User's Guide to the

Kagera Health and Development Survey Datasets

Development Research Group
The World Bank

December 2004

Table of Contents

INTRODUCTION

It has long been established that poverty worsens health. It has been more difficult to establish that poor health worsens poverty or that health improvements can stimulate economic growth. The AIDS epidemic has dramatically raised the mortality rate among adults in their most productive years in Sub-Saharan Africa. What will be the impact on poverty and human capital, in a region where incomes, schooling, and health status are already low? To answer this question and to contribute to improved design of cost-effective programs to deal with the impact on households, the Population and Human Resources Department and the Africa Technical Department of the World Bank jointly launched a research project on "The Economic Impact of Fatal Adult Illness from AIDS and Other Causes in Sub-Saharan Africa" in 1991.[1]

To measure the impact of adult mortality and morbidity on the welfare of individuals and households, the research project launched a longitudinal household survey, known as the Kagera Health and Development Survey (KHDS), in the Kagera region of Tanzania from 1991-94. This region of 1.9 million people, located on the western shore of Lake Victoria adjacent to Uganda and Rwanda and not far from Zaire, is at a political and economic crossroads that is also at the epicenter of the AIDS epidemic in East Africa. The first case of AIDS in Tanzania was identified in Kagera in 1983, and subsequent serological studies have found infection rates among adults as high as 24% in the regional capital of Bukoba in the late 1980s (Killewo and others 1990).

The KHDS interviewed more than 800 households from nearly 50 communities in all five districts of Kagera. Households, community leaders, health facilities, schools, and market vendors were queried in 6-7 month intervals for up to four survey periods. Traditional healers were also interviewed once. Although the KHDS questionnaires were adapted from the World Bank's Living Standards Measurement Study (LSMS) questionnaires, the KHDS was unique because it was fully longitudinal.[2] The panel design offered the researchers the opportunity to measure the change in household consumption and assets between interviews, and thus, to estimate household saving or dissaving—key household-level coping mechanisms.

This document describes the KHDS research design, sampling and survey organization, the questionnaires, and the basic structure of the data for researchers who wish to use the KHDS datasets. Copies of all questionnaires and the interviewer and supervisor manuals can be obtained from the Living Standards Measurement Study web site:

http://www.worldbank.org/lsms/lsmshome.html

---

[1] The research project is also known as the Economic Impact of Adult Mortality (EIAM) Study (Over and Ainsworth 1989). In addition to the authors' sponsoring units, funding was provided by the World Bank Research Committee (RPO674-99, RPO671-75) , the United States Agency for International Development (USAID), and the Danish Agency for Development Assistance (DANIDA).

[2] Although the LSMS has occasionally been administered to a rolling panel of households such that half of the households are unchanged from the previously used sample, the KHDS was the first fully longitudinal LSMS survey.

Figure I.1: Kagera Region, Tanzania

RESEARCH OBJECTIVES AND DESIGN: AN OVERVIEW

A. RESEARCH OBJECTIVES

The primary objective of the Kagera Health and Development Survey (KHDS) was to estimate the economic impact of the death of prime-age adults on surviving household members. This impact was primarily measured as the difference in well-being between households with and without the death of a prime-age adult (15-50), over time. An additional hypothesis was that households in communities with high mortality rates might be less successful in coping with a prime-age adult death. Thus, the research design called for collecting extensive socioeconomic information from households with and without adult deaths in communities with high and low adult mortality rates. Data collected by the KHDS can be used to estimate the "direct costs" of illness and mortality in terms of out-of-pocket expenditures, the "indirect costs" in terms of foregone earnings of the patient, and the "coping costs" in terms of changes in the well-being of other household members and in the allocation on of time and resources within the household as these events unfold.

The KHDS was an economic survey. It did not attempt to measure knowledge, attitudes, behaviors or practices related to HIV infection or AIDS in households or communities. It also did not collect blood samples or attempt to measure HIV seroprevalence; this would have substantially affected the costs and complexity of the research and possibly the willingness of households to participate. Information on the cause of death in the KHDS household survey is based on the reports of surviving household members; the researchers maintained that household coping will respond to the perceived cause of death, irrespective of whether the deceased actually died of AIDS. Lastly, the KHDS did not attempt to measure the psycho-social impact of HIV infection or AIDS deaths.

B. OVERVIEW OF THE RESEARCH DESIGN

The research design called for a longitudinal survey of a sample of households, some of which would experience an adult death and some of which would not, some of them drawn from communities with high adult mortality rates, and some drawn from low-mortality communities. The sampling frame for the survey was based on the 1988 Tanzania Census, which also provided information on adult death rates by ward within Kagera region.

While it was possible to determine which communities had relatively high and low adult death rates from the census data, two additional problems arose that led to the decision for a stratified sample of households based on multiple criteria:

- First, despite the high rates of HIV infection in Kagera and the large number of deaths over time due to AIDS, the death of a prime-age adult is still a relatively rare event over a short time period. This meant that a very large sample would have had to be selected in order to

ensure that the survey could interview enough families suffering our about to suffer the death of a prime-age adult.

- Second, HIV prevalence and adult mortality rates in Kagera were geographically concentrated and thus strongly correlated with different climates and cropping patterns. The highest rural HIV infection rates were in the northeast (10% in Bukoba Rural and Muleba districts and 24% in the town of Bukoba), where tree crops (bananas, coffee) were predominant, while the lowest rates were in the south and west (0.4% in Ngara and Biharamulo districts), where perennial crops and livestock are more common (Killewo and others 1990). A survey design stratified only on mortality rates might confound the effects of high mortality with different agricultural, soil, and rainfall patterns.

Thus, as is described in more detail in the next chapter, the sample of households was selected from a stratified random sample of communities from the 1988 census (stratified on agroclimatic zone and adult mortality rate). Within communities, the household sample was stratified according to the anticipated risk of each household of suffering a prime-age adult death. Households were classified as "high-risk" or "low-risk", based on information obtained from a house-to-house enumeration of all selected communities.

One additional concern was that the high mortality of households might lead to attrition from the sample that is systematically related to household coping. For example, if out-migration is an important coping behavior, then the most severely affected households might leave the sample and the analysis of the remaining households would understate the economic impact of adult deaths. For this reason, at the conclusion of the fieldwork, interviewers attempted to locate and interview all of the individuals who were members of households that dropped out of the longitudinal survey between the first and last interviews, and who were still resident in the region. Individuals were given a specially designed "follow-up questionnaire" that included much of the individual information collected in the household questionnaire, plus information on the reason for leaving the sample and the characteristics of the household were they were now residing.[3]

The final longitudinal household survey followed 816 households at 6-7 month intervals, over a 24-month period from 1991-94. The 816 households were selected from 51 "clusters" of 16 households each located in 49 villages or urban areas representing four economic zones of all districts in Kagera region and, within each zone, representing areas with both high and low adult mortality.

Because household coping behavior is conditioned on local prices, services, and available programs, the KHDS also collected data from the communities from which households were drawn, local markets, the nearest source of modern medical care, and all of the primary schools in the community. This information was collected longitudinally, with the exception of a questionnaire for traditional healers, which was administered only once. While households were drawn from a stratified random sample of households, the health facilities, schools, markets and

---

[3] The attempt to locate and interview those who dropped out is another novel aspect of the KHDS among the other living standards surveys sponsored by the World Bank.

healers interviewed represent those closest to each community and thus are not random samples that are statistically representative of Kagera facilities.

Figure II.1: Location of the KHDS clusters in Kagera Region, Tanzania
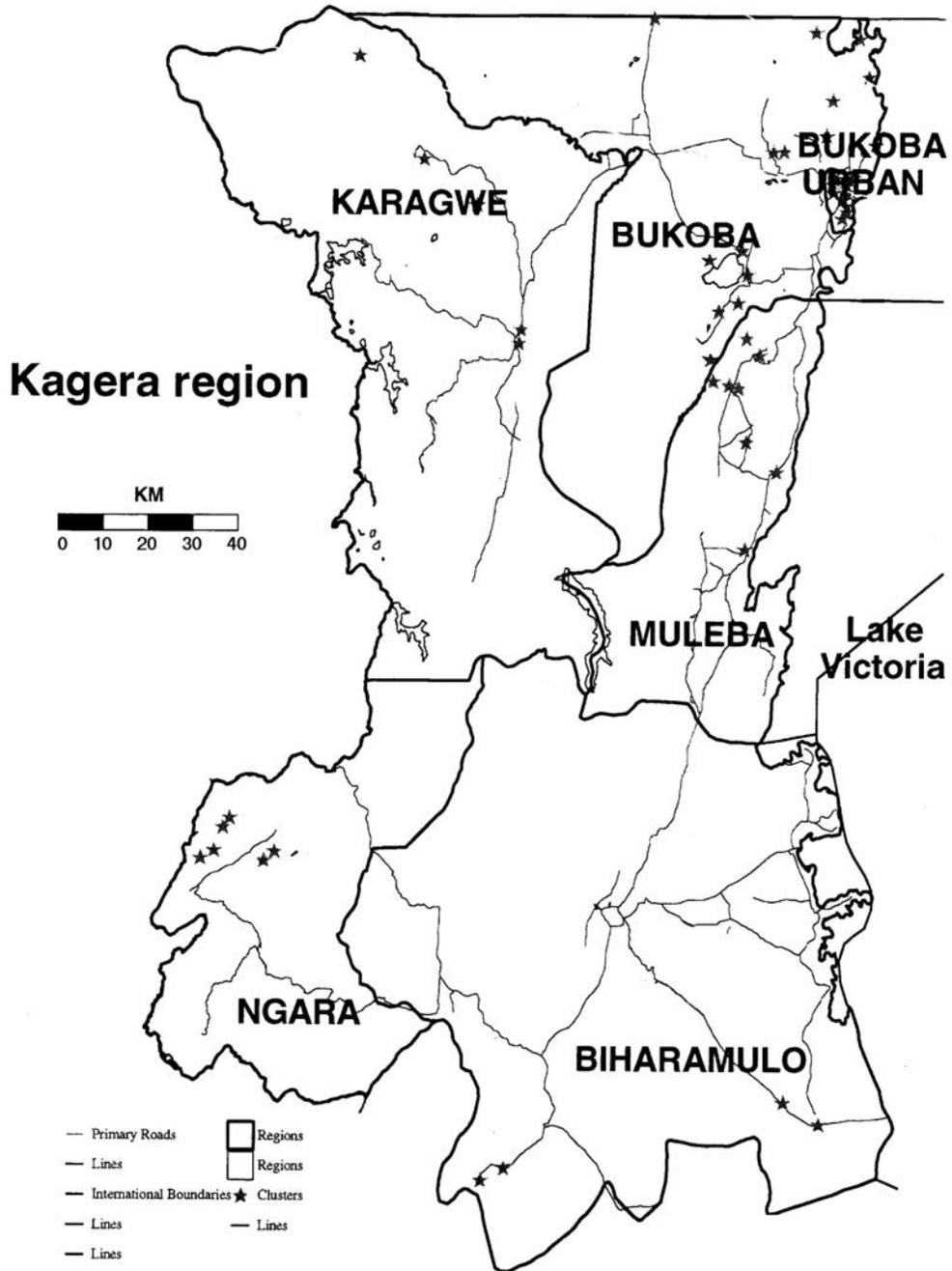
C. THE DEFINITION OF A HOUSEHOLD

In the KHDS, a *household* was defined as a person or group of persons who live in the same dwelling and eat meals together for at least three of the 12 months preceding the date of the survey. There are four exceptions to this definition:

(1) Persons who have recently joined the household, such as spouses, newborn infants, adopted orphans and others who intend to stay until the next interview.

(2) The *head of the household* is identified by the household without any criteria established by the study team and is considered a household member regardless of his/her length of absence.

(3) "Makubaliano" servants, who live with the household without contracts, are considered household members as long as they satisfy the residency requirement.

(4) Tenants and boarders are not household members, regardless of their length of residence.

There were often important changes in household composition between interviews. To be classified as a continuing household (and not a new one), at least one member from the household of the previous interview had to be a continuing household member living in the same dwelling. This requirement was satisfied in all instances, except for cases in which the sole household member of a single-person household died.

D. Types of Questionnaires and the Timing of Fieldwork

Fieldwork was conducted in four distinct time intervals, or *passages*, that lasted 6-7 months each. For example, the first passage of fieldwork took place between September 1991 and May 1992, during which time questionnaires were administered once in all of the households, communities, markets, schools, and health facilities in the sample.

During each passage, interviewers visited each household twice, completing the first half of the household questionnaire in the first visit and the second half of the questionnaire two weeks later. These two household visits within a given passage are called *rounds*.

The *wave* of the household questionnaire corresponds to the number of times that a given household has been interviewed. There are four distinct household questionnaires labeled as *wave 1, wave 2, wave 3,* and *wave 4*. All households interviewed for the first time received a *wave 1* questionnaire, those interviewed for the second time received a *wave 2* questionnaire, a third time a *wave 3* questionnaire, and so forth. When households dropped out mid-survey, they were replaced with new households, which were interviewed for the first time with a *wave 1* household questionnaire, irrespective of the passage. Thus, all households received a wave 1 questionnaire during the first passage, as well as those interviewed for the first time in the second and third passage.[4] Likewise, while most households completed a wave 2 questionnaire during the second passage, households interviewed for the second time during the third and fourth passages also completed a wave 2 questionnaire.

Thus, in the case of the household questionnaire, the wave number of the questionnaire does not necessarily correspond to the passage in which the household was interviewed.[5] However, the questionnaires for communities, markets, health facilities, and schools are also labeled by wave, and for them the wave number of the questionnaire corresponds to the passage in which they were administered. The number of questionnaires of each type completed during each passage is summarized in Table II.1. Traditional healers were interviewed only once, during the third passage.

---

[4] There were no households interviewed for the first time in the fourth passage.

[5] As will be shown in later chapters, this has important implications for the linking of household variables to cluster-level datasets, like data from the price and facility questionnaires: matching wave 1 household data (which can be from any passage) to "wave 1" price and facility data (all of which is from passage 1) does *not* ensure that the variables were collected contemporaneously.

Table II.1: The number and type of questionnaires completed, by passage and wave

Questionnaire	Unit of observation	Wave of the questionnaire	Timing of Fieldwork			
			Passage 1 (9/30/91 - 5/10/92)	Passage 2 (4/23/92 - 11/30/92)	Passage 3 (11/14/92 - 5/25/93)	Passage 4 (6/6/93 - 1/5/94)
Household	Households and individuals	Wave 1	840	46	29	
		Wave 2		803	44	28
		Wave 3			786	43
		Wave 4				759
Follow-up	Individuals	a				169[b]
Community[c]	Community	d	49	49	49	49
Price	Markets or roadside shops	d	53	82	84	90
Health Facility	Nearest health facility	d	42	48	57	61
School	Nearest School(s)	d	62	63	63	64
Traditional Healer	Traditional Healer	e			103	

a. Asked only during the fourth passage.

b. Forty-five of these individuals belonged to 14 households that dropped out in the last passage and moved "intact" between the third and fourth passage. These households were located and administered the household-level sections of the wave 4 questionnaire.

c. There were 51 clusters of households in 49 communities. Some community questionnaires served more than one cluster in the same community.

d. The wave number for these questionnaires corresponds exactly to the passage in which they were administered (i.e., the Wave 1 price questionnaire was in passage 1, the Wave 2 price questionnaire was in passage 2, and so forth.)

e. Asked only in the third passage.

III.   SAMPLE DESIGN AND SELECTION

Qualitative studies of small samples of households can point to hypotheses about the ways in which fatal adult illness affects households. However, policymakers need to know which households are suffering the most, the size of the impact, the extent to which they suffer more than other households in a poor country, and the potential costs and effects of assistance programs. For this purpose, the sample of households must be representative of the population, a random sample for which the probability of selecting each household from the whole population is known.

The KHDS used a random sample that was stratified geographically and according to several measures of adult mortality risk. This strategy allowed the team to ensure an adequate number of households with an adult death in the sample while retaining the ability to extrapolate the results to the entire population.   The results from the household survey show that stratification of the sample on mortality risk at both the community and household level proved to be worthwhile. Among the 816 households in the original sample that began the survey in the first passage, 91 had an adult death in the course of the survey—more than three times the expected number (25) had the households been drawn at random with no stratification.[6]

A.   THE TWO-STAGE STRATIFIED RANDOM SAMPLING PROCEDURE

The KHDS household sample was drawn in two stages, with stratification based on geography in the first stage and mortality risk in both stages.

1.   First Stage: Selection of communities and clusters

In the first stage of selecting the sample, the 550 *primary sampling units* (PSUs) in Kagera region were classified according to eight strata defined over four agronomic zones and, within each zone, the level of adult mortality (high and low). A PSU is a geographical area delineated by the 1988 Tanzanian Census that usually corresponds to a community or, in the case of a town, to a neighborhood. Clusters of households were drawn randomly from the PSUs in each stratum, with a probability of selection proportional to the size of the PSU.

a)   Classification of communities by sampling stratum

The four *agronomic zones* are:[7]

---

[6] The 816 households that began the survey in the first passage were observed, on average, for 1.6 years, generating a total of 1,322.7 years of observation. The average probability of an adult death per household per year, according to the 1988 Tanzania Census, is 0.0188. Thus, the expected number of deaths from a random sample of 816 households observed for 1.6 years is 25. Because households were added to the sample to compensate for attrition, a total of 918 households were eventually interviewed at least once. Between the first and last interview, 102 of these households had an adult death, compared to 27 households that would have been expected to have a death from from a non-stratified sample. Table II.1 indicates 840 households in Passage 1 because 24 "extra" households were interviewed. See page 20 for further information.
[7] Source: Government of Tanzania. 1967. *Tanzania Atlas*.

- Tree Crop Zone: Low fertility soils in areas of high rainfall, where the main crops are bananas, coffee, and tea. This zone is in the northern part of Kagera with communities in Bukoba Rural and Muleba Districts.
- Riverine Zone: Alluvial and colluvial soils of considerable potential, but requiring flood control, where the cropping pattern is mixed and includes cereals, sugarcane, rice, and legumes, as well as tree crops. This zone is in the middle of the region; most of its communities are in Karagwe and Bukoba Rural Districts and a few in Muleba District.
- Annual Crop Zone: Soils of low to medium fertility with moderate potential and lower rainfall, where the cropping pattern is mixed and includes groundnuts, cassava, beans, cotton as well as some cereals (maize, sorghum) and pasture, but few tree crops. This is in the southern part of Kagera in Biharamulo and Ngara Districts.
- Urban Zone: The town of Bukoba, the region's capital, plus an additional 27 communities in Muleba, Karagwe, Ngara, and Biharamulo Districts that were designated as urban by the 1988 census.

The zone labels were chosen for simplicity. They suggest the characteristic, though by no means exclusive, agricultural pattern.

Within each agronomic zone, PSUs were classified according to the level of adult mortality. The 1988 Tanzanian Census asked a 15 percent sample of households about recent adult deaths. Those answers were aggregated at the level of the "ward", an administrative area that is smaller than a district. The adult mortality rate (15-50) was calculated for each ward and each PSU was assigned its ward's adult mortality rate. [8]

Because the adult mortality rates were much higher in some zones than others and the distribution was quite different within zones, "high" and "low" mortality PSUs were defined relative to other PSUs within the same zone. A PSU was classified as in the "high" mortality category if its ward adult mortality rate was at the 90th percentile or higher of the ward adult mortality rates within a given agronomic zone. Table III.1 shows the distribution of PSUs, households, and population across the eight strata.

---

[8] There were 111 wards in Kagera in 1988. We are grateful to Mr. Sylvester A. M. Ngallaba, Senior Statistician and Director of Population, Bureau of Census, Dar es Salaam, Tanzania for providing preliminary results of the Tanzanian census for use as the sampling frame and for stratifying the sample by adult mortality rate.

Table III.1: Distribution of PSUs, Households, and Population
by Adult Mortality Rate and Agronomic Zone, Kagera Region, 1988

| Agronomic Zone | Adult Mortality Rate, 1988 | | Total |
	Low	High	
Tree Crop	174	18	192
	74,952	**6,685**	**81,637**
	365,056	*30,743*	*395,799*
Riverine	151	18	169
	87,414	**8,923**	**96,337**
	420,528	*39,340*	*459,868*
Annual Crop	126	19	145
	56,205	**8,673**	**64,878**
	288,685	*44,136*	*332,821*
Urban	37	6	43
	21,138	**3,745**	**24,883**
	99,162	*15,615*	*114,777*
Total	488	61	549
	239,709	**28,026**	**267,735**
	1,173,431	*129,834*	*1,303,265*

* Figures in regular, bold, and italic typeface are respectively the number of communities (PSUs) , households, and individuals in each stratum. The threshold for the stratum with "high" adult mortality was 29/1000 in the Tree Crop zone, 17/1000 in the Riverine Zone, 8/1000 in the Annual Crop Zone, and 20/1000 in the Urban Zone.

b) Selection of Clusters

Having classified all of the PSUs in Kagera into the eight strata, it remained to select the PSUs from which households would be drawn, and how many households would be interviewed in each of the selected PSUs. To facilitate fieldwork and reduce its costs, the KHDS interviewed households within PSUs in ***clusters*** of 16 households each. Based on experience with other LSMS surveys, this is the number of households that could reasonably be interviewed by a field team of one supervisor, two interviewers, and an anthropometrist in a week. The probability that a PSU was selected within each stratum was proportional to its size (the number of households), according to the following formula:[9]

$$\begin{array}{c} \textit{Probability} \\ \textit{of Selecting} \\ \textit{this PSU} \end{array} = \dfrac{\begin{array}{c}\textit{Number of Clusters} \\ \textit{in this Stratum}\end{array} \ x \ \begin{array}{c}\textit{Number of households} \\ \textit{in this PSU}\end{array}}{\textit{Number of HHs in this Stratum}}$$

---

[9] In the high-mortality urban stratum, eight clusters had to be selected from only six PSUs. In that case, for some PSUs in the stratum the probability exceeded one and more than one cluster was drawn from a PSU.

The research design and the budget called for surveying 50 clusters of 16 households each, for a total sample of 800 households. Divided across the eight strata, this would imply the need to enumerate roughly 6 PSUs in each of six strata and 7 PSUs in two strata (see the first figure in each column of Table III.2). However, to guard against attrition of entire communities and the possibility that actual mortality rates would be found to be quite different from those observed in the census, more PSUs were enumerated than would be needed for the survey. A total of 62 PSUs were selected from the 549 in the region to be enumerated--8 PSUs were selected at random from each of seven strata and all 6 PSUs in the high-mortality urban stratum were selected.

However, the field teams successfully enumerated only 52 PSUs, from which 54 clusters could be drawn (see the second figure in each column of Table III.2). Ten PSUs were not enumerated, generally because they were inaccessible or the teams ran out of time.[10]

Table III.2: Number of PSUs in sample plan, number enumerated, and number selected and interviewed, by stratum

| Agronomic Zone | Adult Mortality Rate*, 1988 | | |
	Low	High	Total
Tree Crop	7/8/8	6/8/7	14/16/15
Riverine	6/6/5	6/7/7	12/13/12
Annual Crop	6/6/6	6/4/4	10/10/10
Urban	7/7/6	6/6/6**	12/13/12**
Total	26/27/25	24/25/24**	48/52/49**

* The numbers in each cell are: The planned number of PSUs to be selected from each stratum/the number of PSUs enumerated per stratum/the number of PSUs interviewed during the survey per stratum.
** Two clusters were to be selected from each of two PSUs in the urban high-risk zone

Of the 52 PSUs that were enumerated, only 48 were needed (allowing for selection of two clusters from each of two PSUs in the urban high-mortality zone). In zone 3, where fewer PSUs were enumerated than were anticipated in the research design, all 10 enumerated PSUs were accepted into the sample. To compensate, it was decided to select a total of 14 PSUs in the tree crop zone, and 12 each in the annual crop and urban zones, for a total of 48. In deciding which PSUs to drop, the PSUs were ordered within each zone, from highest to lowest adult death rate *based on the*

---

[10] Two PSUs were on islands and one was in a game park. The rainy season substantially slowed down the enumeration and made some PSUs inaccessible. Among the 10 PSUs not enumerated, three were in the riverine zone, six were in the annual crop zone, and one was in the urban zone. The distribution of the ten PSUs not enumerated by district is: Bukoba Urban (0); Bukoba Rural (1); Muleba (1); Biharamulo (3); Ngara (3); Karagwe (2).

enumeration results. [11] In order to maximize the differences between PSUs in the high- and low-mortality groupings within a zone (the definition of which remained based on the census), the PSUs dropped from each zone were in the "middle" of the distribution of enumeration adult mortality rates for that zone. For example, in the riverine zone, where 13 PSUs were enumerated, the PSU with the median adult mortality rate from the enumeration was dropped. Using this method, one PSU each was dropped from the riverine and urban zones and two were dropped from the tree crop zone, leaving 48 PSUs from which 50 clusters were selected. A 51st cluster from the high-mortality tree crop stratum was added toward the end of the first passage of field work, to ensure that an adequate sample size would be maintained should an entire cluster drop out later during the panel. Thus, the final KHDS sample included 49 PSUs from which 51 clusters of 16 households each were drawn.

2. Second Stage : Selection of Households

In the second stage, households within each of the selected PSUs were assigned to one of two strata—"sick" or "well"—based on the results of an enumeration of all households in each community. Sixteen households were selected at random per cluster, of which 14 were selected from the "sick" group and 2 from the "well" group.

a) Enumeration of households

Between March 15 and June 13, 1991, 29,602 households were enumerated in 52 primary sampling units. In addition to recording the name of the head of each household, the number of adults in the household (15 and older), and the number of children, the enumeration form asked:

- Are any adults in this household ill at this moment and unable to work? If so, the age of the sick adult and the number of weeks he/she has been too sick to work..

- Has any adult 15-50 in this household died in the past 12 months? If so, the age of each adult and the cause of death (illness, accident, childbirth, other).

The enumeration form asked explicitly about illness and death of adults 15-50 because this is the age group disproportionately affected by the HIV/AIDS epidemic and it is the impact of these deaths that were of research interest.   Since AIDS is sexually transmitted, other adults in the same household with an AIDS patient may also become infected, either through sexual contact with the HIV-infected person or because of similarities in sexual behavior. Thus, AIDS morbidity and deaths are likely to be clustered in households. Information on illness and deaths on the enumeration form could be recorded for a maximum of three people (for each question) per household.

Of the more than 29,000 households enumerated, only 3.7%, or 1,101, had experienced the death of a an adult 15-50 due to illness during the twelve months before the interview and only 3.9%, or

---

[11] The correspondence between the adult mortality rates from the 1988 Tanzanian Census and the rates found by the enumeration was not particularly good. The AMR from the enumeration were often higher for PSUs classified as "low mortality" within a zone, than they were for "high mortality", and vice-versa.

1,145, contained a prime-age adult too sick to work at the time of the interview (see Table III.3). Only 77 households had both an adult death due to illness and a sick adult. This underscores the point that, even with some stratification based on community mortality rates and in an area with very high adult mortality due to an AIDS epidemic, a very large sample would have had to have been selected to observe a sufficient number of households that would experience an adult death during the two-year survey.

Table III.3: Distribution of enumerated households by illness and deaths from illness among adults 15-50

Mortality	Morbidity		Total	(Percent)
	Too sick to work	Not too sick to work		
Death due to illness	77	1,024	1,101	(3.7)
No death due to illness	1,068	27,433	28,501	(96.3)
Total	1,145	28,457	29,602	
(Percent)	(3.9)	(96.1)	(100)	(100)

Note: There was more than one adult death due to illness in 82 households and there was more than one adult too sick to work in 42.

Source: KHDS enumeration data from all 52 PSUs that were enumerated.

b) Selection of households

To further increase the probability of capturing households with adult deaths in the sample, households were stratified according to the extent of adult illness and mortality. It was assumed that in communities suffering from an HIV epidemic, a history of prior adult death or illness in a household might predict future adult deaths in the same household. The households in each enumerated PSU were classified into two groups, based on their response to the enumeration:

- "Sick" households: Those that had either an adult death 15-50 due to illness in the past 12 months, an adult too sick to work at the time of the survey, or both (n=2,169).
- "Well" households: Those that had neither an adult death 15-50 due to illness nor an adult 15-50 too sick to work (n=27,433).

In selecting the 16 households to be interviewed in each PSU from which a cluster was drawn, 14 were selected at random from among the "sick" households in that PSU and 2 were selected at random from among the "well" households. In one cluster, where the number of "sick" households available was less than 14, all available sick households were included in the sample and the balance were from well households. *The final sample drawn for the first passage was therefore 816 households in 51 clusters drawn from 49 PSUs (see Table III.4).*

Table III.4: Distribution of households (number of clusters) selected
for the KHDS sample, by stratum

Agronomic Zone	Adult Mortality Rate*, 1988		Total
	Low	High	
Tree Crop	128	112	240
	(8)	(7)	(15)
Riverine	80	112	192
	(5)	(7)	(12)
Annual Crop	96	64	160
	(6)	(4)	(10)
Urban	96	128	224
	(6)	(8)	(14)
Total	400	416	816
	(25)	(26)	(51)

B. HOUSEHOLD ATTRITION AND REPLACEMENT PROCEDURES

1. Attrition from the household sample

a) Attrition between the enumeration and the first passage

Among the original 816 households selected from the enumeration, 47 (5.8%) could not be interviewed during the first passage, which occurred 7-12 months after the enumeration. The most important reason for attrition was that the household had moved (53% of the cases, see Table III.5). In about a third of these cases, the move was related to the death of a household member. This included five cases in which the household moved following a death and two cases in which the person who died was a single-person household. In nine cases (19%) the household was not interviewed because the head was away.[12] Only 4 households—less than half of a percent of the entire sample of 816 households—refused to participate.

---

[12] This was in fact an error on the part of the interviewers early in the survey, which was subsequently corrected. The presence of the household head was not necessary to conduct the interview, unless a household was a single-person household.

Table III.5: Household attrition from the KHDS sample

Reason	Attrition between the enumeration and first passage		Attrition between the first and last passage	
	Number	Percent	Number	Percent
Moved (not related to death)	18	38.3	46	56.8
Moved (related to death)	7	14.9	19	23.4
Head away	9	19.1	2	2.5
Refused	4	8.5	13	16.0
Illness	2	4.3	1	1.2
Not found	1	2.0		
Reason unknown	6	12.8		
Total	47	100.0	81	100.0
Sample size	816		840	
Percent of sample lost	5.8		9.6	

Note: The 81 households that dropped out after the first passage were replaced with new households; five of the replacements also subsequently dropped out. The reasons for "multiple replacements" are not included in this table
Source: Ainsworth, Ghosh, and Semali (1995), Annex 1.

b) Attrition between the first and fourth passages

During the first passage, a total of 840 households were interviewed. This group included the 816 "original" households selected from the enumeration (or their replacements) and 24 "extra" households. The field teams added these households, taken from the list of replacement households, when they sensed that another continuing household in the sample was likely to drop out or was a poor source of information.[13] By the end of the fourth passage, more than two years later, 81 households (9.6% of the 840 interviewed in the first passage) had dropped out (see Table III.5). In 80 percent of the cases, the reason for attrition was that the household moved; about a third of those moves were related to an adult death in the household, including one case in which a single-person household died. Only 13 households--16% of the household attrition during the panel--refused to participate. Taken over all 840 households interviewed during the first passage, only 1.5% of the households completing a questionnaire in the first passage refused to be interviewed by the end of the survey.

c) Household attrition and adult deaths

While there is no indication that adult deaths were the major reported cause of attrition, it was nevertheless not uncommon for a move to be associated with an adult death. Were households with an adult death more likely to drop out of the sample? In fact, households with an adult death in the 12 months before the enumeration were *less* likely to drop out before the first passage than were households without a death (see Table III.6). On the other hand, the 94 households that had an adult death between passages one and four were half again as likely as

---

[13] Extra households were interviewed during the first passage at the initiative of the field manager. The "extras" were selected from the list of replacement households, however the decision rule for adding extra households was not well documented. An additional 75 households began the survey in later passages, completing a wave 1 questionnaire at the first interview. Their subsequent attrition (5 households) is not studied here.

households without a death to drop out by the end of the fourth passage. Neither of these differences is statistically significant, however.[14]

Table III.6: Attrition of household with and without an adult death (percent)

Type of household and reference period	n	Attrition between the enumeration and first passage	Attrition between the first and last passage
Adult death before the enumeration	347	4.3	
No adult death before the enumeration	469	6.8	
Adult death between first and fourth passage of the panel	94		14.9
No adult death between first and fourth passage	746		9.0

Source: Ainsworth, Ghosh, and Semali (1995), Annex 1.

d) Follow-up of individuals in households that moved

In the course of the household survey, between the first and fourth passages, a total of 86 households left the sample, including the 81 households that began the survey in the first passage and 5 households that replaced them and subsequently dropped out. During the fourth passage, the interviewers attempted to locate the 306 individuals who were members of these households—if they were still alive and living in Kagera region—to be interviewed with a "follow-up questionnaire", described in Chapter V. The time that had elapsed since the last interview was from 6-28 months.

The interviewers were able to locate and interview 169 individuals from 52 households that had dropped out, or 55% of the total (see Table III.7). An additional 10 individuals (3%) were known to have died in that interval, and the remaining 127 (42%) individuals were not interviewed, either because they were outside Kagera (4%) or because the whereabouts of the individual could not be determined (37%). Only 3 individuals known to be residing in Kagera could not be found for a follow-up interview.

---

[14] In a logit regression of the probability that a household would continue in the sample, controlling for geographic zone or district, neither an adult death nor an illness prior to the enumeration was a statistically significant predictor of continuation between the enumeration and the first passage. However, both the urban zone and Bukoba Urban district had a highly significant negative impact. In a logit regression of the 840 households beginning the panel, a death between waves significantly lowered the probability of continuing to the end of the panel (p=.08), as did urban location. For a description of the characteristics of households that dropped out and those that didn't, see Appendix 1 of Ainsworth, Ghosh, and Semali (1995).

Table III.7: Distribution of members of households that dropped out of the sample,
according to whether they completed a follow-up questionnaire

Interview status	Female	Male	Total	%
Found and interviewed	79	90	169	55.2
Died	4	6	10	3.3
Not interviewed	66	61	127	41.5
In Kagera, not found	2	1	3	1.0
Outside Kagera	4	7	11	3.6
Whereabouts unknown	60	53	113	36.9
Total	149	157	306	100.0

2. Procedures for replacing households

In order to guard against a dwindling sample and to eliminate any incentive for interviewers to reduce their workload by not striving to find a household, households that moved, refused, or otherwise dropped out were replaced. At the start of the first passage, the team supervisors were provided with a list of additional households chosen at random from the PSU to be used as replacements.

Beginning in the second passage, the supervisors were to replace a household with another of the same type--"sick" or "well" drawn from the list of replacements. They were provided with the names of 12 additional households from each PSU--six each of "type A" (sick households) and "type B" (well households)--and a new list of sampled households in which the type was indicated. The interviewers and supervisors were not told which type of household (A or B) was a "sick" household.

C. SELECTION OF HEALTH FACILITIES, SCHOOLS, MARKETS AND HEALERS

The sample of health facilities, schools, and markets that were interviewed or visited was selected based on the information provided by community leaders. The facilities interviewed generally represent those closest to the cluster, and thus do not represent a random sample of facilities in Kagera region. Traditional healers were randomly selected within each community.

1. Health facilities

The sample consisted of the nearest health facility (dispensary, health center, or hospital) to each cluster, as indicated on the community questionnaire. Where there was more than one health facility in the cluster (i.e., Bukoba town), all health facilities were to be interviewed. At the same time, some clusters shared the same facilities. The number of facilities interviewed over time increased from 42 in the first passage to 61 by the fourth passage (refer back to Table II.1).

2. Primary schools

The sample consists of the nearest primary school to each cluster. In the event that there were several primary schools in a cluster, a separate questionnaire was completed for each. As a result, 62 primary schools were interviewed in the first passage. This increased by the fourth passage to 64 because of two schools inadvertently omitted in earlier passages.

3. Markets

During the first passage, price data were collected from the nearest market to each cluster. There was no distinction made between whether the data were collected from an open market with several stalls or vendors or whether it was a "duka" or shop of a local merchant, although the type of establishment was noted on the form.[15] For the second through fourth passages, in principle, two price questionnaires were completed for each cluster. One was completed for the nearest marketplace and another was completed for the nearest duka(s).

4. Traditional healers

During the third passage, respondents to the community questionnaire were asked to list all of the traditional healers in the community. A total of 317 healers were listed, with 2-13 recorded per cluster. Two healers were selected at random from the list in each cluster to receive the healer questionnaire. An enthusiastic interviewer in fact interviewed a third healer in one cluster, so 103 of these questionnaires were completed in the third passage—two per cluster in 50 clusters and three in one cluster.[16]

---

[15] In theory, there should have been 51 price questionnaires for the first passage, one per cluster. However, in some clusters the interviewer completed separate questionnaires for markets and dukas, even in the first passage. Further, two PSUs and four clusters were selected from the Hamgembe neighborhood of Bukoba town. However, the interviewers failed to realize that a price questionnaire was to be completed *each time* a Hamgembe cluster was interviewed.

[16] The results of the survey of traditional healers are described in Semali and Ainsworth (1995).

IV. SURVEY ORGANIZATION AND FIELDWORK

The following section describes survey preparation, field procedures, the sequencing of fieldwork, and quality control mechanisms.

A. FIELD PROCEDURES

The Kagera Health and Development Survey (KHDS) was conducted in 51 clusters of households throughout Kagera region. The clusters were distributed across districts as follows: Bukoba Urban (11); Bukoba Rural (17); Muleba (8); Biharamulo (4); Ngara (6); and Karagwe (5).

1. Scheduling Field Work

Four mobile teams based in the KHDS project office in Bukoba, Tanzania conducted the field work. Each team was composed of a supervisor, at least two interviewers, an anthropometrist (who took height and weight measurements), and a driver.

Data collection and data entry for each cluster of 16 households took four weeks. During the first and third week the interviewers collected data and during the second and fourth weeks the data were entered at the Bukoba office. The two visits by the field team to a cluster within a given passage are referred to as *rounds*. The typical schedule for one team interviewing one cluster of households was as follows:

Week one: *Data collection, round one.* The team completed the first ten sections of the household questionnaire in all 16 households and returned to the project office.

Week two: *Data entry, round one.* The data entry operator in the project office entered the data for the 16 households on personal computers and performed range and internal consistency checks on the data. A computer printout of all data and inconsistencies was generated, for use during round two.

Week three: *Data collection, round two.* The field team returned to the cluster to correct errors found by data entry in the round one questions and to complete sections 11-20 of the household questionnaire.

Week four: *Data entry, round two.* The data entry operator corrected errors in sections 1-10 and entered data for sections 11-20. Range and internal consistency checks were again performed to detect data entry errors. A second printout of the entire questionnaire was generated and compared with the data on the original

questionnaire. Data entry errors were corrected and the data were sent to the researchers.

During the weeks that data entry was underway for this typical cluster, the field teams interviewed a second cluster. In this way, each team surveyed two clusters per month, or 32 households, and 12-13 clusters during a passage lasting roughly 6 and a half months. The scheduling of interviews across clusters in the first passage was randomized to assure the maximum coverage of Kagera region during any month. Clusters were interviewed in the same order in subsequent passages.

Before launching the survey, a representative of the Field Manager visited each locale with a representative of the District Office or the District Commissioner. Meetings were held with the village chairman, the village secretary, and the ten-cell leaders to explain the objectives and research methods of the project, and to explain that it was not an assistance project. The collaboration of village leaders was requested to help the team to administer the survey and to assure the safety of the team. Village leaders played an important role by introducing the interviewers to the households, and by assuring a safe and harmonious working environment.

2. Community Service Component

The project gave a small gift to each village, as a sign of appreciation for its cooperation. These gifts were given only as a "token" of appreciation, because the project was not an assistance program. In most cases, this assistance took the form of desks, notebooks, textbooks, and building materials for community schools. Households participating in the survey received small gifts of two kilos of sugar and tea at the end of the second round of each passage. This gift was appreciated by respondents but was unlikely to affect households' coping behavior.

B. QUALITY CONTROL

The KHDS field operations used a number of important supervisory checks and a customized data entry program, which minimized errors by the respondents, interviewers, and data entry operators, and guaranteed high-quality data.

- Sections 1-10 of the household questionnaire, completed in round one of each passage, collected information on each household member. Each individual was interviewed separately, in private. This procedure minimized the use of proxy respondents and ensured greater confidentiality during the interview.

- The supervisors observed one interview per interviewer per week to assure that the correct procedures were being followed and that the interviewer fully understood the questionnaires.

- Before leaving the field site, the supervisors completed a list of twenty or more internal consistency checks on each of the questionnaires and conducted random re-interviews for a

subset of questions in one quarter of all households. In the event of mistakes or incomplete answers, the interviewer returned to the household to complete or correct the questionnaire.

- A customized data entry program in the Bukoba office detected internal inconsistencies, missing information, and erroneous answers after each round.[17] The data entry program also compared the following information between waves: sex, age, activities, durable goods, livestock and housing. Discrepancies were flagged to the supervisor. It also produced a printout of all of the entered data, which was then checked by the supervisors or a group of verifiers for errors. The ability to return to the field between rounds to correct the information directly with respondents was expected to markedly improve the accuracy and timeliness of the data.

- The data entry program for the wave 2, 3, and 4 household questionnaires printed a list of durable goods owned the previous passage (Section 16A1) which allowed the interviewer to confirm that every durable good previously declared could be accounted for during subsequent interviews.
- The field manager and a senior supervisor checked the supervisors' work randomly through unannounced field visits.

The quality controls implemented in the KHDS are discussed in greater detail in the Supervisor Manual.

---

[17] Written by Sistemas Integrales, Santiago, Chile.

V.   KHDS QUESTIONNAIRES

The seven main questionnaires of the KHDS aimed to collect four types of variables, essential to analyzing the impact of fatal adult illness: (1) variables reflecting the well-being of individuals; (2) variables that measure individual and household coping or adjustment; (3) exogenous explanatory variables; and (4) policy instruments (see Table V.1).

Table V.1: KHDS Questionnaires

Questionnaire	Objective	Respondent
Household	To measure individual and household well-being, mortality, and use of services.	Head of household, household members
Community	To measure economic characteristics of communities in which the households resided and the availability of basic services. To identify the sample of schools, health facilities, and traditional healers.	Community leaders
Health Facility	To measure the availability, quality, and price of services for use in the analysis of demand for health services	Medical officer, pharmacist
School	To measure the availability, quality, and price of schooling for use in the analysis of enrollment decisions.	Headmaster, head teacher
Price	To construct a price index of key consumer goods across clusters and over time.	Market vendors, shop keepers
Traditional Healer	To measure the availability, quality, and price of traditional medical services.	Traditional Healer
Follow-up	To measure the well-being of individuals whose households left the sample.	Members of households that left the sample

This chapter briefly describes the contents of these questionnaires by way of introduction, and (except for the healer and follow-up questionnaire) we focus on the Wave 1 version of each questionnaire. It is essential for anyone who intends to use the data to obtain a copy of the questionnaires, on which each question appears verbatim as it was asked and the skip instructions are shown. Further, the wave 1 household questionnaire differs in important respects from the wave 2 questionnaire. This is because the first time that households and facilities were interviewed the information was used as a baseline, against which changes in subsequent interviews could be measured. This is also true to some extent in the other questionnaires. The wave 1 household questionnaire is published in Ainsworth and others (1992), and a detailed description of the changes in all questionnaires between wave 1 and subsequent waves or passages is presented in Appendix 1 to this document.

A. HOUSEHOLD QUESTIONNAIRE

The KHDS Household Questionnaire is the main data collection instrument used to assess the impact of fatal adult illness. To gather information across the four variable types listed above, the household questionnaire collected data on individuals and households in the following subjects:

- Demographic characteristics
- Health status, symptoms, health-seeking behavior and medical expenditures
- Nutritional status
- Mortality and related expenditures
- Human capital, enrollments and education expenditures
- Fertility and contraceptive use
- Time use in the labor force, other productive and health-related activities
- Income levels and sources
- Assets and durable goods, including housing, farm and business assets
- Consumption expenditure, by component
- Savings, debts, transfers and receipt of assistance
- Characteristics of non-resident parents and children, including their mortality

1. Development of the household questionnaire

Rather than design the household questionnaire from scratch, the project team adapted the prototype household questionnaire from the World Bank's Living Standards Measurement Survey (LSMS).[18]

The specific needs of the KHDS required important modifications, primarily to measure changes over time and to enable greater analysis of individual well-being within a household:

- A new module to measure the mortality of household members and of other relatives, the cause of death, health care-seeking behavior before death and expenditures associated with health care and funerals.

- Expanded sets of questions on health status, in terms of both acute and chronic conditions, and on the utilization of health care.

- More detailed information about the time use of each household member across a greater number of activities, including caring for sick household members.

---

[18] See Grootaert (1986) and Ainsworth and van der Gaag (1987) for descriptions of the LSMS prototype. Ainsworth and others (1992) present the English translation of the first wave KHDS questionnaire and describe how it is derived from the LSMS prototype.

- An expanded set of questions on remittances and the flow of resources between household members and persons or organizations outside the household.

- Revision of all modules to better capture inter-personal differences in well-being.

- Adaptation of the questionnaire for a longitudinal research design.

The first draft of the household questionnaire was produced in Swahili in March 1991 and subjected to a two-week field test in Kagera Region from April 20 - May 4. The draft questionnaires were asked of different types of households and communities, to ensure that they were capable of collecting the necessary information efficiently and accurately. Attention was also given to making the questionnaire — which included highly sensitive topics such as income, savings, severe illness, and death — culturally acceptable and as inoffensive as possible. The household questionnaire was tested in Bilele ward of Bukoba town and in the village of Lukindo (located about 15 miles from Bukoba in the surrounding hills).

The wave 1 household questionnaire was tested at two levels. First, individual sections of the questionnaire were tested independently, then the complete household questionnaire with all sections was tested on a small sample of households. While many parts of the questionnaire were tested on urban and rural households in general (education, activities, farming, consumption), it was necessary to seek out specific types of households to evaluate other parts of the questionnaire (for example, households that had experienced a recent adult death due to AIDS for the section on mortality).

One of the major challenges of the field test was to design a module on consumption expenditure that would take into account the seasonality of consumption and production in Kagera, where there are two rainy seasons and two dry seasons in a twelve-month period. Respondents were not able to recall total agricultural production or the quantities of agricultural inputs; all quantity questions in the farming section, except for the total amount of a crop that was sold, were dropped. The health section of the questionnaire was also transformed into separate sets of questions about acute illness (in the past four weeks) and chronic conditions (lasting six months or more). One of the important findings of the field test was that, despite their grief, households were willing to discuss the circumstances surrounding recent deaths with project interviewers.

The wave 2 household questionnaire was tested in 29 households in 4 urban and rural communities in March 1992. The main tasks were to develop ways of updating the rosters of household members (section 1), children living elsewhere (section 2), and household durable goods (section 16A), and to adapt the questionnaire to a six-month reference period. In addition, the field test suggested new, inter-wave checks for the data entry program and for the supervisors in the field.

2. Summary of the household questionnaire and respondents

Table V.2 summarizes the structure, respondents, and contents of the wave 1 household questionnaire. Related variables are often found in more than one part of the questionnaire. For

example, demographic data are collected in seven sections, and expenditures were measured in twelve. Thus, researchers interested in measures such as expenditures, assets, income, and transfers, will need to review many different sections of the questionnaire, not just those sections labeled "expenditure", and so forth. The discussion below is a 'bare bones' review of the contents; those interested in the rationale for collecting these variables should refer to the discussion in Ainsworth and others (1992).

a) Section 1: Household roster

The objective of the Household Roster is identify household members.   For the purposes of the KHDS survey, a household is defined as a person or group of persons who live in the same dwelling and eat meals together for at least three of the 12 months preceding the date of the survey. The household has been "defined" in this way purely to achieve consistency in terms of the field work and to identify those individuals who will answer the in-depth questions of the household questionnaire. Other parts of the household questionnaire collect information on a larger set of individuals to which the household is linked, including non-resident parents, children and other individuals to whom and from whom the household sends and receives transfers of cash or goods.

Section 1 collects the names of household members, their relationship to the head of the household, their age, sex, marital status, and the length of time they have been resident. Household members retain the same identification code throughout all waves of the survey.

***Table V.2: Summary of Sections, content,
and respondents of the KHDS Household Questionnaire***

Section	Title	Respondents	Demographic Characteristics / Mortality	Expenditures	Income	Assets	Transfers	Labor	Schooling	Health	Use of Public Services
1	Household roster	Head of Household*	√								
2	Children residing elsewhere	Head of Household*	√						√		
3	Information on parents	Head of Household*	√						√		
4	Overview of Household Activities	Head of Household*									
5	Education	All household members 7 and older		√			√		√		√
6	Health	All household members		√						√	√
7	Activities of Household Members	All household members 7 and older			√		√	√		√	√
8	Migration	All household members	√								
9	Fertility	All female members who are 15 and older or married	√						√	√	
10	Anthropometry	All household members	√							√	√
11	Farming	Most knowledge-able person		√	√	√		√			√
12	Livestock	Most knowledge-able person		√	√	√	√				
13	Fishing	Most knowledge-able person		√	√	√					
14	Non-farm self-employment	Most knowledge-able person, up to five businesses		√	√	√		√			
15	Housing	Head of Household*		√	√	√	√				√
16	Durable goods	Head of Household*		√	√	√	√				
17	Food Consumption	Most knowledge-able person		√							

Sec-tion	Title	Respondents	Demo-graphic Charac-teristics / Mortality	Expen-ditures	Income	Assets	Transfers	Labor	Schooling	Health	Use of Public Ser-vices
18	Individual Expenditures	All household members		√							
19	Remittances and Credit	All household members		√	√		√				
20	Mortality	Head of Household*	√	√	√		√		√	√	√

* In the head's absence, a principal respondent who is most knowledgeable on the topic was chosen

b) *Section 2: Children residing elsewhere*

Section 2 collects information on all nonresident children (both youngsters and adults) of household members, their relation to members of the household, their age, sex, educational attainment, current area of residence, and type of work. Transfers from these children are captured in Section 19 of the questionnaire on remittances. Each child living elsewhere is assigned a unique identification code that is retained for the entire survey.

c) *Section 3: Information on parents*

Every household member is linked to his or her parents in the household by their identification code. For those whose parent(s) are deceased or living elsewhere, the following information is collected: area of residence of parents who are alive; educational achievement; and primary lifetime work. For orphans and children living away from both parents, the section also obtains information on the length of time that the child has been in the current household. Information on transfers from non-resident parents to household members, as well as transfers from household members to their parents, is collected in Section 19 of the questionnaire, on remittances.

d) *Section 4: Summary of household activities*

This section guides the interviewer in the selection of respondents for the sections of the questionnaire dealing with economic activities and expenditures (Sections 7, 11, 12, 13, 14, 18). It asks about the household's main economic activities and who in the household is most knowledgeable about the household's farming, livestock, fishing, family businesses, and food expenditure.

e) *Section 5: Education*

The section on education has three objectives: (1) to measure the current levels of schooling and training of all household members; (2) to measure the current enrollment of all children in the household; and (3) to measure household expenditure on education in the past 12 months. This last objective includes measuring contributions to the schooling of household members by benefactors outside the household—both individuals and organizations.

The education section collects the following information for each household member 7 and older: literacy; educational attainment; current enrollment and attendance in the seven days before the interview; distance to school; school expenditures in the past 12 months; and scholarships received, in cash and in kind, by type of sponsoring institution. In instances in which a child missed school in the past 7 days, the questionnaire establishes the reasons for absence, including illness of the child and/or illness of other household members.

f) *Section 6: Health*

This section identifies which individuals are suffering from illnesses and details on their health care seeking behavior. The three parts of this section ask about acute illness (part A), chronic

illness (part B), and physical ability and general health (part C, added in the wave 2 questionnaire).

The questions on acute illness record: (1) who in the household was ill in the four weeks before the interview; (2) the symptoms and diagnosis; (3) the health seeking behavior of households as a result of illness; and (4) medical expenditures and the source of finance. For each household member who was ill or injured in the four weeks prior to the interview, the following information is collected: the duration of the illness; the major symptoms; the number of days the patient was unable to perform his/her usual activities; the first three health practitioners consulted; all of the expenditures associated with them, in cash, in kind, and in time; and expenditures on all other consultations for this episode of illness. The acute illness questions also include information about hospitalizations, the costs of medicine and travel to health consultations, and debt incurred due to the illness. The final questions ascertain whether the illness was diagnosed, the actual diagnosis, and whether the patient has recovered.

Chronic conditions, in Part B, are conditions that have existed for 6 months or more. The questions include the symptoms, the duration of the condition, the diagnosis (if a practitioner had been consulted), and expenditures on those conditions in the four weeks before the interview.

Through the skip pattern of this section, all household members, including those who had and didn't have an acute or chronic illness, must respond to questions concerning four symptoms often (but not always) associated with AIDS: recurring diarrhea for a month or more; weight loss; recurring fever: and skin rash.

Part C, which was added in the wave 2 household questionnaire, includes questions on general health, specific disabilities, and activities of daily living, like walking, running, lifting, and performing hard labor.

The consequences of ill health are taken up in many other sections of the questionnaire: economic activities (7); migration (8); fertility (9); anthropometry (10); expenditures (18); remittances (19); and mortality (20).

g) *Section 7: Activities of household members*

The objective of Section 7 is to establish the full scope of economic activities performed by members of the household. In particular, the section aims to: (1) establish the allocation of each household member's time across economic and domestic activities in the days before the interview; and (2) estimate the income of each household member in the past 2 months. It is the section of the questionnaire that will permit the analysis of the impact of adult illness and mortality on the time allocated to individual economic activities, like farming, and on individual and household income.

Section 7 is the largest section of the questionnaire, with 8 parts. It collects information on the number of hours per day spent by each household member in the past 7 days on the following activities, and income from them:

- work as an employee in as many as two jobs (Part B);
- work on an own or family farm, work on a communal farm, processing crops, tending livestock and processing livestock products (Part C);
- work on as many as three own or family businesses (Part D);
- domestic activities (preparing meals, cleaning, doing laundry, shopping), collecting firewood, collecting water, seeking medical care, caring for sick household members, looking for additional work, helping neighbors and attending funerals (Part E).

An elaborate slip pattern sends each respondent to the part of Section 7 that is relevant for his/her activities. Parts F and G capture information on income if the main job last year or the secondary job was different than any of the activities in the seven days before the interview. Finally, part H collects non-earned income for each household member.[19]

This information is obtained from the recall of respondents and is known to he inexact. However, it is intended to represent a rough snapshot of the allocation of each person's time in the 7 days prior to the interview.[20]

Other sections of the questionnaire also collect data on the self-employment activities and income at the *household* level, notably the sections on farming (11), livestock (12), fishing (13) and non-farm self-employment (14). These latter sections collect greater detail on the income and expenditure.

h) Section 8: Migration

The objective of this short section on migration is to establish the length of time each household member has been residing in his/her current place of residence and the circumstances surrounding the last move. Other aspects of migration covered in the questionnaire are on the Household Roster (section 1) and in sections on children living elsewhere (section 2) and non-resident parents of household members (section 3).

i) Section 9: Fertility

The objective of the section on fertility is to ascertain: (1) the number of children ever born to every female household member age 15 and older (or younger, if already married); (2) child mortality; (3) the level of schooling attained by deceased children; (4) current pregnancy status; (5) fetal wastage; and (6) contraceptive use. Children are an asset to the household that will hopefully bring future earnings as the child matures. Section 9 provides information on the potential source of support for elderly survivors in the household who may have lost adult children through fatal illness. Related sections of the questionnaire are the Household Roster (Section 1), children living elsewhere (Section 2), education (Section 5), health (Section 6), anthropometrics (Section 10) and mortality (Section 20).

---

[19] Unlike the LSMS prototype questionnaire, which collects non-labor income at the household level, the KHDS questionnaire obtains this information from each household member.

[20] The KHDS questionnaire seeks to account for *all* activities of each household member in the past 7 days because it is hypothesized that fatal adult illness and mortality would result in shifting responsibilities and allocation of time among household members. In contrast, the LSMS prototype questionnaire was concerned only with the main economic activity and the secondary activity within the past 7 days.

j) Section 10 Anthropometry

In the anthropometry section, the interviewer measures and records the height and weight of all household members. Section 10 also collects information on the immunization status of children. Children under two years of age were weighed in hanging Salter scales, while those who were older were weighted on adult scales. During the first two passages, all adults were measured with spring scales. In the third and fourth passages, the spring scales were replaced with digital scales. This greatly reduced the dispersion in the weight data for the young children weighed on the adult scales.

k) Section 11 : Farming

The objective of the section on farming is to collect information on: (1) annual net income due to cultivation of crops; (2) the number and value of farm assets (land and equipment); and (3) the quantity of crops sold. This section is asked in all households engaged in farming, and the respondent is the person in the household who is most familiar with farm income and expenditure.

The major parts of Section 11 collect information on the number and size of fields (shambas) owned and cultivated by the household, their sale value, and how they were acquired (Part A); the crops cultivated in the past 12 months, the quantity of production sold, income from sale of crops and expenditures on crop inputs (Part B); the number and age of tree crops (Part C); use of and expenditure on farm inputs (Part D); income from the sale of products from homegrown crops, and expenditures on transforming these crops for sale (Part E); and the ownership, value, purchase and sale of agricultural equipment (Parts F and G).

l) Section 12: Livestock

This section assesses: (1) the number and value of livestock owned by the household; and (2) household income in the past 12 months due to livestock activities. The first part of section 12 establishes the household's stock of animals at the time of the survey, the value of the stock and changes in the stock over the past year. Part B collects information on income from processing of livestock products in the past year, such as from milk and egg production. Part C measures the expenditures on livestock production in the past 12 months, for items such as herding, veterinary services, and animal feed.

m) Section 13: Fishing

This section seeks to measure the value of assets of fishermen and their annual income, net of expenses. A separate section on income and assets from fishing was deemed necessary because of the proximity of the project site to Lake Victoria. The three parts of this section include questions on fishing equipment (Part A), income from fishing and smoking or drying fish (Part B) and expenditures on fishing inputs (Part C). The fishermen in this area engage in three types of fishing—with hooks, trawling nets and stationary nets.

n) Section 14: Non-farm self-employment

The objective of Section 14 is to measure the net income and value of assets for small businesses owned or operated by household members. For each business owned by a member of the household, Section 14 collects the expenditures (Parts A and B), income (Part C) and assets (Part D). The reference period for reporting expenditures is the two weeks prior to the interview (that is, the interval between rounds one and two) if the business is functioning or, if the business is not functioning at the time of the interview a typical time unit of the respondent's choosing.

o) Section 15: Housing

This section has the objective of measuring: (1) the value of housing assets; (2) expenditures on housing, water, electricity and other utilities; and (3) the physical condition of the housing, which is a direct measure of well-being.[21] Expenditures on housing and utilities, together with the results of other expenditure sections (16, 17, and 18), are an input into the estimate of total annual household consumption expenditure.

p) Section 16: Durable goods, annual expenditures and income from assistance programs

The objective of this section is to collect information on: (1) the value of durable consumer goods owned by the household (Part A); (2) expenditures on infrequently-purchased items in the past 12 months (Part B); and (3) receipt of cash or in-kind assistance from community organizations (Part C).[22]

For each durable good possessed by the household, information is collected on ownership, year of acquisition, purchase price, and potential sale price. Part C is of central importance to the research, since it is here that receipt of assistance by the household from outside organizations is recorded. Receipt of outside assistance by *individuals is* recorded in Section 19.

q) Section 17: Food consumption

The objective of the section is to collect information on habitual and recent food consumption, on the basis of which an annual measure of food consumption can be constructed. While some households in the Kagera region purchase all of their food, in most cases they consume *both* purchased food and food that they produce at home. Thus, Section 17 had to collect information on the value of home production that was consumed as well as purchased food. The second and more difficult challenge was capturing the seasonality of food consumption. Because different foods are consumed during different seasons of the year, it would be incorrect to ask about food consumption for a recent period and to infer that this pattern was representative of the past 12 months. The seasonality of food production and consumption over the entire 12 months before

---

[21] This section departs in structure from the prototype LSMS questionnaire in Part A, where a series of questions are asked about the construction materials, ownership, and value of each building occupied by the household.

[22] This section differs from the LSMS prototype questionnaire in two ways: (1) it identifies who in the household owns each durable good; and (2) it collects information on transfers from outside organizations, including the Government. Part A, on ownership of durable goods, was substantially modified in the second passage to measure changes in asset ownership across waves (see the Appendix).

the survey had to be considered. This was rendered more difficult for KHDS by virtue of the fact that there are two rainy and two dry seasons in the Kagera region in a 12 month period, and the timing and duration of each season vary according to locale within the region.

The solution to the seasonality problem was to ask every household at the beginning of Section 17 to name the months of the past 12 months during which each wet and dry season took place (Part A). In the remaining parts of Section 17, for each food item that is home-produced (Part B) or purchased (Part C), the respondent must indicate during which months of the year the item was consumed.[23] For home-produced and purchased items that are seasonal (Part B and Part C-1). The respondent must report how often the item was consumed in the rainy season and in the dry season, and the value of the amount consumed on average each time. For non-seasonal purchased items (Part. C-2), the respondent is asked how often the item was consumed in the past 12 months and the value of the amount consumed on average each time. The information on the months of the year that an item was consumed, the seasons represented, the frequency of consumption and the average value will permit computation of an estimate of the value of annual food consumption. As an alternative to this method, the questionnaire also asks for all purchased food items (seasonal and non-seasonal, Parts C-l and C-2) the amount spent since round one (two weeks earlier). Thus, for purchased food it will be possible to estimate annual consumption expenditure by two methods: (1) frequency x average value; and (2) actual expenditures in the past 2 weeks x 26. The structure of Section 17 is summarized in the table below:

Table V.3: The structure of section 17, Food consumption

Section	Title	Questions on Months (Consume?)	Consumption Frequency Questions	Expenditure Since Last Round
17A	The seasons of the past 12 months	N/A	N/A	N/A
17B	Consumption of home production	Yes	Seasonal	No
17CI	Food expenditures, seasonal items	Yes	Seasonal	Yes
17C2	Food expenditures, non-seasonal items	Yes	Annual	Yes

r) *Section 18: Individual expenditures*

While sections 16 and 17 collect expenditures made on behalf of the entire household, section 18 collects information on personal expenditures by individual household members. Section 18A collects expenditures and acquisitions or behalf of individual household members in the past 12 months, while Section 18B collects information on personal expenditures by individual household member since round one (about two weeks previously). The information is gathered for each individual because: (a) it should improve the accuracy of household consumption expenditure estimates; and (b) it will allow the researchers to analyze the distribution of expenditures and acquisitions across different household members, and thus compare levels of individual welfare within households.[24] The results of this section will allow comparison of expenditures and acquisitions across different types of individuals -- for example, orphans vs.

---

[23] By asking for the specific months of the year that an item was consumed, it becomes possible to compute both an annual (12-month) estimate of food consumption expenditure *and a 6-month estimate.* The latter is very important to the KHDS, since the household questionnaire for the subsequent waves will have a 6-month reference period for food consumption expenditures.
[24] Collecting expenditure and consumption information for each individual household member is another way in which the KHDS questionnaire distinguishes itself from the LSMS prototype

non-orphans, boys vs. girls, men vs. women, sick vs. healthy adults and younger vs. elderly adults.

s) *Section 19: Remittances and credit*

Section 19 measures transfers and credit flowing into and out of the household, the reasons for these arrangements, and the level of household savings. Many of the transfers captured in this section will be from nonresident family members - children and parents - whose background is captured in sections 2 and 1 .

Section 19 has three parts. The first two parts collect information on receipt of remittances or borrowing (Part A) and sending of remittances or lending (Part B) in the past 12 months, for each household member.[25] The total number of each type of transaction is recorded, followed by detailed questions on a maximum of three of each type of transaction: the relationship of the lender or borrower to the household member, where the latter lives, the amount received, borrowed or sent, the reason behind the transaction (if any) and the repayment terms (if any). Part C deals with the savings of individual household members: participation in traditional savings organizations in the past 12 months; savings in formal institutions on the day of the interview; and savings kept elsewhere.

t) *Section 20: Mortality*

This section measures: (l) all deaths in the past 12 months and their causes; (2) the health seeking behavior of persons who died in the past 12 months; and (3) household expenditures connected with mortality in the past 12 months. Part A collects information on mortality of relatives living elsewhere. For each death in the household, it collects: the relation of the deceased to the head; age at death; sex; relationship to other household members (including children); marital status; educational background; major work: activity; cause of death; duration of illness before death; symptoms of the fatal illness; consultations before death and expenditures on health care and funerals. A subset of this information is collected on the deaths of non-resident relatives in Part B. As this is perhaps the most sensitive topic in the household questionnaire, it appears at the very end, after the interviewer has established rapport with the household.

---

[25] This is another example of the effort of the KHDS questionnaire to collect information on an individual level so as to better analyze the intra-household distribution of well-being.

Table V.4: Questions in which inheritances received by household members are recorded

Item Acquired	Inheritances from Inside the Household	Inheritances from Outside the Household
Assets		
Land	7H, Q7A,B, 11A, Q4,5,8	7H, Q7A,B, 11A, Q4,5,8, 19A, all Q
Farm Equipment	7H, Q7A,B	7H, Q7A,B, 19A, all Q
Livestock	7H, Q7A,B	7H, Q7A,B, 19A, all Q, 12A, Q11
Fishing Equipment	7H, Q7A,B	7H, Q7A,B, 19A, all Q
Business Assets	7H, Q7A,B	7H, Q7A,B, 19A, all Q
Dwellings	7H, Q7A,B	7H, Q7A,B, 19A, all Q
	15A, Q11,12,12A	15A, Q11,12,12A
		15B, Q11A,B
Durable Goods	7H, Q7A,B	7H, Q7A,B, 19A, all Q
	16A1, Q3,6	16A1, Q3,6, 16A2, Q4A
Goods		
Clothes	7H, Q7A,B, 18A, Q1,3,4	7H, Q7A,B, 18A, Q1,3,4, 19A, all Q
Food	7H, Q7A,B	7H, Q7A,B, 19A, all Q
Other	7H, Q7A,B	7H, Q7A,B, 19A, all Q

B. COMMUNITY QUESTIONNAIRE

The objective of this questionnaire is to elicit community-wide information on:

- demographic characteristics
- economy and infrastructure
- education
- health
- agriculture
- culture

that are common to all households residing in each community. The questionnaire is directed at community leaders who were also specifically asked about the location, distance and identity of the closest market, primary school, dispensary, health center, and hospital. In addition, they were asked to name traditional healers within the community. Information from these questions was used to identify the sample of markets, schools, health facilities, and traditional healers for specific questionnaires in the study. It is important to note that these samples are not randomly selected and therefore, are not representative of the markets, health facilities, schools, and traditional healers in Kagera District. They are representative of the facilities located near the selected sample of households. Changes in the questionnaire across Passages are found in Appendix 1. The questionnaire was shortened between Passage 1 and 2 because communities were being revisited.

C. HEALTH FACILITY QUESTIONNAIRE

The objective of this questionnaire is to establish changes in the demand for health services and the supply of health services offered at a health facility. The Health Facility Questionnaire was administered to the health facility closest to each cluster. Data collected in this section was organized in three parts. Part A was administered to the medical person in charge and had nine sections:

1. Characteristics of the Facility
2. Personnel
3. Equipment
4. Services
5. Immunizations
6. Family Planning
7. Inpatient Services
8. Demand
9. Fee Exemption Policies

Part B was administered to the pharmacist of the facility and asked about the availability of drugs at the facility. Part C, on inpatient and outpatient consultations, was not administered by an interviewer, but was completed by the medical officer in charge.

D. SCHOOL QUESTIONNAIRE

The objective of the school questionnaire is to assist in the analysis of demand for schooling of household members. It was completed for every primary school in a cluster. If there was no school in the cluster, a school questionnaire was completed for the nearest primary school to the cluster.

There are two parts to the school questionnaire. Part A focused on the characteristics, enrollments, and fees for each school and was administered by the interviewer. Part B was left with the headmaster or head teacher of each school so they could refer to school records and inventory to provide information on: (1) the number of textbooks (Kiswahili, math, other) available for the students of each grade; and (2) the number of classes, enrolled students, enrolled female students, students who attended last week, and two-parent orphans enrolled for each grade. A third part, dealing with assistance provided to the school, was added in the second passage (see Appendix 1).

E. PRICE QUESTIONNAIRE

The objective of this questionnaire is to measure prices of key consumption goods throughout the survey area and over time. The price questionnaire contained a list of thirty food items, six pharmaceutical products, and thirteen non-food items. Three prices were collected for each item from three different traders at different locations in the market.

Price data were collected from two types of market: the nearest community market and roadside shops or dukas for each cluster in each passage. In Passage 1 interviewers generally only visited one type of market. For Passages 2 - 4 the Price Questionnaire was fielded twice in each cluster-- once to the nearest community market and once to the roadside shop. It is important to note that each cluster does not have a market that is uniquely associated with that cluster. Different clusters may share markets and the market closest to the cluster may change from passage to passage.

Food items were weighed to the nearest 50 grams for pricing. As an example, 500 grams of sugar at a price of 100 Tsh were recorded as follows:

		OBSERVATION	
		GRAMS	PRICE
06	Sugar	500	100

The only food item in the price questionnaire not measured in grams is the chicken egg - for this, the price of one egg was recorded.

Pharmaceutical products were not weighed. The first four pharmaceutical products in the price questionnaire were measured in tablets (see below) and the last two - liver salts and milk of magnesia - were priced by the packet and the bottle, respectively.

		1ST OBSERVATION	
		TABLETS	PRICE
33	Nivaquine	2	25

Most *non-food items* did not have to be weighed - the pricing unit depended on the item. For example, batteries were always priced for each 1.5 volt battery. Charcoal was the only non-food item that had to be weighed (in grams).

Potential Problems. For items that could not be located, "NA" (not available) was written in all of the columns for that item. If only one person in a market sold an item, this price was recorded in the first column and "NA" was entered into the second and third price columns.

F. TRADITIONAL HEALER QUESTIONNAIRE

This questionnaire documents the prices, types of facilities, services, and referral practices of traditional healers in the survey area. The questionnaire was administered to two healers per cluster, who were randomly selected from those listed by the community leaders.[26] This questionnaire was administered only in Passage 3. This survey includes questions on the number and types of patients seen, the types of health problems encountered, and the healer's knowledge of the etiology of AIDS and of other diseases. There are seven sections of this questionnaire:

---

[26] In one cluster the field workers interviewed a third healer. Because of this, there were 103 completed healer questionnaires at the end of Wave 3 instead of the 102 that were expected.

1. Personal background of respondents
2. Consultations in past seven days
3. Facilities and equipment
4. Knowledge and practices
5. Prescription and referrals
6. Income and prices
7. Childbirth services

Traditional birth attendants were not interviewed. In areas where more than one cluster was selected (for example, Hamgembe), two healers were interviewed for each cluster.

SECTION A: Personal Background

Question 4: Gastrointestinal problems included problems of the mouth, bloating, abdominal pain and swelling, vomiting, jaundice, liver problems, and rectal and anal problems.

STDs (Sexually transmitted diseases) included painful urination, puss from urethra, warts, inguinal swelling, genital ulcers, and genital infections.

Cardiovascular conditions included heart palpitations, high blood pressure, stroke, dyspnea on exercise, and varicose veins.

The total number of patients seen in question 4 had to be at least equal to the number of patients in questions 1-3. Question 4 could have more patients listed if some had more than one condition.

G. FOLLOW-UP QUESTIONNAIRES

If the households most severely impacted by adult deaths disintegrate or move out of the sample, then an analysis of the households that remained might underestimate the severity of the impact. In order to deal with this possibility, during the fourth passage the survey teams obtained information from all individuals who were members of households that dropped out since the first passage, and who were still residing in Kagera region.

- For all households that dropped out at any time after the first passage, the teams attempted to locate all former household members individually, under the assumption that they may now belong to different households. Those individuals who were found were administered a "Follow-up Questionnaire", or FUQ, described below.
- For households that dropped out during the fourth passage, whose members could be located in Kagera region, and who moved *intact*, the household-level sections of the wave 4 household questionnaire were also completed. [27] An *intact household* is one in which no household members have left, unless they died. They include cases in which new members have joined.

---

[27] Households that dropped out in the fourth passage were not replaced.

The survey questionnaires for individuals and for intact households are described below. Detailed instructions on interview procedures and interpretation of questions in the FUQ are provided in the Wave 4 Interviewer Manual.

1. Follow-up questionnaire for individuals

The objective of the FUQ is to obtain information on the well-being of individuals who were in households that moved out of the sample. Using the last questionnaire completed by the household before it dropped out, augmented with data from section 1B concerning where and why the household moved, the field teams attempted to find each member of the households that moved between passages 1 and 4, if they were still in Kagera region. In some instances, household members were no longer living together and had joined other households.

Thus, the FUQ, while organized around the household roster from previous interviews, collected information on individuals who formerly lived together but may not be currently living together. It basically consists of the individual sections of a wave 4 household questionnaire (sections 3, 5, 6, 7, 9, 18, 19) plus some additional questions for each individual concerning the characteristics of the new household in which he/she now resides (sections 00C, 00D, 1C, 1D, 1E). Table V.5 summarizes the sections of the Follow-up Questionnaire and the respondent for each section.

Table V.5: Main sections of the follow-up questionnaire for individuals

Section	Title	Respondent
00C	Survey information sheet*	Survey personnel, about each former household member
00D	Identification of the new household*	Survey personnel, observations for each former household member
1C	Demographic characteristics of the respondent*	All former household members
1D	Economic characteristics of the household*	All former household members
1E	Individual assets*	Former household members 7 and older
3	Information on parents	Former household members
5	Education	Former household members 7 and older
6	Health	Former household members
7	Activities and non-labor income	Former household members 7 and older
9	Fertility	Female former household members 15 and older (or younger if married)
10	Anthropometrics	All former household members
18	Expenditures by household members	All former household members (Part A), and those 15 and older (Part B)
19	Remittances and credit	All former household members (Part A), and those 15 and older (Parts B and C)

* New section that is unique to the Follow-up Questionnaire.

The sections that are unique to the FUQ are discussed briefly below.

a) Section 00C: Survey information sheet

This section identifies which members of the former household moved, when they moved, and where, based on information copied from the household questionnaire for the passage when the household dropped out. In addition, it records the results of the field team's search for the individual within and outside the survey cluster.

b) Section 00D: Identification of the new household

In this section the interviewer records characteristics of the household where the former household member is now residing: the place name, region, district of Kagera (if still within Kagera), and cluster number (if in a survey cluster); the name of the current head of household and whether this is the same as the head at the time the household discontinued; and the current head's sex, age, schooling, religion, tribe, and occupation.

c) Section 1C: Demographic characteristics of the respondent

This section is an abbreviated version of the questions in Section 1 of the KHDS household questionnaire. It asks for the respondent's sex, relationship to the head of the current household, age, marital status, co-residence in the current household of the spouse and his/her ID code, the number of months the respondent has been living with the current household in this dwelling, and the reason why the household member moved. In addition, the interviewer must list the ID code of all people formerly in the respondent's household who are also in the household where the respondent is currently residing.

d) Section 1D: Economic characteristics of the current household

This section collects summary information on the characteristics of the respondent's current household: the number of household members by age (0-14 and 15+) and gender; and the number of members who work on a family farm (question 5), raise animals that they own (question 6), own a business or are self-employed (question 7), or who work as employees for an employer outside the household (question 8).

e) Section 1E: Individual assets

This section collects information on assets and durable goods currently owned by individuals and those sold by individuals since they were last interviewed. The assets and durable goods include: dwellings (questions 1-4A); durable goods (questions 5-8); shambas/fields (questions 9-12A); and cattle (questions 13-16A).

2. Following intact households that moved

A total of 14 households moved intact between the third and fourth passages of fieldwork--that is, all of the household members except those who died were still living together as a unit, even though some new members may have joined the household. All of these households had joined the survey in the first passage and dropped out between the third and fourth passages. The last

complete household questionnaire for all of them was a wave 3 questionnaire. In following these intact households, several different questionnaires were completed:

- As was the case for all households that dropped out at any time during the survey, sections 1A, 1B, and 20a of the household questionnaire were completed and entered with the rest of the household data for wave 4. The information in sections 1A and 1B will indicate that the household members moved.
- The Follow-up Questionnaire for individuals (described above) was completed for all household members who could be located from the 14 intact households. However, the FUQ was not completed for anyone who had joined the household since the third passage.
- Household-level sections of the Wave 4 questionnaire were also completed for these 14 households: sections 1, 2, 4, 11, 12, 13, 14, 15, 16, 17, and 20b.

Thus, except for information on those who joined the household, it is theoretically possible to assemble a complete file of wave 4 data on the 14 households that moved intact between passages 3 and 4, thereby increasing the sample size for the fourth passage.

There was no logical place to put moving expenses or housing construction expenditure in the wave 4 household questionnaire, so interviewers were instructed to record moving expenses for the intact households in Section 16B, item 17, as expenditures on "home services", and construction or purchase of new housing under item 37 of Section 16B.

VI. USING THE DATA

The following section provides information on how the data are organized, what unique identifiers can be used to link data across sections of a questionnaire, across passages, and across different questionnaire types.

A. STRUCTURE OF THE DATA

The KHDS data and documentation are organized into four main directories. These 4 main directories include: KHDS Data; Constructed KHDS Data; Non-KHDS Data; and Documentation. Figure VI.1 describes the directory structure under which the information is stored. The directory and subdirectory titles are constrained to 8 characters or less and are italicized in parentheses where different from the fuller labeling.

Household data are stored by wave of the questionnaire. Thus, in the wave 1 directory of the household data are located the datasets for all households that completed a wave 1 household questionnaire, irrespective of which passage this occurred. All other datasets are stored by passage.

1. File name conventions

Each dataset name is 8 characters long. The first set of characters of the data file name identifies the section of the questionnaire to which the data pertain. The last set of characters describes the level of observation within the data file. The dataset for Section 5 of the household questionnaire, on schooling, for example, is denoted S5___ind. The levels of observation are not limited to the household and/or the individual. Items such as durable goods, livestock, assets, businesses are all levels of observation within the household questionnaire. The conventions for the part of the variable name that denotes the level of observation in the household datasets are in Table VI.1.

Table VI.1: Filename conventions

Level of Observation	Character Identifier
Individual	ind
Household	hh
Children	kid
Durable goods	dur
Expenditure items	exp
Businesses	bus
Other	oth

Figure VI.1: Structure of Directories of Data and Documentation

Main directory	Subdirectories		
	--household ---------------------	--wave1 --wave2 --wave3 --wave4 --zipdata --Follow-up Questionnaire (*fuq*)--	--Intact Households (*intacthh*) --Individuals (*individs*)
	--community --------------------	--pass1 --pass2 --pass3 --pass4	
	--price---------------------------	--pass1 --pass2 --pass3 --pass4	
I. KHDS Data (*KHDSDATA)*---	--school---------------------------	--pass1 --pass2 --pass3 --pass4	
	--healthfacility ------------------	--pass1 --pass2 --pass3 --pass4	
	--healer--------------------------	--pass3	
	--enumeration		
II. Non-KHDS data (*otherKageradata)*			
III. Constructed KHDS data (*KHDSaggr)*			
IV. Documentation---------------- (*KHDS_DOC)*	--Users Guide --questionnaires ---------------- --manuals	--household -- community_facility_price	

Within every data file, variable labels are a maximum of 30 characters in length and usually describe the section of the questionnaire and a question number.

Dataset names do not indicate the wave or passage for which the data apply. A given dataset will have identical names in all four directories by wave. For example, the dataset for section 5 in all four household data directories is labeled S5___ind. However, within every household dataset are variables that identify the wave of the questionnaire and the passage in which it was collected. For this reason, researchers should store datasets into separate directories for each wave or passage.

B. DATASETS

1. Household Datasets

Table VI.2 lists the 49 datasets in the directory for each wave of the household questionnaire. The data are organized by section of the questionnaire. When all of the data in a section of the questionnaire correspond to a common level of observation (e.g., sections 2, 3, 5, 6), the data for that section are in a single data file at the common level of observation. There is more than one data file for some sections of the household questionnaire if the section was too long (e.g., section 7) or if it corresponded to more than one level of observation (e.g., section 11). The answers to solitary household-level screening questions and short household-level parts are grouped into a single file, S_____HH.

Households retained the same identifier across all four passages, and so did individuals. Thus, someone who was in cluster 45, household 10, with id code 03 on the wave 1 household questionnaire would have the same cluster, household, and id code number in every wave/passage. Children on the roster of children living elsewhere (section 2) also kept the same ID codes throughout all waves of the survey questionnaire. Thus, it can be assumed that a person on the roster with the same ID code in adjacent or even non-adjacent passages is in fact the same person.

Some sections of the questionnaire have levels of observation that are sub-household--for instance, the section on businesses (section 14). Data for businesses across different datasets for the same household can be linked by use of the (cluster hh entprsid) codes. There are other examples of other levels of observation that can be studied within households. However, in the case of businesses, the identifier (entprsid) does not necessarily correspond to the same enterprise over time (across passages). The order in which enterprises were listed depended on the order in which respondents reported them. Between passages, old businesses may fold and new ones may be created. Thus, researchers will have to try to match enterprises on other characteristics if they hope to study the same unit over time: "entprsid" should not be used. Likewise, plots of land identified by (shambaid) in Section 11 and dwellings identified by (bldngid) are not necessarily the same units observed across passages or listed in the same order. Except for the rosters of people in Sections 1 and 2, whenever the interviewer was asked to freely list the household's businesses, plots, buildings, etc. in an arbitrary order, it cannot be assumed that the units described were listed in the same order across waves of the questionnaire, or that the same items are on the list.

Data can be linked across sections of the household questionnaire for a given household in the same passage by (cluster hh) and for a given individual by (cluster hh id). Households can be linked across waves/passages by the two variables (cluster hh) and individuals by (cluster hh id). However, variable names for the datasets for each passage or wave are identical, so when linking across passages it is important to reshape the datasets in such as way that the variable names are identified to reflect the passage or wave they were collected.

Table VI.2(a): Household Questionnaire Datasets

Sec.	Section Topic	Filename	Level of Observation	Identification Variable(s)	Number of Observations by Wave			
					1	2	3	4
1	Household Roster	s1___ind	Individual	cluster hh id	5373	5782	5895	5919
2	Children Residing Elsewhere	s2___kid	Individual (children living else-where only)	cluster hh cid	3394	3716	3752	3837
3	Parents	s3___ind	Individual	cluster hh id	5298	5017	4735	4322
4	Overview of Household Businesses	s4___bus	Business	cluster hh seorder	334	469	543	575
5	Education	s5___ind	Individual	cluster hh id	4194	3978	3734	3426
6	Health	s6___ind	Individual	cluster hh id	5297	5022	4734	4320
7	Activities of Household Members	s7a__ind	Individual	cluster hh id	4193	3978	3733	3424
		s7b__ind	Individual	cluster hh id	372	343	345	303
		s7c__ind	Individual	cluster hh id	2739	2760	2778	2461
		s7d__ind	Individual	cluster hh id	238	305	341	348
		s7e__ind	Individual	cluster hh id	4193	3978	3734	3423
		s7f__ind	Individual	cluster hh id	4192	3979	3733	3421
		s7g__ind	Individual	cluster hh id	516	604	775	765
		s7h__ind	Individual	cluster hh id	4195	3978	3733	3421
8	Migration	s8___ind	Individual	cluster hh id	5297	398	431	362
9	Fertility	s9___ind	Individual	cluster hh id	1630	1204	1142	1031
		s9___kid	Individual (children ever born)	cluster hh chorder	5826	243	239	201
10	Anthropometry	s10__ind	Individual	cluster hh id	5299	5016	4732	4313

Table VI.2(b): Household Questionnaire Datasets

Sec.	Section Topic	Filename	Level of Observation	Identification Variable(s)	Number of Observations by Wave			
					1	2	3	4
11	Farming	s11a_oth	Plot of land	cluster hh shambaid	2987	2862	2833	2608
		s11b_oth	Crop	cluster hh cropid	1222	1418	1582	1369
		s11c_oth	Tree crop	cluster hh treeid	1	6978	4	3
		s11e_oth	Processed crop	cluster hh crprodid	5234	426	7321	6565
		s11g_dur	Equipment	cluster hh fequipid	535	790	398	341
					146		977	756
12	Livestock	s12a_oth	Livestock	cluster hh lvstckid	1110	1240	1154	1039
		s12b_oth	Animal product	cluster hh anmprdid	172	156	168	115
		s12c_exp	Expenditure item	cluster hh lvsinpid	596	471	473	359
13	Fishing	s13a_dur	Equipment	cluster hh fisheqid	60	57	55	34
		s13b_oth	Fishing method	cluster hh fshtypid	20	19	13	15
		s13c_exp	Expenditure item	cluster hh fshexpid	76	63	48	34
14	Non-farm self-employment	s14a_bus	Business	cluster hh entprsid	327	465	544	567
		s14b_exp	Expenditure	cluster hh entprsid businp	680	874	887	922
		s14c_bus	Business	cluster hh entprsid	326	463	542	567
		s14d_dur	Asset	cluster hh entprsid bassetid	312	358	397	364
15	Housing	s15a_dur	Dwellings	cluster hh bldngid	1086	1031	973	897
16	Durable goods	s16a_dur	Durable goods	cluster hh dlinenum	5546	842	807	1450
		s16b_exp	Expenditure item	cluster hh hhexpid	3393	2264	2344	2022
		s16c_oth	Assistance organization	cluster hh assistid	305	794	828	742
17	Food Consumption	s17b_exp	Food item	cluster hh typefood	1190	1310	1463	1233
		s17c1exp	Food item	cluster hh typsfood	1	6	8	4
		s17c2exp	Food item	cluster hh typnsfd	8059	6781	6887	5468
					3364	3016	3111	2599
18	Individual Expenditures	s18a_ind	Individual	cluster hh id	5275	5006	4727	4302
		s18b_ind	Individual	cluster hh id	2804	2669	2504	2293
19	Remittances and Credit	s19a_ind	Individual	cluster hh id	5279	4843	4727	4304
		s19b_ind	Individual	cluster hh id	2817	2673	2504	2296
		s19c_ind	Individual	cluster hh id	2795	2670	2504	2294
20	Mortality	s20a_oth	Deceased	cluster hh hhdeadid	556	85	68	75
		s20b_oth	Deceased	cluster hh rdiedid	1436	986	990	887
0	Section verification	s00b_oth	Section	cluster hh sectnum	1825	1769	1673	1558
					8	9	3	8
Misc	Misc.	s_____hh	Household	cluster hh	981	922	846	785

2. Community datasets

Table VI.3 describes the datasets from the community questionnaire. The community questionnaire was asked once per passage in every Primary Sampling Unit (PSU). As noted in the chapter on sampling, there were 49 PSUs and 51 clusters of households—one cluster of

households was selected in each of 47 PSUs and two clusters each were selected from two PSUs. This explains why there are only 49 observations for every passage of community data, even though there were 51 clusters of households. The unique identifier for the community data is the variable *cluster.* In order to map the community data to all clusters and households, the data for cluster 44 should also be attributed to cluster 45, and the data for cluster 46 should also be attributed to cluster 47.

The community data can be linked to households and individuals by a combination of the variables (cluster pass). The community data included 15 datasets in the first passage, but only 13 in the second through fourth passages. This is because the location of secondary schools and the types of crops grown in the community were not expected to change over time and thus asked only in the first passage.

Table VI.3: Community Questionnaire Datasets

Filename	Level of Observation	Identification Variable(s)	Number of Observations by Passage			
			1	2	3	4
c0clustr	Community	cluster	48	49	49	49
c1clustr	Community	cluster	49	49	49	49
c2borrow	Borrower/lender	cluster lender	29	26	45	43
c2clustr	Community	cluster	49	49	49	49
c2social	Services	cluster afaccode	635	683	686	685
c3clustr	Community	cluster	49	49	49	49
c3prisch	Primary school	cluster schno	62	62	62	63
c3secsch (passage 1 only)	Secondary school	cluster sschlid	57	-	-	-
c4clustr	Community	cluster	49	49	49	49
c4hfacil	Health facility	cluster hfaccode	169	147	147	147
c4hprvdr	Health provider	cluster pvdrcode	192	195	196	195
c5clustr	Community	cluster	49	49	49	49
c5cropid (passage 1 only)	Crop	cluster cropcode	291	-	-	-
c6clustr	Community	cluster	49	49	49	49
ccclustr	Community	cluster	49	49	49	49

3. Health facility datasets

Table VI.4 describes the datasets from the health facility questionnaire, which was asked once in each health facility in each passage. The number of facilities interviewed increased over time. During the first passage the interviewers visited only the nearest health facility to each cluster; after the first passage, they visited all nearby health facilities used by households in the cluster. Each health facility has a unique identifier (facnum) that was retained throughout all passages of the survey.

There are 18 health facility datasets for the first passage and 20 datasets for the other three passages. The questions that were asked in section 9 of the questionnaire in passage one were moved into a new section 10 in subsequent passages, and new questions about assistance to the facility were added.

Some health facilities served multiple clusters, and in some clusters more than one health facility was interviewed. Thus, in the first passage fewer health facilities were interviewed than there were clusters. There are two ways of linking health facilities to clusters and households:

(1) In the community questionnaire data file *ccluster* is the variable (facnum) which is the ID code for the nearest health facility to the cluster. This method allows linking of all communities with a single health facility that is the nearest one to the cluster.
(2) In the health facility data file *htp__fac*, there are variables for the four closest clusters serviced by the health facility as cited by respondents to the health facility questionnaire (clu1, clu2 clu3, clu4). Using this method, one can theoretically link more than one facility to a cluster.

Once the health facilities are linked to the correct cluster, they can be linked to households and individuals using the variables (cluster pass). Again, as for the community questionnaire, facilities linked to cluster 44 should also be attributed to cluster 45 and those linked to cluster 46 should also be linked to cluster 47.

Table VI.4: Health Facility Questionnaire Datasets

Filename	Level of Observation	Identification Variable(s)	Number of Observations by Passage			
			1	2	3	4
HTP__FAC	Facility	Facnum	42	48	57	61
ha1__fac	Facility	facnum	42	48	57	61
ha2__per	Personnel	facnum personnl	923	1056	1254	1341
ha3__fac	Facility	facnum	42	48	57	61
ha4__srv	Service	facnum services	291	336	399	427
ha5a_fac	Facility	facnum	41	48	57	61
ha5b_imm	Immunization	facnum immunzn	258	280	301	308
ha6a_fac	Facility	facnum	42	48	57	61
ha6b_fpm	Family planning method	facnum famplncd	218	266	272	286
ha7__fac	Facility	facnum	42	48	57	61
ha8__fac	Facility	facnum	42	48	57	61
ha9__fac*	Facility	facnum	40	-	-	-
ha9a_fac**	Facility	facnum	-	48	57	61
ha9b_ast**	Assistance	facnum asstcode	-	31	45	32
ha10_fac**	Facility	facnum	-	46	57	61
hba__drg	Drug	facnum facdrgcd	1671	1971	2362	2517
hbb__fac	Facility	facnum	40	48	57	61
hc1a_fac	Facility	facnum	41	48	57	60
hc1b_cns	Consultation	facnum illnescd	969	1131	1364	1506
hc2a_fac	Facility	facnum	20	20	28	26
hc2b_ill	Illness	facnum illcode	160	128	411	415

* Passage 1 only. HA9_FAC was renamed as HA10_FAC in passages 2-4.

** Passages 2-4 only

4. Primary school datasets

Table VI.5 lists the datasets from the school questionnaire. Every cluster had at least one primary school and 15 clusters had more than one. All primary schools in the clusters were administered a school questionnaire every passage. Within any given passage, the unique identifier for a school is a combination of (cluster schno), and the schools retained the same identifiers across passages. There was only one dataset for each school during the first passage; a second set of questions on assistance received by the school was added in the second through fourth passages.

The primary school data can be linked to clusters by a combination of (cluster pass). However, more schools were interviewed than there were clusters. Thus, the user can select one of the primary schools in the community to link to each cluster or can try to combine characteristics of all of the schools serving a cluster, and link those combined characteristics to the cluster and the households within. Schools linked to cluster 44 should also be attributed to cluster 45 and those linked to cluster 46 should also be linked to cluster 47.

Table VI.5: School Questionnaire Datasets

Filename	Level of Observation	Identification Variable	Number of Observations by Passage			
			1	2	3	4
school	School	cluster schno	62	63	63	64
asstance (passages 2-4 only)	Donor organization	cluster schno assitmcd	-	71	65	62

5. Price questionnaire datasets

Table VI.6 describes the datasets from the price questionnaire. As noted earlier, during the first passage a price questionnaire was administered to one market per cluster. However, the interviewers were subsequently instructed whenever possible to complete two price questionnaires—one in open markets and one in shops, or *dukas*. Thus, the number of observations on the price questionnaire datasets rises over time. There are six datasets for the price questionnaire in each passage.

The unique identifier for a market within a passage is a combination of (cluster measrplc). However, markets did not retain the same identifiers over time. Household and individual data can be linked to the data from individual price questionnaires by a combination of (cluster pass). Markets linked to cluster 44 should also be attributed to cluster 45 and those linked to cluster 46 should also be linked to cluster 47.

Table VI.6: Price Questionnaire Datasets

Filename	Level of Observation	Identification Variable	Number of Observations by Passage			
			1	2	3	4
prcvrcls	Measuring Place	cluster measrplc	53	82	84	90
pr1__fdp	Food item	cluster measrplc foodtype	864	1052	1007	1047
pr2__ph1	Pharmaceutical item	cluster measrplc phrmctcd	133	200	198	221
pr2__ph2	Pharmaceutical item	cluster measrplc phrmcd1	16	38	34	40
pr3__nf1	Non-food item	cluster measrplc nfoodcd	341	470	471	513
pr3__nf2	Charcoal	cluster measrplc charcoal	6	8	4	8

6. Traditional healer datasets

Table VI.7 describes the three datasets for the traditional healer questionnaire. Two healers were interviewed in each cluster during the third passage, drawn at random from a list provided by

community leaders.[28] The unique identifier for a healer is a combination of (cluster healerid). The characteristics of healers can be linked to households by the (cluster) variable---since it was collected only once, the passage is irrelevant. However, since more than one healer was interviewed per cluster, there are several ways of linking the healer data: (1) selecting only one healer to link; (2) re-labeling the variables to correspond to "healer #1" and "healer #2" and then linking to clusters; or (3) combining the characteristics of the two healers in some form and linking these to the cluster.

Table VI.7: Traditional Healer Questionnaire Datasets

Filename	Level of Observation	Identification Variable	Number of Observations
s____hlr	Healer	cluster healerid	103
sb___cnd	Conditions treated	cluster healerid cndition	3090
se___plc	Treatment place for referrals	cluster healerid plcecode	407

7. Follow-up questionnaire datasets

Individuals who completed follow-up questionnaires retained the same cluster, household, and id variables they had throughout the survey, and thus can be linked to the data from interviews of previous passages before they moved. The same is true of household-level sections of the household questionnaire administered to the 14 households that moved intact between the third and fourth passages. Individuals who completed a follow-up questionnaire can be linked to the 14 intact households in which they were members by the combination of (cluster hh), within the follow-up data.

---

[28] Four healers each were randomly selected from the Hamgembe A and Hamgembe B PSUs, to be attributed to clusters 44 and 45 in the first instance, and to clusters 46 and 47 in the second.

Table VI.8(a): Datasets for individuals who were followed

Section	Section Topic	Filename	Level of Observatiob	Identification Variables	Number of Observations
0	Survey info sheet	s00c_ind	Individual	cluster hh id	192
	Identific. of new household	s00d_ind	Individual	cluster hh id	169
1	Demographic char.	s1c__ind	Individual	cluster hh id	169
		s1d__ind	Individual	cluster hh id	168
	Economic char. of (curr) household	s1e__ind	Individual	cluster hh id	140
	Individual assets				
3	Parents	s3___ind	Individual	cluster hh id	169
5	Education	s5___ind	Individual	cluster hh id	140
6	Health	s6___ind	Individual	cluster hh id	169
7	Activities of Household Members	s7a__ind	Individual	cluster hh id	140
		s7b__ind	Individual	cluster hh id	18
		s7c__ind	Individual	cluster hh id	84
		s7d__ind	Individual	cluster hh id	8
		s7e__ind	Individual	cluster hh id	140
		s7f__ind	Individual	cluster hh id	140
		s7g__ind	Individual	cluster hh id	35
		s7h__ind	Individual	cluster hh id	140
9	Fertility	s9___ind	Individual	cluster hh id	35
		s9___kid	Individual	cluster hh id	4
10	Anthropometry	s10__ind	Individual	cluster hh id	163
18	Individual Expenditures	s18a_ind	Individual	cluster hh id	169
		s18b_ind	Individual	cluster hh id	93
19	Remittances and Credit	s19a_ind	Individual	cluster hh id	169
		s19b_ind	Individual	cluster hh id	93
		s19c_ind	Individual	cluster hh id	93

Table VI.8(b): Datasets for Intact Households that were followed

Section	Section Topic	Filename	Level of Observatiob	Identification Variables	Number of Observations
Misc.	Misc.	s_____hh	Household	cluster hh	14
0B	Summary of survey results	s00b_oth	Section of questionnaire	cluster hh sectnum	237
1*	Household Roster	s1___ind	Individual	cluster hh id	75
2	Children Residing Elsewhere	s2___kid	Individual (Children living else-where)	cluster hh cid	30
4*	Overview of Household Businesses	s4___bus	Business	cluster hh seorder	4
11*	Farming	s11a_oth	Plot of land	cluster hh shambaid	29
		s11b_oth	Crop	cluster hh cropid	149
		s11c_oth	Tree crop	cluster hh treeid	69
		s11e_oth	Processed crop	cluster hh crprodid	2
		s11g_dur	Equipment	cluster hh fequipid	2
12	Livestock	s12a_oth	Livestock	cluster hh lvstckid	9
		s12c_exp	Expenditure item	cluster hh lvsinpid	1
14	Non-farm self-employment	s14a_oth	Business	cluster hh entprsid	4
		s14b_exp	Expenditure	cluster hh entprsid	9
		s14c_bus	Business	businp cluster hh entprsid	4
15*	Housing	s15a_dur	Dwellings	cluster hh bldngid	14
16*	Durable Goods	s16a_dur	Durable goods	cluster hh dlinenum	16
		s16b_exp	Expenditure item	cluster hh hhexpid	24
		s16c_oth	Assistance organization	cluster hh assistid	9
17*	Food Consumption	s17b_exp	Food item	cluster hh typefood	143
		s17c1exp	Food item	cluster hh typsfood	113
		s17c2exp	Food item	cluster hh typnsfd	45
20*	Mortality	s20b_oth	Deceased	cluster hh rdiedid	2

* The parts of these sections asked at the household level are recorded in S_____HH. In addition, data pertaining to sections 11F, 15B and 17A do not have separate data files, but rather appear in S_____HH

8. Enumeration datasets

Table VI.9 describes the two datasets from the enumeration of households—one at the level of the primary sampling unit and one at the household level. The unique identifier for an observation in the first of these datasets is ward and vgno. The data for the 49 PSUs that were selected for the household survey can be linked to the clusters through the variable *cluster*. Note that the data for cluster 44 should be attributed to *both* clusters 44 and 45; data for cluster 46 should be attributed to both clusters 46 and 47. For these two pairs (cluster 44 & 45, and 46 & 47), two clusters each were drawn from those PSUs.

The enum_vlg dataset contains three types of variables: (1) those collected directly from authorities at the level of the PSU; (2) those attributed to the PSU based on the characteristics of the households enumerated in each (labeled with "ENUM"); (3) variables attributed to the PSU from the 1988 census (for example, pop88, the census population in 1988).

The second enumeration data set is enum_hh, which is at the household level and contains the data collected in the Enumeration Survey of all households in the 49 PSUs. The identifier for enum_hh is sequence. They can be aggregated and linked to the PSU-level data through a combination of (ward vgno). These households were also assigned a "sequence number", which is unique and runs from 1 to 29,602. Of the 1,009 households that were ever contacted (or contact attempted), we are able to identify the enumeration data for 983. These are observations with the variables (cluster hh).

Table VI.9: Enumeration Datasets

Dataset	Filename	LEVEL OF OBSERVATION	Identification Variables	Number Of Observations	Description
Enumeration	enum_vlg enum__hh	PSU Household	ward vgno sequence	52 29602	Contains information on illness, deaths, and membership of households fielded in Enumeration Survey

9. Datasets of constructed variables

Four datasets have been constructed to facilitate analysis of the KHDS data:

- Individual and household "key" files that describe the status of each individual and household ever interviewed, for each of the four passages.
- Sample weights for the 816 households in the original sample that were interviewed during the first passage.
- A regional and temporal price index constructed from the price questionnaire data.
- Income, expenditure, and the value of assets aggregated at the household level, based on answers to all relevant sections of the household questionnaire.

The individual and household level "key" files are described in the next section (C). Sample weights can be linked to the 816 households selected for the first passage by c combination of (cluster hh). They can be used to weight the results for the first passage so as to be representative of Kagera region. The price index can be linked to each cluster through a combination of (cluster pass). The methodology used for constructing the sample weights and price index is described in Appendix 2 and 3, respectively. The methodology for constructing the income, expenditure and asset variables at the household level are described in the next chapter.

Table VI.10: Constructed Datasets

Dataset	Filename	Level Of Observation	Identification Variables	Number Of Observations	Description
Key files	key__ind	Individual	cluster hh id	6372	Describe membership status of individuals and attrition status of households during each passage
	key___hh	Household	cluster hh	1009	
	key__cle	Individual	cluster hh cid	4378	
Sample weights	weights	Household	cluster hh	816	Household sampling weights
Income, Expenditure, Assets	inc___hh	Household	cluster hh wave	3376	Household income, assets and expenditures *for each wave of the household survey*
	exp___hh	Household	cluster hh wave	3376	
	capt__hh	Household	cluster hh wave	3376	
	acqa__hh	Household	cluster hh wave	3376	
Price Index	prindex	Cluster	cluster passage	204	

10. Rainfall Data

Monthly rainfall measurements in millimeters over the five districts (Bukoba urban and rural combined) for a 60 month period, January 1989 to December 1993 were assembled with the assistance and data provided by staff in the Department of Meteorology, University of Tanzania, Dar es Salaam and the Climate Prediction Center/Climate Operations Branch, MOAA Science Center, Washington, D.C. There are a handful of missing data points in the rainfall data file: 13 months for Biharamulo, 1 month for Ngara, and 2 months for Karagwe.

Table VI.11: Rainfall Data

Dataset	Filename	Level Of Observation	IDENTIFICATION VARIABLES	Number of Observations	Description
Rainfall	raindata	Month and year	month year	60	Monthly rainfall over 5 year period for 5 districts (with Bukoba urban and rural combined)

C. INDIVIDUAL AND HOUSEHOLD KEY FILES

The household and individual "key" data files provide an overview of the status of households and individuals over the four passages of the KHDS.

1. "Key" dataset for households (key___hh)

The household "key" dataset describes the status of all 1,009 households ever selected for the sample and/or interviewed. This includes:

- 816 households in the original sample actually interviewed in the first passage, including replacements

- 24 "extra" households added by field teams during the first passage

- 46 households that were added in the second passage and 29 households that were added in the third passage as replacements for households that dropped or as "extra" households.

- 94 households that were selected for the sample or as replacement households but that could not be interviewed

The most important variable in this dataset is called "key category", or *keycateg*, and it classifies each household according to whether it was ever interviewed, what passage it joined the sample, and how long it stayed (see Table V.12). By summing together the various categories, it can be shown that 840 households were interviewed in the first passage, and 915 households were ever interviewed during the survey, from one to four times.

Table VI.12: Household Key dataset "ketcateg" Categories

Value	Label	Number of Observations	Description
0	DROP	94	Selected for the sample, but never completed a survey
1	J1ALL4	759	Joined in P1 and completed 4 interviews
2	J1drpP12	37	Joined in P1 and dropped out between P1 and P2
3	J1drpP23	17	Joined in P1 and dropped out between P2 and P3
4	J1drpP34	27	Joined in P1 and dropped out between P3 and P4
5	J2ALL3	42	Joined in P2 and completed 3 interviews
6	J2drpP23	3	Joined in P2 and dropped out between P2 and P3
7	J2drpP34	1	Joined in P2 and dropped out between P3 and P4
8	J3All2	28	Joined in P3 and completed 2 interviews
9	J3drpP34	1	Joined in P3 and dropped out between P3 and P4

In addition to the cluster, household, and key category, there are 13 other variables in the household key file:

hhwt	Household weight (See Appendix 2 for details).
sequence	Sequence number in the enumeration survey which can link the key data with enum_hh. [26 households are missing sequence no.]
num_obs	The number of times the household was interviewed (ranges from 0-4)
lastpass	The last passage that the household was observed.
enumdrop	Dummy variable that flags the 47 households selected for the original sample of 816, but that could not be found or interviewed during the first passage.
replcmnt	Dummy variable that flags households that replaced those in the original sample that couldn't be found during the first passage or subsequently dropped out.
extrahh	Dummy variable that flags the "extra" households interviewed.
orig1	Dummy variable that flags the 816 households of the "official sample" for the first passage actually interviewed, after replacements
orig2	Dummy variable that flags the 816 households originally selected for the first passage, before any households were replaced
intvday1	The date of the wave 1 interview in days elapsed since January 1, 1960.
intvday2	The date of the wave 2 interview in days elapsed since January 1, 1960.
intvday3	The date of the wave 3 interview in days elapsed since January 1, 1960.
intvday4	The date of the wave 4 interview in days elapsed since January 1, 1960.
enumdate	The date of the enumeration in days elapsed since January 1, 1960.
hhreplac	The household id (within the cluster) of the household that is replaced by this household.
replbyhh	The household id (within the cluster) of the household that replaced this household.
extrafor	The household id (within the cluster) of the household for which this household serves as an extra.
w1hhstat	The status of the household in wave 1.
w2hhstat	The status of the household in wave 2.
w3hhstat	The status of the household in wave 3.
w4hhstat	The status of the household in wave 4.

2. "Key" dataset for individuals from household roster (key_ind)

Key_ind tracks the status of all 6,372 individuals ever reported on the household roster for all households ever interviewed.

The variables w1_stat, w2_stat, w3_stat, and w4_stat classify the status of each individual ever observed for interview waves 1, 2, 3, and 4, respectively. An individual's status can fall into one of five categories during the first wave (i.e., in w1_stat):

- "hhmbr" *Household member*: persons listed on the household roster (section 1) and classified as household members.
- "cle" *Child living elsewhere*: persons from the household roster (Section 1) who are ever listed on the child-living-elsewhere roster (section 2)
- "nonmemb" *Non-member*: persons who are on the household roster but do not meet the criteria for household membership
- "unborn" *Unborn*: individuals not yet born who appear on either the household or child living elsewhere rosters in the future waves
- "dieRd1-2" *Died between rounds*: individuals who were interviewed during the first round in a given wave, but who died before the second round of the same wave, two weeks later. In

the household questionnaire datasets, these people will appear *both* in the household roster (section 1) *and* the section measuring the deaths of household members since the last wave (section 20A) of the same questionnaire.

For the second through fourth waves, there are two additional categories:

- "died" *Died*: household members who died since the previous wave (reported in section 20A)
- "hhdrop" *HH dropped*: individuals that dropped out because their household left the survey before it ended

For the third and fourth waves, there is one additional category:

- "survend" *Survey end*: members of households that began the survey after the first wave and who had fewer than four interviews because the survey ended.

There are some important issues to note with respect to the status of household members who died. First, individuals reported as "died" can include persons not reported in the household roster in any previous wave. That is, the individual joined the household and died before they could be recorded on the household roster as a household member. This explains why a person can have a status of "died" in wave 3 (and, therefore, in all following waves) and be listed as a nonmember in waves 1 and 2.

Second, the status of individuals who died between rounds of a wave is reported since the previous passage, even though there may be observations for that individual in sections 1-10 of the household questionnaire of the current passage.

The following "transition" tables describe the changing status of individuals across passages.

Table VI.13(a): Transition Matrix for Individuals in the 915 Households Interviewed At Least Once

STATUS IN PASSAGE 2

STATUS IN PASSAGE 1	HH member	Children living elsewhere	Nonmember	HH dropped	Died	Unborn	Died between Round 1 and 2	Total
HH member	4601	170	295	147	64	0	3	5280
Children living elsewhere	102	136	3	0	0	0	0	241
Nonmember	222	8	399	0	0	0	1	630
Unborn	74	1	0	0	0	141	0	216
Died between Round 1 and 2	0	0	0	0	5	0	0	5
Total	4999	315	397	147	69	141	4	6372

Table VI.13(b): Transition Matrix for Individuals in the 915 Households Interviewed At Least Once

STATUS IN PASSAGE 3

STATUS IN PASSAGE 2	HH member	Children living elsewhere	Non-member	HH dropped	Died	Unborn	Survey End	Died Between Round 1 and 2	Total
HH member	4293	136	284	54	43	0	183	5	4999
Children living elsewhere	108	185	11	4	0	0	7	0	318
Nonmember	239	1	488	10	0	0	9	0	697
HH dropped	0	0	0	147	0	0	0	0	147
Died	0	0	0	0	69	0	0	0	69
Unborn	86	1	0	0	0	69	0	0	141
Died Between Round 1 and 2	0	0	0	0	4	0	0	0	4
Total	4726	324	733	215	116	54	199	5	6372

Table VI.13(c): Transition Matrix for Individuals in the 915 Households Interviewed At Least Once

STATUS IN PASSAGE 4

STATUS IN PASSAGE 3	*HH member*	*Children living elsewhere*	*Nonmember*	*HH dropped*	*Died*	*Survey End*	*Died Between Round 1 and 2*	*Total*
HH member	3942	159	265	86	44	227	3	4726
Children living elsewhere	87	215	6	8	0	7	1	324
Nonmember	230	3	458	27	0	15	0	733
HH dropped	0	0	0	215	0	0	0	215
Died	0	0	0	0	116	0	0	116
Unborn	54	0	0	0	0	0	0	54
End of survey	0	0	0	0	0	199	0	199
Died Between Round 1 and 2	0	0	0	0	5	0	0	5
Total	4313	377	729	336	165	448	4	6372

3. "Key" dataset for children on the Child Living Elsewhere Roster in Section (key__cle)

Key__cle tracks all 4,378 children ever listed on the Child Living Elsewhere (CLE) Roster of Section 2.

Children can be added to the Child Living Elsewhere (CLE) roster between passages because they moved away from the household, because they were already living away from the household but a parent joined the household, or because they were born away from the household since the last passage. They can disappear from the CLE if their parents are no longer in the household, if the children join the household, or if they die. Thus, the status of children living elsewhere is classified as follows (variables w1_cstat, w2_cstat, w3_cstat, w4_cstat):

- *Still away (listed on previous CLE and still away)*
- *Joined household (moved from CLE to household roster)*
- *Died*
- *Parent not household member*
- *Left household (moved from household roster to CLE)*
- *New*

- *Born*
- *Missed in previous wave*
- *Household dropped out of survey*
- *Not yet (will be listed on CLE for future wave)*

D. MISSING DATA

These are two potential sources of missing data: missing data due to the skip pattern; and questions that are incomplete. In the first case, when certain questions remain unanswered as a result of answers to preceding questions, they will appear as missing in the data.

Table VI.14: Missing Observations: Household Questionnaires

Section	Wave 1 Expected	Wave 1 Missing	Wave 2 Expected	Wave 2 Missing	Wave 3 Expected	Wave 3 Missing	Wave 4 Expected	Wave 4 Missing
1	5280	0	4999	0	4726	0	4313	0
3	5280	1	4999	0	4726	1	4313	0
5	4180	1	3968	0	3729	1	3428	6
6	5280	2	4999	0	4726	1	4313	7
7A	4180	2	3968	0	3729	1	3428	7
7B	372	0	344	2	345	0	302	0
7C	2733	1	2757	0	2778	0	2459	1
7D	236	0	304	0	341	0	346	0
7E	4180	2	3968	0	3729	1	3428	11
7F	4180	3	3968	0	3729	1	3428	11
7G	513	0	602	0	774	0	765	0
7H	4180	11	3968	0	3729	1	3428	11
8	5280	2	398	1	433	1	371	10
9	1514	1	1108	10	1063	0	963	4
10	5280	0	4999	0	4726	1	4313	5
18A	5280	10	4999	1	4726	1	4299	14
18B	2797	9	2667	2	2504	1	2293	8
19A	5280	6	4999	1	4726	4	4295	18
19B	2797	5	2667	0	2504	1	2301	7
19C	2797	6	2667	0	2504	1	2301	7

Table VI.15: Price Questionnaires Completed by Wave and Cluster

Passage	Total Clusters With Any Information	Number Of Clusters With 1 Price Questionnaire Completed	Number Of Clusters With 2 Price Questionnaires Completed	Number Of Clusters With Out Information	Total Clusters
1	47	42	4	4	51
2	48	14	34	3	51
3	49	14	35	2	51
4	50	10	40	1	51

VII. AGGREGATED VARIABLES FOR HOUSEHOLD INCOME, CONSUMPTION, AND ASSETS

This section reviews three sets of files: the household current account related to consumption, the household current account related to income and the household capital account.

The household current account includes household income and expenditure. Table VII.1 shows the major components of the current account that have been constructed for the analysis. While we focus here on the definition of household-level aggregates, note that several of the components can be computed at the individual level as well. In the sections that follow, the definition of each of these aggregates of the current account is described according to the components' variable names and the precise questions of the KHDS questionnaire used to construct them.

Table VII.1: Aggregates of the Household Current Account

Income	*Expenditure*
Employment income[a] (incempl)	Food expenditure (expfood)
Income from agriculturalself-employment (incagr)	Consumption of home production (conshome)
Non-farm self-employment income (incbus1h, incbus2h)	Food (conshmfd)
Income from rent (incrent)	Business inputs + outputs (consbus)
Transfer income[a] (inctrans)	Livestock (conslvst)
Other non-labor income[b] (incothnl)	Non-food consumption expenditure[b] (expnfood)
Total household income (inchh1, inchh2)	Remittances sent[a] (expremit)
	Imputed expenditure for wage income in-kind[a] (expkind)
	Total household expenditure (exphh)

a. These aggregates can be computed at the individual level as well. B. Certain components of non-labor income and non-food consumption expenditure can be computed at the individual level.

1. HOUSEHOLD EXPENDITURE (EXP___HH)

Food Expenditure (expfood)

Seasonal food expenditure. There were as many as two rainy and two dry seasons in Kagera region, depending on the locale. Since the types of foods purchased and consumed often vary with seasons, food expenditure information was collected for both seasonal and non-seasonal food items. The definition of "dry" and "rainy" seasons was left to each household; the mapping of the months of the year into each season is found in Section 17A of the household questionnaire. For each of the seasonal food items (in Section 17C1 of the questionnaire), the respondent was asked during which months of the past 12 months the food item was purchased (question 2). He/she was then asked how often the item was purchased during the months of the rainy (dry) season that it was purchased (questions 3 & 5) and the amount usually spent per

transaction (questions 4 & 6). Thus, each food expenditure in the questionnaire was multiplied by the number of times it was bought per month in the relevant season and multiplied by the number of months in the rainy or dry season that the respondent purchased it (question 2), using Section 17A1 to define which months are rainy and which are dry. Expenditures for the two seasons are summed to arrive at expenditures on a single seasonal item, and all items are summed together to arrive at an annual figure for seasonal foods.[29]

Non-seasonal food expenditure. For each food item, the respondent was asked during which months in the past 12 (or 6) months the item was purchased (question 2), how often the item was bought in the months it was purchased (question 3) and the usual amount spent each time it was purchased (question 4). Thus, the amounts in question 4 must be multiplied by the number of months and the frequency from the two previous questions, then summed over all non-seasonal food items.[30]

Expenditure on meals and beverages consumed outside the home. These expenditures are recorded for each individual in Section 1813, for a reference period between rounds of the survey (approximately 2 weeks). The individual expenditures are summed for a 2-week household total and multiplied by 26 or 13, depending whether the aggregate is for 12 or 6 months.

Table VII.2: Components of food expenditure (expfood)

Description of component		Questionnaire
a. Expenditure on food, rainy season	expfrain	S17C1Q4 (if S17C1Q2=1) (use S17AQ1, S17C1Q2 & Q3 to calculate frequency) summed over all food items
b. Expenditure on food, dry season	expfdry	S17C1Q6 (if S17C1Q2=1) (use S17AQ1, S17C2Q2 & Q5 to calculate frequency) summed over all food items
c. Expenditure on non-seasonal food	expfnsn	S17C2Q4 (if S17C2Q2=1) (use S17C2Q2 &Q3 to calculate frequency) summed over all food items
d. Expenditure on meals and beverages consumed away from home	expfdout	S18BQ1A + S18BQ2A summed over all HH members

Consumption Of Home Production (conshome)

Consumption of home produced food (exphmfd) is measured seasonally in Section 17B of the household questionnaire. This calculation uses again the mapping by the respondent of the months of the year into the rainy and dry seasons in Section 17A, question 1. The respondent is asked during which months of the past 12 months the food item was consumed (question 3). He/she was then asked how often the item was consumed out of home production during the months of the rainy (dry) season that it was consumed (questions 4 & 6) and the amount it would

---

[29] An alternative method of computing annual seasonal food expenditures is to multiply the amount spent in the two weeks since the interviewer's last visit (question 8) by 26 (for an annual estimate) or by 13 (for a 6month estimate). However, this method would assume that the amount spent in the past 2 weeks is typical of the amount spent on the item in the previous 12/6 months.

[30] An alternative way of computing non-seasonal food expenditure is to multiply the answer to S17C2Q6 (the amount spent on the item since the last visit (2 weeks earlier) by 26 (for an annual estimate) or 13 (for a 6month estimate). However, infrequently-purchased food items purchased will distort annualized food expenditures for a household that purchased such items in the past 2 weeks, and they will be excluded from aggregates for households that did not purchase them.

have cost to buy the same amount they consumed each time (questions 5 & 7). Thus, the value of home production consumed for each item was multiplied by the number of times it was bought in the stated time period, converted into a monthly figure, and multiplied by the number of months cited by the respondent in question 3 as in the given season (rainy, dry), using the definition of seasons from Section 17A (question 1). Expenditures for the two seasons are summed to arrive at the value of home production consumed on a single seasonal item, and all items are summed together to arrive at an annual figure for seasonal consumption of home production.

Consumption of production and inputs from household business (expbus). The module that measures income and expenditure arising from businesses owned and operated by the household allows for two types of consumption of home production: consumption of inputs for the business by the household (Section 14B) and consumption of the output of the business by the household (Section 14C).

Table VII.3: Components of consumption of home production (conshome)

Description of component	Variable name	Questionnaire
a. Consumption of home-produced food	conshmfd	
Consumption of home-produced food in the rainy season	conshmrn	S17BQ5 (if S17BQ2=1) (use S17AQ1, S17BQ3 & Q4 to calculate frequency) summed over all food items
Consumption of home-produced food in the dry season	conshmdr	S17BQ7 (if S17BQ2=1) (use S17AQ1, S17BQ2 & Q6 to calculate frequency) summed over all food items
b. Consumption of production and inputs from businesses	consbus	
Expenditure on business inputs consumed by the household	consinpt	S14BQ6 (if S14BQ5=1) summed over all business expenditures for all household businesses
Consumption of business output consumed by the household	consotpt	S14CQ10 (if S14CQ9=1) summed over all household businesses
c. Consumption of livestock raised by the household	conslvst	S12AQ12 x S12AQ4 summed over all types of livestock consumed

Non-Food Consumption Expenditure (expnfood)

Non-food consumption expenditure is the sum of six sub-aggregates (see Table VII.4).[31] The calculation of these sub-aggregates is self-explanatory, with the exception of health and funeral expenditure (discussed below). Note that educational expenditure, health expenditure on living household members and part of "other" non-food household expenditure are measured at the individual level and summed over all household members. Expenditure on the health of members who died over the reference period and expenditure on funerals are measured for each person

---

[31] We have not included in this measure the use value of durable goods (nor is it included anywhere in thecalculations of income or expenditure), because of the arbitrariness necessary in making assumptions on which these calculations are based.

who died and summed over all to arrive at a household total. Housing expenditure is measured at the level of each dwelling and summed over all dwellings. Expenditure on utilities and part of the other nonfood expenditure is measured at the household level and requires no aggregation.

Health expenditure. Information on the health expenditure of all *living household members* in the last 12 or 6 months is collected at the individual level in Section 18A, questions 9A and 10A. However, the questions in Section 18A did not probe for specific types of expenditure, beyond "medicines" and "other medical services". Section 6A includes detailed questions for all medical expenditure on episodes of acute illness/injury occurring in the past 4 weeks, including the costs of hospitalization, outpatient costs, drugs, and transportation.[32] In addition, a global question in Section 6B asks for all expenditure on chronic conditions for the past 4 weeks. Since the information in Section 6 is based on a shorter time frame, the sum of expenditures in Sections 6A and 6B (less the amount spent on transport to medical care) should in principle be less than or equal to the medical expenditures in Section 18A. A comparison of the sum of items from Section 6AB (minus transportation) and the medical expenditure in Section 18A found very few instances in which the costs in Section 6AB exceeded those in Section 18A: 3.7 percent of individuals in Wave 1, 2.0 percent in Wave 2, 1.6 percent in Wave 3 and 2.2 percent in Wave 4. For those individuals, annual health expenditure is set equal to the sum of all medical expenditure in Section 6AB (including transportation costs); for all others, annual health expenditure is set equal to the sum of questions 9A and l0A of Section 18, plus transportation costs measured in Section 6A. Health expenditure is then summed over all individuals to arrive at a household total.

In order to calculate annual or 6-month health expenditure on *household members who died* from Section 20A, one must first establish what part of the total medical expenditure before death occurred in the reference period. For this we use Section 20A, question 28 (the time ill before death), question 20 (the date of death) and the date of the round two interview to establish whether any part of the illness occurred outside of the reference period. All of the medical expenditures are included for those whose illness falls entirely within the reference period. For those illnesses occurring outside the reference period, we multiply medical expenditure before death by the proportion of the ill days that fall within the reference period.

Funeral expenditure. Expenditure on the funerals of household members who died is measured in Section 20, question 43. This is supposed to include expenditure by household members and expenditures that were financed by others. Question 44 obtains the value of transfers from others for the funeral, and should be less than the total in question 43. However, in some instances this was not the case. Thus, the variable used for funeral expenditure is either Section 20A, question 43 or question 44, whichever is larger. This is summed over all household members who died.

---

[32] Note that the ordering of questions in Section 6A changed between the Wave 1 and Wave 2 questionnaires, although the variable names remained intact. In Wave 1, hospitalization costs were measured in questions 21, 33, 45 and 53, while in Waves 2-4 they were in questions 23, 35, 47 and 53. Outpatient costs in Wave 1 were measured in questions 23, 35, 47 and 50, while in Waves 2-4 they were captured in questions 20, 32, 44 and 50.

Table VII.4: Components of non-food consumption expenditure (expnfood)

Description of component	Variable name	Questionnaire
a. Education expenditure by members of the household, other individuals and organizations, on household members	expschl	S5Q18H[a] + S5Q22 + S5Q25H[a] + S5Q28, summed over all HH members 7 and older
b. Health expenditure	exphlth	
Expenditure on living household members'	explhlth	Max {(S18AQ9A+S18AQ10A); (S6A + S6B - S6AQ56)} + S6AQ56, summed over all living household members
Expenditure on the health of members who died	expdhlth	(S20AQ31B + S20AQ33 +S20AQ34), summed over all household members who died
c. Funeral expenditure	expfunrl	max {S20AQ43,S20AQ44},summed over all household members who died
d. Housing expenditure (imputed value of owner-occupied housing, rent paid in cash and in kind)	exphsing	S15AQ17 + S15AQ20 + S 15AQ25
e. Utilities (water, garbage disposal, electricity, fuel)	exputil	[S15BQ13 x months] +[S15BQ19 x time unit] +[S15BQ24 x time unit] +[S15BQ27 (6 items)x (26 or 13)]
f. Other non-food household expenditure *Measured at the individual level:* clothing, footwear, jewelry and watches, toys, books, haircuts, handbags, tobacco, gambling, magazines, gasoline and motor oil, writing materials, candles, matches, batteries, sports, personal hygiene products, cosmetics *Measured at the household level:* repairs, kitchen equipment, linens, home services, taxes, dues, wedding, dowries, donations	expotnfd	[S18A (QlB +Q2B +Q3B +Q4B + Q5B +Q6B +Q7B +Q8B)] + [S18B (Q3B +Q4B +Q5B +Q6B +Q7B +Q8B +Q9B +Q10B +Q11B +Q12B) x 26 or 13], summed over individuals S16BQ2 (items 10-13, 16-22)

a. S5QI8H is the total amount spent by household members on each child's schooling, while S5Q25H is the total spent by non-household members. Note that S5Q18A - G may not sum to this total; the amount for specific components were left blank if not known with certainty. The same is true for S5Q25A - G. See the interviewer manual. b. An alternative measure is included in the data set, altlhlth, which is the maximum of (section 18A or section 6A + *annualized* expenditure or chronic illness - S6Q56), computed for living members.

Remittances Sent (expremit)

The remittances sent are calculated as the amount sent, less the amount repaid to the sender, less the amount that will be repaid to the sender, as measured for each household member 15 and older in Section 19B.

Table VII.5: Components of remittances sent

Description of component	Variable name	Questionnaire
Remittances sent	expremit	S 19B (Q7 - Q 11 - Q 12) + S19B (Q18 - Q22 - Q23) + S19B (Q29 - Q33 - Q34) + S19B (Q36 - Q38 - Q39), summed over all HH members

Imputed Expenditure For Wage Income In Kind (expwage)

This includes payment for a wage job in the form of food, crops, animals, subsidized housing, and all other in-kind benefits.

Table VII.6: Components of imputed expenditure for wage income in kind (expkind)

Description of component	Variable name	Questionnaire
Imputed expenditure for wage income in kind, employment in the last 7 days	wagelst7	S7BQ29 + S7BQ31 + S7BQ33, summed over all HH members 7 and older
Imputed expenditure for wage income in kind from main wage job last year/6 months if different from wage job last 7 days	explgtrm	S7FQ25 + S7FQ27 + S7FQ29, summed over all HH members 7 and older

2. HOUSEHOLD INCOME (INCHHL, INCHH2)

Household income is defined as the sum of six components: employment income; income from self-employment in agriculture; non-farm self-employment income; income from rent; transfer income from individuals and organizations; and other non-labor income.[33]

Employment Income (incempl)

Employment income refers to income received as an employee of a private individual or of an institution *other than the household,* for remuneration in cash or in kind.[34] Household employment income is calculated by summing together the employment income of all household members, from Sections 7B, 7F and 7G of the household questionnaire. The components of the household employment income aggregate are described in Table VII.7, below. They are multiplied by the time period over which the individual worked at each job in the past 12 months and aggregated at the household level.

---

[33] Imputed income from the use value of durable goods has not been included because it was felt that unacceptably arbitrary assumptions would have to be made.

[34] Wage income from employment in a family business is cited in Section 7D, question 11, which is the amount received by family workers in cash and in kind. However, this information is also collected at the level of the business in Section 14A, questions 10 (frequency) and 11 (amount) for all household members in all family businesses. To include both would be a double count. We have included the value of payments to household workers as household income, using the results from Section 14A instead of using the amount in Section 7D.

Table VII.7: Components of employment income (incempl)

Description of component	Questionnaire
Annual (or 6-month) income fiom employment in wage jobs in the past 7 days	S7B (Q22+Q25+Q27+Q29+Q31+ Q33 +Q39)
Annual (or 6-month) income from employment in the main job last year (if different from those in the past 7 days, i.e. if S7FQ2=2) and salaried S7FQ7=3	S7F (Q20+Q23+Q25+Q27+Q29)
Annual (or 6-month) income from employment in the secondary job last year (if different from those in the past 7 days, i.e. if S7GQ2=2 and S7GQ3=1)	S7GQ4

Income From Self-Employment In Agriculture (incagr)

Household income from self-employment in agriculture is computed from gross revenues less costs of household-level activities in farming, livestock and fishing, plus the value of home agricultural production that is consumed by the household:

Income from self-employment in agriculture =
Net revenues from agriculture (gross revenues less costs) +
consumption of home production

Depreciation of agricultural equipment should be included as a cost, but it was not available in the questionnaire. Income from renting out land and farm equipment is included in "rental income", a different aggregate below.

In estimating income from crops, respondents were asked to cite the amount sold (plus a unit code) and the price received (plus a unit code). Two problems appear in this data. First, on about 10 percent of all observations the units cited are the same but it is clear that the respondent was reporting a price for a quantity other than that for the unit code. Second, on a small number of crop observations the respondent cited different units for the quantity sold and for the price.

The identification of mis-coded prices is essentially a problem of identifying outlier prices and replacing those outliers with an estimate of the true price that the farmer would have received on that observation. Since prices vary systematically across space, time and rainfall, the approach used for identifying and replacing outliers uses a regression model. In order to pool as many observations as possible in a single regression, we start by assuming that the price per kilogram of any crop will not vary as a function of the units in which it is sold. If this is true, the following identity must obtain:

$$P_{it} = PKG_{it} * Kg/Unit$$

where P_{it} is the price per specified unit of a crop in cluster i at time t, PKG_{it} is the price per kilogram of that crop in that cluster on that date and Kg/Unit is the number of kilograms per unit, which is assumed constant across all clusters and time periods for a specified unit. Taking the logarithm of this expression gives the linear relationship:

$$\log(P_{it}) = \log(PKG_{it}) + \log(Kg/Unit)$$

Assuming that $\log(PKG_{it})$ varies systematically with geographic location, time and rainfall, we can write:

$$\log(PKG_{it}) = B_0 + B_1X_{1i} + B_2X_{2i} + B_3t + B_4RAIN_{it} + E_{it}$$

where X_{1i} and X_2i are the latitude and longitude of cluster i, t is the number of days since the beginning of the survey, $RAIN_{it}$ is the average amount of rainfall in the 12 months preceding the date of interviews in cluster i and E_{it} is a random disturbance term assumed to be independent of these variables. Substituting into the previous equation gives:

$$\log(P_{it}) = Bo + B_1X_{1i} + B_2X_{2i} + B_3t + B_4RAIN_{it} + \log(Kg/Unit) + E_{it}$$

By defining a dummy variable for each unit in which the crop's price is ever specified, $\log(Kg/Unit)$ becomes a vector of dummy variables, the coefficients of which provide estimates of the number of kilograms in each of the other units in which price was specified over the four waves of the survey. These coefficients can be used to construct "conversion factors" for the cases when the quantity unit and price unit do. not match as described below.

For each of the fourteen crops on which there were at least 60 observations, the approach used to identify outliers for replacement was to estimate the above equation by "robust regression" techniques.[35] For twelve of the crops (accounting for 5,019 of the 5,860 observations), our procedure was to estimate the robust regression, to identify an outlier as an observation having an estimate weight of less than 0.1[36], and to replace it with the predicted value for that observation. Of the 5,019 price-quantity observations, 205 have been replaced using this procedure.

For the 22 crops with fewer than 60 observations, plus ground nuts and sugar cane, the regressions fit too poorly to provide plausible identification of the outliers or to permit estimation of the price from a fitted value. For the 841 observations on such crops, we computed the median price by crop and by unit and replaced observations with the median if the observation was less than 10 or more than 190% of the median. This replacement rule identified as outliers and replaced 203 of the 841 observations.

---

[35] Add explanation of "robust regression". The fit was good and the parameter values reasonable on all but two of these equations (those for ground nuts and sugar cane).

[36] Weights are scaled from zero to one. Observations with low weight fit the estimated regression line poorly and are candidates for outliers, whereas those with high weights fit well. See the documentation of the "rreg" command in STATA *Reference Manual,* Release 4, Volume 3, pp. 132-7.

Table VII.8: Components of income from self-employment in agriculture (incagr)

Description of component	Variable name	Questionnaire
a. Net income from crops and crop processing	ninccrop	
Revenues: cash from sale of crops, value of crops kept for seed, value of crops given to workers, value of crops lots, income from the sale of processed crops (trevcrop)		S11B (Q5 + Q7 + Q9 + Q11) summed over all crops grown + S11EQ6 summed over all processed crops
Expenditures: payment for use of land, value of crops used for seed, value of crops given to workers, value of crops lost, expenses on labor, fertilizer, pesticides, transport and processing of crops for sale		S11AQ11 (summed over all plots' of land) + S11BQ9 (summed over all crops) + S11D (Q3 + processing of crops for sale (tcstcrop) + S11EQ8 (summed over all products)
b. Net income from livestock	ninclvst	Revenue: 12BQ3 (income from livestock products) Expenses: 12CQ2 (summed over all inputs)
c. Net income from fishing	nincfish	Revenue: S13BQ5 + S13BQ8 Expenses: S13CQ4
d. Consumption of home production of food and animals	conshmfd + conslvst	See Table VII.3

Turning to the second problem, there were a total of 91 cases where the unit code of quantity was different from the one for which the price was quoted, with 32 cases for coffee and 10 cases for sorghum and maize each.[37] In order to correct this problem we converted the quantities from the given unit to the unit in which price was quoted using the conversion factors estimated from the coefficients of the dummy variables as described above. This procedure was possible for 57 of the 91 cases. In the remaining cases (for which no conversion factor could be estimated) we applied conversion factors from crops judged to be similar in weight per volume.

With respect to income from livestock, note that we considered revenues from the sale of livestock to be the same as the sale of an asset. Thus, this is included in the capital account. The only livestock item with revenue in the current account is the revenue from sale of processed animal products (milk, cheese, etc.)

Non-Farm Self-Employment Income (incbus1h, incbus2h)

Non-farm self-employment income is the sum of net revenues from all businesses, enterprises or other professional self-employment in the household, as measured in Section 14 of the household questionnaire. There are two ways of computing net business income: (1) by subtracting costs from gross revenues; or (2) by using responses to direct questions on net revenues (profit). Below

---

[37] For 7 cases of cassava, 5 cases of groundnuts, 4 cases each of sweet banana and sugarcane, 3 cases each of palm oil, cooking banana and sorghum, 2 cases each of sweet potatoes and tomatoes and 1 case each of Irish potatoes, millet, mangoes, citrus fruits and cabbage the unit of quantity was different from the unit for which the price was quoted.

we describe how each of these two estimates of non-farm self-employment income were calculated (summarized in Table VII.9) and then comment on the problems in using these data. Since family businesses typically do not keep records of their expenditures or revenues, both methods are subject to a great deal of error. The ideal way to collect such information would be by observing the businesses more frequently over an extended period of time. However, in the KHDS (and in other LSMS surveys), this would have been too costly. Income from self-employment or work in a family business is also collected from each household member in Section 7D. Thus, adding together the figures from Section 7D would constitute a third method of estimating household income from non-farm self employment. However, we believe that the more detailed probing in Section 14 should lead to more accurate results. The responses to Section 7D may nonetheless be useful in analyses requiring income at an individual level.

Gross revenues less costs (method #1). The first method involves in-depth probing on the businesses' expenditures for wages and other variable inputs.[38] Expenditures include those on inputs (net of any inputs consumed by the household), as recorded in Sections 14A and 14B. There are two questions on expenditure on business inputs - one that asks for usual expenditure (with a frequency and. time unit) and another that asks for input expenditures for the past week. We used the question for usual expenditure. Estimates by both methods use fairly short recall periods (2 weeks for businesses in operation at the time of the survey and a respondent-determined time period if the business was not operating at the time of the interview) that are expanded to the number of months that the business was in operation in the past 12 or 6 months (S14AQ5). The salaries of household members are both an expenditure to the business and income to the household. In several instances the respondents reported usual expenditure or income on a daily basis, but there was effectively no question about the number of days per week that the business operated. In those cases, we assumed that the business operated 5 days per week or 20 days per month. On the revenue side, there are questions on gross revenues, income in kind and the value of business outputs consumed by the household.[39]

Direct estimate of net revenues (method #2). Within this method there are two ways of arriving at net revenues: (a) from a direct question on the amount of money left after subtracting expenses but before spending on the household; and (b) by summing together the amount of revenue spent on the household and the amount left after expenses on the firm and the household. Presumably they should be the same. We took the larger of the two numbers (a) or (b) as the businesses' profit. To this we have added the value of output consumed, income in kind and payments in cash and kind to household members working in the business. We assume that, in citing the net income after subtracting business expenses, the respondent has probably not excluded the household consumption of inputs before the subtraction. We therefore also add back the value of inputs consumed.[40]

---

[38] Depreciation of business equipment should be included as an expense. However, it was not collected on the household questionnaire.

[39] Note that consumption of business inputs and outputs also appears in the consumption aggregate and are described in Table VII.3 (consinpt, consotpt)

[40] In examining the data, we noticed that households sometimes cited exactly the same amount for the value of "articles for resale" consumed by the household (in the input expenditure section), as they did for "consumption of outputs". This led us to the belief that they were erroneously counting consumption of the same items twice. Therefore, in instances in which the amount cited for consumption of articles for resale exactly equaled the amount cited for consumption of outputs, we set the consumption of outputs equal to zero. There were 44 such businesses in Wave 1.

Table VII.9: Components of non-farm self-employment income[a]

Description of component	Variable name	Questionnaire
Method 1 (revenues less costs)	incbus1h[b]	
Business revenues: Gross revenues + revenue in kind + salaries of household members + value of output consumed		(S14CQ2 or S14CQ15) + S14CQ8 + S14AQ11 + S14CQ10 (summed over all outputs)
Business costs: Salaries of household members + wages for hired workers + cost of inputs - consumption of inputs		S14AQ11 + S14AQ15 + S14BQ3 (summed over all inputs) - S14BQ6 (summed over all inputs)
Method 2 (reported net revenues)	incbus2h	
Maximum of {reported net revenues, (amount of revenue used by household + amount left)}		Max {14CQ6, (14CQ12 + 14CQ14)} for those in operation since last visit, OR Max {14CQ15, (14CQ16 + 14CQ18)} for those not in operation since last visit
Adjustments: add consumption of output, consumption of inputs, in-kind income, payments to household members		S14CQ10 + S 14CQ8 + S14AQ11 + S14BQ6 (summed over all inputs)

a. The variables in this table are also available at the business level of observation (before aggregation to the household level) in inc_bus. b. Business input expenditure can be computed on the basis of usual expenditure or actual expenditure in the past two weeks. Businclh is calculated using usual expenditure.

Issues in the business income data

Most of the businesses owned and operated by the household did not keep records of their expenditures or revenues. Thus, the estimates of business income rely on annualizing respondents' "usual" expenditure and revenues over a short time periods or the amount spent and received in the two weeks before the interview. Any reporting errors have been amplified in the annualization process.

Revenues are often under-reported relative to expenditure. As a result, the first method often produces negative net business income. While this can be expected for some enterprises - especially those that might have gone out of business - it should not be the case for a large number. In the KHDS data set, from 28 to 35 percent of businesses have negative non-farm self-employment income computed by method 1, depending on the wave (see Table VII.10). The second method, since it is based on a question about positive profits, cannot yield a negative number. This is unreasonable for businesses that are no longer in operation. Profits may also be subject to a great deal of reporting error. Thus, 95-99 percent of businesses have positive profits by this method, and the remainder had zero profits, according to method 2.

Finally, there were many cases in which entire sections of the business module of the questionnaire were left blank because the most informed person was not interviewed and/or the respondents simply could not provide the information or could provide only partial information.

Households for which one should assume that net business income has been completely omitted are signaled by a variable "nbusmiss", the number of businesses with missing information. A second variable was created that is the count of the household's businesses that reported no business costs: "bcstmiss". Note that some proxy respondents no doubt did give responses to the business questions even though they didn't know the true values. The table also shows the percent of businesses for which a proxy respondent actually answered the questions: "busproxy". The table below shows the number of households each wave that reported businesses and the number for whom significant amounts of information is known to be missing.

Because of the problems with collecting this information in only a single interview and without records, the problem of missing respondents, and the evident confusion concerning the interpretation of questions in the questionnaire implied by the extreme differences in estimates of net income, we urge analysts to use the aggregates on non-farm self-employment with extreme caution (see Kozel 1990 and Vijverberg 1991 for other LSMS papers that have encountered these problems).

Table VII.10: Issues in the non farm self employment data

Wave	Total businesses[a]	Businesses with missing data (%)[b]	Effective no. of businesses	Percent w/ proxy respondents	Percent of businesses with positive income, method:		Business income: revenues less costs (method 1a)		Business income: reported profits (method 2)	
					1	2	Mean	[min;max]	Mean	[min;max]
1	326	26(8.0)	300	4.0	69.1	99.0	-67,914	$[-1.6 \times 10^7; 6,349,200]$	143,597	[0;6,545,040]
2	463	31(6.7)	432	4.4	70.8	95.4	12,365	[-2,568,500;2,670,000]	45,420	[0;1,800,000]
3	541	53(9.8)	488	3.5	64.5	96.5	-123,636	$[-4 \times 10^7; 1,165,150]$	30,450	[0;1,360,000]
4	563	70(12.4)	493	7.5	64.7	95.7	21,887	[-2,770,700;7,280,000]	43,145	[0;3,536,000]

a. Excludes businesses reported in households 4910 and 4911. b. That is, the number with busmiss=1 or sl4bmiss=1.

Income From Rent (incrent)

The four components of rental income are in Table VII.11 below. Note that the imputed value of owner-occupied housing is obtained directly from the respondents and does not rely on a hedonic prediction. Recall also that this last component is also included in household expenditures.

Table VII.11: Components of income from rent (incrent)

Description of component	Items of the questionnaire
Income from renting land	S11AQ15 (summed over all parcels)
Income from renting farm equipment	S11GQ11 (summed over all equipment)
Income from renting dwellings	S 15BQ6
Rental value of owner-occupied housing	S15AQ17

Transfer Income (inctrans)

This aggregate includes transfers received by all household members from individuals and organizations. The main source of information on remittance income from individuals is Section 19A, while the main source of transfers from organizations is Section 16C.[41] Interviewers were instructed not to record any transfers from organizations in Section 19 and vice versa, so there are no double counts between these two sections. However, there were several questions in other parts of the questionnaire that represent double-counting of transfers from individuals: Section 5 (transfers from individuals for schooling); Section 6 (transfers from individuals or organizations for health care); Section 7H ("other income"); and Section 20A (transfers for funerals). In addition, separate questions in Section 5 ask about transfers for schooling from organizations.

These double-counts afford the opportunity to run internal consistency checks on the data and to include information in the aggregate thought to have been omitted from Sections 19A and 16C. However, remittance income is defined as receipt of transfers from individuals who are not household members, *and that do not have to be repaid.* Remittances that had to be repaid are considered to be loans and are included in the capital account aggregates. While it is possible to compare gross remittances across sections, the other--parts of the questionnaire do not measure the amount of the remittance that has been or will have to be repaid. Thus, the estimate of remittance income from individuals is taken only from Section 19A.

Transfer income from organizations is assumed to be a grant .[42] There is only one clear double count of transfer income from organizations - in Section 5, questions 22 and 28. If transfers for schooling from organizations were reported in Section 5, but none were reported in Section 16C, then that amount has been added to the measure of transfers from organizations.

---

[41] Interviewers were explicitly told not to include transfers from organizations in Section *19A - only* those from individuals.

[42] Section 19A collects information on both transfers and loans; Section 16C is exclusively for transfers from organizations.

Table VII.12: Components of transfer income (inctrans)

Description of component	Variable name	Item in questionnaire
a. Remittances from individuals	incremit	*S 19A (Q7* less *Q11* less *Q12) + (Q18* less *Q22* less *Q23) + (Q29* less *Q33* less *Q34) + (Q36* less *Q38* less *Q39)*
b. Transfers from organizations	incasst	Section *16C, Q2,* aggregated over all organizations

Other Non-Labor Income (incothnl)

Other non-labor income is measured at the individual level in Section 7H. It is the sum of questions 213 through 7B (pension or retirement fund; insurance; interest on bank accounts; income from games; dowry; inheritance), summed over all household members.

Table VII.13: Problems with Non-Farm Self-Employment Data, by Wave (Business Level of Observation)

Problem	Wave 1	Wave 2	Wave 3	Wave 4
Missing data-entire section	5	3	11	12
Missing data – business costs	21	28	42	58
Method 1 problems:				
Consumption of articles for resale = consumption of outputs	44	62	62	46
Consumption of inputs in the past 2 weeks (annualized)				
exceeds all input expenditures annualized:				
"Usual" input expenditures	7	4	1	3
Input expenditures since last visit	16	21	11	6
Method 2 problems:				
Gross revenue reported is less than reported net revenue, method 2	1	0	0	0
Cash profits in relation to amount spent on HH and amount left:				
Cash profits = amt. spent on HH + amt. left	55%	67%	92%	91%
Cash profits > amt. spent on HH+ amt. left	12%	7%	4%	5%
Cash profits< amt. spent on HH+ amt. left	33%	26%	5%	4%

3. THE HOUSEHOLD CAPITAL ACCOUNT (CAPT__HH)

Table VII.14 depicts the household capital account. It is in balance when receipts (increase in savings) equal outlays (net acquisition of physical and financial assets). The increase in savings can be measured as the balance of the current account - that is, income less expenditure. However, as Round and McKay (1992)[43] note, since both income and expenditure are often inaccurately measured, and income is underestimated by more than expenditure, the increase in savings is likely to be underestimated by the current account balance. We thus take advantage of the capital account identity to calculate savings as the net acquisition of physical and financial assets. The sum of these two items can also be thought of as the change in net debt in the household over a given period - which can be either positive or negative.

Table VII.14: Household Capital Account

Receipts	*Outlays*
Savings	Net acquisition of physical assets:
	net land acquisition
	net livestock acquisition
	net acquisition of agricultural equipment
	net acquisition of business assets
	net acquisition of durable goods
	net acquisition of inventories
	Net acquisition of financial assets:
	net repayment of loans

The household capital account aggregates can be calculated in two different ways using the longitudinal KHDS data. Retrospective questions concerning the net acquisition of physical and financial assets over a specific reference period are the basis for estimates of the capital account within a single wave of data. However, a separate set of capital account estimates can be made based on the differences between the value of physical and financial assets in two successive waves. This second method makes better use of the panel design of the KHDS. However, in differencing the data between waves the number of observations per household is automatically reduced by one. Further, the total number of households observed will be reduced to those observed at least twice. Calculations of capital account aggregates using both a single wave of data and consecutive waves are described below.

Estimates Using a Single Wave of Data

Savings in the capital account are the sum of the net acquisition of physical assets and of financial assets, either or both of which can be negative. Round and McKay (1992) note that the net acquisition of physical assets should be estimated after depreciation, but this cannot be done with the KHDS data. Changes in inventories should also be included in the net acquisition of physical assets, but this cannot be done within a single wave of KHDS data. Table VII.15 explains the construction of the capital account aggregates, including the specific questions in the

---

[43] Round, Jeffrey, and Andrew McKay. 1992. "Income-Expenditure Aggregates." In Chapter 7 of Delaine and others, *The Social Dimensions of Adjustment Integrated Survey.* SDA Working Paper No. 14. World Bank, Washington, D.C.

KHDS questionnaire that were used. They are all measured in current terms and have not been adjusted for inflation or regional price variation.

Multi-wave Estimates

An alternative way to obtain estimates of saving is by computing the change in asset holdings across waves. Table VII.16 contains a detailed description of the computation of the stock of assets for a single wave. Given the stocks for each wave, the change in asset holdings can be computed by calculating the difference in asset position from one wave to the next (see Table VII.17). Since the stock of inventories can be computed on a wave by wave basis, the multi-wave method - unlike the single-wave method - yields estimates of the change in inventories that can thus be included in saving.

Table VII.15: Net Change in Assets Using a Single Wave of Data (acqa_hh)

Description of aggregate	Name of aggregate	Sub-components	Name of individual items	Questions
NET ACQUISITION OF PHYSICAL & FINANCIAL ASSETS taqasset = aqpasset + aqfasset				
NET ACQUISITON OF PHYSICAL ASSETS	aqpasset	Net acquisition of all physical assets		
Net acquisition of land	acqsland	Sum of purchase price of all shambas bought in the past 12/6 months	shbuyamt if shbuy 12/shbuy6=1	S11AQ7 if S11AQ6=1
		(Plus) Value of inherited land[b]	shamvalu if shaminhr=1	S11Q9 if S11Q8A=1
		(Minus) revenue from the sale of land	gardamt if garsold=1	S11QE if S11QD=1
		(Minus) value of disinherited land[b]	amtdis if disinlnd=1	S11QH if S11QF=1
Net acquisition of farm equipment	acqseqp	Cost of buying farm equipment minus the receipt of payment from sale of farm equipment	(feqbyamt if feqbuy=1) minus (feqsdamt if feqsold=1)	(S11GQ7 if S11Q5=1) minus (S11GQ10 if S11GQ8=1)
Net acquisition of livestock	acqslvst	Cost of purchasing livestock	lvsbyamt	S12AQ10
		(Plus) Value of livestock born and received	vlvstbrn=lvsslamt x lvstrecd	S12AQ4 x S12AQ11
		(Minus) receipts from sale of livestock	lvssdamt	S12AQ7
		(Minus) value of animals lost, stolen or died	vanmlst=lvsslamt x lvstlost	S12AQ4 x S12AQ13
		(Minus) value of consumption of livestock	vlvsteat=lvsslamt x lvseaten	S12AQ4 x S12AQ12
Net acquisition of business assets	acqsbsas	Cost of buying business assets less receipt of sales of business assets	(basspaid if bassbuy=1) minus (bassinc if basssold=1)	(S14DQ5 if S14Q4=1) minus (S14DQ7 if S14Q6=1)
Net acquisition of fishing equipment	acqsfish	Cost of buying fishing equipment less receipt from sale of fishing equipment	(fsehqp if fsheqbuy=1) minus (fsheqinc if fsheqsld=1)	(S15AQ16 if S15AQ12A=1 and S15AQ12=1) minus (S15BQ11B if S15BQ11A=1)

Table VII.15: Net Change in Assets Using a Single Wave of Data (acqa_hh)
continued

Description of aggregate	Name of aggregate	Sub-components	Name of individual items	Questions
Net acquisition of dwellings	acqsdwel	Net value of inherited dwellings[b]	(bldvalue if acqd6mo=1 and bldacqd=1) minus (amtdsinh if disindwl=1)	(S15AQ16 if S15AQ12A=1 and S15AQ12=1) minus (S15BQ11B if S15BQ11A=1)
		(Plus) Cost of buying dwellings minus receipts from sale of dwellings	(bldvalue if acqd6mo=1 and bldacqd=3) minus (dwelinc if swelsold=1)	(S15AQ16 if S15AQ12A=1 and SS15AQ12=3) minus (S15BQ11 if S15BQ9=1)
Net acquisition of durables	acqsdurb	Cost of buying durable goods less receipts from sale of durable goods	(durexp if dur6mos=1) minus (duramt if durables=1)	(S16AQ6 if S16AQ4=1) minus (S7HQ8B if S7HQ8A=1)
NET ACQUISITION OF FINANCIAL ASSETS	aqfasset	Net acquisition of all financial assets		
Money lent out in past 6/12 months	acqsloan	Value of loans given out within last 6.12 months and still owed by all borrowers	lnt1owed+ lnt2owed+ lnt3owed+ lnt4owed	S19BQ12+ S19BQ23+ S19BQ34+ S19BQ36
debt incurred in past 6/12 months	acqsdebt	Value of debt contracted within last 6/12 months and still owed to all lenders	rec1owed+ rec2owed+ rec3owed+ rec4owed	S19AQ12+ S19AQ23+ S19AQ34+ S19AQ36

[a]These operations are carried out over all shambas, types of livestock, etc.. [b] These can be calculated beginning in Wave 2; not included in Wave 1 component.

Table VII.16: Value of Household Assets (capt_hh)

Asset type	Name of aggregate	Sub-components[a]	Name of individual items	Questions
STOCK OF PHYSICAL & FINANCIAL ASSETS totstock=phystock+finstock				
HOLDINGS OF PHYSICAL ASSETS	phystock			
Land	landvalu	Total value of shamba/garden owned	shamvalu	S11AQ9
Farm equipment	eqipvalu	Total value of farm equipment, excluding buildings	feqslamt if fequipid not equal 10	S11GQ4 if S11GQ1 not equal 10
		Total value of hand tools[b] (hndtvalu)	hoesamt+axesamt+machamt+pickamt+shovamt+whbamt+ sickamt +pangamt +munduamt +shearamt+othramt	S11F1 to S11F11
Farm buildings	fbldvalu	Total value of farm buildings[b]	feqslamt if fequipid equal 10	S11GQ4 if S11GQ1 equal 10
Livestock	lvstvalu	Total value of livestock owned	lvsslamt x lvsttno	S12AQ4 x S12AQ3
Business Assets	busnvalu	Total value of business assets	bassvalu	S14DQ3
Fishing equipment	fishvalu	Total value of fishing equipment	fsheqval	S13AQ2
Owner-occupied dwellings	odwlvalu	Value of dwellings owned & occupied by households	bldvalue	S15AQ16
Other dwellings	udwlvalu	Value of other dwellings not occupied by household	dwelvalu	S15BQ8
Durables	durbvalu	Value of durable goods owned by household	durvalue	S16AQ7
Farm inventories	invfvalu	Value of crop inventories	crstkamt	S11BQ13
Business inventories	invbvalu	Value of unsold products	bsselall	S14CQ22
FINANCIAL ASSETS	finstock	Savins + loan – debt		
Cash Savings	Saving	Total savings	totsavng	S19CQ12
Total loans	loan	Value of loans still owed by all borrowers	acqsloan[c] + lnowedat	acqsloan[c] S19AQ41
Total debts	debt	Value of debt still owed to all lenders	acqsdebt[c] - repowed	acqsdebt[c] S19BQ41

[a] These operations are carried out over all shambas, types of livestock, etc. [b] Not available for Wave 1. [c] For definition of afasset1 see Table VII.15 above.

Table VII.17: Change in Net Asset Holdings Between Waves (cassethh)

Description of component	Variable name	Time Period
CHANGE IN PHYSICAL & FINANCIAL ASSETS: ctotstk=cphystk+cfinstk		
Total change in stock of physical assets	cphystk	change in value of stock of physical assets[a] (phystock) between: wave2-wave1, wave3-wave2, wave4-wave3
Total change in stock of financial assets	cfinstk	change in value of stock of financial assets (finstock) between: wave2-wave1, wave 3-wave2, wave4-wave3

[a] The value of farm buildings and of hand tools was not collected until Wave 2. In computing the change in physical assets holdings from Wave 1 to Wave 2, they were excluded from Wave 2 so as not to artificially inflate the change in the physical asset position between these two periods. For all subsequent intervals the value of farm buildings and hand tools was included.

References

Ainsworth, Martha, Susmita Ghosh, and Innocent Semali. 1995. "The impact of adult deaths on household composition in Kagera Region, Tanzania." August draft.

Ainsworth, Martha, Godlike Koda, George Lwihula, Phare Mujinja, Mead Over, and Innocent Semali. 1992. "Measuring the Impact of Fatal Adult Illness in Sub-Saharan Africa: An Annotated Household Questionnaire." LSMS Working Paper 90. World Bank, Washington, DC.

Ainsworth, Martha and Jacques van der Gaag. 1987. "Guidelines for Adapting the LSMS Living Standards Questionnaires to Local Conditions." LSMS Working Paper 34. World Bank, Washington, DC.

Grootaert, Christiaan. 1986. "Measuring and Analyzing Levels of Living in Developing Countries: An Annotated Questionnaire." LSMS Working Paper 24. World Bank, Washington, DC.

Government of Tanzania. 1967. *Tanzania Atlas*. Dar Es Salaam.

Round, Jeffery and Andrew McKay. 1992. "Income-Expenditure Aggregates." In Ghislaine Delaine and others, The Social Dimensions of Adjustment Integrated Survey: A Survey to Meaure Poverty and Understand the Effects of Policy Change on Households. SDA Working Paper 14. World Bank, Washington, DC.

Semali, Innocent and Martha Ainsworth. 1995. "A Profile of Traditional Healers in an Area Hard Hit by the AIDS Epidemic: Kagera Region, Tanzania." Mimeo.

APPENDIX I: CHANGES TO KHDS QUESTIONNAIRES ACROSS WAVES AND PASSAGES

The following appendix details the changes in questionnaires of all types across waves and passages. Questionnaires applied across passages differed by passage due to improvements in the survey tool or due to references to information in past passages. This appendix details the changes between waves so researchers can understand how these changes may have affected the data collected by KHDS. These changes are discussed in this section so users-of KHDS data will not need to possess all four waves of the package of KHDS questionnaires. Descriptions of the content of each of the questionnaires are found in Section V: KHDS Questionnaires of the Documentation Report. Supplements to this discussion are the KHDS Interviewer's Guide and Supervisor's Guide.

QUESTIONNAIRE SCHEMES - USING THE QUESTIONNAIRES

The basic schemes of KHDS conventions and reference periods are discussed below.

1. Conventions

When survey questions from the questionnaires are used in the discussion of changes - either to illustrate a new wording scheme or to help the reader to understand a point - the conventions are as follows (and are the same as they were in the questionnaires for the interviewers):

- Questions are written verbatim in the manner that they were asked;
- Everything written in lower case letters was read aloud to the respondent.
- Everything written in capital (block) letters was an instruction to the interviewer, usually in English, and was not read to the respondent;
- Questionnaires were almost entirely pre-coded - for most questions all possible answers were anticipated and were listed and numbered after the question;
- The order of asking questions was controlled by skip instructions - for example, "> 10" meant "skip to question 10" (some skip instructions changed between waves);
- Skip instructions in parentheses applied only to the particular response that they followed (usually in the manner of "if no, skip to question 11 ") while skip instructions in a box were global skips, meaning the interviewer skipped to the indicated question regardless of the response from the interviewee.

2. Reference Periods

Throughout the survey questionnaires, a major change between Wave 1 and the subsequent three waves was the length of the reference period used in most questions. Reference periods that were twelve or twenty-four months in Wave 1 questions were usually reduced to six months in

questions in Waves 2, 3, and 4.[44] This six-month period was usually an approximation, referring to the period of time that had elapsed since the interviewer's last visit, which in some cases was closer to seven months. Many questions in Waves 2, 3, and 4 thus began with, "In the last six months or since I was here last…".

However, for sections that gathered information on expenditure or consumption, the six-month reference period meant precisely six months. For example, Section 17 (Food Consumption) explicitly asked the interviewer to put an X in a small box for each of the last six months. If the respondent made a large food purchase just after the last visit, but before the first month containing an X on the questionnaire, the interviewer excluded this purchase from the estimated amounts. The precise six-month reference period also applies to certain questions in the follow-up questionnaire, administered as part of Wave 4.

KHDS QUESTIONNAIRES

The several different types of questionnaires used in this study (and the waves in which they were administered) are the following[45]:

- Household Questionnaire (all waves)
- Community Questionnaire (all waves)
- Price Questionnaire (all waves)
- Health Facility Questionnaire (all waves)
- School Questionnaire (all waves)
- Healer Questionnaire (Wave 3 only)
- Follow-Up Questionnaire (Wave 4, only to fragments of households lost to the survey)

This section details the changes made from one wave to the next for each of the KHDS questionnaires. Because the household questionnaire was by far the longest and most important questionnaire, most of this section is devoted to that particular questionnaire.

1. Household Questionnaire

There are twenty sections in the household questionnaire.

---

[44] See Section II: Research Objectives of Documentation Report for clarification of Wave and Passage distinctions
[45] See Section V: KHDS Questionnaires of Documentation Report for description of each KHDS Questionnaire

SECTION 1: HOUSEHOLD ROSTER

Box A1.1 Major Changes to the Wave 2 Household Questionnaire

Important changes to the Wave 2 household questionnaire were in the following sections:

Section 1: Household Roster. Section 1B, "Location of Members Who Have Moved", was added.
Section 8: Migration. This section was asked only of new household members.
Section 9: Fertility. Continuing females over 50 years old were not asked this section again.
Section 16: Durable Goods, Annual Expenditures, and Income from Assistance Programs. Part A, on durable goods, was divided into two sections (16A1 and 16A2) in order to measure the changes in ownership of durable goods between Waves *1* and *2*.

Wave 2
At the beginning of waves 2, 3, and 4, interviewers used Section 1 to identify continuing household members, new household members, and household members from the previous wave who were no longer with the household.[46] Part A was administered to continuing and new household members and Part B was administered to persons no longer in the household. Part B was not asked for Wave 1 because, as the first wave for which a household roster was made, there was no point of reference.

For Waves 2, 3, and 4, Question 1A of Part A asks about persons on the roster from the previous wave: "Is..[NAME]..still normally living and eating with you in this dwelling?". If YES go to Question 2, if NO skip to the next person. When Question 1A had been asked of all persons from the previous wave's household roster, interviewers then proceeded to Part B, "Location of Members Who Have Moved" and asked four questions about all persons who were no longer with the household. The four questions in Part B follow.

PART B. Location of Members Who Have Moved

Question 1:
　　　"Has…[NAME]…moved away from this household?"
　　　　　　(IF NO, PROBE TO SEE IF THE PERSON DIED.)
　　　　　　YES, MOVED…1
　　　　　　NO, DIED……...2 (GO TO NEXT PERSON)
Question 2:
　　　"When did ..[NAME].. move away from your household?"
　　　　　　(LIST MONTH AND YEAR)

---

[46] Within the sub-sections for each questionnaire, the changes made between waves are detailed in this documentation in the sub-section for the latter wave. Within each sub-section, when nothing is mentioned for a particular wave this means that there were no changes to the questionnaire for that wave.

Question 3:

    "Where did ..[NAME].. move to? Was it ....."

Somewhere else in this place?	1
A village elsewhere in Kagera?	2
A town elsewhere in Kagera?	3
Dar es Salaam?	4
An urban area elsewhere in Tanzania?	5
A rural area elsewhere in Tanzania?	6
A foreign country?	7
DON'T KNOW	8

Question 4:

    "What was the main reason that ..[NAME].. moved away from your household?"

        MOST IMPORTANT REASON:

EMPLOYMENT ....……… ………...1
BUSINESS OPPORTUNITIES ...... ……………2
POSTED TO A NEW AREA ...…... ………..3
LAND NOT AVAILABLE……...…………4
SCHOOLING ……………….…..…………5
MARRIAGE ………………………………6
DIVORCE .…………………………. …………7
WIDOWHOOD ……………….…………..8
DEATH OF PARENTS….……………………9
ILLNESS OF HOUSEHOLD MEMBERS….10
ILLNESS OR DEATH OF FAMILY
MEMBER LIVING ELSEWHERE ……….11
OWN ILLNESS/TO OBTAIN
MEDICAL CARE.………………………….12
OTHER FAMILY PROBLEMS ……….…..13
POLITICAL/ECONOMIC PROBLEMS ….....14
NATURAL DISASTERS ……..……………15
OTHER   ………………………….…..16

Wave 3

Section 1B

A new code was added to Question 1 for persons who were household members in Wave 2, but who were non-members in Wave 3 and had not moved or died. For example, a household member may have been a makubaliano servant during Wave 2 and was a boarder in Wave 3. The new code was:

    STILL HERE, NOT A MEMBER ………   3
    (> NEXT PERSON)

SECTION 2: CHILDREN RESIDING ELSEWHERE

The purpose of this section was to trace the movement of children of household members into and out of the household as a result of adult mortality in the household or in other households outside the sample.

Wave 2

In Wave 2, the interviewers updated the roster of children living elsewhere. See Questions 2A-17. Question 2A asks for the status of children living elsewhere, and the possible responses were:

- CHILD STILL AWAY
- CHILD JOINED HOUSEHOLD
- CHILD DIED
- BOTH PARENTS NO LONGER IN HOUSEHOLD
- CHILD LEFT HOUSEHOLD SINCE WAVE 1
- CHILD OF NEW HH MEMBER
- CHILD BORN ELSEWHERE
- CHILD MISSED IN WAVE 1

SECTION 3: INFORMATION ON NON-RESIDENT PARENTS OF HOUSEHOLD MEMBERS

This section was relevant to the study because (1) living nonresident parents could be an important source of assistance to a household and (2) it helped determine if some household members were orphans.

Wave 2

As was the case during Wave 1, this section was asked about all household members. However, Questions 5A and 13A were added so that the questions on parents' education were not repeated for continuing household members. All other questions remained the same. Both new questions ask: "Is ...[NAME]... a new member of the household this wave?" When the answer was NO in Question 5A, interviewers skipped to Question 9 and when NO in Question 13A, interviewers skipped to Question 17.

SECTION 4: MAIN ACTIVITIES OF THE HOUSEHOLD

This short section helped guide the interviewer in the selection of respondents for subsequent sections of the questionnaire dealing with economic activities and expenditures (sections 7, 11, 12, 13, 14, and 18) .

Wave 2

The reference periods for all of the activities in Questions 1, 3, 5, 7, and 8 were changed from the past twelve months to the past six months (that is, since the Wave 1 interview).

Question 7 and 8: The Swahili translations of these questions were improved, based on interviewer recommendations from Wave 1.

SECTION 5: EDUCATION

Schooling Expenditures. The recording of some schooling expenditures was tricky for interviewers, so the same explanation that they had in their interviewer manual is included here.

There were four sets of questions in Section 5 where interviewers were asked to record information on schooling expenditures for a member of the household. They were:

- Question 18 Expenditures by household members for the student's schooling, in the past twelve months
- Questions 19-22 Value of sponsorships made by outside organizations for the student in the past twelve months.
- Questions 23-25 Expenditures by persons outside the household for the student's schooling in the past twelve months.
- Questions 26-28 Value of in-kind contributions to student's schooling from organizations in the past twelve months.

If a person (instead of an organization) gave the student an in-kind contribution (for example, an uncle in Nairobi sent textbooks as a contribution), then interviewers recorded that information in questions 23-25, on individual contributions from outside the household.

No expenditures were entered in more than one of these groups in Section 5. For example, interviewers were instructed not to enter the same expenditures in questions 19-22 as in questions 26-28.

The expenditures included those for the current school year and for the previous school year, provided that they fell within the previous twelve months. For example, if the interview took place in November of 1991, the school expenses for the first term of the 1991-92 school year were included, along with those for the second half of the 1990-91 school year.

Questions 18 and 25. The schooling expenditures made by household members in the twelve months previous to the interview for children living in the household who attended school were recorded in Question 18. The schooling expenditures made by other persons on the pupil's behalf were recorded in Question 25. In each of these questions, there are seven columns for recording itemized expenditures, and an eighth column for recording the total amount spent by household members in the previous twelve months. The rules for recording information were as follows:

1. If the amount spent on an item was known, it was recorded in the correct column (columns A-G).
2. If nothing was spent on an item, interviewers entered "0".
3. If the respondent could not itemize expenditures, the non-itemized amount was included in the total (column H) and the columns for itemizing were left blank.
4. Interviewers always wrote a TOTAL amount in column H. This represented the sum of all items A-G and all other non-itemized amounts.
5. If there was an expenditure on an item but the respondent didn't know the amount, interviewers wrote "DK" in the column. This item could not be included in the TOTAL column, since the respondent didn't know what the amount was.

EXAMPLE 1. The following expenditures were made by household members in the previous twelve months on behalf of a student with ID code 03: 300 Tsh for the UPE fund; 2,000 Tsh

(Tanzanian Schilling)for a uniform; 3,200 Tsh for books and school supplies; nothing for transportation to school; nothing for boarding and lodging; 8,000 Tsh for school fees; and 5,000, Tsh for extra tutoring. The answers to question 18 are:

A	B	C	D	E	F	G	H
UPE fund?	Uniforms/sports cl	Books/sch supplies	Transport	Board/Lodging	School fees	Other	TOTAL
300	2000	3200	0	0	8000	5000	18500

EXAMPLE 2. In another household, members spent the following for a student's school expenses (student is a household member): 300 Tsh for the UPE, 8000 Tsh for the school fees, and a total of 6500 Tsh on uniforms and school supplies (they cannot itemize the 6500 Tsh). The student lives at home, so the household paid no transportation costs and no lodging or board costs for him or her. The answers to question 18 are:

A	B	C	D	E	F	G	H
UPE fund?	Uniforms/sports cl	Books/sch supplies	Transport	Board/Lodging	School fees	Other	TOTAL
300			0	0	8000	0	14800

EXAMPLE 3. In a third household, the head gave the student 15,500 Tsh with which to pay all school expenses for the next school year. In this example, the respondent has not been able to itemize, so we write the total amount in column H.

A	B	C	D	E	F	G	H
UPE fund?	Uniforms/sports cl	Books/sch supplies	Transport	Board/Lodging	School fees	Other	TOTAL
							15500

Question 22. A SPONSORSHIP is a cash contribution to the schooling of a household member from an organization. The value of a sponsorship in the previous twelve months may include part of one or two school years. If the student in question received a sponsorship during the previous school year but did not receive one for the current year, interviewers were instructed to ask how much the monthly scholarship payments were and how many months in the previous twelve months that the sponsorship was received. From this they calculated the total.

If during the previous twelve months the person had two sponsorships of different amounts, the total amount received from each was calculated, taking into account the number of months in each case.

Other sections where interviewers recorded schooling expenditure

Interviewers were instructed not to count twice any expenditure *within* Section 5. However, there were two other places in the household questionnaire where interviewers recorded these expenditures again:

- Cash *sponsorships or in-kind contributions* for schooling from an organization or institution (questions 19-22 and 2629) were also recorded in section 16, part C, "Receipt of Assistance" (page 59).
- *Expenditures by persons outside the household for* the student's schooling were also recorded as an incoming gift or loan for the student in section 19, part A.

The expenditures made by household members for the student's schooling were NOT recorded a second time. NO EXPENDITURES FOR SCHOOLING WERE RECORDED IN SECTION 18A OR 18B.

EXAMPLES:

Item	*Where Recorded:*
• School uniform donated by Social Welfare Office	Section 5, questions 26-28 Section 16C, line 30
• Contribution to UPE fund by uncle living in Dar	Section 5, questions 23-25 Section 19A
• Exercise books paid for by a household member	Section 5, question 18C
• School lunches contributed by World Vision	Section 5, questions 26-28 Section 16C, line 33
• Textbooks purchased by a relative in Nairobi	Section 5, question 25C Section 19A

SECTION 6: HEALTH

This section was divided into two parts, Part A covering acute conditions (illness in the four weeks preceding the interview) and Part B covering chronic conditions (those that have existed for 6 or more months). Note that in the Wave 1 questionnaire in Ainsworth and others (1992), the page for Section 6B (page 64) was mistakenly put after the first page of Section 7 (page 63).

Wave 2
Because the reference period for Section 6A was the last four weeks and for Section 6B the past six months, there were no changes in the reference period for this section.

Questions 1, 4, and 6: The skip instructions were changed so that respondents skipped to Question 78 instead of going to Section 6B.
Questions 19-23 were reworded and the order of the questioning was changed so that the respondent could better understand the difference between in-patient consultations and out-patient consultations.

Question 19:

"How many times did you visit this establishment for this illness or injury when you were not hospitalized?"

IF NONE, WRITE ZERO AND GO TO 21.

Question 20:

"How much did you pay for all of these visits to these establishments for this illness or injury?

IN CASH AND IN KIND, EXCLUDE COST OF MEDICINES, IF FREE WRITE ZERO

Question 21:

"Did you have to spend a night in this establishment because of this illness or injury?"

YES…1; N0…2 (GO TO 24).

Question 22:

"How many nights?"

Question 23:

"How much have you paid or will you pay altogether for the stay at this establishment?"

Questions 31-35 and 43-47 were also re-ordered to ask for outpatient visits first.

Question 31:

"How many times did you visit this establishment for this illness or injury when you were not hospitalized?"

IF NONE, WRITE ZERO AND GO TO 33.

Question 32:

"How much did you pay for all of these visits to these establishments for this illness or injury?"

IN CASH AND IN KIND, EXCLUDE COST OF MEDICINES, IF FREE WRITE ZERO.

Question 33:

"Did you have to spend a night in this establishment because of this illness or injury?"

YES ...l; N0...2 (GO TO 36).

Question 34:

"How many nights?"

Question 35:

"How much have you paid or will you pay altogether for the stay at the establishment?"

Question 43:

"How many other times did you visit this establishment for this illness or injury when you were not hospitalized?

IF NONE, WRITE ZERO AND GO TO 45.

Question 44:

"How much did you pay for all of these visits to these establishments for this illness or injury?

IN CASH AND IN KIND, EXCLUDE COST OF MEDICINES, IF FREE WRITE ZERO.

Question 45:

"Did you have to spend a night in this establishment because of this illness or injury?

YES ...1; N0...2 (GO TO 48)."

Question 46:
> "How many nights?"

Question 47:
> "How much have you paid or will you pay altogether for the stay at this establishment?"

Questions 72 and 78 were added to Section 6A to find out how many times the respondent has been ill or injured in the past six months (since Wave 1).

Question 72:
> "In addition to these illnesses and injuries in the past 4 weeks, how many times have you been ill or injured in the last 6 months?
> > DO NOT COUNT ILLNESS/INJURY IN PAST 4 WEEKS

Question 78:
> "How many times have you been ill or injured in the last 6 months?"
> > GO TO SECTION 6B

Section 6C: General Health was was added to Waves 2, 3, and 4 (see copy of this section below).

Wave 3

Two new illness codes were added to Questions 65 and 66 of Section 6A and Question 4 of Section 6B:

> COMMON COLD   26
> INFLUENZA/FLU   27

Wave 4

Section 6A. In Question 68, the reference period was changed to the past six months.

SECTION 7: ACTIVITIES AND NON-LABOR INCOME

Section 7 collected information on economic activities performed by household members in order to reveal the breakdown of income or time across household members. The information that this section collected on income was less detailed than in other sections, notably the sections on farming (11), livestock (12), fishing (13), and non-farm self-employment (14).

Wave 2

All twelve-month reference periods in this section were changed to six-month reference periods (or since Wave 1). Questions 19 and 35 in section 7D, which were blank in Wave 1, were dropped. The questions that followed the dropped questions were not renumbered.

Wave 3

Section *7C.* New Question 8A asks: "How many of these hours (listed in Question 8) were spent working on land that is collectively owned by your age group or initiation group? PROBE FOR ACTUAL HOURS EACH DAY."

Section *7E.* Three new questions were added to Part E - Questions 13A, 13B, and 17. Questions 12 and 13 ask whether the respondent had spent any time caring for or visiting with a sick *household member.* New Questions 13A and 13B ask the same for non-household members.

Question 13A:

"In the past seven days, have you spent any time caring for sick persons who are not household members?"

Question 13B:

"How many hours did you spend each day caring for sick persons who are not household members?"

Questions 16 and 17 ask about the amount of time that the respondent spent helping neighbors or relatives, who were not household members, with work on their shambas, without receiving payment. Question 17A was added.

Question 17A:

"How many of these hours (mentioned in Question 17) were spent helping members of your age group or initiation group?"

Question 17A is the second example of intentionally double-counting the time use of the respondent. This means that the answers to Question 17A must always be smaller than or equal to the answers to Question 17.

Wave 4

Section *7B*. For Question 12, code 2, interviewers were instructed to skip to new Question 13A.

Question 13A:

"How many days in the past six months were you unable to work at this job due to illness?"

Interviewers wrote the number of days that the person was employed at the *job* but was unable to work. The maximum number of days that could appear as an answer to this question was 180.

Question 13B:

"For how many of these days that you were unable to work did you get no pay?"

The answer to Question 13B could not be larger than the answer to Question 13A.

Question 23A:

"Is the salary that you have just cited net or gross of taxes?"

Question 23B:

"How much is subtracted from your salary for taxes?"

Section 7F. Questions 11A and 11B were added. Question 11A asks how many days the person was unable to work and 11B asks how many of these were without pay. The wording of these questions is identical to Questions 13A and 13B in Section 7B, respectively.

Questions 21A and 21B were added. Question 21A asks if salary was net or gross and 21B asks how much was deducted. The wording of these questions is identical to Questions 23A and 23B in Section 7B, respectively.

SECTION 8: MIGRATION

This section established the length of time each household member had been residing in the current place of residence and the circumstances surrounding the last move. The first three

sections of the household questionnaire also asked questions on migration. For all waves, none of the questions in this section changed. However, the *respondents changed* because this section was only administered to new household members. Continuing household members were not asked the migration questions again.

SECTION 9: FERTILITY

The objectives of this section were to determine: (1) the number of children ever born to every female household member who responded in Wave 1 age 15 and older; (2) child mortality; (3) the level of schooling attained by deceased children; (4) current pregnancy status; (5) fetal wastage; and (6) contraceptive use.

Wave 2
The number of female respondents for this section was reduced. Respondents for Wave 2 were:
> 1. ALL NEW FEMALE HOUSEHOLD MEMBERS 15 AND OLDER (AND MARRIED 14 YEAR-OLDS).
> 2. ALL CONTINUING FEMALE HOUSEHOLD MEMBERS 15 TO 50 YEARS OLD (AND MARRIED 14 YEAR-OLDS)

Unfortunately, this was incorrectly listed as:

> 1. ALL NEW FEMALE HOUSEHOLD MEMBERS
> 2. ALL CONTINUING FEMALE MEMBERS 15 TO 50.

Continuing household members over 50 years old were not asked these questions again.

Question 19 was reworded to ask whether the respondent has already lived "with a husband or boyfriend". In the Wave 1 questionnaire the wording was "with a man".

Wave 3
Because potential respondents for this section were incorrectly identified in the Wave 2 questionnaires, the Wave 3 questionnaire was corrected to read:

1. ALL NEW FEMALE HOUSEHOLD MEMBERS 14 AND OVER
2. ALL CONTINUING FEMALE MEMBERS 1450.

SECTION 10: ANTHROPOMETRICS

Anthropometric and immunization data were collected in this section as objective indicators of health and nutritional status. All household members were to be weighed and measured regardless of their age.

Wave 2
The instructions and procedures were the same as for Wave 1.

Question 10: the coding for the number of immunizations was changed to:

       NONE   0
       ONE   1
       TWO   2
       THREE   3

Respondents no longer answered YES or NO, but stated how many times they had been immunized for a particular disease.

Passage 3 and Passage 4

Digital scales were introduced for data collected during these two passages, therefore affecting comparisons of data across Waves 1-4.[47]

SECTION 11: FARMING

This section collected information on: (1) annual net income due to cultivation of crops; (2) the number and value of farm assets (land and equipment); and (3) the quantity of crops sold.

Wave 2
All twelve-month reference periods in this section were changed to six-month reference periods. Income from all crops or products sold by all household members in the past six months should be included, even if a member left or died between waves.

Introduction page. Questions D and E were added to find out if the household had sold any shambas or gardens in the past six months (Question D) and, if so, how much was received (Question E). Questions F, G, and H were added to find out if anyone in the household lost shambas or gardens from disinheritance. Question G asks about the number of shambas or gardens that were disinherited and Question H asks about the value of these lands.

Question D:
     "Have the members of your household sold any shambas or gardens since my visit six
     months ago?"
          YES ...1, N0...2 (GO TO QUESTION F)
Question E:
     "How much did the household receive from the sale of shambas or gardens since my visit
     six months ago?"
          AMOUNT:...

---

[47] The measurement of weight for household members changes between Passage 2 and Passage 3 due to a change in measurement tools (i.e. digital scales) used. Therefore, changes in individual weights over the survey period are affected by the change in scales, yet relative comparisons across individuals within a passage are not affected by this change since all individuals were weighed using the same scale during each passage.

Question F:

"Have the members of your household disinherited any shambas or gardens since my last visit 6 months ago?"

YES ...1, N0...2 (GO TO PART A)

Question G:

"How many shambas or gardens were disinherited?"

SHAMBAS:..., GARDENS:...

Question H:

"If you had wanted to sell this/these shamba(s) or garden(s), how much would you have gotten for them?"

AMOUNT:...

Section 11A. Code 2 of Question 5 ("BOUGHT ON CREDIT") was dropped and code 1 was altered so it only says, "BOUGHT" instead of "BOUGHT WITH CASH". Questions 5A and 5B were added.

Question 5A:

"Was this shamba bought on credit?"

YES ...1, N0...2 (GO TO 6)

Question 5B:

"How much money is left to repay?"

IF IN KIND, CITE VALUE; IF COMPLETELY REPAID, WRITE 0."

Section 11D. Questions 14, 15, and 16 were added to find out if the household owed any money for farm inputs obtained on credit or for advances that it received (Question 14) and, if so, how much was owed by the household on the day of the interview (Question 15).

Question 14:

"DID THE HOUSEHOLD PURCHASE ANY INPUTS?

CHECK THE ANSWERS TO Q1, 4, 6, 8, 10, 12..

YES ...1, N0...2 (GO TO PART E)"

Question 15:

"Does your household currently owe money for any of these farm inputs or for advances that must be repaid?

YES ...1, N0...2 (GO TO PART E)"

Question 16:

"How much does your household owe for all farminputs or advances? AMOUNT:..."

Section 11F.   HAND TOOLS.

This section was re-organized so that the interviewers could obtain an estimate of the value of all hand tools owned by the household in each category.

A. How many of the following tools are owned by members of your household? WRITE THE NUMBER OF EACH TOOL IN THE BOX. IF NONE, WRITE 0. ASK QUESTION A FOR ALL TOOLS BEFORE ASKING QUESTION B.	B. For how much could you sell all of these tools today? IF ANSWER TO QUESTION A IS ZERO, WRITE 0 IN QUESTION B. AMOUNT	
1. Hoes?		
2. Axes?		
3. Machetes?		
4. Picks ?		
5. Shovels ?		
6. Wheel-barrows ?		
7. Sickles?		
8. Pangas?		
9. Mundu?		
10. Pruning shears?		
11. Others?		

Section 11G. A new line (10) was added to ask about farm buildings that were owned or had been bought or sold since Wave 1. Other lines were added for equipment used in processing homeproduced crops for sale. These pieces of equipment included: grinders (item 11); banana beer "boats" (item 12); gericans and drums (item 13); and other equipment for processing crops (item 14) .

Wave 3
Section 11A. Question 16 was added to indicate whether each shamba owned during the previous six months was still owned by the household.

Question 16:
        "Does a member of your household still own this shamba?"
                YES ...1, N0...2, GO TO NEXT SHAMBA.

SECTION 12: LIVESTOCK

The objective of this section was to assess: (1) the number and value of livestock owned by the household; and (2) household income in the past twelve months due to livestock activities.

For Waves 2 through 4, the twelve-month reference period was changed to a six-month reference period.

Wave 2
Section 12A. Question 2 was slightly altered.

Question 2:

"Do any members of your household own or are they raising ..[ ].. at the present time?"

The interviewer instructions were changed in Questions 11 and 13 to account for the possibility that animals were inherited or disinherited.

Question 11:"NUMBER BORN, INHERITED OR RECEIVED"
Question 13:"NUMBER LOST, STOLEN, DIED OR DISINHERITED"

Wave 3
Questions 1 and 2 were altered to ask only about animals *owned by* members of the household in the previous six months (or since the last wave). In the previous wave these questions asked about animals that were raised or owned in the previous six months.

SECTION 13: FISHING

This section measured the value of assets of fishermen and their annual income. The only change was that after Wave 1 the reference period was changed to the past six months.

SECTION 14: NON-FARM SELF-EMPLOYMENT

This section measured the net income and value of assets for small businesses owned or operated by household members.

Wave 2
The reference period was changed to the past six months, and the "since-my-last-visit" reference period was changed to "two weeks ago" (referring to round one interviews for Wave 2).

SECTION 15: HOUSING

This section measured: (1) the value of housing assets; (2) expenditures on housing, water, electricity, and other utilities; and (3) the physical condition of the housing, which is a direct measure of well-being.

Wave 2
Section 15A. Question 12A was added.

Question 12A:

"Was this dwelling acquired in the past 6 months?"
YES ...1, N0...2

Section 15B. In Questions 5 and 9, the reference period was changed to six months. Questions 11A and 11B were added to find out if the household had disinherited a dwelling or building in the past six months, and to determine the value of the disinherited buildings.

Wave 3
The changes for this wave were in the removable Part C. In Question 5 of Section II, researchers expanded the list of diagnoses for inpatient attendances to include the following: 10-pneumonia; 11-other respiratory illness; 15-complications of pregnancy; 09accidents and injuries; 12-intestinal parasites/worms; 13-measles; 14-neonatal conditions; and 16other.

Wave 4
Question 11A:
> "Did anyone in the household disinherit a dwelling in the past 6 months?"
> > YES ...1, N0...2 (GO TO 12)

Question 11B:
> "If you had wanted to sell this dwelling, how much would you have received?"
> > AMOUNT:...

SECTION 16: DURABLE GOODS, HOUSEHOLD EXPENDITURES AND ASSISTANCE

The objective of this section was to collect information on the following: (1) the value of durable consumer goods owned by the household (16A); (2) expenditures on infrequently-purchased items in the past twelve months (1613); and (3) receipt of cash or in-kind assistance from community organizations (16C).

Wave 2
Section 16A. One of the main objectives of the Wave 2 household questionnaire was to measure the changes in ownership of durable goods between Waves 1 and 2. To accomplish this, Section 16, Part A, on durable goods, was divided into two parts:

Section 16A1, which asks about the disposition of the durable goods owned by the household during Wave 1.

Section 16A2, which asks about other durable goods owned by the household or acquired since Wave 1.

Section 16A1 was not included in the Wave 2 main questionnaire. It was produced at the end of Wave 1 by the data entry program, and listed the most important durable goods owned by the household members during that wave. Section 16A1 was unique for every household, depending on what they owned during Wave 1 (see example of this section below). So that respondents' answers to Question 5 were not influenced, interviewers were instructed not to remind the respondent of the durable goods described in Question 1.

Double-counting income from the sale of durable goods.
Section 16A1 was not the only part of the questionnaire that asked about the sale of durable goods. Each household member 15 and older was also asked whether they received income from the sale of durable goods in Section 7H, Questions 8A and 813. Income from the sale of durable goods was recorded in both Section 7H and Section 16A1.

Section 16A2 is similar to Section 16A of Wave 1, except that it asks about all durable goods acquired since Wave 1 (that is, in the past six months) and any other durable goods not mentioned in Section 16A1. However, pots and pans, furniture, and lanterns no longer appeared on the list of goods. Interviewers were instructed not to ask about these items during Wave 2. Income from the sale of these goods was recorded in Section 7H, and purchase of these goods since Wave 1 was recorded in Section 16B.

Sections 16B and 16C. The reference period for Sections 16B and 16C was changed to six months. In Section 16C, Question 1, assistance from the KHDS has been added to the list of organizations giving assistance.

Wave 3
Section 16B.

Section 16D. Question 4 was added: "How many of these mandays (from the response to Question 3) were contributed by members of age groups of household members?" The answer to Question 4 must be less than or equal to the answer to Question 3.

Wave 4
Section 16A1. Question 6 was added to the printout of durable goods from the previous waves. It asks, for all goods that were not mistakenly on the list from a previous round:

"How was this ...[ITEM]... originally acquired?"
BOUGHT ………………..…….….1
INHERITED………….………….2
GIFT ……………….…....3
OTHER (Specify) ………….….….4

The purpose of this question was to find out how the household acquired each good that they currently owned in Wave 4 or had owned in the previous wave.

Section 16A2. In Question 1, two new durable goods were added to the list on the left side of the page:
Watches (saa) . ………….............…………...56
Jewelry (pete, heleni, mikufu, bangili, shanga, n.k)....57

If an item was acquired in the last six months, the new Question 6 in section 16A1 did not record how it was acquired. Therefore a question was placed after Question 4 concerning recently acquired items. New Question 4A asks:
"How was this ...[ITEM]... acquired?"
BOUGHT …………………. 1
INHERITED……………..... 2
GIFT/OTHER …………...…3

SECTION 17: FOOD CONSUMPTION

Section 17 was conceptually the most difficult part of the KHDS questionnaire to design. The objective was to collect information on habitual and recent food consumption, and on this construct an annual measure of food consumption.

Section 17 explicitly asks the interviewer to put an X in a small box for each of the last six months. If the respondent made a large food purchase just after the last visit, but before the first month containing an X on the questionnaire, the interviewer excluded this purchase from the estimated amounts.

Wave 2
The only changes to Wave 2 are that twelve-month reference periods were changed to six-month reference periods, and that references to "my last visit" were changed to "two weeks ago".

Wave 3
Throughout this section, the TIME UNIT codes were expanded to include a new code (6 MONTHS ...7)

SECTION 18: EXPENDITURES BY HOUSEHOLD MEMBERS

Section 18A collects expenditures and acquisitions on behalf of individual household members in the past twelve months, while Section 18B collects information on personal expenditures by individual household members since round one (or about two weeks earlier). All twelve-month reference periods were changed to six-month reference periods after Wave 1.

SECTION 19: REMITTANCES AND CREDIT

The objective of this section was to measure transfers and credit flowing into and out of the household, the reasons for these arrangements, and the level of household savings. All twelve-month reference periods were changed to six-month reference periods after Wave 1.

Wave 2
Sections 19A and 19B. In Questions 4, 15, and 26, an instruction was added: "OR 99 IF CHILD IS DECEASED". This instructed interviewers to use 99 for the ID code of children who sent remittances but had since died. There was also a global skip instruction to go to Question 7, 18, and 29, respectively, from these questions.

Wave 3
Section 19C. For Question 1, during waves 1 and 2, many persons answered "NO" because they didn't belong to a , bujunill or "upatu". However, these were merely examples - researchers wanted to know if the respondent belonged to any informal savings organization. Because these examples seemed to be affecting answers, they were removed from the question.

In Questions 2 and 5, a new code (6 MONTHS ...7) was added to the time unit codes.

Wave 4
Section *19A*. To identify items received by inheritance, a new code (INHERITED ...10) was added to Questions 9, 20, and 31.

SECTION 20: MORTALITY

This objective of this section was to measure: (1) all deaths in the past twenty-four months and their causes; (2) the health seeking behavior of persons who died in the past twelve months; and (3) household expenditures connected with mortality in the past twelve months. After Wave 1, the reference period for deaths was changed from twenty-four months to the past six months (or since the previous wave).

Wave 2
Question 3A of Section 20A was added to record the ID code of the deceased household member.

Question 3A:
"ID CODE OF DECEASED; IF PERSON WAS NEVER CLASSIFIED AS A HH MEMBER BUT DIED IN THE HOUSEHOLD, USE ID CODE 98."

Two questions that provided instructions to the interviewer - Question 29 of section 20A and Question 24 of section 20B were dropped.

If a household had disintegrated and members died, only Section 20A was completed and the interviewer had to find respondents in the neighborhood to help them answer the questions.

Wave 3
Section 20A. In Question 4, a new code was added:
NA/HOUSEHOLD DISINTEGRATED ...16.

In Question 31, a new Part B was added:
"Was this traditional or modern treatment?"
MODERN .... 1, TRADITIONAL ....2.
The former Part B became Part C. In Part C, interviewers were asked to probe for all payments in cash *and in kind.*

In Questions 41 and 42 two new codes were added to the illness codes:
COMMON COLD   26
INFLUENZA/FLU   27

Wave 4
The rule applied by the interviewers for determining which deaths were recorded in 20A and which in 20B was to include a death in section 20A in any of three situations: (a) the person was a household member at the time of death; (b) the person had come to live in the household

shortly before dying; and (c) the person died elsewhere but was shipped as a corpse to the household for burial. Thus the only deaths recorded in section 20B have been those of relatives of household members who died *and were buried* away from the household.

In section 20A, the practice of including the deaths of individuals who died elsewhere but were buried by the household was continued in Wave 4. However, an additional code was added to Question 3A in order to identify these deaths. Question 3A was altered to read:

3A. ID CODE OF DECEASED. USE ONE OF THE FOLLOWING CODES FOR SPECIAL CASES:

      LIVING WITH HH AT TIME OF DEATH, BUT NEVER CLASSIFIED AS
      HOUSEHOLD MEMBER……………………………………………….98

      NON-MEMBER, DIED ELSEWHERE, BURIED BY HH IN CLUSTER…..97

2. Community Questionnaire

The Community Questionnaire had six sections. It was administered to the community leaders in each of the forty-nine sample communities in order to gain community cooperation with the interview team and also to elicit some community-wide information.

Researchers dropped many questions in the Community Questionnaire after Wave 1 because the answers to these questions were unlikely to change between waves (for example, section *2* of the Wave 1 questionnaire had fifty-three questions while in Wave *2* it had only twenty-five questions). References period of twelve months or longer were reduced to six months for Waves 2 through 4.

SECTION 1: DEMOGRAPHIC INFORMATION

This section was almost entirely changed after Wave 1 - the only repeat question is Question 1, "How many people are living in this community." Questions 2 through 7 of Wave 1 were dropped, and were replaced by eleven questions on migration, births, and deaths since Wave 1 (see Wave *2,* Section 1 on following pages - pp. 4 and 5 of Wave *2* questionnaire - or copy from diskette).

SECTION 2: ECONOMY AND INFRASTRUCTURE

Wave 2
Questions 1, 4-13, 15-18, 24-36, and 44 from Wave 1 were dropped. Of the fifty-three questions in Wave 1, only twenty five were used in Wave 2. The reference period for questions throughout this section was changed to six months.

SECTION 3: EDUCATION

Wave 2
As suggested above, non-changing information gathered in Wave 1 was not gathered again. Therefore, 19 questions were dropped in this section for this wave. The only questions that were kept were numbers 1-7 and 2628 (the latter three were renumbered as 8-10).

SECTION 4: HEALTH

Wave 2
Questions 15-17 (about traditional birth attendants) and *20-22* were dropped, leaving eighteen questions of the original twenty-four.

Wave 3
Question *13. A* new column was added to record the gender of the healer (male ... 1, female ... *2).*

SECTION 5: AGRICULTURE

Wave 2
Only sixteen questions of the original twenty-five were used in Wave 2. Questions 2-4, 13-16, 24, and 25 from Wave 1 were dropped.

SECTION 6: CULTURE

Wave 2
Questions 2, 7, and 9 from Wave 1 were dropped. The reference period for Wave 2 Questions 1, 4, and 7-10 was changed to six months.

Wave 3
Question 5. This question was split into two - Questions 5A and 5B. Question 5A asks: "How long is the traditional mourning period for ...[ ]... in this community?". Question 5B asks: "How long is the actual mourning period in this community for...[ ]...?". The unit codes and the categories (middle aged adult, child, old person, and infant) remained the same as in Wave 2.

Question 6. Item (d) was added: "Can the widow be inherited as a wife by the brother of the deceased".

3. Health Facility Questionnaire

For an explanation of how the Health Facility Questionnaire was administered, see pages 64-68 in the Wave 1 interviewer's manual.

Part A of the questionnaire, administered to the medical person in charge, has nine sections:

I.   Characteristics of the Facility
II.   Personnel
III.   Equipment
IV.   Services
V.   Immunizations
VI.   Family Planning
VII.   In-patient Services
VIII. Demand
IX.   Exemptions

Part B asked the pharmacist of the facility about the availability of drugs at the facility.

Part C was left with the medical officer in charge, so that he or she could consult official documents. The two sections are outpatient consultations and in-patient services.

SECTION 9: ASSISTANCE

Wave 2
A new section called "Assistance" (page 12, part IX) was added to Part A of the health facility questionnaire for Wave 2 (see attached copy of this section). The "Exemptions" section was then renumbered to be Section X. The reference period for Part B, Question 11 was changed to six months.

Wave 3
The changes for this wave were in the removable Part C. In Question 5 of Section II, researchers expanded the list of diagnoses for inpatient attendances to include the following: 09-accidents and injuries; 10-pneumonia; 11-other respiratory illness; 12-intestinal parasites/worms; 13-measles; 14-neonatal conditions; 15-complications of pregnancy; and 16-other.

Wave 4
Part B. Question 12 was added to the end of Part B of the Health Facility Questionnaire, (page 17):

"For how many days in the past six months has this facility been completely without:
A. Analgesics?
B. Antibiotics?
C. Anti-malarial drugs?
D. Anti-helminth drugs?"

The response to this question could not exceed 180 days. *Analgesics* are pain-killers, like aspirin and paracetamol. *Antihelminth drugs* kill various parasites, like worms. This question is different from Question 4 of Part B because it does not mention individual drugs. For analgesics, the respondent was asked to provide the number of days that the facility had no analgesics whatsoever - not just those listed in Part B, Question 1.

Part C. The response categories were reduced in the response tables for Sub-Section I (Out-Patient Consultations), Question 3 and Sub-Section II (In-Patient Services), Question 5. There are now only five categories in which consultations were recorded:

      MALE CHILDREN <15
      MALE ADULTS 15+
      FEMALE CHILDREN <15
      FEMALE ADULTS 15+
      TOTAL

4. School Questionnaire

For a detailed explanation of how the Primary School Questionnaire was administered, see the Interviewer Manual. The school questionnaire was completed for every primary school in a cluster. When there was no school in the cluster, a school questionnaire was completed for the primary school nearest the cluster.

Part A has three sections: I. Characteristics; II. Enrollments; and III. Fees.
Part B was left with the headmaster or head teacher of the school so they could refer to school records to answer the questions.

Wave 2
The codes for grades P8 and beyond were dropped throughout the questionnaire (specifically, in Part A, Section I, Question 3; Part A, Section III, Question 2; and Part B, Question 1 and 2).

Part B. Information on the number of orphans who have lost two parents (Question 2) was collected by gender - that is, for boys and girls separately.

Part B. Question 3 was added:
      "How many teachers are employed at this school with the following certificates?"
            GRADE A...; GRADE B...; GRADE C...; GRADE D...; OTHER...;
            TOTAL TEACHERS...

Part B. Questions 4 through 6 were add, asking about the types, sources, and value of assistance received in the past six months from community organizations.

Wave 4
Part A. Section I. The skip instruction Question 7 was deleted.

Part A. Section II.
Former Question 2 was replaced with Questions 2A, 2B, and 2C:
Question 2A:
      "Why does the school not admit some children?"
            AGE……………………………………………1
            NOT ENOUGH SPACE………………….…2
            CHILD NOT RESIDENT IN VILLAGE ….....…3

          BAD BEHAVIOR...........................………......4
          POOR GRADES...................………................5
          STUDENT CAN'T PAY………………….....6
          OTHER (Specify)…………………………...…7

Question 2B:

    "Is there a waiting list for admission to this school?"
        YES   1,   NO   2 (>3)

Question 2C:

    "How many children are on the waiting list?"

Part B. Questions 2A and 2B were added:
    "2A.   How many children completed P7 during the last academic year?"
    "2B.   How many of the P7 graduates gained admission to secondary school?"

5. Price Questionnaire

The price questionnaire did not change in any of the passages.

6. Healer Questionnaire

This questionnaire was administered only in Passage 3.

7. Follow-Up Questionnaire

During each of the first three waves of the KHDS, several households in the sample were replaced because the household moved or the head died and the household disintegrated. The objective of the follow-up survey was to obtain information on the well-being of individuals who were members of these "replaced" households.:

The Follow-Up Questionnaire, a subset of the Wave 4 questionnaire, was designed to be administered to fragments of households that had moved between Waves 1 and 4 and therefore had been lost to the household surveys. [48] The Follow-Up Questionnaire differs from the Wave 4 Household Questionnaire because it omits all the household level sections, since most of the respondents are no longer in the same household in which they began the survey. This special questionnaire includes all or portions of sections 1, 3, 5, 6, 7, 9, 10, 18, and 19. These sections collect information on parents, schooling, health, activities and time use, fertility outcomes, anthropometry, expenditures, and remittances.

Section 1, the household roster, was modified to allow the respondent to identify the other members of his or her current household. In addition, the following five pages were added:

1. Section OOC: Survey information sheet;
2. Section OOD: Identification of the new household;

---

[48] Households that moved intact received the entire Wave 4 questionnaire even if they had never received Waves 2 or 3.

3. Section 1C: Demographic characteristics of the respondent;
4. Section 1D: Economic characteristics of the respondent;
5. Section lE: Individual assets.

The *individual* was the unit of observation for the follow-up questionnaire. However, the questionnaire was designed so that information on all of the individuals in the former household could be collected with the same questionnaire, even if they were currently residing in different households. Information about the respondents was collected about their current households, not the household to which they previously belonged.

Reference period. As explained earlier, for the other KHDS questionnaires the six-month reference periods in Waves 2, 3, and 4 usually meant "since my last visit." For the Follow-Up Questionnaire, however, researchers wanted to compare households to one another, even if one had last been seen six months ago and another seventeen months ago. Therefore the six-month reference period for Sections 3, 5, 6, 7, 10, 18, and 19 in this questionnaire means "since exactly six months ago..." and the exact dates were noted by the interviewers. Section 9 on fertility was the only exception to this rule - in this case the reference period was for the entire period of time since this section was last completed for the respondent.

Other members of the respondent's current household. other members of the new household to which former KHDS household members moved were not interviewed. The Follow-Up Questionnaire was only for former KHDS household members.

APPENDIX II: HOUSEHOLD WEIGHTS

The results from the KHDS can be extrapolated to the whole population by weighting the result for each household by that household's probability of being selected out of the whole population. Sample weights for households are includes in the data set key___hh. The weights are calculated as a function of two processes: the PSU being selected from the sample of PSUs in the region and the household being selected among all the households in the PSU once the PSU is selected.

- Let P_1 represent the probability of selecting a given primary sampling unit out of the set of all PSUs in the region. See Section III for details about P_1, the probability of selecting the PSU.   There are 49 different values for P_1.

- Let P_2 represent the probability of selecting a household, given that its primary sampling unit has been selected. Because of the stratification of the households within a PSU by their morbidity and mortality status, households within the same PSU can have different weights. There is a common weight for households that were categorized as sick and a common weight for households that were categorized as well based on the enumeration survey. Thus, there are a total of 98 different values for P_2: two values of P_2 for each of the 49 PSUs in the sample.

The sampling weight to be applied to each household is given by:

$$W \; = \; \frac{1}{P_1 \; \bullet \; P_2}$$

The weight on a given household can be interpreted as the number of households in the entire population that are represented by this particular sample household. ***Sample weights were constructed only for 816 households in the first passage.***[49]

Table AII.1 shows the average sample weight for sick and well households in each of the 8 strata. Because high-mortality communities were over-sampled, the household weights in the high mortality communities are smaller than those in the low-mortality communities by a factor of 5 to10. Over-sampling of the "sick" households within a primary sampling unit makes the weights on sick households smaller than those on "well" households in the same cell, by a factor of as much as 100.

---

[49] Of the 816 "original" households interviewed in the first passage, 47 were in fact replacements necessitated by the move or refusal of one of the 16 households on the list provided by the researchers. This complication was ignored in the computation of the sample weights.

Table AII.1: Mean Sample Weight for "Well" and "Sick" Households by Agronomic Zone and 1988 Adult Mortality Rate

		Adult Mortality Rate, 1988		
Agronomic Zone	**Household type**	**Low**	**High**	**Total**
Tree Crop	Well	4,509	429	2,539
	Sick	63	6	37
Riverine	Well	9,161	615	4,111
	Sick	59	9	50
Annual Crop	Well	3,347	598	2,106
	Sick	6	6	21
Urban	Well	1,612	179	775
	Sick	7	2	10
Total	Well	4,262	443	2,283
	Sick	54	6	30

APPENDIX III: PRICE INDEX

A. METHODOLOGY ADOPTED TO CONSTRUCT PRICE INDICES

We used quantity weighted Laspeyre's price index using the following formula:

$$I_{tv} = \sum_{i=1}^{M} w_i \left(\frac{P_i^{tv}}{P_i^{00}} \right)$$

$$where \quad w_i = \frac{P_i^{00} Q_i^{00}}{\sum_{i=1}^{M} P_i^{00} Q_i^{00}}$$

I_{TV} is the price index in village V in time T; w_i is the mean share of expenditure in total expenditure for each good in wave 1 of the survey (P_i^{00} and Q_i^{00} are price and quantity in wave 1, which in this case would be the amount spent on the item), and the sum of these weights would equal 1. The term in parenthesis in the first equation is the relative price ratio of good i in village v and time t. This method yields a price index for each village in each wave.

The above procedure involved the following steps:

(1) A matching of goods in the price questionnaire with goods that have expenditure reported in the household questionnaire. Unfortunately, not all the commodities in the household questionnaire had matching prices in the price questionnaire, so the weights had to be adjusted upwards so that they add up to 1. The implicit assumption here is that the variation in prices of goods for which we do not have prices is the same as that observed for commodities for which we do have price information.

(2) Since many of the prices were missing, we had to impute the missing prices using a model that attempted to explain price variation across villages and time.

(3) The next step involved calculating the relative prices using the average price in Kagera region in wave 1 as the base.

(4) Finally, the ratios of the relative prices were multiplied by the corresponding expenditure weights for each commodity and summed over goods to arrive at the price index for each cluster.

We explain these steps in detail below after presenting a brief overview of the KHDS questionnaire.

B. THE KHDS PRICE QUESTIONNAIRE & CALCULATION OF THE PRICE INDICES

The survey collected price data independently of the household survey, which is very useful since it removes some of the heterogeneity in the quality of goods implicit in the unit values derived from household surveys. However, the disadvantage is that there is no simple one-to-one matching of the categories used in the price questionnaire and the household questionnaire, a problem which is common to most surveys. To build spatial price indices, the commodities mentioned in the household questionnaire had to be matched with the commodities that had prices in the price questionnaire.

Table AIII.1 is a list of food items listed in the household questionnaire. A blank indicates it has a price in the price questionnaire and N indicates no price. For three items, yams, citrus fruits and wheat flour the matches were not exact, but we used these prices as very close approximations. The household questionnaire had a total of 58 food items which matched 26 items in the price questionnaire, and if we add the five non-food items (firewood, charcoal, linen, battery and kerosene) mentioned in the household questionnaire that had prices in the price questionnaire, it gives us a total of 31 items.[50]

The KHDS price questionnaire collected information on 48 goods from shops and markets in each cluster/village. Out of these 48 goods 31 goods had expenditure information in the household questionnaire[51].

---

[50] Expenditure on all the food items in the household questionnaire were reported in section 17B, C1 or C2. Expenditures on firewood, charcoal and battery were reported in section 15, question 27; Expenditure on linen and battery were reported in sections 18A and 18B respectively.

[51] The goods in the price questionnaire that did not appear in the household questionnaire are: Maize flour, Tea leaves, Peas, Aspirin, Paracetemol, Nivaquine, Liver Salt, Milk of Magnesia, Iron coal pot, Hurricane lamp, Mini-kerosene, Matches, Bar soap, Machete, Metal bucket and Plastic bucket.

Table AIII.1: Match between items in KHDS household
questionnaire and price questionnaire

Item from HOUSEHOLD QUESTIONNAIRE	Status on price questionnaire	Item from household questionnaire	Status on price questionnaire
Cooking Banana	On price questionnaire	Pumpkin	Not on price questionnaire
Sweet Banana	On price questionnaire	Cabbage	Not on price questionnaire
Other Banana	Not on price questionnaire	Other Vagatables	Not on price questionnaire
Cassava	On price questionnaire	Spices	Not on price questionnaire
Dry Cassava	On price questionnaire	Fish	On price questionnaire
Other Cassava	On price questionnaire	Edible Insects	Not on price questionnaire
Yams	Approximate price	Eggs	On price questionnaire
Irish Potatoe	On price questionnaire	Alcohol	On price questionnaire
Maize	Not on price questionnaire	Milk	On price questionnaire
Millet	On price questionnaire	Milk Products	Not on price questionnaire
Rice	On price questionnaire	Other Food	Not on price questionnaire
Beans	On price questionnaire	Wheat Flour	Approximate price
Ground Nut	On price questionnaire	Macaroni	Not on price questionnaire
Sunflower Seeds	Not on price questionnaire	Biscuits	Not on price questionnaire
Mamba nuts	Not on price questionnaire	Other Oil	Not on price questionnaire
Palm Oil	On price questionnaire	Jams	Not on price questionnaire
Butter	Not on price questionnaire	Salt	On price questionnaire
Margarine	Not on price questionnaire	Poultry	On price questionnaire
Avocado	Not on price questionnaire	Beef	On price questionnaire
Mangoes	Not on price questionnaire	Mutton	On price questionnaire
Paw Paw	Not on price questionnaire	Pork	Not on price questionnaire
Citrus Fruits	Approximate price	Baby Food	Not on price questionnaire
Pineapple	Not on price questionnaire	Sodas	Not on price questionnaire
Other Fruits	Not on price questionnaire	Alcohol	Not on price questionnaire
Fruit Juice	Not on price questionnaire	Yeast	Not on price questionnaire
Sugar	On price questionnaire	Milk Powder	On price questionnaire
Tomatoes	On price questionnaire	Tinned Meat	Not on price questionnaire
Onions	On price questionnaire	Other prepared food	Not on price questionnaire
Eggplant	Not on price questionnaire	Other Food	Not on price questionnaire

Table AIII.2 below lists the 31 goods and the percentage of clusters that had reported a price, along with the average weight (over clusters) of each good in total expenditure. About 58% of total mean expenditure was accounted for by these 31 goods, and the shares reported in table AIII.2 were scaled upwards to reflect the rest of the 42% so that the sum of the weights add up to 1.[52]

Table AIII.2: Reported Prices And Expenditure Weights By Good Type

	% Reporting Price	Expenditure Weights
Raw Cassava	34	2.5
Dry Cassava	17	2.6
Cassava Flour	11	4.3
Millet	17	1.7
White Bread	13	9.3
Rice	59	3.8
Sugar	67	4.2
Sweet Potato	28	1.9
Irish Potato	7	2.3
Kidney Beans	27	3.1
Salt	76	1.0
Groundnut	33	1.5
Tomato	41	1.7
Cooking Banana	20	8.8
Sweet Banana	41	1.8
Orange	8	1.0
Cooking Oil	61	1.1
Local Brew	38	4.5
Onions	66	1.1
Chicken Eggs	62	1.6
Chicken	42	1.8
Beef	30	3.3
Goat's Meat	18	2.1
Fish	35	5.3
Fresh Milk	10	3.2
Milk Powder	6	3.2
Kerosene	68	2.5
Firewood	12	6.3
Battery	68	5.5
Linen	20	1.5
Charcoal	6	5.6

---

[52] These are the unadjusted expenditure weights, i.e. the shares have not been scaled upwards to account for the omitted goods.

As the table reveals, many of the items had missing prices that needed to be imputed for building price indices. Since the methodology of imputations is very important, this is explained in greater detail below.

C. IMPUTING MISSING PRICES

The commodities were divided into two groups: food and non-food items.

For each group, the first step was to eliminate extreme values and outliers. For this and for the purposes of imputation the following variables were used as explanatory variables in the price regressions:

(a) Spatial variables: this included the latitudes and longitudes of the market or the shop.

(b) Time variables: days elapsed since January 1, 1960 till the interview day. Since different markets were interviewed on different days, this variable controls for temporal variation.

(c) Market variables: two variables were used to characterize the unit that was administered the market questionnaire: whether or not it was a market, and the distance to the nearest capital, the latter to capture possible effects of transportation costs.

(d) Rainfall: finally, the availability of rainfall data by district enabled controlling for rainfall variation.

Different specifications were experimented with that explained the variation in the prices the best, and the final sets of variables included the latitudes and longitudes (linear, squared and cubed), latitude interacted with the square of the longitudes and vice versa, and the product of the two. For the time variables a linear and a squared term were used. The market variables described above were used linearly, and the best variable for rainfall was a moving average of rain in the last 12 months.

The Welsch Distance test was performed to exclude outliers. Formally, Welsch's Distance W_i is

$$W_i = DFITS_i \sqrt{\frac{n-1}{i-h_i}}$$

where $DFITS_i$ is defined as

$$DFITS_i = \frac{r_i}{\sqrt{\dfrac{h_i}{(1 - h_i)}}}$$

where r_i are the studentized residuals, h_i are the leverage measures obtained from the diagonal elements of the hat matrix, and n is the total number of observations. The value of W_i will be affected by the size of the residuals and the size of the leverage.

About 1.8% of the food prices and 1.3% of the non-food prices were classified as outliers and omitted from the imputations.

Before predicting the prices, one last step was to try and see if some of the items had similar patterns of movements in their prices. This was necessary because of the lack of sufficient observations for some items. For this purpose a Chow test was used, which tested whether for equality of coefficients for groups of good type. On the basis of these tests, done separately for food and non-food, 8 food and 7 non-food items were combined together for the predictions.

Finally, log of price was regressed on the spatial, time, market and rainfall variables for each good (that was not combined with other goods based on the Chow test) and predicted values were generated. For goods that were combined based on the Chow test, similar regressions were used except that dummies for used for each item, and again predicted values were generated.

Once all the missing prices were estimated, the formula for Lasperye's index mentioned above was applied to arrive at the indices. Two sets of indices were calculated, one based on all predicted prices, and the other on predicted prices for the missing prices, and the actual prices for the ones that were not missing.

APPENDIX IV: RESEARCH PAPERS – KAGERA HEALTH AND DEVELOPMENT SURVEY 1991-1994

Ainsworth, M., K. Beegle, and G. Koda. 2002. "The Impact of Adult Mortality and Parental Deaths on Primary School Enrollment in Northwestern Tanzania." *Africa Region Human Development Working Paper Series*. Washington, D.C.: The World Bank. *Background research paper prepared for the African Economic Forum*, December 2000, Addis Ababa, Ethiopia. Forthcoming *Journal of Development Studies*
http://www.worldbank.org/afr/hd/wps/tz_adult_mortality.pdf

Ainsworth , M. and J. Dayton. 2003. "The Impact of the AIDS Epidemic on the Health of Older Persons in Northwestern Tanzania. *World Development* 31(1): 131-148. *World Bank Policy Research Working Paper,* no. 2649. Washington, D.C.: The World Bank. *Background research paper prepared for the African Economic Forum*, December 2000, Addis Ababa, Ethiopia.

Ainsworth, M., D. Filmer and I. Semali. 1998. "The Impact of AIDS Mortality on Individual Fertility: Evidence from Tanzania" in M. Montgomery and B. Cohen, eds. *From Death to Birth: Mortality Decline and Reproductive Change*. Washington, DC: National Academy Press.

Ainsworth, M., G. Koda, G. Lwihula, P. Mujinja, M. Over and I. Semali. 1992. "Measuring the Economic Impact of Fatal Adult Illness in Sub-Saharan Africa: An Annotated Household Questionnaire". *Living Standards Measurement Study Working Paper*, no. 90. Washington, D.C.: The World Bank.

Ainsworth , M.and M. Over. 1997. *Confronting AIDS: Public Priorities in a Global Epidemic*. A World Bank Policy Research Report. Washington, D.C.: Oxford University Press. 355 pages. Revised edition, 1999. (See Chapter 4)
http://www-wds.worldbank.org/servlet/WDS_IBank_Servlet?pcont=details&eid=000009265_3980219162747

Ainsworth, M. and I. Semali. 2000. "The Impact of Adult Deaths on Child Health in Northwestern Tanzania". *World Bank Policy Research Working Paper no. 2266*. Washington, D.C.: The World Bank.
http://www-wds.worldbank.org/servlet/WDS_IBank_Servlet?pcont=details&eid=000094946_00012505525064

Ainsworth, M.and I. Semali. 1998. "Who dies from AIDS? Socioeconomic Correlates of Adult Deaths in Kagera Region, Tanzania" in Ainsworth, Fransen and Over, eds. (1998). *Confronting AIDS: Evidence from the Developing World* Background Papers from the World Bank Policy Research Report, *Confronting AIDS: Public Priorities in a Global Epidemic*. Brussels: European Commission.

Alderman, Harold, Hans Hoogeveen, Mariacristina Rossi. 2004. "Reducing Child Malnutrition in Tanzania. Combined Effects of Income Growth and Program Interventions" mimeo, The World Bank.

Beegle, K. 2003. "Labor Effects of Adult Mortality in Tanzanian Households." *World Bank Policy Research Working Paper No 3062*. Washington, D.C.: The World Bank. forthcoming *Economic Development and Cultural Change*

Beegle, K,. R. Dehejia, R. Gatti. 2003. "Do Households Resort to Child Labor to Cope with Income Shocks?" *World Bank Policy Research Working Paper No 3075*. Washington, D.C.: The World Bank..

Burke, K. and K. Beegle. 2003. "Why Children Aren't Attending School: The Case of Northwestern Tanzania." *Journal of African Economies* 13: 333-355.

Dayton, J. and M. Ainsworth. 2002. "The Elderly and AIDS: Coping with the impact of adult death in Tanzania." Social Science and Medicine 59: 2161-2172. *Population Council Working Paper*, no. 160. New York. http://www.popcouncil.org/pdfs/wp/160.pdf

Lundberg, M., and M. Over. 2001. "Transfers and Household Welfare in Kagera, Tanzania." Prepared for UNAIDS. Presented to research seminar, University of East Anglia; the IAEN Economics of AIDS Symposium; the XIII International AIDS conference, Durban, South Africa, July 2000; and the International Health Economics Association Conference, July 2001, York, England.
paper: http://www.iaen.org/conferences/durpapers/38lundbergtext.pdf
diagrams: http://www.iaen.org/conferences/durpapers/38lundbergdata.pdf

Lundberg, M., M. Over, and P. Mujinja. 2002. "Do Savings Predict Death? Precautionary Savings during an Epidemic." Manuscript, prepared for UNAIDS, Geneva, December, 2002.

Lundberg, M., M. Over, and P. Mujinja. 2000. "Sources of Financial Assistance for Households Suffering an Adult Death in Kagera, Tanzania," *South African Journal of Economics*, 68:5:947-984. *World Bank Policy Research Working Paper*, no. 2508.
http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2001/01/26/000094946_0101170531277/Rendered/PDF/multi_page.pdf

Mujinja, P., V. Dorabawila, and M. Over. 1995. "A Profile of Health Facilities in the Areas hard hit by the AIDS Epidemic in the Kagera Region, Tanzania."

Semali, I. and M. Ainsworth. 1995. "A Profile of Traditional Healers in an Area Hard-hit by the AIDS Epidemic: Kagera Region, Tanzania". University of Dar es Salaam and The World Bank.

World Bank and University of Dar es Salaam. 1993. "Report on a Workshop on: The Economic Impact of Fatal Adult Illness in Sub-Saharan Africa."

World Bank and University of Dar es Salaam. 1995. "Coping with AIDS: The Economic Impact of Adult Mortality Form AIDS and Other Causes on Households in Kagera, Tanzania."