Institute of Business and Economic Research

Center for International and Development Economics Research

(University of California, Berkeley)

Year 2008 *Paper C*08-151

Tracking, Attrition and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1)

Sarah Baird
UC San Diego

Joan Hamory
UC Berkeley

Edward Miguel
UC Berkeley and NBER

Tracking, Attrition and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1)

Abstract

Understanding the possible pitfalls of survey data is critical for empirical research. Among other things, poor data quality can lead to biased regression estimates, potentially resulting in incorrect interpretations that mislead researchers and policymakers alike. Common data problems include difficulties in tracking respondents and high survey attrition, enumerator error and bias, and respondent reporting error. This paper describes and analyzes these issues in Round 1 of the Kenyan Life Panel Survey (KLPS-1), collected in 2003-2005. The KLPS-1 is an innovative longitudinal dataset documenting a wide range of outcomes for Kenyan youths who had originally attended schools participating in a deworming treatment program starting in 1998. The careful design of this survey allows for examination of an array of data quality issues. First, we explore the existence and implications of sample attrition bias. Basic residential, educational, and mortality information was obtained for 88% of target respondents, and personal contact was made with 84%, an exceptionally high follow-up rate for a young adult population in a less developed country. Moreover, rates of sample attrition are nearly identical for respondents who were randomly assigned deworming treatment and for those who were not, a key factor in the validity of subsequent statistical analysis. One vital component of this success is the tracking of respondents both nationally and across international borders (in our case, into Uganda), thus we discuss in detail the costs and benefits of tracking movers. Finally, we study KLPS-1 data quality more broadly by examining enumerator error and bias, as well as survey response consistency. We conclude that the extent of enumerator error is low, with an average of less than one recording error per survey. Errors decrease over time as enumerator experience with the survey instrument increases, but increase over the course of multiple interviews within a single day, presumably due to fatigue. We do find some evidence that the enumerator-respondent match in terms of gender, ethnicity, and religion correlates with responses regarding trust of others and religious activities, suggesting some field officer bias on sensitive questions. Reporting reliability is analyzed using respondent re-surveys. These checks show high levels of consistency across survey/re-survey rounds for the respondent's own characteristics and personal history,with lower reliability rates on questions asked about others' characteristics. The steps taken in the design of KLPS-1 to avoid common errors in survey data collection greatly improved the quality of this panel dataset, and provide some valuable lessons for future field data collection projects.

Tracking, Attrition and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1)*

Sarah Baird, University of California, San Diego
Joan Hamory, University of California, Berkeley
Edward Miguel, University of California, Berkeley and NBER

August 2008

Abstract: Understanding the possible pitfalls of survey data is critical for empirical research. Among other things, poor data quality can lead to biased regression estimates, potentially resulting in incorrect interpretations that mislead researchers and policymakers alike. Common data problems include difficulties in tracking respondents and high survey attrition, enumerator error and bias, and respondent reporting error. This paper describes and analyzes these issues in Round 1 of the Kenyan Life Panel Survey (KLPS-1), collected in 2003-2005. The KLPS-1 is an innovative longitudinal dataset documenting a wide range of outcomes for Kenyan youths who had originally attended schools participating in a deworming treatment program starting in 1998. The careful design of this survey allows for examination of an array of data quality issues. First, we explore the existence and implications of sample attrition bias. Basic residential, educational, and mortality information was obtained for 88% of target respondents, and personal contact was made with 84%, an exceptionally high follow-up rate for a young adult population in a less developed country. Moreover, rates of sample attrition are nearly identical for respondents who were randomly assigned deworming treatment and for those who were not, a key factor in the validity of subsequent statistical analysis. One vital component of this success is the tracking of respondents both nationally and across international borders (in our case, into Uganda), thus we discuss in detail the costs and benefits of tracking movers. Finally, we study KLPS-1 data quality more broadly by examining enumerator error and bias, as well as survey response consistency. We conclude that the extent of enumerator error is low, with an average of less than one recording error per survey. Errors decrease over time as enumerator experience with the survey instrument increases, but increase over the course of multiple interviews within a single day, presumably due to fatigue. We do find some evidence that the enumerator-respondent match in terms of gender, ethnicity, and religion correlates with responses regarding trust of others and religious activities, suggesting some field officer bias on sensitive questions. Reporting reliability is analyzed using respondent re-surveys. These checks show high levels of consistency across survey/re-survey rounds for the respondent's own characteristics and personal history, with lower reliability rates on questions asked about others' characteristics. The steps taken in the design of KLPS-1 to avoid common errors in survey data collection greatly improved the quality of this panel dataset, and provide some valuable lessons for future field data collection projects.

* Acknowledgements: We would like to thank our collaborators on the broader KLPS project for their comments and work that made this paper possible, including Chip Bury, David Evans, Jenna Johnson-Hanks, Matthew Jukes, Michael Kremer, and Karen Levy. Chris Blattman, Blastus Bwire, Lorenzo Casaburi, Jamie McCasland, Owen Ozier, Sarah Y. Siegel, Peter Wafula, and Ethan Yeh provided excellent research assistance. Moshe Buchinsky, Duncan Thomas, and Jeff Kling have provided valuable comments on the survey and tracking research design. This project would not have been possible without the generous financial support of the U.S. National Institutes of Health (NICHD), National Science Foundation, Social Science Research Council, and the University of California, Berkeley.

1. Introduction

High quality survey data is essential for answering many key research questions in development economics. To reduce omitted variable bias and gather reliable information on geographic and socioeconomic mobility, survey data efforts must involve a panel dimension and also successfully track individuals over time. Currently, few long-term longitudinal data collection efforts have been undertaken in less developed countries, and most existing efforts make limited attempts to track individuals that have migrated, often because earlier survey rounds failed to gather or maintain the necessary contact information. The additional costs associated with locating movers frequently leads survey enumerators to track only those residing in nearby areas, or to ignore such respondents altogether. Tracking individuals who migrate across international borders has rarely been attempted in a less developed country context.[1] Yet, failure to follow these respondents can result in high rates of sample attrition, and potentially lead to substantial biases in estimates in populations where migration is widespread.

This paper describes respondent tracking, attrition, migration and other data quality issues in a recent panel (longitudinal) survey of Kenyan youth, the Kenyan Life Panel Survey Round 1 (KLPS-1), and highlights lessons for future panel data collection efforts in less developed countries. The KLPS-1 sought to track and interview 7,500 individuals over a seven year period. These individuals attended primary school in Busia, a rural district of western Kenya bordering Uganda, in 1998, and took part in the Primary School Deworming Project (PSDP), which provided medical treatment for intestinal worm infections to 75 Busia primary schools between 1998 and 2001. Treatment was randomly phased in, with 25 schools receiving treatment beginning in 1998, another 25 beginning in 1999 and the final 25 beginning in 2001.[2] In 1998, these individuals ranged in age from 6 to 19 and were, on average, 12 years

[1] Collected at the same time as the KLPS-1, the Kagera Health and Development Survey 2004 tracked Tanzanian migrants who moved to Uganda (visit http://www.edi-africa.com/research/khds/tracking.htm for more information). Note that this study tracked households, not individuals. The Mexican Family Life Survey, fielded in 2002 and 2005, is a nationally representative survey that tracked individuals migrating from Mexico to the United States (Arenas and Herrera 2007).

[2] For more information on the PSDP, refer to Miguel and Kremer (2004).

old. The KLPS-1 traces these pupils through their teenage years, a time of dramatic life changes as well as potentially high geographic mobility.

Migration over 1998-2005 was significant: among the 88 percent of target respondents for whom we obtained residential location information, nearly 19 percent were living outside of Busia district at the time of the follow-up survey. If those we were unable to survey were disproportionately likely to have migrated, as seems plausible, rates could be even higher. Most migrants moved elsewhere within Kenya, with approximately half moving to an urban location. A considerable number also moved to Uganda, creating an added difficulty of tracking individuals across international borders. These cross-sectional "snap-shot" figures further understate total migration among our rural Kenyan youth sample: nearly 32 percent of surveyed individuals report living outside of Busia District at some point during the seven-year survey period, and 8 percent lived outside of Kenya (at least temporarily).

Despite the great deal of geographic mobility, tracking adolescents in the KLPS-1 was very successful: contact was effectively made with 84 percent of target respondents, and 83 percent were surveyed (the remaining 1% either had died or refused to participate in the survey). This success stems from a number of specific strategies undertaken by the field team, including the use of a two-stage tracking methodology, a team-oriented approach, and a willingness to embrace increasingly widespread cell phone technology. This paper describes the strategies employed to successfully locate target respondents, and highlights those we consider most useful. We hope these lessons will prove valuable for others implementing panel survey data collection in less developed countries.

Along with a discussion of the logistics of tracking, this paper also explores survey attrition in the KLPS-1 and its implications for future analysis. Because tracking is costly and time consuming, it is essential to understand the extent to which data quality improved as a result of tracking efforts, including whether or not these efforts were successful in reducing attrition bias concerns. We estimate whether individual baseline characteristics as well as outcomes differ between those who migrated and others. If, for example, we were to find that migrants have higher wages, a failure to track these individuals could

lead us to underestimate wages among our sample and bias regression analyses of the determinants of wages. We also provide a discussion of the monetary costs of tracking.

This first set of analyses follows in the footsteps of Thomas *et. al* (2001), which examines tracking and attrition in the first three waves of the Indonesia Family Life Survey (IFLS, fielded in 1993, 1997 and 1998). The authors emphasize the importance of tracking movers in panel data collection. Indeed, the IFLS followed households and a subset of individuals from each household, providing a rare glimpse into issues associated with locating movers in a less developed country context. The authors conclude that there are indeed significant differences in observable characteristics between those who left their baseline home and those who did not, with differences increasing in distance moved.

There are three aspects of the KLPS-1 that are exceptionally well-suited to the study of individual migration, and allow us to build on this earlier work in order to shed more light on issues of tracking and attrition. First, KLPS-1 tracked individuals rather than households. In most existing panel datasets, tracking is household-based, implying that a household is considered successfully tracked if any member of the household is located. Focusing on the individual greatly increased the difficulty of the tracking exercise. Second, the KLPS-1 follows school-age children and young adults, a group that is very geographically and socially mobile. This mobility further complicates data collection. Adolescents frequently migrate for schooling, marriage, or labor market opportunities, and we explore the unique challenges survey efforts may face when tracking this important demographic group. Finally, the KLPS-1 involved tracking individuals to any and all destinations, including across international borders (mainly into Uganda)—a practice which remains unusual in longitudinal data collection.[3] This component of the study required obtaining necessary research permissions from a second country and managing the logistics of international field work, both of which require extensive planning.

[3] Due in part to the rich ethnic and geographic diversity of Indonesia, not to mention a population dispersed across thousands of different islands, tracking of movers in the IFLS was limited to those who had migrated within the thirteen study provinces. However, data collected from contacts left behind suggests that only a very small part of the sample actually did migrate outside the thirteen IFLS provinces, and only a minute fraction went abroad.

Survey data can potentially suffer from other quality issues beyond attrition—including data entry error, field enumerator bias and recording errors, and respondent recollection mistakes—and we also explore those issues here. The design of the KLPS-1 incorporates several elements that allow us to explore these concerns in great detail. For instance, to both correct for and better understand data entry errors, a coding process was devised to allow data entry mistakes to be easily corrected using survey hard copies, and to discover which question formats lead to the most errors.

A record of survey enumerator completion errors was generated during these data quality checks, allowing us to study which field officers made the most mistakes. To supplement this analysis, a field enumerator survey gathered individual field officer characteristics which were then correlated with error rates. This analysis indicates that, while overall error rates are very low, more educated enumerators tend to make somewhat fewer data errors, while females on our team committed slightly more errors. Furthermore, we find evidence of enumerator fatigue (error rates increase with the number of interviews completed in a given day) and enumerator learning (error rates decrease as experience administering surveys increases). We also find some indication that the gender, ethnic group or religion "match" between respondents and enumerators were sometimes associated with survey responses, especially on sensitive questions related to sexual assault and trust.

Finally, we estimate response reliability using data on the subset of KLPS-1 respondents who were re-surveyed shortly after their original survey was administered (three months later on average). We find high rates of consistent responses, especially for questions asked about the respondent's own characteristics or personal experiences.

The remainder of the paper proceeds as follows. Section 2 briefly reviews tracking and sample attrition in other panel data collection efforts in both wealthy countries and in less developed countries. Section 3 introduces the KLPS-1, discusses execution of the survey, and examines attrition and migration patterns in the data. Sections 4 and 5 explore other data quality issues through analysis of data entry mistakes, enumerator characteristics and survey-resurvey comparisons. The final section concludes.

2. Attrition Rates in Longitudinal Datasets

Large sample panel datasets allowed for major advances in applied microeconomic research in wealthy countries since as early as the 1960s, and began playing a role in less developed countries with the onset of the World Bank's Living Standards Measurement Surveys (LSMS) in the late 1980s.[4] One of the best known longitudinal data collection efforts is the Panel Survey of Income Dynamics (PSID), launched in the United States in 1968. The PSID was originally designed to study poverty dynamics, but has since been used to analyze a wide range of outcomes. By 1981, 40 percent of the original respondents had left the sample, potentially leading to significant biases in analysis using the remaining respondents (Becketti *et al.* 1988). Part of the reason for this high attrition rate was that the PSID did not attempt to survey individuals who had moved.

Thomas *et al.* (2001) highlight a number of surveys in developing countries that have also suffered from substantial attrition due to failure to track movers. The Cebu Longitudinal Health and Nutrition Survey in the Philippines, which surveyed pregnant women 14 times over two years, suffered an attrition rate of nearly two-thirds through completion of all survey rounds. Alderman *et al.* (2001) explore attrition in two other studies that did not track movers—the Bolivian Pre-School Program Evaluation Household Survey and Kenyan Ideational Change Survey—and note that the attrition rates in these surveys are also high, at 35 percent for the Bolivian study and 28 percent for women (and 41 percent for couples) in the Kenyan study.[5]

Panel datasets that *do* explicitly follow individuals and households when they move are somewhat more common in wealthy countries than in less developed countries. One survey of particular relevance to our work is a 2002 assessment of the U.S. Moving to Opportunity (MTO) program. MTO was administered by the Department of Housing and Urban Development, and offered vouchers to low

[4] Smaller panel dataset efforts began slightly earlier in less developed countries, most notably, the ICRISAT's household surveys in Indian villages from 1975-1985.

[5] It is important to note that although attrition rates in these surveys are high, this does not necessarily imply that subsequent analyses using the data will be biased. In fact, research addressing attrition in these surveys finds that the attrition do not significantly bias regression results for many outcomes of interest (see Fitzgerald *et al.* 1998; Alderman *et al.* 2001; Falaris 2003).

income households to move out of high poverty neighborhoods in Baltimore, Chicago, Los Angeles and New York, making tracking of households essential to the project impact evaluation. The 2002 survey followed individuals, including both adults and children aged 5-19, in the 4,500 households who enrolled in the program from 1994-1997 (Kling and Liebman 2004). The project achieved an impressive "effective response rate" of 90 percent for adults and 88 percent for children (Orr *et al.* 2003).[6]

Another relevant U.S. survey is the National Longitudinal Survey of Youth 1997 (NLSY97), comparable to the KLPS-1 in that it focuses on tracking adolescents. Following in the footsteps of the successful NLSY79, which tracked 14-21 year olds starting in 1979 with an attrition rate of only 22 percent through 2002, the NLSY97 followed a representative sample of 8,984 U.S. adolescents aged 12-17. By 2003, six years after the beginning of the study, only 14 percent of the sample had attrited (Olsen 2005). Although the MTO and NLSY97 surveys were very successful in limiting attrition, many of the techniques used to guarantee success, such as phone interviews and mail-in responses, are not always feasible in a developing country context.

In less developed countries, most long-run panel data sets track households rather than individuals, and few follow respondents who have migrated. The gold standard of recent less developed country panel data sets is the Indonesia Family Life Survey (IFLS), which tracked both households and individuals in three waves in 1993, 1998, and 2000.[7] This effort was extremely successful, locating almost 88 percent of original household members in 2000 and 91 percent of individuals specifically targeted for tracking in cases of separation from the baseline household.[8] Among household members aged 5-19 at baseline, an age group comparable to the KLPS sample, 80 percent were found. This somewhat lower tracking rate is attributed to the fact that individuals in this age range are highly mobile

[6] The MTO survey followed a two-phase tracking design. In the first phase, the entire sample was tracked. At the close of this effort, search resources were focused on a random sub-sample of the remaining unfound target respondents in order to reduce non-response bias. Evaluators thus report the cumulative survey response rate in the form of the "effective response rate" (ERR), which is a function of the response rates of the main sample (MRR) and the more intensively-tracked sub-sample (SRR): ERR=MRR+SRR*(1-MRR). For a more detailed explanation, see Orr *et al.* (2003). The KLPS employed a similar tracking design (see Baird *et al.* 2007 for more information.)

[7] An additional wave was conducted in 1998 with 25% of the sample. IFLS4 has not yet been made public.

[8] Only a subset of individuals were selected for tracking in cases where members of a baseline household had separated from the original household.

and thus more difficult to find. In addition, only some of these children were specifically targeted for tracking if not found within their original household (Strauss *et al.* 2004).[9]

The KwaZulu-Natal Income Dynamics Study (KIDS) followed a sample of South African households between 1993 and 1998. This survey successfully located 84 percent of the 1,393 households in the original sample after five years, where success was defined as re-interviewing at least one member of the 1993 household. Without tracking movers, the rate would have been less than 80 percent (Maluccio 2000). However, only 78 percent of surveyed children between the ages of 6 months and 6 years were successfully re-interviewed after the five year hiatus (Alderman *et al.* 2001).[10]

The Institute of Nutrition of Central America and Panama (INCAP) implemented a nutritional supplement program to a sample of 2,393 children between 1969 and 1977. Between 2002 and 2004, INCAP attempted to re-survey these participants. Enumerators interviewed 84 percent of those individuals alive and known to be living in Guatemala, though this constituted only 66 percent of the baseline sample. Attrition would have been much higher if movers had not been tracked within Guatemala: 8 percent of the original sample migrated to nearby villages, 23 percent to Guatemala City, and 9 percent to other cities and towns in Guatemala (Grajeda *et al.* 2005).

A comparison between the surveys noted above that do track migrants and those that do not suggests that following movers is critical to reducing sample attrition. Drawing from the lessons learned from this literature, we now turn to tracking in KLPS-1 and issues of attrition, migration and data quality.

3. Tracking, Attrition, and Migration in the KLPS-1

3.1 Background to the KLPS

[9] It is important to note that although the IFLS tracked the majority of movers, it did place some restrictions on who was tracked. In particular, respondents who had moved outside the 13 IFLS provinces to other provinces in Indonesia, or who moved outside Indonesia, were not tracked.

[10] It is not immediately clear whether this age group of children is more or less mobile than the teenage individuals on which we focus in the KLPS-1 data.

The Primary School Deworming Project (PSDP) was started in early 1998 by International Child Support (ICS) Africa, a non-governmental organization located in the western Kenyan district of Busia.[11] This program provided medical treatment for intestinal worms to over 30,000 children enrolled in 75 primary schools in the Funyula and Budalangi divisions of Busia District. Treatment was randomly phased in to the schools between 1998 and 2001, with twenty-five schools beginning treatment in 1998, another twenty-five in 1999, and the final twenty-five in 2001. Since this area of western Kenya is characterized by particularly high intestinal helminth infections rates, this intervention could potentially lead to significant improvements in the health and education of these children.[12] Using the school-level random assignment to generate treatment and comparison groups, previous work has examined the effect of the health intervention on short-term child educational and health outcomes (Miguel and Kremer 2004). Deworming treatment was found to have substantial health and school participation benefits, though academic test performance was not significantly affected.

The Kenyan Life Panel Survey Round 1 (KLPS-1) was launched in mid-2003 in an effort to re-interview a representative sub-sample of 7,500 of these children who were initially in grades 2-7 at the start of the PSDP in 1998. All respondents were administered a general survey instrument (known as the Household, or H, Module) that collected information on household composition, education, health, nutrition, fertility, labor market outcomes, migration, religion, social capital, crime victimization, and political attitudes. Other modules detailing household consumption expenditures (the Expenditures, or E, module), marriage and contraception practices (the Family, or F, module), and cognitive aptitude (Cognitive, or C, module) were administered to subsets of these individuals.[13] The main objective of the KLPS-1 survey effort was to create a panel database of information on these adolescents for the period 1998-2005, allowing researchers to undertake a detailed analysis of longer term deworming treatment effects, as well as other issues involving a temporal component, such as migration.

[11] International Child Support Africa is the NGO formerly known as Internationaal Christelijk Steunfonds Africa.
[12] Infection rates were over 90% in 1998 according to Kenya Ministry of Health parasitological surveys.
[13] In this paper, we focus on data quality issues in the KLPS-1Household (H) Module.

3.2 Respondent Tracking

It was essential to locate as many of the 7,500 selected target respondents as possible. Needless to say, searching for individuals in rural Kenya is an onerous task. Tracking migrants is particularly problematic in the absence of basic tools such as forwarding addresses and phone books, which rarely exist in rural Africa. This difficulty was especially salient for the KLPS-1 since it followed young adults during their teens and early twenties, an age group that is likely to be quite mobile due to evolving marriage, schooling, and labor market opportunities. In the survey area it is also fairly common for children to move back and forth between households of relatives and acquaintances, sometimes relocating great distances in the process. Furthermore, transportation infrastructure is lacking in many regions, making it difficult to reach many locations. For instance, enumerators had to hire small boats to reach respondents living on remote Kenyan and Ugandan islands in Lake Victoria, and traveled to some homes by foot or bicycle. While telecommunications infrastructure is also underdeveloped, as we discuss below, increasing numbers of respondents had access to cellular phones during the KLPS-1 survey period, and the expansion of this technology greatly aided respondent tracking.

Enumerators followed a two-part strategy to locate KLPS-1 target respondents.[14] In the "local tracking" phase, the survey team visited respondents' former primary schools and family homes within Busia District. This phase was itself split into two components, School-Based Tracking (SBT) and Home-Based Tracking (HBT). Given that the PSDP program had administered deworming treatment through schools, and thus the primary contact between researchers and children occurred there, teachers and administrators were the principal means of locating respondents at the start tracking work. During SBT, enumerators visited PSDP schools and established the status of each KLPS-1 target respondent, for instance, among the possibilities: (i) currently attending a PSDP school; (ii) currently attending a non-PSDP school within Busia District; (iii) attending school outside of Busia District; (iv) not in school, lives

[14] For additional information on tracking procedures, refer to Baird *et al.* (2007).

nearby, at home during the day; (v) not in school, lives nearby, not expected to be at home during the day; (vi) not in school, lives away from home; (vii) unknown; or (viii) deceased.

In some cases, especially for younger respondents still of primary school age, individuals were easily found and surveyed at school, and no further tracking was needed. For the rest of the target respondents, HBT aimed to identify their current residential location or, for the small numbers of deceased pupils, to collect some basic information on their outcomes before death. In situations where the individual had moved away, local relatives and acquaintances were interviewed for additional information on the target respondent's new residential location as well as details on when the individual might visit Busia (facilitating the scheduling of a relatively low cost local interview).

Once all 75 former PSDP schools had been visited, as well as local homes, the enumeration team embarked on the second phase, "long-range tracking", which entailed traveling outside of Busia District within Kenya, and to neighboring Uganda. In practice this phase was much more arduous than local tracking. The survey team was generally welcomed in homes and communities throughout Busia District and other nearby districts and areas, due to its affiliation with ICS, a well-known local NGO. However, this name recognition did not extend to other parts of Kenya and Uganda, and field officers frequently found it necessary to authenticate themselves through official documentation and even phone calls to the target respondents' friends and family back in Busia (whom they had met during local tracking).

Tracking in urban areas was especially difficult, as locating an individual without the assistance of the tight-knit communities more characteristic of rural areas required a great deal of time and footwork. To locate respondents who had moved to the sprawling housing projects surrounding Nairobi, field officers were forced to work in pairs to navigate potentially dangerous urban slums. Other respondents had relocated to islands in Lake Victoria, requiring enumerators to hire boats to trek around remote areas of the lake, and often to obtain separate local permissions from officials on each island visited. Still other individuals had migrated to mainland Uganda, and in addition to needing new permissions for conducting surveys in such cases, enumerators had to adapt to working under foreign laws and customs. One

respondent was even found and surveyed in London, England by a KLPS co-investigator working there at the time (Matthew Jukes).

The 7,500 target respondents were randomly divided into two tracking "waves" to more easily allow for improved tracking methods and survey instruments to be incorporated over time. Wave 1 began in August 2003. By early June 2004, the main phase of Wave 1 local and long range tracking were complete, and 64 percent of the Wave 1 sample respondents had been found. To minimize sample attrition bias within existing time and financial constraints, a representative sample of the remaining unfound target respondents was selected to be tracked in the "intensive" phase of Wave 1, consisting of both local and long range visits. Approximately one quarter of the still unfound Wave 1 target respondents were randomly sampled for intensive tracking, and additional resources were brought to bear to find these individuals during June and July 2004. Fifty-six percent of this subsample was found. Thus, the effective Wave 1 tracking rate was (0.64) + (0.36)*(0.56) = 84 percent.[15]

Wave 2 of KLPS-1 began in August 2004, and continued through June 2005, by which time local and long-range tracking efforts had found 68 percent of the target respondent sample. Approximately one quarter of the remaining unfound target respondents were again sampled for the intensive tracking phase, and were sought from June through August 2005, by which time 52 percent of the intensive sub-sample had been found. The Wave 2 effective tracking rate was (0.68) + (0.32)*(0.52) = 85 percent, almost identical to Wave 1.

Panel A of Table 1 presents the population-weighted figures for final tracking status of targeted individuals.[16] Eighty-four percent of target respondents were effectively located, and 83 percent were surveyed; the remaining one percent were either deceased or refused to be surveyed.[17] Older individuals

[15] Similar to the methodology explained above for the MTO study, the formula used here is effective tracking rate ETR=ITR+(1-ITR)*STR, where ITR is the tracking rate from the initial tracking phase, and STR is the tracking rate from the sub-sample (intensive) tracking phase.

[16] Results on tracking are presented using the sample of individuals who were either found, or not found but searched for during intensive tracking, in order to properly weight the latter individuals using the "effective tracking rate" methodology explained above.

[17] The sample used for these figures includes all individuals surveyed, otherwise located but not surveyed, and sampled for intensive tracking. This is a total sample of 5,602 individuals.

are somewhat harder to locate: this observation is supported by the nearly monotonically increasing attrition rates across initial 1998 grade and age group (see Tables 2 and 3). Tracking rates are very similar across the three deworming treatment groups, although slightly higher for Group 1 students. Tracking rates differ significantly by gender, with 86 percent of males but only 82 percent of females located. One reason for this difference could be that females rather than males typically move to their new spouse's home upon marriage. This is consistent with the migration figures presented in Panel B of Table 1.[18]

Refusals by target survey respondents (or their parents) were rare, at less than one percent. Rates are similar across the two tracking waves, treatment groups and by gender, although they do vary slightly across age and grade groups. Less than one percent of target respondents were found to be deceased, and even for these individuals enumerators often collected some basic information from relatives on their schooling and health outcomes before their death.

The overall tracking rate of 84 percent compares favorably to other recent panel data collection efforts in less developed countries discussed in Section 2 above, particularly when compared to studies that also tracked adolescents and young adults. In Section 3.4 below we discuss what we believe to be the most important elements behind KLPS's respondent tracking success.

3.3 Migration Patterns

One of the biggest challenges faced in tracking was, of course, the geographic mobility of the target respondents. The concerted multi-stage tracking process, however, yielded high rates of tracking success, as described above. Even when migration made it impossible to directly survey target respondents, residential location information was often gathered through interviews with relatives and friends at the target respondent's former school or home. Combining the H-module data with information gathered from relatives and friends, we have detailed residential location information for 88 percent of the KLPS-1 target sample (this is an effective tracking rate taking into account the intensive stage, as above).

[18] Migration patterns are explored further in Hamory and Miguel (2008).

Panel B of Tables 1 through 3 details the residential location data for various respondent subgroups. Consistent with the attrition rates described above, residential information is more likely to be missing for older respondents: the fraction of target respondents with residential location information is, for the most part, monotonically decreasing with age and grade. The collection rate of residential location information is similar across the three PSDP experimental treatment groups. Males are more likely to have residential location information than females, consistent with H-module attrition patterns.

Nineteen percent of those with residential location information were no longer living in Busia District at the time of KLPS-1 data collection, suggesting a great deal of migration, with five percent living in neighboring districts within Kenya as well as Uganda's Busia district (right across the border) and 14 percent living further afield. Of those who moved outside of Busia and its neighboring districts, nearly three quarters were living in one of Kenya's five large urban areas – Nairobi, Mombasa, Kisumu, Nakuru and Eldoret.[19]

These figures are likely an underestimate of the true migration rate. First, and most importantly, individuals missing residential information are also disproportionately likely to have migrated. As evidence of this, the higher H-module attrition rates characterizing older respondents, females and those interviewed in Wave 2 are accompanied by higher recorded migration rates for these groups. Second, above we describe migration at a single snapshot, which excludes individuals who had migrated at some earlier point during 1998-2005 but had returned to Busia before KLPS-1 was collected. Including those who moved at any time during 1998-2005, the total migration rate outside of Busia and neighboring areas jumps from 14 to nearly 25 percent.

Figure 1 maps residential locations at the time the KLPS-1 was collected, based on the H-module data as well as interviews with family and friends; Figure 2 contains only H-module information. While

[19] We define urban areas as those with populations of greater than 150,000. Our measure of urban residence is imperfect in that there are some districts considered "rural" even within these metropolitan areas. We ignore this distinction, which in any case is often outdated, since outskirts of areas that were rural in the past are today often integral parts of the city, or its expanding suburbs.

much migration was confined to areas in Kenya and Uganda bordering Busia, many respondents moved to distant urban centers, especially Nairobi, or even international locations like Kampala and London.

3.4 Factors Contributing to High Tracking Rates

The KLPS-1 effective tracking rate of 84 percent compares favorably to other well-known panel data collection efforts. We attribute this tracking success to several factors, namely: (i) the decision to track individuals who had migrated (of course); (ii) the careful design of tracking protocols, (iii) the quality, training, and teamwork of the enumerators; (iv) the increasing availability of cellular phone technology; and (v) positive word-of-mouth about the survey experience through respondents' social networks.

We described previously the use of School Based Tracking and Home Based Tracking methodologies. These tools allowed field officers to utilize the local community's knowledge in their search for target respondents. The information obtained through these two methods allowed enumerators to more accurately identify target respondents' current residential locations and schedule interviews in the most efficient way possible. The field team leader then compiled this information into a single summary "tracking sheet" for each target respondent. This sheet provided the field teams with a centralized database of precise locations where target respondents were thought to live, study and work, and was updated whenever new information was gathered. Constant communication between the field team members and the team leader ensured information was kept current and could be immediately incorporated into team travel and interview appointment plans.

The teamwork and dedication of the field officers was also invaluable. The field leaders showed tremendous judgment and composure under often trying field circumstances, given the inevitable logistical challenges that emerged. All field officers were born in Kenya, and were fluent in one or more of the Luhya ethnic group dialects widely spoken in the study area. Familiarity with the local language and customs increased respondents' and their relatives' comfort levels during interviews.

Given the high mobility of the age group targeted in the KLPS-1, it was impractical to assign the search for particular respondents to a single field officer. Rather information gathered on respondents'

whereabouts was shared immediately among the enumeration team, allowing the field officers to work together to locate a single individual; for instance, if we discovered during Home Based Tracking in Busia that a target respondent had recently moved to Nairobi, a field officer currently in Nairobi to survey another respondent could quickly schedule an additional interview.

The project was run in collaboration with a respected local non-governmental organization (NGO), and their reputation combined with positive word of mouth among respondents was also critical. Community leaders and residents in Busia District were almost always eager to assist in locating respondents, likely in large part because they were familiar with the NGO's work, and may have even received some assistance form the NGO in the past. In one case, a father brought his son, a target respondent, to the main NGO office in Busia town after hearing that enumerators sought to interview him. Although it was repeatedly made clear to respondents that they would receive no additional financial assistance by taking part in the survey, we cannot rule out the possibility that some individuals chose to participate in anticipation of some future assistance, particularly since the NGO had provided extensive assistance to local youth in the past, and continued to do so during the study period.

The use of cellular telephone technology also contributed to KLPS-1's tracking success. Official estimates show a massive increase in mobile phones in Kenya between 1998 and 2004, from just 9,000 phones to over 1.5 million (Republic of Kenya 2007), and the numbers have continued to soar since.[20] KLPS-1 H-module surveys indicate that over 12% of respondents had cell phone access, but this more than doubled over the course of the survey, from under 8 percent in Wave 1 (2003-2004) to over 16 percent in Wave 2 (2004 -2005). Cell phone usage among respondents living in urban areas is even higher: 8 percent of rural KLPS-1 respondents report a cellular phone contact number, in contrast to 62 percent of those living in urban areas.

This increase in mobile phone usage proved to be very important in locating target respondents. In the absence of straightforward and accessible home address information, cell phone contact numbers

[20] According to Mbiti (2008) 80% of Kenya is now covered by a cell-phone network and the number of subscribers has increased from 6.48 million in June 2006 to 9.30 million in June 2007.

enabled survey enumerators to track down and set up meetings with many respondents, allowing for tracking to more closely resemble that undertaken in a rich country context. Using enumerator phone call logs from the last quarter of the survey period (spring and summer of 2005), we find that over 130 respondents were tracked via contact phone information, and this led to successful survey enumeration in 95 cases.[21] This method was extremely useful for respondents who had left Busia district, especially those who had moved to urban areas, and saved enumerators a great deal of time in tracking down respondents in person, by asking around for respondents in their neighborhood, for instance. In addition, cell phone technology came in handy for meetings with target respondents who were suspicious of enumerators, as field officers were able to call the relative or friend back in Busia who had provided the target respondent's tracking information and allow them to confirm the enumerator's identity directly. The continued spread of mobile phones in Kenya makes us optimistic about success in Round 2 of KLPS data collection, which is currently in the field.

Finally, the most important determinant of KLPS-1 tracking success stemmed from the decision to track movers, including those that migrated locally, elsewhere in Kenya, and internationally. Failure to track these movers would have increased attrition from 16 to 30 percent. Tracking those who lived outside Busia increased the tracking rate by 8.5 percentage points, and tracking those who moved internationally (nearly all to Uganda) increased the tracking rate by 3.1 percentage points.

3.5 Empirical Analysis of Attrition and Migration

Even though the KLPS-1 survey completion rate was quite high, it remains less than one hundred percent. To the extent that attrition is non-random, in other words, unsurveyed individuals differ from those surveyed, analysis using the KLPS-1 data could potentially be biased. In particular, if attrition was specifically correlated with the 1998 PSDP intervention (i.e., if certain treatment groups had more attrition than others) then estimated deworming impacts could be subject to bias.

[21] Unfortunately we do not have such detailed information for Wave 1 or the first half of Wave 2 of tracking.

We use a probit framework to assess whether any individual or household characteristics (including assignment to deworming) are related to survey attrition in the KLPS-1. Table 4 provides descriptive statistics of the key regression variables, and Table 5 displays the regression results.[22] Columns 1-4 include the full tracking sample, while columns 5 and 6 are run on a sub-sample for which we have a richer set of 1998 baseline data.[23] In an important result for later analysis, assignment to deworming through the initial intervention (in particular, years assigned to deworming treatment and school assignment to deworming treatment group) is not significantly related to attrition. Females and children who were older at baseline are significantly less likely to be surveyed, on the order of 6 percent less for females and more than one-third less for the oldest individuals at baseline, relative to their younger counterparts. Gender and school average mock test score interacted with treatment group are also significant, suggesting that females as well as children from better schools who received more years of deworming treatment were more likely to be surveyed; however, we cannot reject the null hypothesis that years of deworming treatment and all of its interactions are jointly equal to zero.

We next explore individual and household characteristics related to migration, the survey costs associated with locating the interviewing these movers, and the differences in various life outcomes associated with moving versus remaining in the baseline location. First, we repeat the attrition analysis in Table 5, this time using an indicator for whether the individual was living outside of Busia District at the time of the KLPS-1 survey as the dependent variable. Table 6 displays the same analysis using only H-module data. Deworming treatment assignment is not significantly related to whether the individual migrated. The coefficient estimate on the female indicator is statistically significant and positive, indicating that females are 20 percent more likely to migrate outside of Busia District than males, a substantial difference driven by local marriage practices. The sub-sample specification in column (5)

[22] Note that only children attending primary school in 1998 were included in the KLPS sampling frame. This population is still generally representative of the adolescent population in western Kenya, however. According to 1998 DHS data, 85% of children in Western Province aged 6-15 were enrolled in school. The data indicate that the pupil population drops by roughly one third between grades 3 and 7, with the largest decreases in cohort size occurring between grades 1 to 2, 2 to 3 and 7 to 8.

[23] Baseline survey data exist for individuals who were present on the pre-announced day the test or survey was administered, and includes only students in grades 3 through 7 in 1998.

further suggests a positive relationship between individuals' baseline academic test scores and later migration, such that the change from one standard deviation below the mean test score to one standard deviation above the mean increases the probability of migration outside of Busia District by 31 percent.[24] Focusing specifically on the characteristics of individuals who were living internationally at the time of enumeration (Table 7), we find once again that years of deworming treatment is not statistically significant. In general, individual characteristics appear to be much more weakly associated with decision to leave Kenya, gender is unrelated, although older adolescents are slightly more likely to move abroad. Taken together, the results in Tables 6 and 7 indicate that age, gender, and academic test scores are significantly associated with migration within Kenya.

It is important to weigh the financial costs of KLPS-1 tracking with its benefits. Table 8 provides estimates of KLPS-1 data collection and tracking costs. Costs increased significantly between the first and second tracking waves, due to both an increase in survey modules administered to each individual (only the H-module was administered in Wave 1, while a subset of respondents were also administered the E, F, and C modules in Wave 2) and an increase in the difficulty of locating respondents as they age. Further, costs of the "intensive" tracking phase per individual found were roughly four times the costs of regular tracking. This finding falls remarkably in line with the sampling scheme used in intensive tracking, in which intensive tracking individuals each stand in for approximately four target respondents. Most importantly, nearly 64 percent of the intensive sample had migrated outside Busia, as opposed to just over 17 percent of the regular sample, suggesting that the higher costs of traveling to locate and interview migrants was a key determinant of the higher intensive phase costs.

Given the high cost of tracking migrants, the question becomes what was gained in sample size and sample representativeness from following those who left Busia District, and to do so, it is important to understand whether including these individuals in the analysis in any way changes our results. Table 9 summarizes the results of estimating the impact of deworming treatment assignment on KLPS-1 health

[24] This relationship is explored more thoroughly in Hamory and Miguel (2008). Note that the definition of migration used in that paper is different from the one employed here.

and education outcomes for the full sample versus the sample of migrants.[25] We focus on the specification in column (2) from Table 5, but include an indicator for whether the pupil migrated and an interaction between this indicator and years of deworming treatment. Several interesting differences emerge between migrants and the full sample, both in the characteristics of the population that migrated, and in the impact of years of deworming treatment. Migrants are 1.7 centimeters shorter than non-migrants on average. Treatment has a significantly larger correlation with height for this group – for migrants, one additional year of treatment increased height by 0.4 centimeters, while for the full sample it only increased height by 0.1 centimeters, a difference that is significant at the 90 percent confidence level. These patterns are consistent with the idea that the less healthy individuals (at baseline) gained more from deworming treatment. Treatment also had a larger effect on weight gains for migrants, such that migrants with one additional year of deworming treatment are nearly one half kilogram heavier. Significantly different treatment effects also appear in outcomes related to education and labor. Migrants are less likely to be enrolled in school, yet more likely to be "not idle" (i.e., either in school or working), pointing to the fact that many individuals migrate for work. The deworming treatment effect on being "not idle" is significantly smaller (3 percent) for migrants. In sum, the substantial monetary cost associated with tracking KLPS-1 movers was worthwhile since migrants have substantially different characteristics than non-migrants and these appear to interact with deworming treatment in non-trivial ways.

This section provided a perspective on the challenges, innovations, and successes of KLPS-1 tracking. We now turn to other data quality issues.

4. Other Data Quality Issues: Survey Enumerators

4.1 Background

Another important set of data quality concerns include enumerator recording error or bias. Field officer bias, or "role-independent interviewer effect" as it is called in the demography literature (Bignami-Van Assche et al. 2003), refers to the possible effects of interviewer characteristics on respondents' answers.

[25] These results are presented in greater detail in Baird et al. (2007).

The idea is that respondents may be more likely to personally identify with particular enumerators, and this rapport may determine how he/she chooses to respond to survey questions. To illustrate, a female respondent may be more likely to trust and thus open up to a female enumerator on a sensitive topics related to sexual behavior or assault. Field officer bias could also arise if enumerators impose his or her attitudes or views on the respondent; for example, a religious enumerator may intentionally (by using a religious greeting or metaphor) or unintentionally (by wearing a visible cross necklace) influence the respondent to also portray herself or himself as devout.

There is an extensive literature on field officer bias in rich countries, dating back at least to Hyman *et al.* (1954). In *Interviewing in Social Research*, the authors provide the groundwork for successful interviewing strategies. Fowler and Mangione (1990) summarize more recent evidence and provide advice for future survey efforts. The literature on field officer bias in less developed countries, however, is more recent and more limited. A number of papers focus on the impact of enumerator gender on answers to fertility and sexual behavior questions, but these studies vary in their conclusions. Axinn (1991), using data from Nepal, and Blanc and Croft (1992) using Ghanaian data, conclude that field officer gender can affect responses to sensitive questions, with male enumerators generally gathering under-reports of many sexual behaviors. In contrast, Becker *et al.* (1995) find that Nigerian respondents were more likely to open up to male enumerators about contraception methods, including female respondents. Bignami-Van Assche *et al.* (2003) find mixed results for Kenya and Malawi, depending on the gender of respondent and the question addressed, with stronger evidence of bias in Malawi then in Kenya. Note that none of this research focuses on adolescents.

4.2 Enumerator Characteristics

We summarize enumerator characteristics in Table 10. Over the 2003-2005 survey period, the enumeration team was composed of twenty field officers in total, roughly evenly split between men and women, although at any given point in time the field team was typically composed of ten to fourteen enumerators. Just over half of the team originated from Busia district, and most are of Luhya descent (the

numerically dominant local ethnic group). Nearly three-quarters of enumerators were married, all had at least a secondary school degree, and most have also obtained either a college or university level diploma. Many had previous experience working for a non-governmental organization, though only a third of the team had surveying experience prior to work on the KLPS-1.

4.3 Field Officer Bias

Turning now to an examination of field officer bias in the KLPS-1, we focus our analysis on whether responses to sensitive survey questions vary depending on whether the gender of the field officer and the respondent match, whether the tribe of the field officer and respondent match, and finally whether the religion of the field officer and respondent match. With the exception of female respondents randomly selected for the F-module in Wave 2, there was no systematic matching between enumerators and respondents along gender lines, or any other dimension, so these matches should be largely arbitrary.

The questions we use are listed in Table 11, and these include inquiries related to pregnancy, contraceptive use, sexual assault, trust of others, religion, and voting in a recent national election. In addition, we examine six other questions that we think should be relatively non-sensitive as controls, namely, whether the individual attended school in 2003, whether anyone in the household farmed, whether the respondent had held a job in the last year, number of members in the family, whether the respondent received any professional training, and the number of meals the respondent ate yesterday (the day before the interview).

Table 12 displays the results from a regression of each outcome on the gender of the respondent, a variable indicating whether the gender of the respondent and the enumerator match (our proxy for when field officer bias is likely to be a concern), and the interaction of these two.[26] The interaction is included to reveal whether the gender match makes more of a difference to females or males. Panel A contains the questions we consider to be of a sensitive nature. Looking first at the sexual behavior questions, we see

[26] For the gender match analysis we dropped those individuals that were administered a fertility module since in these cases there was always a female enumerator interviewing a female respondent.

that gender match is not a significant predictor overall. The gender match of the enumerator and the respondent is important for the trust variables, where respondents are significantly more likely to indicate that they trust members of their own tribe (22 percent) or church (11 percent) if the genders match, an effect that is stronger for male respondents. The magnitude of the effects is equally large for questions related to political and religious participation. Males are more likely to admit that they do not go to church regularly, or vote, to male enumerators, while for females the impact is reversed. It is interesting that the impact of the gender of the respondent and the enumerator matching goes in very different directions across these variables.

Panel B contains our control outcomes, variables that a prior reasoning suggest should be less sensitive than those just discussed. As expected, we see no direct impact of enumerator-respondent gender match on farming, work, training, household size and number of meals outcomes. The gender match does affect one of these six control variables, namely, whether the respondent responds that he/she attended school in 2003. This result may reflect the fact that the respondent wants to portray him or herself as educated to the enumerator, who is relatively well-educated. However, the impact is fairly small in comparison to those found for the sensitive questions. In addition, in the case of farming and number of meals, the gender match and interaction term are found to be jointly significant at the 90% level—a fairly weak result, but one that may call for further investigation

We focus on tribe in Table 13 and include as independent variables whether the respondent was Luhya or not, a variable indicating whether the tribe of the respondent and enumerator match (again defining tribe as Luhya or not) and the interaction of these two. Tribe match seems to matter for the reporting of pregnancy and trust of one's own church/mosque members, but only with fairly weak statistical significance. The interaction and the respondent-enumerator tribe matching are jointly significant across the trust variables, again pointing to the sensitive nature of these questions in a country such as Kenya where ethnic divisions are highly politicized. Interestingly, across all the trust variables we see that those of the Luhya tribe (the majority in the study area) seem to be more trusting than those of other tribes, although there is no decisive explanation for this result. Turning to the control variables in

Panel B of Table 13, we find that the respondent-enumerator tribe match is a significant predictor for several variables, suggesting that there may be cultural norms in Kenya regarding ethnicity that cause the respondent-enumerator tribe match to matter even for answers to standard socioeconomic variables. Fortunately for the analysis of KLPS-1, this issue is of minor concern since the vast majority of both enumerators (85%) and respondents (95%) are ethnic Luhyas.

Turning to the religion results in Table 14, we see that the respondent-enumerator religion match seems to affect reports of trust of one's own tribe (individuals are 17% more likely to report trust) and church/mosque members (individuals are 12% more likely to report trust), and in both cases the signs are positive and significant at 99% confidence. The religion result is intuitive since respondents may be reluctant to report mistrust of their religious group to another group member.

In sum, the gender, tribe and religion "match" of the enumerator and the respondent affects responses to certain sensitive survey questions in the KLPS-1, with the gender match being potentially the most important. Effects on non-sensitive control questions–on socioeconomic and demographic matters–are generally less consequential, although the tribe match of the enumerator and the respondent does seem to matter even for some of these variables. While the findings are broadly reassuring about data quality in KLPS-1, it may nonetheless be worth controlling for the respondent-enumerator match along these key demographic dimensions during statistical analysis with this data.

There is one additional issue related to survey data biases worth mentioning. In 4.5 percent of KLPS-1 H-module surveys, individuals other than the enumerator and the respondent were present during the interview. In the majority of cases, these additional observers were a parent, guardian or village elder. It is possible that the presence of this person could make the respondent less likely to reveal sensitive information, or otherwise influence the answers that he/she might give. That said, the presence of such a person may be endogenous to respondent characteristics or answers, making this a less convincing test of response bias than the respondent-enumerator analysis presented above.

Table 15 compares responses in surveys where only the enumerator and the respondent were present to those where additional individuals were in the room. We control for age in this analysis, since

younger respondents are generally those that are more likely to have an extra individual present These results indicate that respondents are more likely to report trusting members of their own tribe and church/mosque, more likely to report regular church attendance and voting in the 2002 election, and less likely to report sexual assault, when someone else is present at the interview. However, the presence of others does not seem to be related to reporting of contraception practices or pregnancy, suggesting the bias is not universal across all sensitive questions. The respondent is less likely to report school attendance in a recent year, as before.[27] Overall, it does appear that having someone else present at the time of enumeration could potentially affect survey responses, reinforcing the importance of consistently holding private interviews, and suggesting it may be useful to control for the presence of other individuals at the interview. That said, it is worth re-emphasizing that we cannot confidently give these correlations a causal interpretation since unobserved respondent characteristics could be correlated with the presence of others during the interview.

4.4 Survey Enumeration Mistakes

Simple mistakes in the process of survey enumeration can also reduce data quality. Given this possibility, regular checks on the quality of field enumeration were an integral part of the quality control process in the KLPS-1. As completed household surveys were returned to the field office, a representative subset for each enumerator were selected by field research managers and evaluated according to key criteria (described below). Enumerator-specific reports on survey quality were then generated periodically (usually monthly) to provide them with constant feedback on their job performance.

Table 16 displays summary statistics of these performance reviews. On average, each field officer enumerated 260 surveys, 20 (or 8 percent) of which were chosen for a quality check. The average time to complete a household survey (H-module) was 62 minutes, and each field officer completed approximately 2.5 surveys per day of field work, the rest of their day being taken up with preparations and travel time.

[27] The reports of female respondents appear to be more sensitive to the presence of others at the interview (results available from authors upon request), although, once again, the possibility that the presence of others is endogenous to respondent characteristics cannot be ruled out.

Enumeration quality was very good overall. Surveys typically had only 1.42 blank fields, 2.03 "don't know" responses (enumerators were trained to probe respondents in order to minimize these, although not all can be considered errors if the respondent genuinely did not know an answer), and 0.81 other errors (for example, skip pattern violations). Enumeration quality was fairly consistent across the two waves, with some improvements over time.[28]

We examine how error rates vary as a function of enumerator characteristics in columns 1 and 2 of Table 17, where column 1 includes the full set of characteristics and column 2 only includes those characteristics collected for all twenty field officers. The two variables that appear to most consistently drive the error rate are enumerator education and age: those with more education have significantly lower error rates, while errors increase with age in column 1 (the age range of enumerators was 20 to 38). In column 2, however, when age is excluded as a control, enumerator education level is no longer statistically significant at traditional confidence levels. In both specifications female enumerators make slightly more errors than males.

We also have a measure of the total number of mistakes made in each survey from checks of reported values in the entered data, and consistencies across survey questions, carried out in the electronic database; this "mistake dataset" was briefly mentioned above (in Section 1). Mistakes identified in the database were checked against survey hard copies (for the entire KLPS-1 database), and then corrected electronically when appropriate.[29] Columns (3) and (4) of Table 17 repeat the specifications of columns (1) and (2), using this alternate measure of enumeration error as the dependent variable. This time the only significant explanatory variable is gender, where again females make slightly more errors than males. Taking all four regressions into account, field officer characteristics do not appear to be important determinants of survey error in the KLPS-1, with the possible exception of gender. We believe this

[28] Note that 70% of the re-survey checks were in Wave 2 of KLPS-1.
[29] Field officer mistakes from this "mistakes dataset" include "hard copy is missing the entry" (code 4), "hard copy is incorrect" (code 5), or "hard copy and cleaned data entries are the same for entries that were flagged as erroneous"(code 6).

finding could be related to the careful training and constant evaluation of field officers, as well as the conscientious design and pre-testing of the KLPS-1 survey module.

We next examine error rates and other enumerator characteristics, in particular the relationship with the number of surveys enumerated. For instance, there could be a learning curve in survey administration, such that experienced field officers make fewer mistakes. However, fatigue could be a factor over the course of a single day. Table 18 provides an analysis of number of mistakes by survey to capture these potential effects. The coefficient estimates do provide evidence of a learning curve in survey administration, albeit one that quickly flattens out with further experience (note the positive quadratic term in surveys completed). There is evidence of fatigue increasing survey error rates later in the day, but the magnitude of this association is rather small, given that field officers complete an average of 2.5 interviews per day.

Although error rates in the KLPS-1 H-module are low overall, this average could potentially conceal greater variation across survey sections. Table 19 summarizes error rates by section, and shows that there in fact minimal variation across sections, with low error rates nearly across the board. The exception is Section 21, which is somewhat different than the rest of the survey since it serves as an interview summary filled in by enumerators after the conclusion of the survey, and thus does not contain information collected from the respondent. The reason for the large number of mistakes here stems from the failure of some enumerators to fill in the section at all, since it apparently was not always adequately emphasized by field managers. The two sections with the next highest error rates are Sections 4 (family characteristics) and Section 13 (migration). These two sections require very detailed reporting from respondents, and reasonably complicated skip patterns for enumerators, two plausible reasons explanations why error rates are slightly higher.

Overall, enumerator bias and survey completion mistakes were minimized in the KLPS-1 due to intensive enumerator training and continuous quality checks throughout data collection. Enumerators were trained for at least three weeks in the office before beginning KLPS-1 field work, and then most for an additional three weeks of survey piloting. Regular meetings were held throughout the process to

ensure that the field team was comfortable with every single question in the various KLPS-1 survey modules. Team meetings were also held throughout KLPS-1 data collection to discuss issues that arose in the field, as well as to clarify common errors found during enumerators' performance evaluations. Along with this work in Kenya, U.S.-based researchers combed through the data to document mistakes and correct them in the database, often after extensive survey hard copy checks when discrepancies arose. These efforts worked to ensure that KLPS-1 data is of consistently high quality.

5. Other Data Quality Issues: Respondents

5.1 Overview

While the PSDP dataset and the KLPS-1, taken together, contain longitudinal information along many dimensions on the same individuals over time, several key KLPS-1 variables rely on retrospective questions. Retrospective questions require respondents to recall events that happened in the past, and the elapsed time opens up the possibility of recall errors, and thus bias in statistical analysis.

A large literature has emerged on the quality of retrospective information, mainly focused on surveys undertaken in wealthy countries. Although there is disagreement as to which factors impact recall most, much of this literature points to a relationship between data quality and respondent characteristics – with some evidence that more educated and higher socioeconomic status respondents provide more accurate responses – as well as the importance of the length of time since the event in question and its salience (Sudman and Bradburn 1973, Mathiowetz and Duncan 1988, and Beckett *et al.* 2001). A smaller body of work has concentrated on less developed country survey responses. Yet since individuals living in less developed countries are generally poorer, less educated, and subject to different cultural norms regarding surveys than people in wealthy countries, there are plausible reasons to expect the quality of retrospective information to be systematically worse in poor countries (Beckett *et al.* 2001).

The majority of retrospective data quality analyses have focused on whether the details of a particular event are remembered over quite long horizons, generally periods of several years. The gold

standard in this line of research compares respondent reports to a reliable outside source.[30] This is an excellent strategy as long as the benchmark source is more reliable than the individual report; Beckett *et al.* (2001) argue this is not always the case. Furthermore, this type of analysis is a tall order for retrospective data collected in less developed countries, where benchmark information from official sources may not exist (or may not be easily accessible).

A second strategy used in the literature compares responses from the same individual collected at different points in time to gauge reliability. Beckett *et al.* (2001) uses repeated retrospective histories collected in two waves of the Malaysian Family Life Survey enumerated twelve years apart to study recall of health events, and Smith and Thomas (2004) use this same data to assess the quality of reported migration histories. This methodology allows for a comparison of consistency in reports over time, though we cannot be assured that either captures the "truth".

The analysis of KLPS-1 that follows contains a number of innovations. We focus on variables that play a key role in many studies of individuals and households in development economics, including age, educational attainment, parent education, and the migration indicator variable. We explore survey recall reliability through comparisons of responses among individuals who were surveyed twice; resurveys were typically conducted within months of the original interview. While Beckett *et al.* (2001) attribute some of the test-retest discrepancies they find to changes in social norms over the twelve year gap in surveying, that is not a concern for us; inconsistencies are rather much more likely to be linked to respondent reporting errors, or enumerator and data entry errors. Since we have greatly reduced entry error through the rigorous data cleaning process described previously, and the analysis in Section 4 suggests a limited degree of field officer recording error or bias, most test-retest inconsistencies can plausibly be attributed to respondent error. [31]

[30] For instance Bowman *et al.* (1997) compares self-reports and laboratory records to study medical information for a sample of Australian women, and Kornfeld and Bloom (1999) compares self-reported wages to those of state unemployment insurance records for a sample of low-income individuals in the National Job Training Partnership Act study.

[31] The timing of our test-retest is closer to the work of Pierret (2001), who uses two rounds of the NLSY 1979 collected one year apart to compare retrospective responses to questions regarding food stamp receipt and

Another noteworthy dimension is our focus on adolescents and young adults. KLPS-1 respondents are entering a period of great change in their lives, many are leaving their homes, some to continue with their studies and others to embark on a new job or marriage. Such individuals, whom Pierret (2001) calls the most "interesting" and "complex" because they have many important events to recall, were found in his work to be characterized by the lowest response reliability.

5.2 Survey-Resurvey Comparison

A representative sub-sample of completed H-module questionnaires was chosen for re-interview during Wave 2, consisting of nearly 5 percent of all completed surveys.[32] Resurveys consisted of a subset of H-module questions. A field officer other than the one who carried out the original survey conducted the resurvey. Approximately 81 percent of sampled resurvey individuals were successfully resurveyed. In addition, several other respondents were mistakenly interviewed more than once, both in Waves 1 and 2. Usually a field officer other than the original enumerator performed the second survey, however, in contrast to the randomly chosen resurveys, all questions on these questionnaires were re-enumerated.[33] Together, we have survey-resurvey information for 208 individuals (4 percent of all KLPS-1 Household Modules).[34]

The design of the KLPS-1 H-module survey incorporated elements that we believe are beneficial to the analysis of survey response reliability. First, any errors due to enumeration or data entry should be minimized, allowing us to attribute test-retest inconsistencies to respondent recall error. Also, the time elapsed between survey and resurvey in the KLPS-1 was kept quite short, indeed, the average lag time is approximately three months.

---

employment history. Bignami-Van Assche (2003) is also related, but focuses mainly on "sensitive" questions such as contraception and HIV/AIDS among respondents in rural Malawi.

[32] 179 individuals were originally chosen for the resurvey sample, of which 122 had completed H-module surveys (or 4.6 percent of the Wave 2 completed surveys).

[33] In only one case is the survey enumerator the same for both the Wave 2 survey and resurvey. In only seven of these cases is the survey enumerator the same for both the Wave 1 survey and Wave 2 resurvey.

[34] Ninety-nine were randomly chosen for the re-survey, and 109 were not. To increase sample size in the analysis, below we include all resurveys (208), rather than just focusing on the 99 randomly selected Wave 2 resurveys.

Table 20 compares the baseline characteristics of those individuals resurveyed to all other Household Module respondents. These results suggest a weak positive relationship between age and being resurveyed, but few significant associations with other observable individual characteristics; the one partial exception is the final specification (column 5), which shows a negative relationship between the resurvey indicator and the interaction of assigned years of deworming with the pupil test score. Still, these findings suggest that our reliability analysis is broadly applicable to the full KLPS-1 sample.

Table 21 contains several measures of survey-resurvey response reliability for a range of variables. Sub-tribe is a variable coding the self-reported ethnicity of the respondent, and is used here as a sort of control in the analysis. Ethnic identity is very important in the survey area and should not typically change over time, thus in the absence of other types of error we expect to see exact correspondence between reports. Although the means of the sub-tribe measure are not meaningful in themselves, the difference between these means suggests almost no response inconsistency: 95 percent of responses report the same sub-tribe over time.

Moving across the table, we examine several other variables widely used in empirical development economics – age, own and parent education, and migration. Questions about the respondent herself/himself all show high reporting consistency across the two surveys. The proportion of "Don't know" responses are nil in almost all cases, the fraction of survey-resurvey response pairs that exactly match are very high, and there is no evidence of a statistically significant difference in test-retest means.[35]

Results for parent education tell a different story. The proportion of "Don't know" responses is much higher, at 12-15 percent, and the fraction of survey-resurvey matches is much lower at 53 percent for fathers and 51 percent for mothers; even allowing for reports to count as a match if they differ by just one year only increases the response match rate to roughly three quarters. There are no statistically significant difference in means in reported parent education across the survey and resurvey; in other words, it does not appear that reported education was systematically inflated (or deflated) over time. Still

[35] Respondent age does have a somewhat lower match rate than education, likely due to the frequent confusion over date of birth among survey respondents themselves. Note that allowing the age responses to count as a match if they are within one year of each other increases the match rate to over ninety percent.

the extent of this noise suggests that coefficient estimates on parent education variables (and possibly other parent variables) reported by children might be substantially attenuated in many regression analyses.

Figure 3 plots survey and resurvey reports against each other for respondent age and education, and the parent education variables, with data points sized to signify the number of observations. This figure shows graphically that reports of own age and educational attainment are highly consistent, while response matches for parent educational attainment are more widely dispersed around the 45 degree line.

We further examine whether the consistent recall of parent education is systematically related to respondent characteristics (Tables 22 and 23). The only respondent characteristic that is strongly associated with consistent recall of father's education is average primary school participation (from the PSDP dataset): respondents who attended school more themselves are more likely to consistently report the same value for their fathers' educational attainment in both the survey and resurvey. This result either suggests that awareness of father's schooling might be related to the importance placed on education in the household, or on the respondent's own cognitive ability (as reflected in their education).

Table 23 suggests a somewhat different set of determinants for consistent recall of mother's education. Individuals who are older and completed primary school are more likely to report their mother's education consistently, those who weighed more at baseline are less likely to; we do not have a convincing explanation for the association with weight. Interestingly, the physical presence of a parent living in the home with the respondent is not significantly related to reliable reporting of the parent's educational attainment.

In sum, this section has shown that although survey responses relating to the respondent's own characteristics are highly reliable, reports on the outcomes of other family members (here, parents) are far less reliable, with less educated respondents having the least consistent reporting. This suggests that care should be taken in analyses that rely on respondent reports about others using the KLPS-1, and plausibly also using other datasets where one individuals provides reports on other household members.[36] In

[36] Goldstein and Udry (1999) collect survey information in Ghana on own and spouse income in order to examine knowledge of outcomes of other household members.

particular, these findings are worthy of serious consideration by development economists who utilize LSMS style household surveys, which are often conducted among poorly educated respondents and where one household member typically reports information (on consumption expenditures, among other variables) for many other household members. These appear to be precisely the conditions in our setting where unreliable reports are most likely.

6. Conclusion

The KLPS-1 is a useful dataset for examining many data quality issues, given the careful attention paid to attrition and tracking, data entry and enumerator mistakes, enumerator characteristics and the existence of resurveys. In this paper we utilize all of these tools to examine data quality and its possible impact on subsequent analysis using KLPS-1, and also present more general lessons on survey data quality issues in development economics.

One of the main concerns with any panel data analysis is the extent and impact of attrition. In the KLPS-1, basic residential, educational and mortality information for 2003-2005 were obtained for 88% of the target respondents from 1998, and direct contact was made with 84%. These rates compare favorably to other panel datasets in the less developed country context, and are particularly high given the young adult respondent sample. The finding of no differential sample attrition across the treatment arms of the initial randomized deworming intervention is crucial for future analysis that will assess the impact of deworming on later life outcomes.

A key reason for this low attrition rate was the field team's success in tracking movers, nationally and even internationally. Field teams tracked movers composing more than 14% of our overall sample, with 3.1% of these living outside of Kenya. The reasons for their success include: (i) the decision to track individuals who had migrated (of course); (ii) the careful design of tracking protocols, (iii) the quality, training, and teamwork of the enumerators; (iv) the use of cellular phone technology; and (v) positive word-of-mouth about the survey experience through respondents' social networks. Although it was costly to track these individuals over long distances, movers and non-movers look quite different along a

number of observable characteristics, reinforcing the importance of respondent tracking for credible econometric inference.

Additional data collection efforts on mistakes and enumerator characteristics allowed us to carefully analyze additional components of data quality. Overall, the quality of survey enumeration was very high, with only 0.81 errors on average per survey. Moreover, there is little evidence that enumerator characteristics influenced error rates, though there is some suggestive evidence that higher enumerator education is associated with fewer errors and that female enumerators commit somewhat more errors on average. There is some evidence that the enumerator-respondent gender or religion match affects survey responses, particularly on issues of trust, while female respondents are more open in discussing sexual behavior with female respondents, but for most non-sensitive socioeconomic questions, the respondent-enumerator match is less influential.

Finally, we examined response consistency using several variables commonly employed in the development economics empirical literature. Respondents appear to recall facts about their own lives much more reliably than others' characteristics, perhaps as expected, and this is particularly salient for measures of parent education. More educated respondents provide more consistent reports on their parents' education.

A carefully designed field data collection strategy meant that sample attrition and data errors were minimized in the Kenya Life Panel Survey, Round 1. Although costly in time and other resources, the end result of the KLPS-1 field work is a panel dataset that we believe can be used with confidence by researchers in economics, demography, and other social sciences. Through our field work we have learned invaluable lessons on what makes and breaks successful data collection efforts, and this paper sought to illuminate these issues in the hope of contributing to successful future survey data collection in less developed countries.

7. References

Alderman, Harold, Jere Behrman, Hans-Peter Kohler, John A. Maluccio and Susan Watkins (2001). "Attrition in Longitudinal Household Survey Data." *Demographic Research*, Max Planck Institute for Demographic Research, 5(4), 79-124.

Arenas, Erika and Christian Herrera (2007). "Tracking Migration Across Borders: The Mexican Family Life Survey Experience." Presented at the "U.S.-Mexico Binational Conference on Migration and Social Policy Reform" at the Center for U.S.-Mexican Studies, University of San Diego, California. Presentation slides archived at usmex.ucsd.edu/programs/documents/ChristianHerrera.pdf.

Axinn, William (1991). "The Influence of Interviewer Sex on Responses to Sensitive Questions in Nepal." *Social Science Research*, 20(3), 303-318.

Baird, Sarah, Christopher Blattman, Lorenzo Casaburi, David K. Evans, Joan Hamory, Michael Kremer, Edward Miguel, Owen Ozier, Sarah Y. Siegel, Peter Wafula, and Ethan Yeh (2007). "The Kenyan Life Panel Survey (KLPS) Round 1: Survey Data User's Guide."

Becker, S., Feyistan, K. and P. Makinwa-Adebusoye (1995). "The Effect of Sex of Interviewers on the Quality of Data in a Nigerian Family Planning Questionnaire." *Studies in Family Planning*, 26(4), 233-40.

Beckett, Megan, Julie DaVanzo, Constantijn Panis, Christine Peterson, and Narayan Sastry (2001). "A Review of Retrospective Data Quality Evaluations in the Malaysian Family Life Survey." *Journal of Human Resources*, 36, 593-625.

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

Blanc, A.K. and T.N. Croft (1992). "The Effect of the Sex of Interviewer on Responses in Fertility Surveys: The Case of Ghana." Paper presented at the Annual Meeting of the Population Association of America, Denver, Apri1 30-May 1.

Bignami-Van Assche, Simona, Georges Reniers and Alexander A. Weinreb (2003). "An Assessment of the KDICP and MDICP Data Quality," *Demographic Research Special Collections*, Max Planck Institute for Demographic Research, 1(2), 31-76.

Bowman, Jennifer Ann, Rob Sanson-Fisher and Sally Redman (1997). "The Accuracy of Self-reported Pap Smear Utilization." *Social Science & Medicine*, 44(7), 969-976.

Celtel Kenya website. Retrieved March 2008: http://www.ke.celtel.com.

Economic Development Initiatives. Kagera Health and Development Survey 2-Tracking. Retrieved January 2008: http://edi-africa.com/research/khds/tracking.htm.

Falaris, Evangelos M. (2003). "The Effect of Survey Attrition in Longitudinal Surveys: Evidence From Peru, Cote d'Ivoire and Vietnam." *Journal of Development Economics*, 70, 133-157.

Fitzgerald, John, Peter Gottschalk and Robert Moffitt (1998). "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." NBER Technical Working Papers #0220.

Fowler, F.J. and T.W. Mangione (1990). “Standardized Survey Interviewing: Minimizing Interviewer-related Error.” California: Sage Publications.

Goldstein, Markus and Christopher Udry. (1999). “Agricultural Innovation and Resource Management in Ghana.” Final Report to IFPRI Under MP17.

Grajeda, Rubén, Jere R. Behrman, Rafael Flores, John A. Maluccio, Reynaldo Martorell and Aryeh D. Stein (2005). "The Human Capital Study 2002-04: Tracking, Data Collection, Coverage, and Attrition." *Food and Nutrition Bulletin,* 26(2) (Supplement 1), S15-S24.

Hamory, Joan and Edward Miguel (2007). “Individual Ability and Selection into Migration.” Working Paper, U.C. Berkeley.

Hyman, H.H., Cobb, W.J., Feldman, J.J., Hart, C.W. and C.H. Stember (1954). *Interviewing in Social Research.* Chicago, IL: University of Chicago Press.

Kling, Jeffrey R. and Jeffrey B. Liebman (2004). "Experimental Analysis Of Neighborhood Effects On Youth." Working Papers 1, Princeton University, Department of Economics, Industrial Relations Section.

Kornfeld, Robert and Howard Bloom (1999). "Measuring Program Impacts on Earnings and Employment: Do Unemployment Insurance Wage Records Agree with Survey Reports of Individuals?" *Journal of Labor Economics*, 17(1), 168-197.

Maluccio, John (2000). “Attrition in the KwaZulu Natal Income Dynamics Study, 1993-1998.” FCND Discussion Paper No. 95.

Mathiowetz, N. and G. Duncan (1988). “Out of Work, Out of Mind: Response Errors in Retrospective Reports of Unemployment.” *Journal of Business and Economic Statistics*, 6(2), 221-229.

Mbiti, Isaac (2008). “The Impact of Cell Phones in Kenya.” Online presentation. http://www.authorstream.com/presentation/Dario-60455-Mbiti-presentation-Impact-Cell-phones-Kenya-phone-Growth-Mobile-vs-Fixed-presentat-Education-ppt-powerpoint/.

Miguel, Edward, Sarah Baird and Michael Kremer (2007). "The Long Run Impacts of a School Health Intervention in Kenya.” Working Paper.

Miguel, Edward and Michael Kremer (2004). “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.” *Econometrica*, 72(1), 159-217.

Olsen, R. (2005). “The Problem of Survey Attrition: Survey Methodology is Key.” *Monthly Labor Review*, 128, 63-70.

Orr, Larry, Judith D. Feins, Robin Jacob, Erik Beecroft, Lisa Sanbonmatsu, Lawrence F. Katz, Jeffrey B. Liebman, and Jeffrey R. Kling (2003). “Moving to Opportunity Interim Impacts Evaluation.” Washington, DC: U.S. Department of Housing and Urban Development.

Pierret C.R. (2001). “Event History Data and Survey Recall.” *The Journal of Human Resources*, 36(3): 439–466.

Republic of Kenya: Central Bureau of Statistics. (2007). *Statistical Abstract: 2007*. Government Printer: Nairobi.

Safaricom Kenya website. Retrieved March 2008: http://www.safaricom.co.ke.

Smith, James P. and Duncan Thomas (2003). "Remembrances of Things Past: Test-Retest Reliability of Retrospective Migration Histories." *Journal of the Royal Statistical Society, Series A*,166(1), 23-49.

Strauss, J., K. Beegle, B. Sikoki, A. Dwiyanto, Y. Herawati and F. Witoelar (2004). “The Third Wave of the Indonesia Family Life Survey (IFLS3): Overview and Field Report”. WR-144/1-NIA/NICHD.

Sudman, S. and Bradburn N. M. (1973). “Effects of Time and Memory Factors on Response in Surveys.” *Journal of the American Statistical Association*, 68, 805-815.

Thomas, Duncan, Elizabeth Frankenberg and James P. Smith (2001). “Lost But Not Forgotten: Attrition and Followup in the Indonesia Family Life Survey.” *Journal of Human Resources* 36(3), 556-592.

Table 1: Summary statistics on sample attrition and residential location

Means		Tracking Wave		Treatment Group			Gender	
	All	1	2	1	2	3	Females	Males
Panel A: Sample attrition, KLPS H-module								
Found[a]	0.844	0.852	0.836	0.850	0.842	0.839	0.824	0.863
Surveyed	0.827	0.835	0.818	0.833	0.822	0.825	0.805	0.847
Not surveyed, dead	0.008	0.010	0.007	0.008	0.010	0.007	0.008	0.009
Not surveyed, refused	0.008	0.006	0.010	0.009	0.010	0.006	0.010	0.006
Panel B: Residential location information								
Have residential location information	0.879	0.888	0.871	0.880	0.871	0.887	0.857	0.901
Among those with residential location information:								
Residence in Busia district	0.812	0.825	0.799	0.810	0.828	0.800	0.790	0.832
Residence in districts neighboring Busia district[b]	0.049	0.046	0.052	0.049	0.038	0.060	0.066	0.034
Residence outside of Busia and neighboring districts	0.139	0.129	0.149	0.141	0.135	0.141	0.144	0.134
In Nairobi	0.049	0.040	0.058	0.043	0.057	0.047	0.057	0.041
In Mombasa	0.024	0.021	0.028	0.029	0.023	0.021	0.030	0.019
In Nakuru	0.011	0.012	0.009	0.014	0.008	0.010	0.010	0.011
In Kisumu	0.007	0.007	0.006	0.006	0.008	0.007	0.010	0.004
Number of Observations	5602	2730	2872	1905	1777	1920	2741	2861

Notes: The sample used here includes all individuals surveyed, found deceased, refused participation, found but unable to survey, and not found but searched for during intensive tracking. All figures are weighted in order to maintain initial population proportions.

[a] The proportion found is the combined rates of pupils surveyed, found deceased, refused and found but unable to survey.

[b] Districts neighboring Busia include Siaya, Busia (Uganda), and other districts in Kenya's Western Province.

Table 2: Summary statistics on sample attrition and residential location: by grade

Means	1998 Grade					
	2	3	4	5	6	7
Panel A: Sample attrition, KLPS H-module						
Found[a]	0.866	0.876	0.860	0.845	0.816	0.772
Surveyed	0.849	0.858	0.846	0.831	0.799	0.751
Not surveyed, dead	0.006	0.010	0.012	0.007	0.006	0.010
Not surveyed, refused	0.011	0.009	0.001	0.006	0.011	0.011
Panel B: Residential location information						
Have residential location information	0.883	0.880	0.898	0.904	0.866	0.833
Among those with residential location information:						
Residence in Busia district	0.887	0.878	0.822	0.776	0.724	0.723
Residence in districts neighboring Busia district[b]	0.034	0.049	0.045	0.067	0.054	0.051
Residence outside of Busia and neighboring districts	0.079	0.073	0.133	0.157	0.221	0.226
In Nairobi	0.011	0.018	0.049	0.055	0.111	0.080
In Mombasa	0.013	0.015	0.027	0.041	0.023	0.031
In Nakuru	0.005	0.006	0.014	0.008	0.017	0.018
In Kisumu	0.008	0.003	0.002	0.003	0.009	0.019
Number of Observations	990	1018	967	910	870	847

Notes: The sample used here includes all individuals surveyed, found deceased, refused participation, found but unable to survey, and not found but searched for during intensive tracking. All figures are weighted in order to maintain initial population proportions.

[a] The proportion found is the combined rates of pupils surveyed, found deceased, refused and found but unable to survey.

[b] Districts neighboring Busia include Siaya, Busia (Uganda), and other districts in Kenya's Western Province.

Table 3: Summary statistics on sample attrition and residential location: by age group

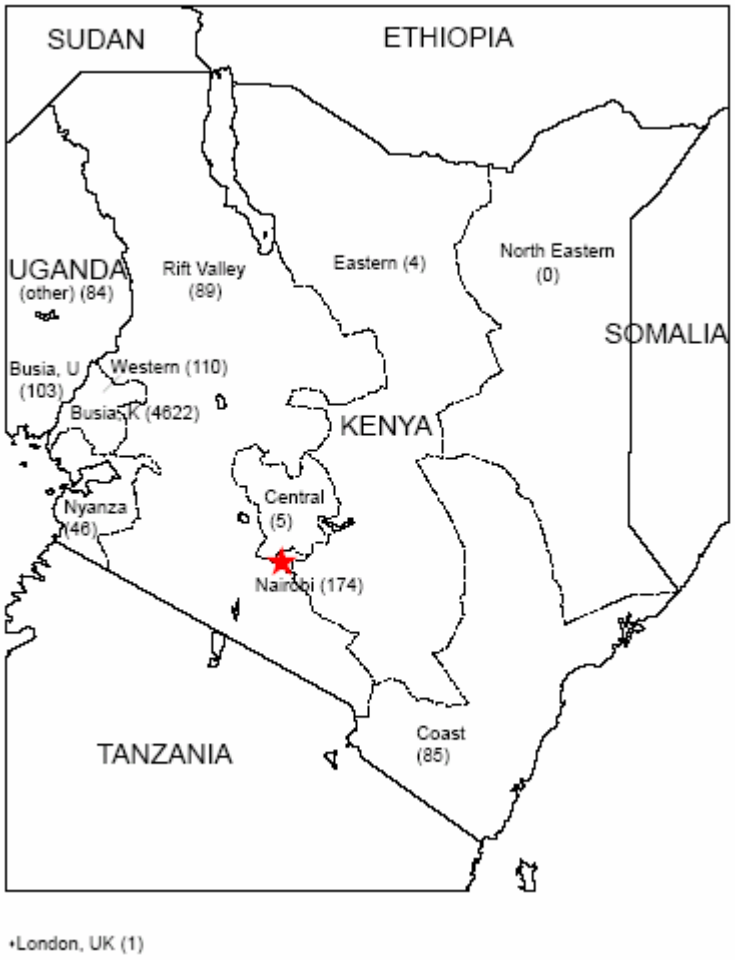
Means	1998 Age					
	6-9	10-11	12-13	14-15	16-20	Missing Age
Panel A: Sample attrition, KLPS H-module						
Found[a]	0.968	0.923	0.870	0.814	0.779	0.751
Surveyed	0.957	0.908	0.854	0.803	0.760	0.724
Not surveyed, dead	0.001	0.006	0.008	0.005	0.012	0.016
Not surveyed, refused	0.010	0.009	0.008	0.005	0.007	0.010
Panel B: Residential location information						
Have residential location information	0.960	0.927	0.900	0.885	0.874	0.784
Among those with residential location information:						
Residence in Busia district	0.926	0.900	0.816	0.764	0.722	0.754
Residence in districts neighboring Busia district[b]	0.034	0.025	0.045	0.045	0.032	0.095
Residence outside of Busia and neighboring districts	0.041	0.075	0.139	0.191	0.247	0.152
In Nairobi	0.010	0.016	0.054	0.072	0.119	0.043
In Mombasa	0.003	0.016	0.023	0.032	0.033	0.032
In Nakuru	0.001	0.004	0.012	0.013	0.016	0.015
In Kisumu	0.006	0.007	0.003	0.012	0.008	0.007
Number of Observations	431	1104	1500	1207	405	955

Notes: The sample used here includes all individuals surveyed, found deceased, refused participation, found but unable to survey, and not found but searched for during intensive tracking. All figures are weighted in order to maintain initial population proportions.

[a] The proportion found is the combined rates of pupils surveyed, found deceased, refused and found but unable to survey.

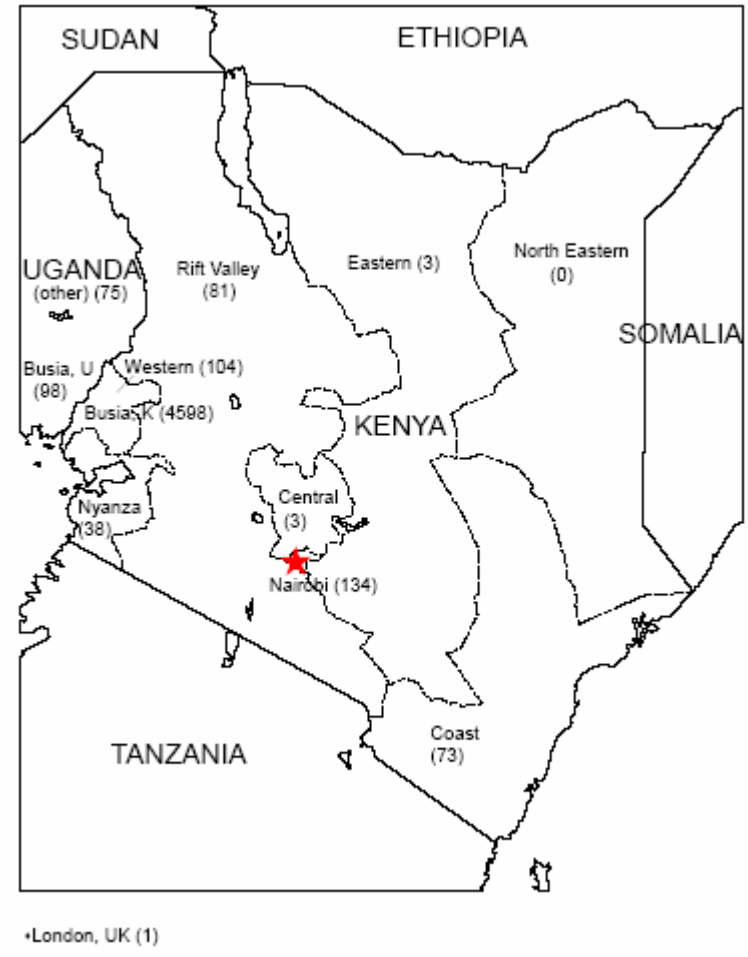
[b] Districts neighboring Busia include Siaya, Busia (Uganda), and other districts in Kenya's Western Province.

Figure 1: Residential locations by province, among those surveyed or tracked



Notes: The sample in this figure includes all individuals with residential location information who were either directly surveyed or where there is proxy survey information from the tracking survey.

Figure 2: Residential locations by province, among those surveyed



Notes: The sample in this figure includes all individuals with residential location information who were directly surveyed.

Table 4: Summary statistics for key variables

Variable	Mean	Std Dev	# Obs
Years of assigned deworming treatment during 1998-2003[a]	3.31	1.84	5602
Female	0.486	0.500	5602
Grade (1998)	4.25	1.69	5602
Age (1998)	12.48	2.19	4647
Group 1 school	0.363	0.481	5602
Group 2 school	0.312	0.463	5602
Budalangi division school	0.382	0.486	5602
School average test score (1996)[b]	0.045	0.439	5602

Notes: The sample used here includes all individuals surveyed, found deceased, refused participation, found but unable to survey, and not found but searched for during intensive tracking. All figures are weighted in order to maintain initial population proportions.

[a] Years of assigned deworming treatment is calculated using treatment group of school and child's standard, and is not adjusted for females over the age of 13.

[b] The average school test score is based on the 1996 Busia District mock exams, and has been converted to units of individual standard deviations.

Table 5: Determinants of survey attrition

	Dependent Variable: Indicator for Individual Surveyed					
	(1)	(2)	(3)	(4)	(5)	(6)
Years assigned deworming	-0.001	-0.001		-0.018	-0.009	-0.021
	[0.008]	[0.008]		[0.012]	[0.012]	[0.018]
Group 1 school (1998)			0.001			
			[0.023]			
Group 2 school (1998)			-0.004			
			[0.023]			
Female	-0.051					
	[0.015]***					
Age (1998)	-0.028	-0.028	-0.028	-0.03	-0.031	-0.023
	[0.005]***	[0.005]***	[0.005]***	[0.009]***	[0.008]***	[0.013]*
School average mock score (1996)	-0.017	-0.018	-0.016	-0.062		
	[0.016]	[0.016]	[0.019]	[0.024]***		
Pupil test score (1998)					-0.001	-0.015
					[0.010]	[0.018]
Child falls sick often, self-report (1998)					-0.021	-0.021
					[0.021]	[0.021]
Household has a latrine (1998)					-0.021	-0.019
					[0.029]	[0.029]
Household owns cattle (1998)					0.008	0.006
					[0.025]	[0.025]
Child weight, kg (1998)					-0.002	-0.003
					[0.002]	[0.002]
Yrs assigned deworming * Female				0.024		0.026
				[0.013]*		[0.021]
Yrs assigned deworming * Age				0.000		-0.002
				[0.003]		[0.004]
Yrs assigned deworming * School avg mock score				0.013		
				[0.008]*		
Yrs assigned deworming * Pupil test score						0.005
						[0.006]
Controls for gender, grade, tracking wave	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5602	5602	5602	5602	3440	3440
Mean [std dev] of dependent variable	0.827	0.827	0.827	0.827	0.814	0.814
	[0.378]	[0.378]	[0.378]	[0.378]	[0.389]	[0.389]

Notes: Probit specifications, with marginal effects evaluated at mean values. The sample employed in columns (1)-(4) includes all individuals who were surveyed, found deceased, refused participation, found but unable to survey, or not found but searched for during intensive tracking. The sub-sample employed in columns (5) and (6) additionally includes only those with 1998 pupil questionnaire and individual test score data. All regressions are weighted in order to maintain initial population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. Years assigned deworming is calculated using treatment group of school and individual's standard in 1998, and is not adjusted for females over the age of 13. Age measures are demeaned. School average test score data has been converted into units of individual standard deviations. All specifications include controls for missing age data, and (4) and (6) include an interaction with years assigned deworming. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. In columns (4) and (6), we cannot reject the null that deworming treatment and all of its interaction terms are jointly zero.

Table 6: Determinants of migration

	Dependent Variable: Indicator for Individual Migrated					
	(1)	(2)	(3)	(4)	(5)	(6)
Years assigned deworming	-0.001	-0.001		0.003	0.012	0.012
	[0.009]	[0.009]		[0.010]	[0.010]	[0.015]
Group 1 school (1998)			0.001			
			[0.024]			
Group 2 school (1998)			-0.017			
			[0.025]			
Female	0.030					
	[0.013]**					
Age (1998)	0.007	0.007	0.007	0.011	0.006	0.019
	[0.005]	[0.005]	[0.005]	[0.008]	[0.007]	[0.013]
School average mock score (1996)	-0.010	-0.010	-0.006	0.003		
	[0.020]	[0.020]	[0.020]	[0.033]		
Pupil test score (1998)					0.025	0.014
					[0.010]**	[0.017]
Child falls sick often, self-report (1998)					0.015	0.014
					[0.024]	[0.024]
Household has a latrine (1998)					0.000	0.002
					[0.025]	[0.025]
Household owns cattle (1998)					0.021	0.019
					[0.019]	[0.018]
Child weight, kg (1998)					0.002	0.001
					[0.002]	[0.002]
Yrs assigned deworming * Female				-0.005		0.003
				[0.010]		[0.021]
Yrs assigned deworming * Age				-0.001		-0.004
				[0.002]		[0.003]
Yrs assigned deworming * School avg mock score				-0.004		
				[0.011]		
Yrs assigned deworming * Pupil test score						0.004
						[0.007]
Controls for gender, grade, tracking wave	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5208	5208	5208	5208	3186	3186
Mean [std dev] of dependent variable	0.150	0.150	0.150	0.150	0.161	0.161
	[0.357]	[0.357]	[0.357]	[0.357]	[0.367]	[0.367]

Notes: The indicator for migration takes on a value of one if the individual left Busia District, and zero otherwise. Probit specifications, with marginal effects evaluated at mean values. The sample employed in columns (1)-(4) includes all individuals who were surveyed. The sub-sample employed in columns (5) and (6) additionally includes only those with 1998 pupil questionnaire and individual test score data. All regressions are weighted in order to maintain initial population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. Years assigned deworming is calculated using treatment group of school and individual's standard in 1998, and is not adjusted for females over the age of 13. Age measures are demeaned. School average test score data has been converted into units of individual standard deviations. All specifications include controls for missing age data, and (4) and (6) include an interaction with years assigned deworming. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. In columns (4) and (6), we cannot reject the null that deworming treatment and all of its interaction terms are jointly zero.

Table 7: Determinants of international migration

	Dependent Variable: Indicator for Migration Out of Kenya					
	(1)	(2)	(3)	(4)	(5)	(6)
Years assigned deworming	-0.001	-0.001		-0.002	0.005	0.005
	[0.003]	[0.003]		[0.005]	[0.004]	[0.005]
Group 1 school (1998)			-0.002			
			[0.010]			
Group 2 school (1998)			0.004			
			[0.012]			
Female	-0.001					
	[0.008]					
Age (1998)	0.003	0.003	0.003	0.006	0.006	0.015
	[0.002]	[0.002]	[0.002]	[0.004]	[0.002]**	[0.006]***
School average mock score (1996)	-0.016	-0.016	-0.016	0.009		
	[0.013]	[0.013]	[0.013]	[0.014]		
Pupil test score (1998)					-0.001	0.002
					[0.004]	[0.006]
Child falls sick often, self-report (1998)					0.012	0.012
					[0.009]	[0.009]
Household has a latrine (1998)					-0.007	-0.007
					[0.012]	[0.012]
Household owns cattle (1998)					0.010	0.008
					[0.007]	[0.007]
Child weight, kg (1998)					0.000	0.000
					[0.001]	[0.001]
Yrs assigned deworming * Female				0.005		0.001
				[0.007]		[0.007]
Yrs assigned deworming * Age				-0.001		-0.003
				[0.001]		[0.001]*
Yrs assigned deworming * Pupil test score						-0.001
						[0.002]
Yrs assigned deworming * School avg mock score				-0.007		
				[0.005]		
Controls for gender, grade, tracking wave	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5210	5210	5210	5210	3188	3188
Mean [std dev] of dependent variable	0.038	0.038	0.038	0.038	0.037	0.037
	[0.191]	[0.191]	[0.191]	[0.191]	[0.188]	[0.188]

Notes: Probit specifications, with marginal effects evaluated at mean values. The sample employed in columns (1)-(4) includes all individuals who were surveyed. The sub-sample employed in columns (5) and (6) additionally includes only those with 1998 pupil questionnaire and individual test score data. All regressions are weighted in order to maintain initial population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. Years assigned deworming is calculated using treatment group of school and individual's standard in 1998, and is not adjusted for females over the age of 13. Age measures are demeaned. School average test score data has been converted into units of individual standard deviations. All specifications include controls for missing age data, and (4) and (6) include an interaction with years assigned deworming. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. In column (4), we cannot reject the null that deworming treatment and all of its interaction terms are jointly zero. In column (6) we can reject this hypothesis at the 1% level of significance.

Table 8: Summary of tracking costs

	# Pupils	Unit Cost
Panel A: Costs per pupil looked for		
Total	7,530	40.41
Wave 1 (2003-2004)	3,745	34.04
Wave 2 (2004-2005)	3,785	44.11
Panel B: Costs per pupil found		
Total	5,309	57.32
Wave 1 regular	2,396	40.77
Wave 1 intensive	186	158.37
Wave 2 regular	2,570	53.94
Wave 2 intensive	157	179.31

Notes: All costs are reported in nominal US dollars, converted using average daily interbank exchange rates over the tracking period specified. Total costs include piloting costs, while costs broken out by wave do not.

Table 9: Difference in treatment impacts: migrants versus non-migrants

Dependent Variable	Migrant Indicator	Years of Treatment	Tyears* Migrant	Mean [std dev] of dep variable	N
Height (cm)	-1.710	0.110	0.394	164.104	5052
	[0.675]**	[0.141]	[0.233]*	[8.848]	
Weight (kg)	-0.635	0.082	0.465	55.920	5180
	[0.799]	[0.110]	[0.264]*	[9.040]	
BMI (weight (kg)/height(m)^2)	0.223	0.010	0.071	20.693	5046
	[0.250]	[0.032]	[0.086]	[2.668]	
General Health (=1 if Self Perceived Health Very Good)	0.044	0.010	-0.011	0.660	5172
	[0.052]	[0.008]	[0.017]	[0.474]	
Summary HIV Index [0,1] (=1 if No Symptoms)	-0.004	0.001	0.002	0.911	5198
	[0.011]	[0.003]	[0.003]	[0.126]	
Highest Grade Attended	0.058	0.000	0.032	7.638	4088
	[0.196]	[0.028]	[0.051]	[2.010]	
Speed of Attainment (# Grades passed/Total Years)	-0.020	0.001	-0.001	0.640	5209
	[0.028]	[0.004]	[0.007]	[0.242]	
Indicator for Currently In School	-0.133	0.009	-0.004	0.614	5080
	[0.052]**	[0.008]	[0.013]	[0.487]	
Indicator for In School or Working (Not Idle)	0.103	0.010	-0.024	0.795	5005
	[0.042]**	[0.007]	[0.011]**	[0.403]	
Indicator for In School or Finished Primary School (Not Dropout)	-0.007	0.013	-0.026	0.798	5152
	[0.054]	[0.006]**	[0.015]*	[0.402]	
Grade Attained by Age 16	0.021	-0.020	0.003	6.821	4088
	[0.130]	[0.024]	[0.038]	[1.513]	

Notes: All dependent variables are as measured at time of enumeration, with the exceptions of "speed of attainment" and "grade attained by age 16". Coefficient on years of deworming treatment is from a weighted least squares regression using the controls in specification (2) from Table (7). In addition, we include an indicator for whether the pupil migrated and interact this with treatment. Robust Huber-White standard errors in brackets. Disturbance terms are allowed to be correlated among students in the same primary school (in 1998). All observations are weighted in order to maintain initial population proportions in the sample. Years assigned deworming is calculated using treatment group of school and individual's standard in 1998, and is not adjusted for females over the age of 13. Age measures are demeaned. School average test score data has been converted into units of individual standard deviations and is based on the 1996 Busia District mock exams. All specifications include controls for missing age data. Additional controls include gender-standard interactions, tracking wave, and parent education measures. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 10: Enumerator characteristics

Variable	Mean	# Obs
Female	0.55	20
Age (2003)	28.47	19
Tribe is Luhya	0.85	20
Place of birth: Busia, Kenya	0.55	20
Place of birth: elsewhere in Kenya	0.45	20
Highest educational attainment: some post-secondary, no degree	0.25	20
Highest educational attainment: college degree	0.40	20
Highest educational attainment: university degree	0.35	20
Marital status during 2003-5 data collection: single	0.30	20
Marital status during 2003-5 data collection: married	0.70	20
Number of children	1.60	20
Had previous experience working for an NGO	0.53	19
Had previous experience working with on a survey project	0.35	17
Average monthly income prior to work at ICS (USD)	135	17

Notes: This table describes characteristics of the twenty enumerators who administered the KLPS survey. At the time this information was collected, three field officers were no longer available. Thus, characteristics of these three enumerators were estimated for all but the last two variables (age and previous NGO experience could not be estimated for one of these enumerators).

Table 11: Outcomes used for field officer bias analysis

Panel A: Sensitive Variables	
Pregnant	Has the respondent or partner ever been pregnant?
Contraception	Did the respondent or partner do anything to avoid pregnancy?
Sexual Assault	Has the respondent been a victim of sexual assault in the past 12 months?
Trust	Does respondent think most people can be trusted?
Trust Tribe	Does respondent trust most members of his/her tribe?
Trust Other Tribes	Does respondent trust most members of other tribes?
Trust Church/Mosque	Does respondent trust most members of his/her church/mosque?
Trust Other Church/Mosque	Does respondent trust most members of other church/mosque?
Vote	Did respondent vote in the 2002 National election?
Church Regularly	Does the respondent attend church regularly?
Panel B: Control Variables	
School Attendance (2003)	Did the respondent attend school at anytime during 2003?
Work	Did the respondent have a job in the past 12 months?
Farming	Did the respondent's household farm in the past 12 months?
Household Size	Number of people that live/eat meals in respondent's household
Training	Did the respondent receive professional training?
Meals	Number of meals eaten respondent yesterday.

Table 12: Field officer bias analysis-gender

	Female respondent	Respondent-enumerator gender match	Interaction	Joint significance (gender match and interaction)	Mean [std dev] of dep variable	N
Panel A: Sensitive Variables						
Pregnant	0.207	-0.007	0.015	No	0.167	4831
	[0.015]***	[0.014]	[0.021]		[0.373]	
Contraception	-0.014	0	0.015	No	0.041	795
	[0.024]	[0.029]	[0.033]		[0.197]	
Sexual Assault	0.016	-0.001	0.004	No	0.019	4798
	[0.006]***	[0.005]	[0.008]		[0.138]	
Trust	-0.079	-0.074	0.121	Yes	0.151	4827
	[0.015]***	[0.014]***	[0.021]***		[0.358]	
Trust Tribe	0.175	0.166	-0.303	Yes	0.745	4830
	[0.018]***	[0.017]***	[0.025]***		[0.436]	
Trust Other Tribes	-0.041	-0.021	0.031	No	0.228	4830
	[0.017]**	[0.016]	[0.025]		[0.419]	
Trust Church/Mosque	0.088	0.087	-0.178	Yes	0.806	4829
	[0.016]***	[0.015]***	[0.023]***		[0.396]	
Trust Other Church/Mosque	0.014	0.03	-0.148	Yes	0.397	4831
	[0.020]	[0.019]	[0.028]***		[0.489]	
Vote	-0.033	-0.011	0.032	Yes	0.029	4826
	[0.007]***	[0.007]*	[0.010]***		[0.168]	
Church Regularly	0.024	-0.125	0.201	Yes	0.750	4819
	[0.018]	[0.017]***	[0.025]***		[0.433]	
Panel B: Controls						
School Attendance (2003)	-0.115	-0.048	0.095	Yes	0.706	4734
	[0.019]***	[0.018]***	[0.027]***		[0.456]	
Work	-0.201	-0.016	-0.003	No	0.287	4818
	[0.018]***	[0.017]	[0.026]		[0.452]	
Farming	-0.005	-0.002	-0.035	Yes	0.844	4828
	[0.015]	[0.014]	[0.021]*		[0.363]	
Household Size	-0.016	0.142	-0.048	No	6.780	4838
	[0.142]	[0.134]	[0.202]		[3.458]	
Training	0.001	0.006	0.013	No	0.075	4822
	[0.011]	[0.010]	[0.015]		[0.263]	
Meals	0.172	0.01	-0.083	Yes	2.224	4831
	[0.027]***	[0.026]	[0.039]**		[0.667]	

Notes: Each row represents a separate linear regression specification. The sample includes all surveyed individuals, except those who were also sampled for a Family Module (these individuals were all females who were paired with a female enumerator). All regressions are weighted in order to maintain initial population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. Dependent variables are defined in Table 11. Female refers to the gender of the respondent. The variable Gender Match takes on a value of one if the enumerator and the respondent are the same gender. The interaction variable is the simple interaction of the female indicator with the gender match indicator. The significance of the joint test of the gender match and the interaction is shown in the final column. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 13: Field officer bias analysis-tribe

	Luhya Respondent	Respondent-enumerator tribe match	Interaction	Joint significance (tribe match and interaction)	Mean [std dev] of dep variable	N
Panel A: Sensitive Variables						
Pregnant	0.007	0.105	-0.114	No	0.187	5177
	[0.026]	[0.062]*	[0.064]*		[0.390]	
Contraception	-0.017	-0.027	0.057	No	0.033	946
	[0.028]	[0.055]	[0.057]		[0.180]	
Sexual Assault	0.006	0.014	-0.003	No	0.021	5144
	[0.010]	[0.023]	[0.024]		[0.144]	
Trust	0.085	0.005	-0.071	Yes	0.146	5172
	[0.024]***	[0.056]	[0.058]		[0.354]	
Trust Tribe	0.084	0.085	-0.117	Yes	0.733	5175
	[0.030]***	[0.070]	[0.072]		[0.443]	
Trust Other Tribes	0.090	0.007	-0.052	Yes	0.22	5175
	[0.028]***	[0.066]	[0.068]		[0.414]	
Trust Church/Mosque	0.146	0.114	-0.203	Yes	0.796	5174
	[0.027]***	[0.064]*	[0.065]***		[0.403]	
Trust Other Church/Mosque	0.085	-0.059	-0.012	Yes	0.386	5176
	[0.033]***	[0.077]	[0.079]		[0.487]	
Vote	0.038	0.007	-0.021	No	0.028	5171
	[0.011]***	[0.026]	[0.027]		[0.166]	
Church/Mosque Regularly	-0.003	0.032	-0.015	No	0.754	5171
	[0.029]	[0.069]	[0.070]		[0.431]	
Panel B: Controls						
School Attendance (2003)	0.039	-0.078	0.055	No	0.689	5076
	[0.032]	[0.075]	[0.077]		[0.463]	
Work	0.161	0.076	-0.223	Yes	0.284	5162
	[0.030]***	[0.071]	[0.073]***		[0.451]	
Farming	0.118	0.084	-0.173	No	0.839	5172
	[0.025]***	[0.058]	[0.060]***		[0.368]	
Household Size	0.596	1.106	-1.410	Yes	6.767	5181
	[0.233]**	[0.547]**	[0.561]**		[3.439]	
Training	0.004	0.084	-0.093	Yes	0.081	5163
	[0.018]	[0.043]*	[0.044]**		[0.272]	
Meals	-0.207	-0.440	0.633	Yes	2.238	5178
	[0.045]***	[0.106]***	[0.109]***		[0.671]	

Notes: Each row represents a separate linear regression specification. The sample includes all surveyed individuals. All regressions are weighted in order to maintain initial population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. Dependent variables are defined in Table 11. The variable Luhya refers to the tribe of the respondent. The variable Tribe Match takes on a value of one if the enumerator and the respondent are members of the same tribe. The interaction variable is the simple interaction of the Luhya indicator with the tribe match indicator. The significance of the joint test of the tribe match and the interaction is shown in the final column. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 14: Field officer bias analysis-religion

	Religion of respondent	Respondent-enumerator religion match	Mean [std dev] of dep variable	N
Panel A: Sensitive Variables				
Pregnant	0.007	0.017	0.186	4821
	[0.001]***	[0.018]	[0.389]	
Contraception	0.001	-0.039	0.041	879
	[0.001]	[0.022]*	[0.199]	
Sexual Assault	0.001	-0.001	0.023	4792
	[0.000]**	[0.007]	[0.148]	
Trust	-0.001	0.002	0.148	4816
	[0.001]	[0.017]	[0.355]	
Trust Tribe	-0.002	0.129	0.747	4819
	[0.001]*	[0.020]***	[0.435]	
Trust Other Tribes	0.002	-0.023	0.207	4819
	[0.001]*	[0.019]	[0.405]	
Trust Church/Mosque	0.002	0.097	0.813	4818
	[0.001]***	[0.018]***	[0.390]	
Trust Other Church/Mosque	0.004	0.014	0.376	4820
	[0.001]***	[0.023]	[0.484]	
Vote	0.001	0.002	0.028	4815
	[0.000]**	[0.008]	[0.166]	
Church/Mosque Regularly	0.005	0.009	0.775	4814
	[0.001]***	[0.020]	[0.418]	
Panel B: Controls				
School Attendance (2003)	-0.007	0.012	0.692	4724
	[0.001]***	[0.022]	[0.462]	
Work	0.005	-0.041	0.291	4805
	[0.001]***	[0.021]*	[0.454]	
Farming	0	0.021	0.842	4816
	[0.001]	[0.017]	[0.365]	
Household Size	-0.001	-0.361	6.828	4825
	[0.008]	[0.163]**	[3.467]	
Training	0	-0.018	0.082	4808
	[0.001]	[0.013]	[0.274]	
Meals	-0.004	-0.003	2.235	4822
	[0.001]**	[0.032]	[0.683]	

Notes: Each row represents a separate linear regression specification. The sample includes all surveyed individuals. All regressions are weighted in order to maintain initial population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. Dependent variables are defined in Table 11. The variable Religion refers to the religion of the respondent. The variable Religion Match takes on a value of one if the enumerator and the respondent are members of the same religion. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 15: Effect of having other people in the room

	Indicator for Other People Present in Room	Mean [std dev] of dep variable	N
Panel A: Sensitive Variables			
Pregnant	-0.001	0.188	5101
	[0.027]	[0.391]	
Contraception	-0.027	0.037	937
	[0.029]	[0.189]	
Sexual Assault	-0.025	0.021	5072
	[0.010]**	[0.143]	
Trust	0.032	0.145	5096
	[0.024]	[0.352]	
Trust Tribe	0.084	0.732	5099
	[0.030]***	[0.443]	
Trust Other Tribes	0.007	0.222	5099
	0.007	[0.416]	
Trust Church/Mosque	0.041	0.796	5098
	[0.027]	[0.403]	
Trust Other Church/Mosque	0.044	0.387	5100
	[0.033]	[0.487]	
Vote	0.033	0.028	5095
	[0.011]***	[0.166]	
Church Regularly	0.063	0.755	5090
	[0.029]**	[0.430]	
Panel B: Controls			
School Attendance (2003)	-0.224	0.687	5005
	[0.032]***	[0.464]	
Work	-0.011	0.285	5086
	[0.031]	[0.452]	
Farming	-0.05	0.838	5097
	[0.025]**	[0.368]	
Household Size	-0.001	6.763	5107
	[0.233]	[0.3443]	
Training	0.045	0.082	5090
	[0.019]**	[0.274]	
Meals	0.10	2.238	5102
	[0.046]**	[0.672]	

Notes: Each row represents a separate linear regression specification. All regressions are weighted in order to maintain initial population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. Dependent variables are defined in Table 11. The variable other people takes on a value of one if an individual other than the enumerator and the respondent was present during the interview. Age of respondent is a control in each regression. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 16: Quality of field officer enumeration

Variable	Mean	Std Dev	Min	Max	# Obs
Number of household surveys enumerated by field officer	260.25	173.14	29	560	20
Number of household surveys enumerated by field officer, per day	2.46	0.63	1.14	3.26	20
Number of surveys reviewed in quality checks, per field officer	20.20	12.02	3	36	20
Time to complete interview (minutes)	61.84	7.32	52.33	83.33	20
Handwriting score[a]	1.20	0.28	1	2	20
Number of "don't know" responses per survey	2.03	0.99	0.75	4.63	20
Number of fields left blank per survey	1.42	1.62	0.25	6.45	20
Number of errors per survey	0.81	0.51	0.17	2	20
Number of logical inconsistencies per survey[b]	0.26	0.26	0	0.82	20

Notes: The information in this table comes from reviews of a subset of surveys per field officer. The first and third rows report number of surveys. The remaining rows report field officer averages, averaged over the entire field team.
[a] Handwriting score was measured on a scale of 1-4, with 1 being very clear and legible.
[b] An example of a logical inconsistency is a skip pattern violation.

Table 17: Error rates across field officer characteristics

	Dependent variable: overall error rate			
	(1)	(2)	(3)	(4)
Enumerator married during 2003-5 data collection	-0.465	0.276	-0.571	0.443
	[1.054]	[1.346]	[1.907]	[1.420]
Enumerator total number of children ever born to you	-0.671	0.321	-0.121	-0.321
	[0.494]	[0.597]	[0.893]	[0.631]
Enumerator highest education level attained by 2003-5	-1.548	-0.765	-0.438	-0.645
	[0.563]**	[0.633]	[1.018]	[0.668]
Enumerator had NGO experience prior to work at ICS	-1.226		-1.645	
	[0.977]		[1.766]	
Enumerator had survey experience prior to work at ICS	1.578		-0.564	
	[0.918]		[1.661]	
Enumerator average monthly income in previous job (Ksh)	0.000		0.000	
	[0.000]		[0.000]	
Enumerator number of surveys completed	-0.002	0.001	0.005	0.004
	[0.002]	[0.003]	[0.004]	[0.003]
Indicator for Female Enumerator	2.071	2.376	2.505	2.794
	[0.925]*	[1.020]**	[1.672]	[1.077]**
Enumerator Age in Fall 2003	0.437		0.050	
	[0.105]***		[0.191]	
Indicator for Enumerator Born in Busia District	-1.252	-0.643	-0.731	-0.304
	[0.737]	[1.014]	[1.333]	[1.070]
Constant	-0.178	5.657	7.389	8.932
	[3.210]	[3.078]*	[5.804]	[3.248]**
Observations	17	20	17	20
R-squared	0.889	0.443	0.577	0.385

Notes: Ordinary least squares specifications. The sample employed in columns (1) and (3) includes only enumerators where we do not have missing information. The sample employed in columns (2) and (4) include all enumerators. All regressions are weighted in order to maintain initial population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. The overall error rate is calculated in two ways: in columns (1) and (2) the error rate from the sample of surveys that were carefully evaluated for errors; in columns (3) and (4) the error rate is calculated by tabulating the total number of mistakes across all surveys as defined by the mistake dataset. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 18: Number of survey mistakes and enumerator fatigue, learning

	Dependent variable: number of mistakes in survey	
	(1)	(2)
Number of surveys administered previously, during KLPS-1	-0.040	
	[0.005]***	
Number of surveys administered previously, during KLPS-1, squared (divided by 1000)	0.092	
	[0.015]***	
Number of surveys administered on day of enumeration		0.715
		[0.093]***
Observations	5205	5205
R-squared	0.054	0.014
Mean [std dev] of dependent variable	7.704	7.704
	[6.945]	[6.945]

Notes: Ordinary least squares specifications. The sample employed includes the full set of KLPS-1 Household surveys, minus three for which the enumerator was not identified, and three that were administered by a field officer not part of the regular KLPS team (and therefore did not administer more than a single survey). Robust standard errors in brackets, corrected for clustering at the enumerator level. Total number of mistakes is calculated by tabulating the across all sections of the survey as defined by the mistake dataset. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 19: Data entry errors by Household (H) Module section

Average Errors	Mean	Std	Min	Max
Section 1	0.14	0.38	0	4
Section 2	0.14	0.47	0	9
Section 3	0.28	1.36	0	12
Section 4	0.64	1.27	0	16
Section 5	0.05	0.23	0	2
Section 6	0.41	1.86	0	31
Section 7	0.31	1.70	0	18
Section 8	0.08	0.34	0	7
Section 9	0.06	0.63	0	16
Section 10	0.05	0.76	0	19
Section 11	0.05	0.81	0	27
Section 12	0.32	1.62	0	28
Section 13	0.62	4.01	0	35
Section 14	0.31	1.63	0	41
Section 15	0.02	0.55	0	15
Section 16	0.44	1.13	0	7
Section 17	0.07	0.59	0	12
Section 18	0.05	0.33	0	12
Section 19	0.11	0.55	0	13
Section 20	0.08	0.63	0	13
Section 21	4.12	2.84	0	11
Overall	8.35	7.65	0	168

Table 20: Regressions for individual resurveyed

	Dependent Variable: Indicator for Individual Resurveyed				
	(1)	(2)	(3)	(4)	(5)
Years assigned deworming	-0.003	-0.003	0.000	-0.004	-0.001
	[0.002]	[0.002]	[0.003]	[0.003]	[0.004]
Female	0.002				
	[0.004]				
Age (1998)	0.000	0.000	0.004	0.005	0.007
	[0.002]	[0.002]	[0.002]*	[0.003]*	[0.003]**
School average mock score (1996)	-0.005	-0.004	-0.013		
	[0.007]	[0.007]	[0.012]		
Dropped out before completing primary school	-0.012	-0.012	-0.014	-0.008	-0.010
	[0.006]*	[0.006]*	[0.006]**	[0.009]	[0.008]
Father alive	0.004	0.004	0.004	0.005	0.007
	[0.007]	[0.007]	[0.007]	[0.010]	[0.008]
Mother alive	0.010	0.010	0.009	0.005	0.004
	[0.006]	[0.006]	[0.006]	[0.011]	[0.010]
Living with father	0.003	0.003	0.003	0.003	0.000
	[0.007]	[0.007]	[0.006]	[0.009]	[0.008]
Living with mother	0.005	0.004	0.005	0.010	0.012
	[0.006]	[0.006]	[0.006]	[0.008]	[0.008]
Average school participation (1998)	0.000	0.000	0.000	-0.003	-0.004
	[0.010]	[0.010]	[0.010]	[0.017]	[0.017]
Pupil test score (1998)				-0.004	0.006
				[0.004]	[0.006]
Child falls sick often, self-report (1998)				0.002	0.004
				[0.010]	[0.008]
Household owns cattle (1998)				-0.002	-0.003
				[0.004]	[0.004]
Child weight, kg (1998)				0.000	0.000
				[0.001]	[0.001]
Yrs assigned deworming * Female			-0.005		-0.003
			[0.004]		[0.005]
Yrs assigned deworming * Age			-0.001		0.000
			[0.001]		[0.001]
Yrs assigned deworming * School avg mock score			0.003		
			[0.003]		
Yrs assigned deworming * Pupil test score					-0.004
					[0.002]**
Controls for gender, grade, tracking wave	Yes	Yes	Yes	Yes	Yes
Observations	4901	4901	4901	3051	3051
Mean [std dev] of dependent variable	0.033 [0.178]	0.033 [0.178]	0.033 [0.178]	0.036 [0.187]	0.036 [0.187]

Notes: Probit specifications, with marginal effects evaluated at mean values. The sample employed in columns (1)-(3) includes all individuals who were surveyed and have information on school attendance in 1998, parent death, parent residential location, and dropout status at the time of interview (from the main

survey). The sub-sample employed in columns (4) and (5) additionally includes only those with 1998 pupil questionnaire and individual test score data. All regressions are weighted in order to maintain initial population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. Years assigned deworming is calculated using treatment group of school and individual's standard in 1998, and is not adjusted for females over the age of 13. Age measures are demeaned. School average test score data has been converted into units of individual standard deviations. All specifications include controls for missing age data, and (3) and (5) include an interaction with years assigned deworming. Unless otherwise noted, variables are measured at time of KLPS-1 interview. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. In column (3), we can reject the null that deworming treatment and all of its interaction terms are jointly zero at the 10% level of significance. In column (6) we cannot reject this hypothesis.

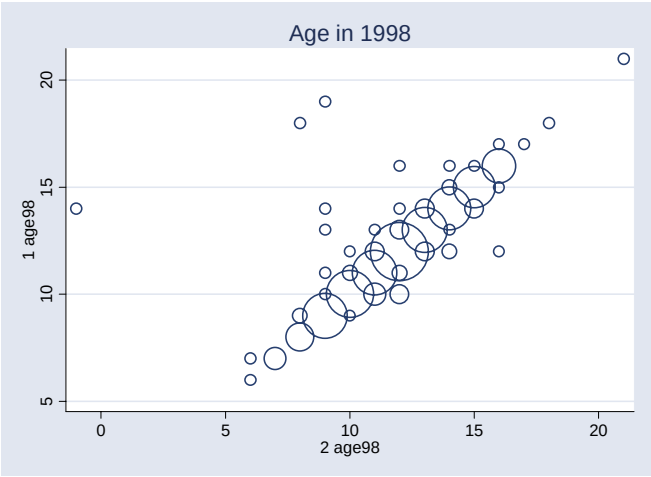
Table 21: Survey-resurvey response comparison

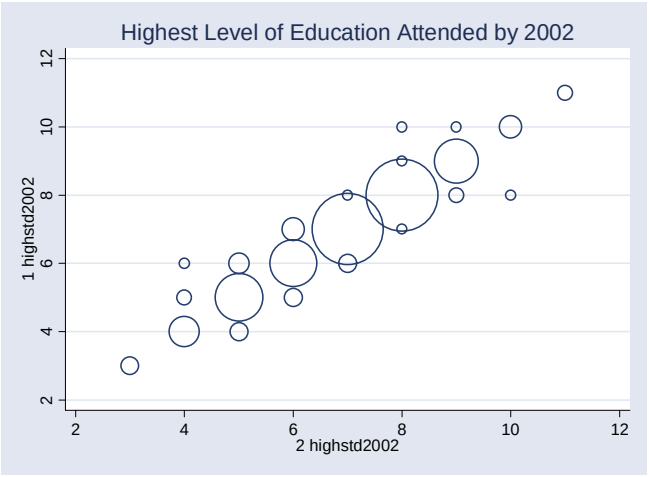
	Sub-Tribe	Age in 1998	Highest Grade Attended by 2002	Indicator for Ever Left Local Areas	Educational Attainment of Father	Educational Attainment of Mother
Mean [std dev] of survey response[a]	1.40	12.10	6.95	0.09	8.72	5.84
	[0.83]	[2.55]	[1.59]	[0.29]	[3.55]	[3.84]
Mean [std dev] of resurvey response[a]	1.39	11.94	6.93	0.09	8.62	5.77
	[0.82]	[2.50]	[1.60]	[0.29]	[3.30]	[3.77]
Difference [std err][a]	0.01	0.16*	0.02	0.00	0.11	0.07
	[0.05]	[0.09]	[0.03]	[0.02]	[0.22]	[0.22]
Fraction pairs with matching responses[a]	0.948	0.756	0.859	0.914	0.529	0.505
Fraction pairs with responses within one year[a]	--	0.922	0.985	--	0.750	0.705
Fraction survey responses of "don't know"	0	0	0	0	0.154	0.127
Fraction resurvey responses of "don't know"	0	0.005	0	0	0.154	0.119
Fraction survey responses that are impossible values	0	0.005	0	0	0	0
Fraction resurvey responses that are impossible values	0	0	0	0	0.007	0.008
Number survey-resurvey pairs	135	207	206	208	136	134
Pairwise correlation coefficient[a]	--	0.869	0.964	--	0.797	0.819

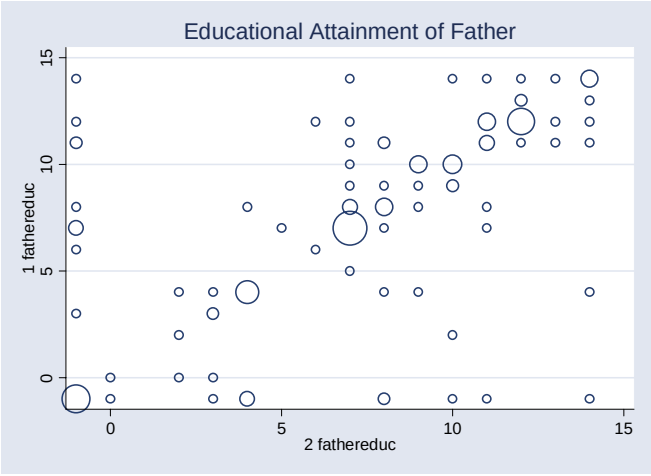
Notes: Statistics presented in this table come from a comparison of survey and resurvey responses to KLPS Household Module questions, except for highest grade attended by 2002 and indicator for ever left local areas, which were created by combining responses to several questions. Modules were not re-administered in their entirety to each respondent intentionally selected for a resurvey. Instead, each selected respondent was randomly assigned particular sections of the survey for a second round of enumeration. For this reason, the number of observations (survey-resurvey pairs) varies across variables.

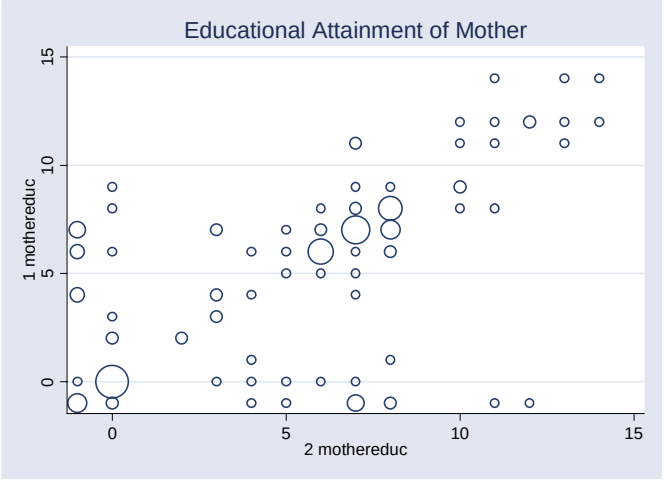
[a] These figures are only calculated among the pairs of survey-resurvey responses that do not contain responses of "don't know" or responses with impossible values.

Figure 3: Reliability of survey data









Notes: These figures plot survey values against resurvey values. Points are weighted to denote number of observations included. A value of "-1" denotes a response of "don't know". Responses with impossible values are excluded.

Table 22: Recall consistency in father's education and individual characteristics

	Dependent Variable: Indicator for Survey-Resurvey Match in Report of Father's Education				
	(1)	(2)	(3)	(4)	(5)
Years assigned deworming	-0.05	-0.051	-0.062	-0.026	-0.102
	[0.033]	[0.034]	[0.053]	[0.038]	[0.078]
Female	0.101				
	[0.100]				
Age (1998)	-0.026	-0.03	0.01	-0.018	0.078
	[0.036]	[0.038]	[0.068]	[0.041]	[0.106]
School average mock score (1996)	-0.072	-0.097	-0.477		
	[0.106]	[0.114]	[0.287]*		
Dropped out before completing primary school	-0.208	-0.214	-0.26	-0.172	-0.094
	[0.135]	[0.138]	[0.151]*	[0.186]	[0.206]
Father alive	0.015	-0.004	-0.004	0.019	-0.009
	[0.126]	[0.130]	[0.135]	[0.199]	[0.226]
Mother alive	0.065	0.058	0.021	-0.004	-0.058
	[0.198]	[0.197]	[0.205]	[0.266]	[0.272]
Living with father	-0.002	0.018	0.038	-0.021	0.029
	[0.148]	[0.150]	[0.152]	[0.211]	[0.238]
Living with mother	-0.058	-0.101	-0.094	-0.078	-0.073
	[0.164]	[0.174]	[0.182]	[0.233]	[0.263]
Average school participation (1998)	-0.033	0.022	0.036	0.564	0.853
	[0.175]	[0.188]	[0.186]	[0.337]*	[0.354]**
Pupil test score (1998)				0.071	0.112
				[0.063]	[0.149]
Household has a latrine (1998)				0.073	0.16
				[0.171]	[0.151]
Household owns cattle (1998)				-0.038	-0.156
				[0.145]	[0.151]
Child weight, kg (1998)				0.005	0.003
				[0.010]	[0.012]
Yrs assigned deworming * Female			0.061		0.115
			[0.080]		[0.105]
Yrs assigned deworming * Age			-0.009		-0.038
			[0.017]		[0.027]
Yrs assigned deworming * School avg mock score			0.117		
			[0.071]		
Yrs assigned deworming * Pupil test score					-0.006
					[0.045]
Observations	129	129	129	88	88
Mean [std dev] of dependent variable	0.488 [0.502]	0.488 [0.502]	0.488 [0.502]	0.500 [0.503]	0.500 [0.503]

Notes: Probit specifications, with marginal effects evaluated at mean values. The sample employed in columns (1)-(3) includes all individuals who were resurveyed and have information on school attendance in 1998, parent death, parent residential location, and dropout status at the time of interview (from the main survey). The sub-sample employed in columns (4) and (5) additionally includes only those with 1998 pupil questionnaire and individual test score data. All regressions are weighted in order to maintain initial

population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. Years assigned deworming is calculated using treatment group of school and individual's standard in 1998, and is not adjusted for females over the age of 13. Age measures are demeaned. School average test score data has been converted into units of individual standard deviations. All specifications include controls for missing age data, and (3) and (5) include an interaction with years assigned deworming. Unless otherwise noted, variables are measured at time of KLPS-1 interview. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. In columns (3) and (5), we cannot reject the null that deworming treatment and all of its interaction terms are jointly zero.

Table 23: Recall consistency in mother's education and individual characteristics

	Dependent Variable: Indicator for Survey-Resurvey Match in Report of Mother's Education				
	(1)	(2)	(3)	(4)	(5)
Years assigned deworming	0.043	0.064	0.095	0.015	0.147
	[0.035]	[0.036]*	[0.063]	[0.044]	[0.083]*
Female	0.017				
	[0.100]				
Age (1998)	0.019	0.035	0.037	0.104	0.298
	[0.031]	[0.033]	[0.072]	[0.044]**	[0.099]***
School average mock score (1996)	-0.172	-0.264	0.197		
	[0.112]	[0.145]*	[0.386]		
Dropped out before completing primary school	-0.185	-0.21	-0.257	-0.121	-0.263
	[0.112]*	[0.101]**	[0.103]**	[0.174]	[0.133]**
Father alive	0.074	0.024	0.059	0.08	-0.196
	[0.167]	[0.163]	[0.157]	[0.181]	[0.227]
Mother alive	-0.028	0.01	0.017	0.142	0.259
	[0.182]	[0.191]	[0.193]	[0.219]	[0.188]
Living with father	-0.083	-0.038	-0.063	-0.162	0.115
	[0.125]	[0.125]	[0.124]	[0.169]	[0.226]
Living with mother	-0.052	-0.159	-0.161	-0.079	-0.242
	[0.131]	[0.143]	[0.150]	[0.186]	[0.266]
Average school participation (1998)	-0.005	0.159	0.12	0.429	0.066
	[0.190]	[0.235]	[0.229]	[0.289]	[0.337]
Pupil test score (1998)				0.064	-0.036
				[0.069]	[0.157]
Child falls sick often, self-report (1998)				0.137	0.096
				[0.141]	[0.157]
Household has a latrine (1998)				-0.169	-0.062
				[0.150]	[0.195]
Household owns cattle (1998)				-0.169	-0.136
				[0.125]	[0.160]
Child weight, kg (1998)				-0.026	-0.043
				[0.012]**	[0.015]***
Yrs assigned deworming * Female			-0.097		-0.154
			[0.083]		[0.106]
Yrs assigned deworming * Age			-0.001		-0.035
			[0.018]		[0.026]
Yrs assigned deworming * School avg mock score			-0.128		
			[0.105]		
Yrs assigned deworming * Pupil test score					0.028
					[0.047]
Observations	128	122	122	85	81
Mean [std dev] of dependent variable	0.430 [0.497]	0.430 [0.497]	0.430 [0.497]	0.430 [0.497]	0.447 [0.500]

Notes: Probit specifications, with marginal effects evaluated at mean values. The sample employed in columns (1)-(3) includes all individuals who were resurveyed and have information on school attendance in 1998, parent death, parent residential location, and dropout status at the time of interview (from the main survey). The sub-sample employed in columns (4) and (5) additionally includes only those with 1998 pupil questionnaire

and individual test score data. All regressions are weighted in order to maintain initial population proportions, and standard errors are corrected for clustering at the 1998 school level. Robust standard errors in brackets. Years assigned deworming is calculated using treatment group of school and individual's standard in 1998, and is not adjusted for females over the age of 13. Age measures are demeaned. School average test score data has been converted into units of individual standard deviations. All specifications include controls for missing age data, and (3) and (5) include an interaction with years assigned deworming. Unless otherwise noted, variables are measured at time of KLPS-1 interview. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. In column (3) we cannot reject the null that deworming treatment and all of its interaction terms are jointly zero. In column (5) we can reject this hypothesis at the 10% level of significance.