0	13	0.9	8	0.9	5	1.4	3	0.6	6	0.7	3	0.8
N	2654		1143		1511		908		370		538		882		377	

NOTE: Numbers presented are N and (column) percentages.

Table x.2a 2004 Descriptive Statistics by 2006 Recruitment Status – Women

<u>Women</u> Variable	<u>Reinterviewed</u>		<u>Not Reinterviewed</u>		<u>Difference</u>		
	Mean	Std Dev	Mean	Std Dev	Mean		t-test
Age	37.08	(11.19)	32.80	(10.88)	4.28	**	5.28
District							
Balaka	0.35	(0.48)	0.39	(0.49)	-0.04		-1.00
Mchinji	0.32	(0.47)	0.40	(0.49)	-0.08	*	-2.23
Rumphi	0.33	(0.47)	0.22	(0.41)	0.11	**	3.31
Religion							
None	0.01	(0.08)	0.01	(0.07)	0.00		0.17
Catholic	0.16	(0.37)	0.26	(0.44)	-0.10	**	-3.24
Muslim	0.28	(0.45)	0.26	(0.44)	0.02		0.45
Mission Prot	0.25	(0.43)	0.19	(0.39)	0.06	*	1.72
AIC	0.17	(0.37)	0.13	(0.34)	0.03		1.12
Pentecostal	0.09	(0.29)	0.09	(0.28)	0.00		0.17
Other	0.05	(0.22)	0.07	(0.25)	-0.02		-0.80
Household Owns							
Bed w/matress	0.27	(0.44)	0.25	(0.44)	0.02		0.62
Radio	0.68	(0.47)	0.71	(0.46)	-0.03		-0.75
Bicycle	0.51	(0.50)	0.53	(0.50)	-0.02		-0.48
Pit latrine	0.84	(0.37)	0.76	(0.43)	0.08	**	2.87
Education							
Secondary	0.06	(0.24)	0.12	(0.32)	-0.06	**	-3.19
Primary	0.65	(0.48)	0.67	(0.47)	-0.02		-0.65
None	0.29	(0.46)	0.21	(0.41)	0.08	*	2.47
Lived elsewhere 6+ mos	0.30	(0.46)	0.32	(0.47)	-0.02		-0.47
Number living children	4.08	(2.22)	3.31	(2.22)	0.77	**	4.58
Ever used contraception	0.51	(0.50)	0.49	(0.50)	0.02		0.48
Lifetime sexual partners	1.88	(1.81)	2.05	(2.19)	-0.17		-1.26
AIDS Worry							
Not Worried	0.32	(0.47)	0.32	(0.47)	-0.01		-0.20
Worried a Little	0.23	(0.42)	0.24	(0.43)	-0.01		-0.42
Worried a Lot	0.46	(0.50)	0.44	(0.50)	0.02		0.54

N^a 1288 85.24 % 223 14.76 %

NOTE: Two-sample T-test with unequal variance; ** p<0.01, * p<0.05

^a The difference in attrition status by gender (panel A vs. B) is significant at p <0.05.

Table x.2b 2004 Descriptive Statistics by 2006 Recruitment Status – Men

<u>Variable</u>	<u>Reinterviewed</u>		<u>Not Reinterviewed</u>		<u>Difference</u>		
	<i>Mean</i>	<i>Std Dev</i>	<i>Mean</i>	<i>Std Dev</i>	<i>Mean</i>		<i>t-test</i>
Age	43.43	(12.10)	41.05	(13.08)	2.38	*	2.29
District							
Balaka	0.32	(0.47)	0.49	(0.50)	-0.18	**	-4.63
Mchinji	0.34	(0.47)	0.35	(0.48)	-0.01		-0.35
Rumphi	0.35	(0.48)	0.15	(0.36)	0.19	**	5.13
Religion							
None	0.02	(0.15)	0.02	(0.14)	0.00		0.21
Catholic	0.20	(0.40)	0.23	(0.42)	-0.03		-0.83
Muslim	0.26	(0.44)	0.37	(0.48)	-0.11	**	-2.71
Mission Prot	0.19	(0.39)	0.14	(0.35)	0.05		1.36
AIC	0.18	(0.39)	0.09	(0.28)	0.09	**	2.81
Pentecostal	0.08	(0.27)	0.09	(0.28)	-0.01		-0.28
Other	0.07	(0.26)	0.07	(0.25)	0.00		0.14
Household Owns							
Bed w/matress	0.29	(0.45)	0.24	(0.43)	0.05		1.28
Radio	0.79	(0.41)	0.77	(0.42)	0.02		0.46
Bicycle	0.60	(0.49)	0.57	(0.50)	0.03		0.71
Pit latrine	0.88	(0.33)	0.84	(0.37)	0.04		1.44
Education							
Secondary	0.15	(0.35)	0.08	(0.28)	0.06		1.82
Primary	0.70	(0.46)	0.71	(0.46)	-0.01		-0.15
None	0.16	(0.36)	0.21	(0.41)	-0.06		-1.52
Lived elsewhere 6+ mos	0.46	(0.50)	0.48	(0.50)	-0.02		-0.43
Number living children	5.46	(3.45)	4.52	(3.10)	0.94	**	3.37
Ever used contraception	0.51	(0.50)	0.50	(0.50)	0.01		0.29
Lifetime sexual partners	4.40	(5.12)	4.59	(4.80)	-0.19		-0.49
AIDS Worry							
Not Worried	0.37	(0.48)	0.29	(0.45)	0.09	*	2.25
Worried a Little	0.26	(0.44)	0.35	(0.48)	-0.10	**	-2.69
Worried a Lot	0.37	(0.48)	0.36	(0.48)	0.01		0.25
N AIDS Discuss Partners							
N ^a	951	83.20 %	192	16.80 %			

NOTE: Two-sample T-test with unequal variance; ** p<0.01, * p<0.05

^a The difference in attrition status by gender (panel A vs. B) is significant at p <0.05.

Table x.3 Attrition Status (2006) by HIV Status (2004) and Receipt of HIV-Test Results (2004)

<u>All</u>				<u>Men</u>				<u>Women</u>			
<u>All</u>	<u>HIV Status</u>	<u>Reinterviewed?</u> <u>(2006)</u>			<u>Reinterviewed? (2006)</u>			<u>Reinterviewed? (2006)</u>			
		<u>Yes</u>	<u>No</u>		<u>Yes</u>	<u>No</u>		<u>Yes</u>	<u>No</u>		
	Positive	126	71	Positive	39	38	Positive	87	33		
		63.96	36.04		50.65	49.35		72.50	27.50		
	Negative	2283	447	Negative	1091	230	Negative	1192	217		
		83.63	16.37		82.59	17.41		84.60	15.40		
	Pearson chi2(1) = 48.793			Pearson chi2(1) = 47.901			Pearson chi2(1) = 11.836				
	Pr = 0.000			Pr = 0.000			Pr = 0.001				
<u>(Received Results) 2004</u>											
	Positive	83	42	Positive	26	24	Positive	57	18		
		66.40	33.60		52.00	48.00		76.00	24.00		
	Negative	1607	249	Negative	751	134	Negative	856	115		
		86.58	13.42		84.86	15.14		88.16	11.84		
	Pearson chi2(1) = 38.072			Pearson chi2(1) = 36.387			Pearson chi2(1) = 9.271				
	Pr = 0.000			Pr = 0.000			Pr = 0.002				
<u>(Did NOT Receive Results) 2004</u>											
	Positive	32	24	Positive	7	10	Positive	25	14		
		57.14	42.86		41.18	58.82		64.10	35.90		
	Negative	557	170	Negative	270	78	Negative	287	92		
		76.62	23.38		77.59	22.41		75.73	24.27		
	Pearson chi2(1) = 10.579			Pearson chi2(1) = 11.743			Pearson chi2(1) = 2.524				
	Pr = 0.001			Pr = 0.001			Pr = 0.112				

NOTE: Numbers presented are N and (row) percentages.

Table x.4 OLS and Logit Models Predicting Key Outcome Variables, Conditional on Attrition between 2004-2006 – Coefficients

	AIDS Worry		N AIDS Discussion Partners		N Sexual Partners		Ever Used Contraception
Attrition	2.63 (1.79)		^a		^a		0.56 (0.42)
Balaka ^b	1.96 ** (0.22)		-1.15 ** (0.36)		0.28 (0.19)		0.68 ** (0.08)
*attrition	0.72 (0.20)		-0.32 (0.92)		-0.04 (0.48)		1.10 (0.35)
Rumpi ^b	2.00 ** (0.24)		0.42 (0.38)		-0.63 ** (0.20)		0.98 (0.13)
*attrition	0.65 (0.24)		1.09 (1.21)		0.72 (0.64)		0.68 (0.28)
Gender (female=1)	1.29 * (0.13)		-1.76 ** (0.33)		-2.1 ** (0.17)		0.93 (0.10)
*attrition	0.77 (0.21)		0.04 (0.90)		-0.07 (0.47)		1.13 (0.35)
Age	0.98 ** (0.00)		0.00 (0.01)		-0.01 (0.01)		0.96 ** (0.01)
*attrition	0.99 (0.01)		0.00 (0.04)		0.02 (0.02)		1.01 (0.02)
Owns bed w/ Mattress	1.04 (0.12)		0.32 (0.36)		0.17 (0.19)		0.98 (0.12)
*attrition	1.26 (0.41)		0.55 (1.05)		-0.82 (0.55)		1.46 (0.53)
Owns bicycle	0.97 (0.09)		0.06 (0.29)		-0.1 (0.15)		1.40 ** (0.14)
*attrition	0.85 (0.21)		-0.17 (0.80)		-0.35 (0.42)		1.21 (0.33)
Seocndary Education ^c	0.92 (0.18)		0.82 (0.63)		0.5 (0.33)		1.54 * (0.33)
*attrition	0.44 (0.23)		-1.08 (1.72)		-0.93 (0.90)		1.17 (0.70)
Primary Education ^c	1.02 (0.12)		0.11 (0.37)		0.22 (0.19)		1.25 (0.16)
*attrition	0.66 (0.21)		2.09 * (1.01)		-0.36 (0.53)		0.87 (0.30)
N living children	1.03 (0.10)		0.37 (0.30)		0.29 (0.16)		1.16 (0.12)
*attrition	1.25 (0.31)		1.24 (0.81)		0.29 (0.43)		1.23 (0.35)
Lived outside district	1.03 (0.02)		0.13 * (0.06)		0.17 ** (0.03)		1.18 ** (0.03)
*attrition	1.07 (0.06)		-0.15 (0.18)		-0.23 * (0.09)		1.05 (0.07)

Table x.4 (cont'd) OLS and Logit Models Predicting Key Outcome Variables, Conditional on Attrition between 2004-2006 – Model Fit Statistics

Constant (Cut 1)	0.46 (0.11)	^a ***	5.6 (0.79)	**	3.36 (0.41)	**	2.35 (0.63)	**
Constant, Cut 2	1.32 (0.32)	^c						
-2 Log Likelihood	2307.63	-					-1458.87	
χ^2	98.44						168.51	
Pseudo R ²	0.02						0.05	
Adjusted R ²			0.05		0.13			
N	2194		2241		2217		2231	

NOTES: ** p<0.01, * p<0.05

Model I is an ordered-logistic regression, with “No Worry” as the omitted category.

Models II & III are OLS regressions and Model IV is a logistic regression.

Coefficients presented in Models I and IV are Odds Ratios and unadjusted beta's for Models II and III, and standard errors in parentheses.

^a Omitted because of multi-collinearity.

^b Mchinji is the omitted category.

^c No formal education is the omitted category.

^d The first constant presented for Model I is the constant for the first cut (between “No Worry” and “Little Worry”).

^e The second constant presented for Model 1 is the constant for the second cut (between “Little Worry” and “Very Worried”).

References

adams, j., P. Anglewicz, S.p. Helleringer, C. Manyamba, J. Mwera, G. Reniers, S.C. Watkins, and A. Weinreb. 2006. "A Rose by Any Other Name: Identifying Elusive and Eager Would-be Respondents in Longitudinal Data Collection." Philadelphia: Population Studies Center, University of Pennsylvania.

Alderman, H., J.R. Behrman, H.-P. Kohler, J.A. Maluccio, and S.C. Watkins. 2001. "Attrition in Longitudinal Household Survey Data." *Demographic Research* 5(4):79-124.

Beckett, S., W. Gould, L.A. Lillard, and F. Welch. 1988. "The Panel Study of Income Dynamics after Fourteen Years: An Evaluation." *Journal of Labor Economics* 6(4):472-492.

Bignami-Van Assche, S., G. Reniers, and A.A. Weinreb. 2003. "An Assessment of the KDICP and MDICP Data Quality." *Demographic Research* S1(2):31-76.

Bignami-Van Assche, Simona. 2003. "Are we measuring what we want to measure? Individual consistency in survey response in rural Malawi," *Demographic Research* S1(3): 77-108.

Bignami-Van Assche, Simona et al. 2004. "Protocol for biomarker testing in the 2004 Malawi Diffusion and Ideational Change Project." SNP Working Paper No.7, Philadelphia: University of Pennsylvania.

Doctor, H.V. and A.A. Weinreb. "Mortality among married men in rural Kenya and Malawi." *Journal of African Population Studies*.

Ford, K. and V. Hosegood. 2005. "AIDS Mortality and the Mobility of Children in Kwazulu Natal, South Africa." *Demography* 42(4):757-768.

Grassly, N.C., M. Morgan, N. Walker, G. Garnett, K.A. Stanecki, J. Stover, T. Brown, and P.D. Ghys. 2004. "Uncertainty in estimates of HIV/AIDS: the estimation and application of plausibility bounds." *Sex Transm Infect* 80 Suppl 1:i31-38.

Levy, Paul S. and Stanley Lemeshow. 1999. *Sampling of Populations: Methods and Applications*, Third Edition. New York: John Wiley & Sons, Inc.

Maluccio, J.A. 2000. "Attrition in the Kwazulu Natal Income Dynamics Study, 1993-1998." Pp. 50 in *Food Consumption and Nutrition Division*. Washington, D.C.: International Food Policy Research Institute.

Reniers, G. 2001. "The post-migration survival of traditional marriage patterns: Consanguineous marriages among Turks and Moroccans in Belgium." *Journal of Comparative Family Studies* 32(1):21-+.

—. 2003. "Divorce and Remarriage in Rural Malawi." *Demographic Research* Special Collection 1:Article 6.

Timaeus, I.M.and M. Jasseh. 2004. "Adult Mortality in Sub-Saharan Africa: Evidence from Demographic and Health Surveys." *Demography* 41(4):757-772.

Watkins, Susan C., Eliya M. Zulu, Hans-Peter Kohler and Jere R. Behrman. 2003. "Introduction to Social Interactions and HIV/AIDS in Rural Africa." *Demographic Research* S1(1): 2-29.

Weinreb, Alexander A. 2006. "The Limitations of Stranger-Interviewers in Rural Kenya." *American Sociological Review* 71(4): 1014-1039.