

Description of 2004 MDICP adolescent sampling process

Alex Weinreb

awein@mscc.huji.ac.il

In 2004, the MDICP added an adolescent panel of men and women aged 15-24 to its existing sample of ever-married women and their husbands. This panel can be broadly characterized as a *disproportionate stratified random sample* with stratification on sex and age. Here I describe the basic steps taken to generate the sample.

Introductory steps & considerations

During the 2004 round of MDICP data collection, two household rosters were collected in each sampled community. The first was collected from all households in the sampled villages—that is, MDICP and non-MDICP households—by a preliminary field-team of "household listers." Across the three research sites, these teams visited households 1-3 weeks before the main teams of survey interviewers. The second roster was collected by these main survey interviewers. That is, it was incorporated into the primary survey instrument administered to all women respondents.

Although the two rosters differ somewhat in their details, both asked the respondent for names, ages and marital status of all *de facto* household residents. As described below, the adolescent sample draws on both those rosters.

Our overall aim was to interview approximately 350 adolescents in each of the three research sites, with an even breakdown by sex. Based on prior experience in African rural areas, and informal conversations in Malawi, we expected adolescents to be more mobile than people in the main MDICP sample. Consequently, in generating our sample lists, we assumed a relatively low 80% interview completion rate. This demanded that we sample approximately 450 adolescents per site.

It was immediately apparent that a simple random sample of all adolescents in the sample frame would not address important analytic needs (and desires) of leading MDICP personnel. We therefore adopted a slightly more complicated sampling strategy. We subsequently had to modify this approach a little more once fieldwork began. In general, then, sampling took place in three roughly chronological stages. I describe each in turn.

Stage 1

Since we wanted to maximize the number of adolescents from existing MDICP households—this would allow for interesting and potentially important within-household linkages—we sampled 100% of adolescents listed in the main survey interviews.¹ These interviews accounted for approximately 32% of all adolescent interviews.

¹ The mechanics of this process was as follows. In each site, the interviewer teams began fieldwork with questionnaires for main survey respondents. As the questionnaires came in from the field each day, adolescents listed on the household roster were tagged and their details entered into a spreadsheet. As the pace of interviewing main survey respondents slowed—ie., after reaching approximately 3/4 of those that we would eventually reach—selected interviewers were assigned to interview adolescents listed on the rosters of MDICP households.

Stage 2

In order to generate the remaining 75% of the adolescent sample, we drew on the prior household rosters collected by listing teams. At this stage we encountered our second problem, highlighted in Table 1. Simply, because the difference in mean age at marriage between men and women expresses itself most significantly at these very ages, drawing a simple random sample of adolescents would have generated a dataset with high levels of multicollinearity between age and marriage. In the Rumphi sample, for example, 29.9% of all listed male adolescents were married, as opposed to 65.4% of all women adolescents.²

Table 1. Proportion of adolescent listees married, by sublocation, gender and age

Age	MCHINJI		RUMPHI	
	Proportion Married	Proportion Married	Proportion Married	Proportion Married
	MEN	WOMEN	MEN	WOMEN
15	0.01	0.03	0	0.05
16	0.01	0.10	0.15	0.15
17	0.04	0.15	0.12	0.36
18	0.03	0.39	0.11	0.51
19	0.06	0.56	0.10	0.70
20	0.16	0.68	0.18	0.75
21	0.37	0.75	0.30	0.82
22	0.44	0.76	0.50	0.90
23	0.52	0.66	0.55	0.96
24	0.56	0.82	0.75	0.88

Notes: Data are from MDICP general household roster

Since marriage and transitions into marriage were key substantive interests of a number of MDICP-related researchers (e.g., Reniers, Poulin, Kohler, Clark), I adopted a sampling strategy which would help minimize the threat these high levels of multicollinearity between age and marriage posed to subsequent analysis. Specifically, I used different sampling fractions by age for men and women.

The main aim of the fractions was to push the samples of men and women adolescents toward a more similar profile of married and unmarried. Relative to their proportions in the actual population, this meant: (1) *increasing* the number of married adolescent men in the sample by weighting the men's sample to the higher adolescent ages; and (2) *decreasing* the number of married adolescent women in the sample by weighting the women's sample to the lower adolescent ages.

A subsidiary aim of the fractions took into account the longitudinal nature of the project. That is, knowing that the MDICP planned to collect data from these adolescents in 2006 or 2007, I wanted to generate a sample that was more unmarried than married, since we could expect many of the unmarried individuals

² Data on marital status of household members was missing in a small proportion of the listers' household roster. In the Rumphi sample lists, it was missing in a much larger proportion of rosters – for 38% of all male adolescents and 23% of female adolescents (during the translation into Tumbuka the column was inexplicably dropped). Whether or not we had data on adolescents' marital status from the household rosters, they were included in the sampling frame. If and when they were selected for the actual sample lists (procedures discussed below), interviewers assigned to them were simply outfitted with both a married and not-married questionnaire, and instructed to ask their marital status prior to beginning the formal part of the interview, that is, after greetings and introduction.

to marry in the intersurvey period. In other words, I wanted to generate a sample that, following data collection in 2006/7, would allow for a prospective study of marriage among a cohort of young Malawian adults. The relative dearth of empirical studies on marriage in rural sub-Saharan Africa would make this prospective study quite valuable.

The specific sampling fractions by age, sex, and site are presented in Table 2. So, too, are the numbers of adolescents selected from the general household rosters. Note that both the proportion married and overall number of adolescents appearing on the lists varied across the sites, but since we also needed to generate a specific number of adolescents in each site, the sampling fractions had to be specific to each site.

Table 2. Sampling fractions used to identify the 2004 MDICP adolescent respondents from general household lists, and overall size of sample lists (N), by sublocation, gender and age .¹

Ages	Mchinji		Balaka		Rumphi	
	Men	Women	Men	Women	Men	Women
15	0.09	0.34	0.60	0.75	0.50	0.70
16	0.09	0.34	0.60	0.75	0.50	0.70
17	0.09	0.34	0.60	0.75	0.50	0.70
18	0.11	0.17	0.45	0.15	0.50	0.35
19	0.11	0.17	0.45	0.15	0.50	0.35
20	0.11	0.17	0.45	0.15	0.40	0.35
21	0.11	0.07	0.45	0.10	0.40	0.18
22	0.40	0.07	0.30	0.10	0.40	0.18
23	0.40	0.07	0.30	0.10	0.40	0.18
24	0.40	0.07	0.30	0.10	0.40	0.18
N	228	230	212	197	198	194

Notes: ¹ These sampling fractions cannot be used to weight the representativeness of adolescents selected from the MDICP households, since 100% of the latter were selected for interviews, irrespective of age (and marital status).

Stage 3

As hinted at by the variation in size of sample across research sites in the final row of Table 2, once fieldwork began, it became apparent that our questionnaire completion rate was lower than the 80% that we expected. In order to reach our target of ~350 individuals per site – or at least get closer to it – we consequently added some individuals to the original sample lists. We did this by drawing them at random from the household rosters. In other words, among these final additions, there was an equal probability of being selected into the sampling list.