

# **Kagera Health and Development Survey 2004**

## **Basic Information Document**

Kathleen Beegle  
Joachim De Weerd  
Stefan Dercon

March, 3 2006

## Table of Contents

1	Overview.....	1
2	Survey Questionnaires .....	1
2.1	Household Questionnaire: Review of Sections.....	2
2.2	Household Questionnaire: Highlights of Substantial Differences .....	3
2.3	Tracking Modifications to the Household Questionnaire.....	16
2.4	Community Questionnaire: Review of Sections.....	17
2.5	Community Questionnaire: Highlights of Substantial Differences .....	17
2.6	Price Questionnaire: Review of Sections.....	20
2.7	Price Questionnaire: Highlights of Substantial Differences .....	21
2.8	School Questionnaire: Review of Sections.....	22
2.9	School Questionnaire: Highlights of Substantial Differences .....	22
3	Sample.....	22
3.1	KHDS 91-94 Household Sample: First Stage.....	22
3.2	KHDS 91-94 Household Sample: Second Stage .....	23
3.3	KHDS 2004 Households.....	24
3.4	Community Survey .....	29
3.5	Primary School Survey .....	30
3.6	Price Survey .....	30
4	Organization of Field Work .....	30
4.1	Recruitment.....	30
4.2	Pre-Fieldwork Tracking .....	31
4.3	Main Field Work.....	31
4.4	Tracking .....	32
5	Data.....	33
5.1	Data Processing.....	33
5.2	Data File Structures.....	33
5.3	Linking 2004 Data Sets.....	37
5.4	Linking Individuals Over Time .....	37
	Appendix A: List of KHDS 2004 Data Sets .....	40
	Appendix B: List of Related Documents .....	42
	Appendix C: Comparison of Key Variables with other Kagera Data Sets.....	43

## **ACKNOWLEDGEMENTS**

This survey was funded by the Knowledge for Change Trust Fund at the World Bank and the Danish Agency for Development Assistance (DANIDA) and implemented by E.D.I (Economic Development Initiatives). The principle investigators of the KHDS 2004 project are Kathleen Beegle (World Bank), Joachim De Weerd (E.D.I., Tanzania) and Stefan Dercon (Oxford University), with co-collaborators Flora Kessy (ESRF, Tanzania), Godlike Koda (University of Dar es Salaam), the late George Lwihula (IPH Muhimbili), Gideon Kwesigabo (IPH Muhimbili), Phare Mujinja (IPH Muhimbili), and Innocent Semali (IPH Muhimbili). Fieldwork activities were managed by E.D.I (Economic Development Initiatives) and data entry activities were supervised by Christopher Ksoll (Yale University). We would like to express appreciation to the research committee, regional administration, field staff, data entry operators and, finally, to all KHDS 2004 respondents. We are extremely grateful for their continued cooperation and willingness to participate in this longitudinal study.

# 1 Overview

The Kagera Health and Development Survey 2004 (KHDS 2004) took place in 2004 as a fifth round following on the four rounds of the baseline Kagera Health and Development Survey 1991-1994 (KHDS 91-94). The KHDS 2004 was designed to provide data to understand economic mobility and changes in living standards of the sample of individuals interviewed 10-13 years ago. The KHDS 2004 attempted to reinterview all respondents ever interviewed in the KHDS 91-94. This entailed attempting to track these individuals, even if they had moved out of the village, region or country.

This document is designed to give information to enable proper and effective use the KHDS 2004 data. Since much of the KHDS 2004 is based on the original baseline 1991-1994 KHDS, data users are encouraged to carefully review “User’s Guide to the Kagera Health and Development Survey Datasets.” (World Bank, 2004), which serves to complement this KHDS 2004 basic information document.

For papers using the KHDS 2004 data, we recommend that this document be cited as follows:

Beegle, Kathleen, Joachim De Weerd, and Stefan Dercon. (2006). Kagera Health and Development Survey 2004 Basic Information Document. mimeo. The World Bank.

## 2 Survey Questionnaires

The KHDS 2004 mainly consists of a household survey covering a wide range of topics. The KHDS 2004 also includes three community questionnaires to accompany the household survey (community, price, and primary school questionnaires).

The KHDS 2004 project used the original questionnaires from the KHDS 91-94 as the foundation of the survey instruments. The household questionnaire collects information on a wide range of topics, including: housing amenities, consumption, income, assets, time allocation of individuals, business activities, remittances, support from organizations, education, and health, including anthropometric measures. The community questionnaire collects data on the physical, economic and social infrastructure of the baseline communities. The primary school questionnaire collects information on the amenities at schools, composition of the student body, and assistance to schools. Finally, up to three price observations are collected in each community from local markets/stalls on a list of commonly purchased food and non-food items.

Where possible, comparability is maintained with the KHDS 91-94 survey instruments. However, the questionnaires for the KHDS 2004 were revised to reflect changes in the region since 1994. Further, the household questionnaire was redesigned in an effort to capture key transitions that have occurred since the previous survey. These revisions included:

- Inclusion of a module on the incidence of economic shocks from the last 10 years (both positive and negative) for all panel respondents.

- Inclusion of a module on migration for respondents who relocated since the KHDS 91-94.
- Inclusion of a module on informal insurance groups.
- Expansion of questions on the circumstances of deaths.
- Inclusion of information on the remittances, loans, bride price payments, social communication and labor transfers between previous members of the KHDS 91-94.

This section of the Basic Information Document reviews the 4 surveys of the KHDS 2004. For each survey, substantial differences are highlighted between the survey instrument used in the 1991-1994 rounds and in 2004.

**Users are encouraged to use this as a general guide to understand the questionnaires; however, this should not substitute for looking at the actual questionnaires directly. Users are encouraged to look directly at the survey instruments for literal question wording and to identify differences between survey instruments. The household questionnaires are available in Swahili (as used in the field) and English (a translated version of the Swahili field questionnaire); the community surveys were produced only in English.**

## ***2.1 Household Questionnaire: Review of Sections***

The household questionnaire is divided into numerous sections, each of which covers a fairly distinct aspect of household activities. Anthropometric measurements and the questionnaire on mortality of household members are administered in separate forms attached to the household questionnaire.

Each section of the household questionnaire has four types of respondents selected according to the content of the section: the interviewer, household head, most knowledgeable person in the household and individual household members. The only section for which household members are not respondents is the first section covering basic survey information (household location, GPS Coordinates, interviewing language, completion status of section, etc...).

Table 1 lists the sections of the KHDS 2004 household questionnaire. For sections that require only one respondent, the identification code of the respondent is listed in the introductory part of each section. The respondent in this section was either the household head or the most knowledgeable person. In sections where the information collected pertains to all individuals, each member of the household was asked to respond for himself or herself. However, if the individual was not capable of responding (for example if sick or too young), a proxy respondent was used to collect these data. This would be the person most knowledgeable to answer on behalf of the unavailable/incapacitated respondent.

**Table 1: Section Allocation in the Household Questionnaire**

<b>Section</b>	<b>Topic</b>	<b>Respondent</b>
Section 00	Basic Survey information	Interviewer & Household Head (or most knowledgeable)

**Table 1: Section Allocation in the Household Questionnaire**

Section	Topic	Respondent
Section 1	Household Roster	Household Head (or most knowledgeable)
Section 1	Household & Network Roster Cards	Interviewer
Section 2	Previous Children Residing Elsewhere	Household Head (or most knowledgeable)
Section 3	Main Activities of the Household	Household Head (or most knowledgeable)
Section 4	Information on Parents	Individual Household Members
Section 5	Education	Individual Household Members
Section 6	Health	Individual Household Members
Section 7	Activities and Non-Labor Income	Individual Household Members
Section 8	Individual Expenditures	Individual Household Members
Section 9	Migration	Individual Household Members
Section 10	Shocks Experienced in the Past 10 Years	Previous Household Member > 19
Section 11	Farming	Most knowledgeable person
Section 11	Agriculture	Most knowledgeable person
Section 12	Livestock	Most knowledgeable person
Section 13	Non-Farm Self-Employment	Most knowledgeable person
Section 14	Housing	Household Head (or most knowledgeable)
Section 15	Durable Goods, Expenditures, Inheritance, and Bride Price	Household Head (or most knowledgeable)
Section 16	Food Consumption and Expenditures	Most knowledgeable person
Section 17	Informal Organizations, Ability to Cope, Assistance from Organizations	Most knowledgeable person
Section 18	Gifts and Loans Received/Sent	Most knowledgeable person
Separate Form	Anthropometry	All household members
Separate Form	Mortality of Previous Household Members	Most knowledgeable person

## ***2.2 Household Questionnaire: Highlights of Substantial Differences***

Many changes were made in the KHDS 2004 household questionnaire compared to the KHDS 91-94 household questionnaire. Some questions were added and some dropped. Section

arrangement was also revised to provide better continuity during interviews. The following are the main changes included in the 2004 questionnaire:

- Section 9 (Fertility) and Section 11C (Age of tree crops) from the 1991-1994 questionnaire were dropped in 2004.
- Section 13 (Fishing) from 1991-1994 was incorporated in Section 13 (non-farm business) in 2004.
- In 1991-1994, information in Section 7 was collected on the main job and secondary job done in the past 12 months, while in 2004 data was collected only on the main job done.
- In 1991-1994, Section 20b collected information on deaths of deceased relatives who were not household members. In 2004, the mortality questionnaire covered only deceased household members from the 1991-1994 survey.
- The following were new sections introduced in 2004: Section 10 (Shocks experienced in the past 10 years), Section 15C (Household Two-Week Expenditures), Section 15D (Inheritance and Bride Price), Section 17B (Ability to cope) and Section 18A (Interactions with Network Members).

The remainder of this section details some of the key changes, section-by-section, between the KHDS 91-94 household questionnaire and the KHDS 2004 household questionnaire.

Table 2 provides information on the modifications made in the 2004 household questionnaire.

**Table2: Highlights of substantial differences household questionnaires**

<b>KHDS 2004</b>	<b>KHDS 1991-1994</b>	<b>Topic</b>	<b>Note</b>
Sec 00	Sec 00	Basic Survey Information	
Sec 1	Sec 1	Household Roster	
Sec 2	Sec 2	Children Residing Elsewhere	
Sec 3	Sec 4	Main Activities of the Household	
Sec 4	Sec 3	Information on Parents of Household Members	
Sec 5	Sec 5	Education	
Sec 6	Sec 6	Health	
Sec 7A	Sec 7A	Activities	
Sec 7B	Sec 7B	Employment During the Past 7 Days	
Sec 7C	Sec 7C	Self-employed Farmers During the Past 7 Days	
Sec 7D	Sec 7D	Self-employed Businessmen During the Past 7 Days	
Sec 7E	Sec 7E	Other Activities in the Past 7 Days	
Sec 7F	Sec 7F	Main Job in the Last 12 Months	

**Table2: Highlights of substantial differences household questionnaires**

<b>KHDS 2004</b>	<b>KHDS 1991-1994</b>	<b>Topic</b>	<b>Note</b>
	Sec 7G	Secondary job in the past 12 months	Section 7G dropped
Sec 7G	Sec 7H	Non-Labor Income	
Sec 8	Sec 18A	Individual Expenditures	
Sec 9	Sec 8	Migration	Not comparable
	Sec 9	Fertility	Section 9 dropped
Sec 10		Shocks experienced in the last 10 years	Section 10 new
Separate Form	Sec 10	Anthropometry	
Sec 11	Sec 11	Farming	
Sec 11A	Sec 11A	Land	
Sec 11B	Sec 11B	Crops	
Sec 11C	Sec 11D	Farm Inputs	
	Sec 11C	Age of tree crops	Section 11C dropped
Sec 11D	Sec 11E	Sales of Products From Home Grown Crops	
Sec 11E	Sec 11F	Hand Tools	
Sec 11F	Sec 11G	Farm Equipment	
Sec 12	Sec 12	Livestock	
Sec 12A	Sec 12A	Animals	
Sec 12B	Sec 12B	Sale of Animal Products	
Sec 12C	Sec 12C	Livestock Expenditures	
	Sec 13A	Fishing	Section 13 dropped
	Sec 13B	Fishing	
	Sec 13C	Fishing	
Sec 13A	Sec 14A	Non-Farm Self-Employment	Revised
	Sec 14B	Non-Farm Self-Employment Expenditures	
Sec 13B	Sec 14C	Non-Farm Self-Employment Income	
Sec 13C	Sec 14D	Non-Farm Self-Employment Assets	
Sec 14	Sec 15	Housing	

**Table2: Highlights of substantial differences household questionnaires**

<b>KHDS 2004</b>	<b>KHDS 1991-1994</b>	<b>Topic</b>	<b>Note</b>
Sec 14A	Sec15A	Type and Ownership of Dwelling	
Sec 14B	Sec15B	Housing Expenditures	
Sec 15A	Sec16A	Durable Goods	
Sec 15B	Sec 16B	Household Annual Expenditures	Some items moved from Sec 18A 1991-1994
Sec 15C		Household Two-Week Expenditures	Section 15C new
Sec 15D		Inheritance or Bride Price received	Section 15D new
Sec 16	Sec 17	Food Consumption	
Sec 16A	Sec 17A	Seasons of the Past 12 Months	
Sec 16B	Sec 17B	Food Consumption of Home Production	
Sec 16C-1	Sec 17C-1	Food Expenditures, Seasonal Foods	
Sec 16C-2	Sec 17C-2	Food Expenditures, Non-Seasonal Foods	
Sec 17A	Sec 19C	Informal Organizations	
Sec 17B		Ability to Cope	Section 17B new
Sec 17C	Sec 16C	Receipt of Assistance from Outside Organizations	
Sec 18A		Interactions with Network Members	Section 18A new
Sec 18B	Sec 19A	Gifts and Loans Received from Others	
Sec 18C	Sec 19B	Gifts and Loans Given to from Others	
	Sec 18B	Expenditures in last two weeks	Section 18B dropped
Separate Form	Sec 20	Mortality of household members	

***Section 00: Survey Information***

The household identification number, household location, GPS coordinates, religion, ethnic group, language used during the interview and other technical information related to the interview are noted. Detailed instructions on how to find the household and special information related to the questionnaire are also noted. The survey information indicates the completion status of each section, dates in which the interview took place and updates to the household tracking form.

*Differences between the 1991-1994 and 2004 questionnaires:*

- The household location is divided into 6 subdivisions: country, region, district, ward, village and sub-village. The selection of relevant details of the location in 1991-1994 was left to the discretion of the interviewer.
- In 2004, the GPS coordinates of each household visited were taken to mark the location of the household. This information is not in the public use data for privacy protection purposes.
- In 2004, the questionnaire identified if the household was in the same location (in the same village) as it was 10 years ago.
- In 2004, in addition to noting the GIS data, interviewers also wrote detailed instructions on how to locate the household.

### ***Section 1: Household Roster***

The household roster lists all persons who are currently residing in the household. The household head is listed first and receives the id code 01. Household members are generally defined as include “all people who normally sleep and eat their meals together in the household during at least three (3) of the twelve (12) months preceding the interview”. However, there are four exceptions to this definition:

- The following persons are household members, even if they have spent fewer than 3 months in the household in the past 12 months:
  - (1) The person identified as the head of the household.
  - (2) Persons who just joined the household and expect to be long-term residents (i.e. expected to be residing in the household in the next 6 months), such as newborn infants aged less than three months or new spouses.
- The following persons are not household members, even if they have slept in the same dwelling and taken their meals with the rest of the household for the entire 12 months before the survey:
  - (3) Tenants and boarders and their dependents.
  - (4) Contract servants and their dependents.

In addition to the household roster, there is also a Network Roster Card which is completed for all households. The Network Roster Card lists two groups of persons:

- 1) Respondents who were ever household members in KHDS 91-94 and do not reside as a current member in the household being interviewed
- 2) based on the KHDS 91-94 roster of Children Residing Elsewhere, those on this roster who have at least one parent in the household being interviewed and for whom their area of residence is known.

The Network Roster Card is later used to complete Section 18A. After completion of the Household Roster in Section 1, the first part of the Network Roster Card is completed (all non-

resident panel respondents). The second part of the Network Roster Card is completed after Section 2.

*Differences between the 1991-1994 and 2004 questionnaires:*

- KHDS 91-94 determined age by asking the age in years and date of birth while in 2004 age was determined by asking the age only.
- The 2004 household roster identified persons with multiple partners residing in the same household while in 1991-1994 only one partner for each member was noted.
- In 2004, question 10 in Section 1 identifies the roster identification code from 1991-1994 roster for panel respondents.<sup>1</sup> See Section 5.4 on using this variable to link individuals across surveys.

***Section 2: Children Residing Elsewhere***

In Section 2, information is collected for all children of household members who were recorded in Section 2 of KHDS 91-94 as children of household members living elsewhere. There are no age restrictions to be listed in this section. Further information is recorded for children who have at least one parent in the household and are still alive. This information includes sex, age, current residence, place of employment, educational attainment and parent's identification code. These children are also added to the Network Roster Card of the household to observe how frequently they interact with household members. This section was only administered to households which had children residing elsewhere in 1991-1994.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, this section is updated for all children residing elsewhere in 1991-1994. It is not a full roster of the children of current (2004) household members.
- In 2004, if the child does not reside either in the household, same village or nearby village the information on the district, region and country was collected.

***Section 3: Main Activities of the Household***

This section lists main economic activities undertaken by the household. The name of person who knows most about each activity is recorded for activities such as farming, livestock, family businesses, and food expenditure is recorded. This section serves as a guideline to interviewers on which sections to administer and their appropriate respondents. The sections that apply here are Sections 11, 12, 13 and 16.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, handicraft business was added as a main activity option, while businesses that provide services were dropped.
- In 2004, the separate fishing section administered in 1991-1994 was incorporated into the non-farm self-employment section.

***Section 4: Information on Parents***

---

<sup>1</sup> Panel respondents are individuals ever listed on the KHDS 91-94 household roster as a household member.

This section collects information on the parents of all household members. For those whose parent(s) are living elsewhere, information on area of residence of parents and educational achievement of parents is collected. If a child's parent was a respondent in 1991-1994 and is not living in the household, the identification code of the parent from 1991-1994 roster is noted. For children under 15 years living away from both parents (either due to fostering or orphanhood), the section also obtains information on the length of time that the child has been living in the current household.

*Differences between the 1991-1994 and 2004 questionnaires:*

- The 1991-1994 survey asked if the parent died in the past 12 months. The 2004 survey asked the specific year in which the parent died.
- In 2004, the question on parent's occupation was dropped.

### ***Section 5: Education***

The education section collects the following information for each household member 6 years 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.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, household members aged 6 or above were interviewed regarding schooling. In 1991-1994, household members aged 7 and above were interviewed.
- In 2004, questions on whether household members could do written calculations were dropped.
- In 2004, information on the age at which the respondent started school and name of school was acquired.
- In 2004, the codes for reasons of absence at school were modified.
- The 2004 survey did not acquire any information on vocational training.
- In 2004, only distance from the household to the school was asked. In 1991-1994 both distance from the household to the school and the time it takes to get to the school were asked.
- In 2004, the question on means used to get to school was dropped.
- In 2004, scholarships received in cash and in-kind were combined together and the total value was computed, while in 1991-1994 they were computed separately.

### ***Section 6: Health***

This section identifies individuals suffering from illnesses and collects details on their health care seeking behavior. The three parts of this section ask about acute illness (Part A), chronic illness (Part B), and general health (Part C).

The questions on acute illness (Part A) record information on the use of mosquito nets, whether the respondent was ill in the four weeks prior to the interview, the symptoms and diagnosis if ill, and the health seeking behavior related to the illness. For each household member who was ill or

injured in the four weeks prior to the interview, the following information is collected: duration of the illness, major symptoms, number of days the patient was unable to perform his/her usual activities and type health practitioner consulted. The final questions ascertain the actual diagnosis and the ailments the patient thinks he is suffering from.

Chronic conditions (Part B) are defined as conditions that have existed for 6 months or more. The questions asked in this section inform on the symptoms and the duration of the condition, as well as the diagnosis (if a practitioner was consulted). A respondents are asked to report on four symptoms: recurrent diarrhea (for a month or more), weight loss, recurring fever, and skin rash.

General health (Part C) consists of questions on self-assessed health status, disability status and activities of daily living (for household members 15 years or older).

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, respondents were asked whether they use mosquito nets and if the mosquito nets had ever been impregnated.
- In 1991-1994, respondents were asked to provide information on the first three practitioners they visited. In 2004, information was collected on the first practitioner only.
- In 2004, questions on health expenses incurred during illness were moved to Section 8 (Individual Annual Expenditures).
- In 2004, additional information on what the respondent thought s/he was suffering from besides the health practitioner's diagnosis was collected.
- In 2004, a traditional healer was added as an option of where the respondent sought help.
- In 2004, questions on assistance from household members or elsewhere during illness were dropped.
- In 1991-1994, respondents provided information regarding qualifications of the practitioner, distance to the health facility, means of transport, time used to visit health facilities, and the number of times they sought help. In 2004, these questions were dropped.

***Section 7: Activities and Non-Labor Income***

Section 7 is the largest section of the questionnaire, consisting of seven parts (Parts A-G). This section has been greatly modified compared to the 1991-1994 questionnaire. All individuals aged 7 and older were asked to answer questions on the economic activities listed in this section. For persons who did not work in the seven days preceding the survey, data were collected on job search and reasons for not seeking employment. For those who had worked seven days prior to the survey, information was collected on hours worked, type of employer, wage, and benefits associated with employment.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, illness of family member was added as a reason for not working in the past 7 days.
- In 2004, the ruling party was dropped as an employer.

- In 1991-1994, respondents provided information on their gross and net salaries. In 2004, only the net salary was inquired about.
- In 2004, detailed information was collected only for the first main job. In 1991-1994, information was collected for up to two jobs.
- In 2004, information on health coverage from employers and the main job done in the past 12 months was not collected.

### ***Section 8: Individual Expenditures***

Section 8 collects information on personal expenditures by individual household members. Questions 1 - 4 collect information on expenditure on clothing and medical coverage twelve months preceding the survey. The expenditure accounted for include the respondent's own input, as well as presents from other people. Questions 5 -10 inform on individual expenditures in the past 2 weeks on the following items: food and drinks consumed outside the household, cigarettes and tobacco, gambling, movies and make-up.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, structure and recall period are slightly revised from 1991-1994 Section 18A.

### ***Section 9: Migration***

This section collects information on the panel respondents' residence in the past 10 years. Only panel household members age 10 or older respond to this section. Information is collected on: year in which the respondent left the household he was living in 10 years ago, reasons for leaving, duration of the period in which he has lived in the current household, value of cash and in-kind taken during migration.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 1991-1994, the section was administered to all household members. In 2004, it was only administered to panel respondents.
- Migration codes slightly differ between the 1991-1994 and 2004 questionnaires.
- In 2004, questions on place of birth and nationality were dropped.

### ***Section 10: Shocks experienced in the past 10 years***

Section 10 is a new section introduced in 2004. The section collects information on shocks experienced in the past 10 years. This section is administered to panel respondents aged 20 years or above. For each year in the past 10 years, the panel respondent is asked whether the year was: very good, good, normal, bad or very bad. For every year that was classed as either very good or very bad, the respondent is asked to give reasons. Information on the coping methods used by the respondent is asked for years that are reported to be very bad.

### ***Section 11: Agriculture***

This section consists of 7 parts (Parts A-F). Part A informs on the number and size of fields (shambas) owned and cultivated by the household, as well as their sale value. Part B focuses on

the crops cultivated in the twelve months preceding the survey, the quantity of production sold, income from sale of crops and expenditure on crop inputs. Information on expenditure on farm inputs is collected in Part C, while Part D refers to income generated from the sale of home grown crops, as well as expenditure on transforming these crops for sale. Part E establishes the number of hand tools owned and their sale value. Part F concludes the section by informing on the ownership value, purchase and sale of agricultural equipment. The respondent in this section is the household member identified in Section 3 as the most knowledgeable about the household's farm. However, if the most knowledgeable person is not available for interview, the next most knowledgeable person available is interviewed. It is noted in Section 11 if the person interviewed is different from the person identified in Section 3.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, inheritance questions were not asked in this section. All inheritance questions were moved to Section 15D.
- In 1991-1994, the owner of the shamba in the household was identified by recording the owner's identification code, while in 2004 information on the individual who owns the shamba was not collected.
- In 2004, crops like coffee, trees and bananas were sub-divided to specific types farmed by the household. For example, coffee was sub-divided into arabica and robusta.
- In 2004, information on crop allocation among the shambas was not collected.
- Section 11C from 1991-1994 was completely omitted.
- In 2004, the respondents reported if an agricultural advisor had ever visited them in the past 12 months.
- Codes for home grown products, hand tools and farm equipment were modified in 2004.

***Section 12: Livestock***

This section assesses the number and value of livestock owned by the household and household income in the past 12 months generated by livestock related activities. Part A 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. The value of animals consumed or lost in the past 12 months is also established. Part B collects information on income from processing livestock products in the past year, such as milk and eggs. Part C measures the expenditures on livestock production in the past 12 months, for items such as herding, veterinary services, and animal feed.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, goats and cows were subdivided into dairy and non-dairy as opposed to 1991-1994 where no specifications were made.
- In 2004, information on the participation of household members in preparing animal products was not collected.

***Section 13: Non – Farm Self - Employment***

Section 13 gathers information on household businesses. Data are collected on the 3 most important enterprises operated by the household. The respondent for each enterprise is the household member most familiar with its operation (identified in Section 3). Part A informs on the type of business and the best informed person. The name and identification code of the person interviewed in Section 13 is recorded to compare with the person identified in Section 3. For each business, data on employees and expenditures on wages and raw materials are collected for the twelve months preceding the interview. The respondent is asked how much profit was made and the value of enterprise products that were consumed by the household (Part B). Information on assets owned by the business and their current value if they were to be sold is collected last (Part C).

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, this section include fishing enterprises which were covered in a separate section in 1991-1994.
- In 2004, enterprise expenditures are combined together with income, while in 1991-1994 they were separate categories.
- In 2004, the respondent is asked to report the net income from the past two weeks; in 1991-1994, the respondent was asked about gross income as well.

#### ***Section 14: Housing***

Section 14 contains information on type and ownership of dwelling (Part A) and housing expenses and housing characteristics (Part B). Information is collected on the number of rooms in the dwelling, ownership status, rental costs if rented, toilet type, source of lighting, source of energy used for cooking and distance to water source. This section also contains information on materials for walls, flooring, roofing and windows. Respondents who owned their dwellings were asked to report the resale price of their dwellings.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, questions relating to how the dwelling was acquired and whether the dwelling had ever been mortgaged were dropped.
- In 2004, questions relating to which household member owns the dwelling were dropped.

#### ***Section 15: Durable Goods, Household Expenditures, Inheritance and Bride Price***

Section 15 collects information on various non-food household expenditures. In Part A, information is collected on the durable goods owned by household members and the resale price if the goods were to be sold. Part B includes questions asked relate to household expenditures from the twelve months preceding the survey such as rent, dwelling renovations, taxes, jewelry and others. Information on two-week expenditures on newspapers, petrol, pens, kerosene, matches, batteries and soap is collected in Part C. Part D consists of questions regarding inheritance and bride price. In order to link the deceased person to the household member who received inheritance, the identification code of the deceased is noted. The inheritance received whether in cash, land or in-kind is recorded in terms of its retail price value. For every bride

price received, the name and the identification code of the person associated with the bride price is recorded.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, a question was added to determine whether any of the household members have a bank account.
- In 2004, Section 15D (Inheritance and bride price received) was added.

### ***Section 16: Food Consumption and Expenditures***

Section 16 gathers information on food consumption and expenditures in the past 12 months. In Part A, seasons are marked for every month of the year. For every crop produced by the household, the following information is gathered: month in which the crop was grown, number of times the crop was consumed by the household and the value of the crop consumed, if it were to be bought (Part A). The same information is also collected for animals, fish, birds and insects. Data on seasonal food expenditure is collected for twelve months preceding the survey (Part C-1). For seasonal food bought information is collected on the months in which the crop was bought and the costs of buying the crops during the wet season and dry season. Further the following information on non-seasonal food expenditures is collected in Part C-2 (the month in which each item was purchased, the number of times the item was purchased and the cost incurred each time the item was purchased).

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, fruit and vegetables in consumption of home production were grouped together. In 1991-1994, data was collected separately for each item.
- In 2004, local brew made at home was added.
- In 2004, new questions were added on: the distance to the nearest market, the number of meals the household has per day, the number of times the household consumed meat in the past week and the incidence of food shortages in the past year.

### ***Section 17: Informal Organizations, Ability to Cope, Assistance from Organizations***

Section 17 consists of three subsections. Part A is a new section in 2004. It covers participation of household members in informal insurance organizations. This section collects information such as name of group, number of members, identification codes of household members registered in the group, type of assistance given, contributions of the household members twelve months preceding the interview and information on the assets owned by the group. Part B informs on coping strategies used by household members when they are in need of financial assistance. Assistance from outside organizations given to the household in the past 10 years is recorded in Part C: the name of the organization, amount received in the past 10 years and amount received in the past 12 months are noted.

*Differences between the 1991-1994 and 2004 questionnaires:*

- Part A is a new sub-section in 2004.
- In 2004, Part C organization codes were modified and questions were added to collect information on assistance received in the past 10 years and the past 12 months.

### ***Section 18: Gifts and Loans Received/Sent***

Section 18 collects information on the interaction with non-household members, focusing on cash and in-kind gifts sent/received. Non-members are divided into two groups: those who are on the Network Roster Card and others. The Network Roster Card consists of panel respondents who do not live in the household interviewed but were previously residing with the panel respondents in this household in at least one round of 1991-1994, or were listed in 1991-1994 as children living elsewhere (see description of Section 1 above).

In Part A household members are asked to give information on whether they have received from or have given gifts or loans to network members listed on the Network Roster Card. Other information collected concerns reasons for assistance, total value of the transfers and receipt of physical assistance. Further data on the interaction of members is collected such as whether they have ever lived together, communication frequency and the possibility of getting assistance when needed. Part B collects information on gifts and loans received from others (that is, people not listed on the Network Roster Card); Part C informs on gifts and loans given to others. For each transfer made the following information is collected: the name of the person associated with the transfer, relation to the household head, place of residence, type of transfer, total value, the reason for the transfer and provision of physical assistance provided.

*Differences between the 1991-1994 and 2004 questionnaires:*

- Section 18A is a new section in 2004.
- In 2004, questions on receipt of physical assistance were added and follow-up questions on outstanding loans were dropped.

### ***Anthropometry***

Anthropometric measurements are completed for every household member. Data were collected on the age (re-asked as it is included in Section 1 also), gender, date of measurement, weight and height of each household member. It was also noted if female respondents were pregnant or breast-feeding. A clinic card was requested for every child less than 6 years to ascertain whether the child has received injections for measles, polio, tetanus and tuberculosis.

*Differences between the 1991-1994 and 2004 questionnaires:*

- There are no differences.

### ***Mortality***

This instrument measures all deaths of panel respondents from the baseline household that the household being interviewed originates since the last baseline survey. The respondent for this questionnaire is the person who knows best about the circumstances of the death(s). There can be different respondents for multiple deaths from the baseline household. For every deceased person, data is collected on the year of death, members the deceased resided with at the time of death and two years preceding the death, cause of death and places where the deceased sought help during illness.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 1991-1994, information was collected on deaths of household members (Section 20A) and deaths of other relatives (Section 20B). In 2004, this questionnaire collects information on deaths of panel respondents (household members from the 1991-1994 roster). There is up to one mortality questionnaire for every baseline (1991-1994) households (if at least one person had died). So, in 2004, some households share the same mortality questionnaire (in cases where two more surviving respondents now reside in separate households).

### **2.3 Tracking Modifications to the Household Questionnaire**

In order to ease some of the work load for the tracking phase of the KHDS 2004, modifications were made for the administration of the household questionnaire during tracking. These modifications are outlined in Table 3 below.

**Table 3: KHDS 2004 Household Questionnaire Tracking Modifications**

<b>Section</b>	<b>Topic</b>	<b>Modification for Tracking</b>
Sec 00	Survey Information	no change
Sec 1	Household Roster (including Household Roster Card)	no change
Sec 2	Previous Children Residing Elsewhere	no change
Sec 3	Main Activities of the Household	no change
Sec 4	Information on Parents	Complete for PHHM and current HHM 10 years and under
Sec 5	Education	no change
Sec 6	Health	Complete for PHHM
Sec 7	Activities and non-Labor Income	Complete for PHHM
Sec 8	Individual Expenditures (2 week and annual)	no change
Sec 9	Migration	no change
Sec 10	Shocks Experiences in the Last 10 Years	no change
Sec 11	Agriculture	Sec 16B: Q1 only. Sec 11C-D: dropped
Sec 12	Livestock	Sec 12B-C: dropped
Sec 13	Non-Farm Self-Employment	Sec 13B: Q1-Q4 only
Sec 14	Housing	
Sec 15	Durable Goods, Expenditures, Inheritance and Bride Price	no change
Sec 16	Food Consumption and Expenditures	no change

**Table 3: KHDS 2004 Household Questionnaire Tracking Modifications**

Section	Topic	Modification for Tracking
Sec 17	Informal Organizations, Ability to Cope, Assistance from Organizations	Sec 17A: dropped, except in Dar es Salaam and Mwanza City
Sec 18	Gifts and Loans Received/Sent	Sec 18B-C: dropped
Separate Form	Anthropometry	Complete for panel respondents and current household membrs 10 years and under
Separate Form	Mortality of Previous Household Members	no change

## 2.4 Community Questionnaire: Review of Sections

An expanded community questionnaire was developed for the KHDS 2004. The community questionnaire was administered by the team supervisor and respondents included the village chairmen, development officers, education officers, other government officials and health care workers. The questionnaire was administered in all of the baseline enumeration areas for a total of 49 community interviews (see section on sampling). GPS coordinates of the area where the questionnaire was administered were taken for each community. Table 4 lists the sections of the KHDS 2004 community questionnaire.

**Table 4: Section Allocation in the Community Questionnaire**

Section	Topic	Respondent
Survey Information	Enumeration area number/Location where the community questionnaire was administered, GPS coordinates, date.	Interviewer
Section 0	Selecting respondents	Village Chairman
Section 1	Demographic information	Village Chairman/Secretary
Section 2	Economy and Infrastructure	Community Development Officer
Section 3	Education	Chairman of the Education Committee
Section 4	Health	Chairman of the Health Committee
Section 5	Agriculture	Agricultural Officer
Section 6	Culture	Village Chairman
Section 7	Shocks in the past 10 years	Most Knowledgeable Village Elder

## 2.5 Community Questionnaire: Highlights of Substantial Differences

The substantial changes to the community questionnaire include:

- A new section was included on shocks experienced in the past 10 years (Section 7).
- Data was collected on population share of ethnic groups.
- GPS coordinates were taken in each community for all enumeration areas.

- Questions on access to roads, electricity and water were introduced.
- Questions on the culture of mourning were asked for three different periods: the time of the survey, 10 years prior to the interview and 20 years prior to the interview.
- Information was collected on access to vocational training and secondary education.
- Information was also collected on temporary migration and seasonal employment of community members.

### ***Survey Information***

This section collects the following information: Enumeration area name, enumeration area number, date of interview and GPS coordinates.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, GPS coordinates were taken for each enumeration area at the location where the community questionnaire was administered but is not publicly available for confidentiality reasons.

### ***Section 0: Selecting Respondents***

This section contains background information on respondents including their occupation, age, sex and the number of years they have resided in the community.

*Differences between the 1991-1994 and 2004 questionnaires:*

- There are no differences with the 1991-1994 questionnaire.

### ***Section 1: Demographic Information***

Section 1 collects information on population, migration and village sub-divisions. Data is also collected on proportion of different ethnic and religious groups in the community, migration movements to and from the village, as well as reasons for the migration.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, the village population was recorded according to three different age groups.
- In 2004, questions regarding population share of ethnic groups and religious groups in the community were added.
- In 2004, the recall period was changed to 10 years.
- In 2004, questions were added to determine the name of areas which immigrants come from and move to.
- In 2004, up to four reasons for migration were recorded compared to only three reasons in 1991-1994 .
- In 2004, information was collected on village sub-divisions.

### ***Section 2: Economy and Infrastructure***

This section records information on the major economic activities of people in the community, village infrastructure and access to electricity and water. It also informs on temporary migration,

availability of social services, disasters in the past 10 years, organizations in the community and the impact of the Rwanda refugee crisis on the village.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, the recall period was changed to 10 years.
- In 2004, new questions on access to roads, electricity and water were added.
- In 2004, additional data was collected on temporary migration and seasonal employment within village and outside the village.
- In 2004, questions regarding the impact of refugees on villages, informal insurance groups, collective action activities and communal assets were also added.

### ***Section 3: Education***

This section records information on nursery schools, primary schools, secondary schools and vocational training centers. The following information is collected for each school in the community: the number of schools at each level, distance to the school, school ownership, number of classes, the year the school started operating, services rendered and tuition fees.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, questions on adult literacy programs were dropped.
- In 2004, new questions were added to collect information on nursery schools, secondary schools and vocational training centers.

### ***Section 4: Health***

Section 4 gathers information on the availability of health services in the community such as dispensaries, health centers, hospitals, drug shops, private laboratories, and village health workers. Data were also collected on the major health problems in the community, access to health facilities, birth, death, and HIV/AIDS testing centers.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, questions concerning traditional healers and their qualifications, major problems with health services in the community and major causes of death in the community among children and adults were dropped.
- In 2004, new questions on HIV/AIDS testing, access to drug shops (these are not pharmacies; they are local shops which sell medication), and private laboratories were added.

### ***Section 5: Agriculture***

Section 5 asked basic questions regarding agricultural activities conducted in the community such as type of crops farmed, the existence of agriculture extension centers and cooperative societies, access to farming equipment, irrigation, rainfall patterns, and prices of land. In addition data was collected on hourly wage rates for various agricultural work for adults and children between the ages of 10 and 14.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, new questions on the main crops grown in the community were added.
- In 2004, questions that collected information on sharecroppers were dropped.
- In 2004, questions on the access of farmers to implements such as tractors, fertilizers and insecticides were added.

### ***Section 6: Culture***

In this section, the community reports information on orphans under the age of 15, orphanages, mourning customs and inheritance. Data collected relates to the community's typical mourning period and mourning customs at the time of the survey compared to the past two decades. The same comparison is made for inheritance of land, housing and other property.

*Differences between the 1991-1994 and 2004 questionnaires:*

- In 2004, questions on actual mourning periods for different age groups in the society were dropped.
- In 2004, questions on inheritance settlement in case of a male death were added.
- In 2004, the recall period was changed to 10 years.

### ***Section 7: Shocks in the past 10 years***

Section 7 is a new section introduced in 2004. It collects information on the shocks experienced by the community in past 10 years. The section acquires information on community hardships in terms of wealth and living conditions. It also asks about periods that were very good for the majority of community members.

## ***2.6 Price Questionnaire: Review of Sections***

Supervisors in all enumeration areas administered the price questionnaire. Price data were collected from two types of markets for each enumeration area: the nearest community market and roadside shop. The questionnaire contained a list of thirty food items, five pharmaceutical items, and thirteen non-food items. Three observations were made for each item from three different traders at different locations in the market. A village elder escorted the supervisor to administer the questionnaire in order to insure accurate price quotations. Table 5 gives an overview of the sections of the questionnaire.

**Table 5: Section Allocation in the Price Questionnaire**

<b>Section</b>	<b>Topic</b>	<b>Respondent</b>
Survey Information	GPS coordinates, date.	Interviewer
Part I	Food Prices	Sellers of the products in question
Part II	Pharmaceutical Prices	Sellers of the products in question
Part III	Non-Food Prices	Sellers of the products in question

## ***2.7 Price Questionnaire: Highlights of Substantial Differences***

Overall there were no substantive changes; a few items were added to the list.

### ***Survey Information***

This section collects the following information: enumeration area name, enumeration area number, date of interview and GPS coordinates of the market place. This section also specifies whether the observations were done at a roadside stall or market.

#### ***Differences between the 1991-1994 and 2004 questionnaires:***

- In 2004, GPS coordinates of the market and shops were recorded, as well as how frequently it is organized. GPS coordinates are not publicly available for confidentiality reasons.

### ***Part I: Food Prices***

This part collects information on prices of food items such as tea leaves, onions, eggs, chicken, goat meat, cow meat, garden peas, fresh milk and powdered milk. All food items were measured in grams except chicken eggs - for which the price of one egg was recorded.

#### ***Differences between the 1991-1994 and 2004 questionnaires:***

- The chicken price taken in 2004 was for live chickens, whereas in 1991-1994 the status of the chicken was not specified.

### ***Part II: Pharmaceutical Prices***

Part II collects information on pharmaceutical products. These products are measured in tablets; they are not weighed.

#### ***Differences between the 1991-1994 and 2004 questionnaires:***

- Observations were made separately for packed and unpacked pharmaceutical products in 2004, while the 1991-1994 questionnaire did not make this distinction.
- The 2004 survey replaced the Cloroquine option with Metakelfin, which is currently the most popular malaria drug used in Tanzania.

### ***Part III: Non-Foods Prices***

Non-food items did not have to be weighed as the price unit depends on the item. If only one person in a market sold an item, this price was recorded in the first column.

#### ***Differences between the 1991-1994 and 2004 questionnaires:***

- In 2004, observations were made for two brands of batteries, whereas in 1991-1994 observations were made for one brand.
- In 2004, for all non-food items that come in different sizes, the volume was specified whereas in 1991-1994 the volumes were not specified.

- In 2004, in recording the unit price for dry cell batteries and hurricane lamps, data was collected for specific brands, whereas in 1991-1994 the brands were not specified.

## ***2.8 School Questionnaire: Review of Sections***

The school questionnaire collects information on primary schools in the enumeration area. The school questionnaire was administered in each school within the enumeration area. If there were no schools in the enumeration area, the questionnaire was administered in the nearest school from the enumeration area. In total seventy-two school questionnaires were administered during the survey, with a maximum of three schools per enumeration area. Part A of the questionnaire is completed by the supervisor and it concentrates on the characteristics of the school, as well as the school fees and enrollment rates. Part B is left with the head teacher, who records information on the number of textbooks available for students of each grade, the number of classes, number of enrolled students, number of students who completed Standard 7, number of students admitted to secondary school and the number of teachers employed at the school. Table 6 gives an overview of the sections of the school questionnaire.

**Table 6: Section Allocation in the School Questionnaire**

<b>Section</b>	<b>Topic</b>	<b>Respondent</b>
Cover page	Survey Information	Interviewer
Part A	School characteristics, enrollment and fees	Head teacher
Part B	Text books, Standard 7 completion, number of teachers employed and assistance or contributions	Head teacher

## ***2.9 School Questionnaire: Highlights of Substantial Differences***

The content of the questionnaire is the same as 1991-1994 with one exception. In the 2004 survey GPS coordinates and school statistical identification number provided by the government were recorded for every school in the enumeration area but these data are not publicly available for confidentiality reasons.

# **3 Sample**

## ***3.1 KHDS 91-94 Household Sample: First Stage***

The KHDS 91-94 household sample was drawn in two stages, with stratification based on geography in the first stage and mortality risk in both stages. A more detailed overview of the sampling procedures is outlined in “User’s Guide to the Kagera Health and Development Survey Datasets.” (World Bank, 2004).

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. Enumeration areas of households were drawn randomly from the PSUs in each stratum, with a probability of selection proportional to the size of the PSU.

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", which is an administrative area that is smaller than a district. The adult mortality rate (ages 15-50) was calculated for each ward and each PSU was assigned the mortality rate of its ward.

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 allocated to 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.

The KHDS 1991-1994 selected 51 communities as primary sampling units (also referred to as enumeration areas or clusters). In actuality, 2 pairs of enumeration areas were within the same community (in the sense of collecting community data on infrastructure, prices or schools). This, for community-level surveys, there are 49 areas to interview.

### ***3.2 KHDS 91-94 Household Sample: Second Stage***

The household selection at the second stage (with enumeration areas) was a stratified random sample. That is, households expected to experience an adult death were over-sampled. In order to stratify the population, an enumeration of all households was undertaken. Between March 15 and June 13, 1991, 29,602 households were enumerated in the 51 areas. 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 were also noted.
- 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) were also noted.

The enumeration form asked explicitly about illness and death of adults between the ages of 15-50 because this is the age group disproportionately affected by the HIV/AIDS epidemic; it is the impact of these deaths that was of research interest. Out of over 29,000 households enumerated, only 3.7 percent, or 1,101, had experienced the death of an adult aged 15-50 caused by illness during the twelve months before the interview and only 3.9 percent, or 1,145, contained a prime-age adult too sick to work at the time of the interview. Only 77 households had both an adult death due to illness and a sick adult. This supports the point that, even with some stratification

based on community mortality rates and in an area with very high adult mortality caused by an AIDS epidemic, a very large sample would have had to have been selected to ensure a sufficient number of households that would experience an adult death during the two-year survey.

Using data from the enumeration survey, 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 enumeration area were classified into two groups, based on their response to the enumeration:

- “Sick” households: Those that had either an adult death (aged 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 (aged 15-50) due to illness nor an adult (aged 15-50) too sick to work (n=27,433).

In selecting the sixteen households to be interviewed in each enumeration area from which a enumeration area was drawn, fourteen were selected at random from the "sick" households in that enumeration area and two were selected at random from the "well" households. In one enumeration area, where the number of "sick" households available was less than fourteen, all available sick households were included in the sample; the numbers were balanced using well households. The final sample drawn for the first passage consisted of 816 households in 51 enumeration areas.

### ***3.3 KHDS 2004 Households***

KHDS 2004 sampling strategy was to reinterview all individuals who were household members in any round of the KHDS 1991-1994 and who were alive in the last interview.<sup>2</sup> The household in which these individuals live would be administered the full household questionnaire. For all household members alive during the last interview in 1991-1994, but found to be deceased by 2004, information about the deceased would be collected in the mortality questionnaire. This questionnaire intended to collect data on the circumstances of their death, as well as on their living arrangements and limited information on health seeking behavior prior to death. The respondents for this questionnaire were typically panel respondents who were previous household members with the deceased, other relatives, neighbors or close friends.

Although the KHDS is a panel of respondents and the concept of a ‘household’ after 10-13 years is a vague notion, it is common in panel surveys to consider recontact rates in terms of households. Table 7 shows the rate of recontact of the baseline households, where a recontact is defined as having interviewed at least one person from the household.<sup>3</sup> In this case, the term household is defined by the baseline KHDS survey which spans a period of 2.5 years. Due to movements in and out of the household, some household members may have not, in fact, lived

---

<sup>2</sup> One serious problem that is side-stepped by this approach is constructing a definition of what makes a household the same household as 10 years ago, especially if there are individuals who have migrated, split-off or the household has dissolved.

<sup>3</sup> All statistics from the KHDS 2004 are subject to revision if a beta version is released pursuant to corrections or edits are made upon further review.

together in the household at the same time in the 1991-1994 rounds (for example, consider one sibling of the household head moving into the household for 1 year and then moving out, followed by another sibling moving into the household).

**Table 7: KHDS Households**

KHDS 1991-1994		KHDS 2004 Re-interview Rates		
Number of interviews during 1991-1994	Households interviewed	Re-interviewed	Deceased	Untraced
1	39	22 (56%)	2 (5%)	15 (39%)
2	45	38 (84%)	1 (2%)	6 (13%)
3	69	59 (86%)	1 (1%)	9 (13%)
4	759	713 (94%)	13 (2%)	33 (4%)
Overall	912	832 (91%)	17 (2%)	63 (7%)

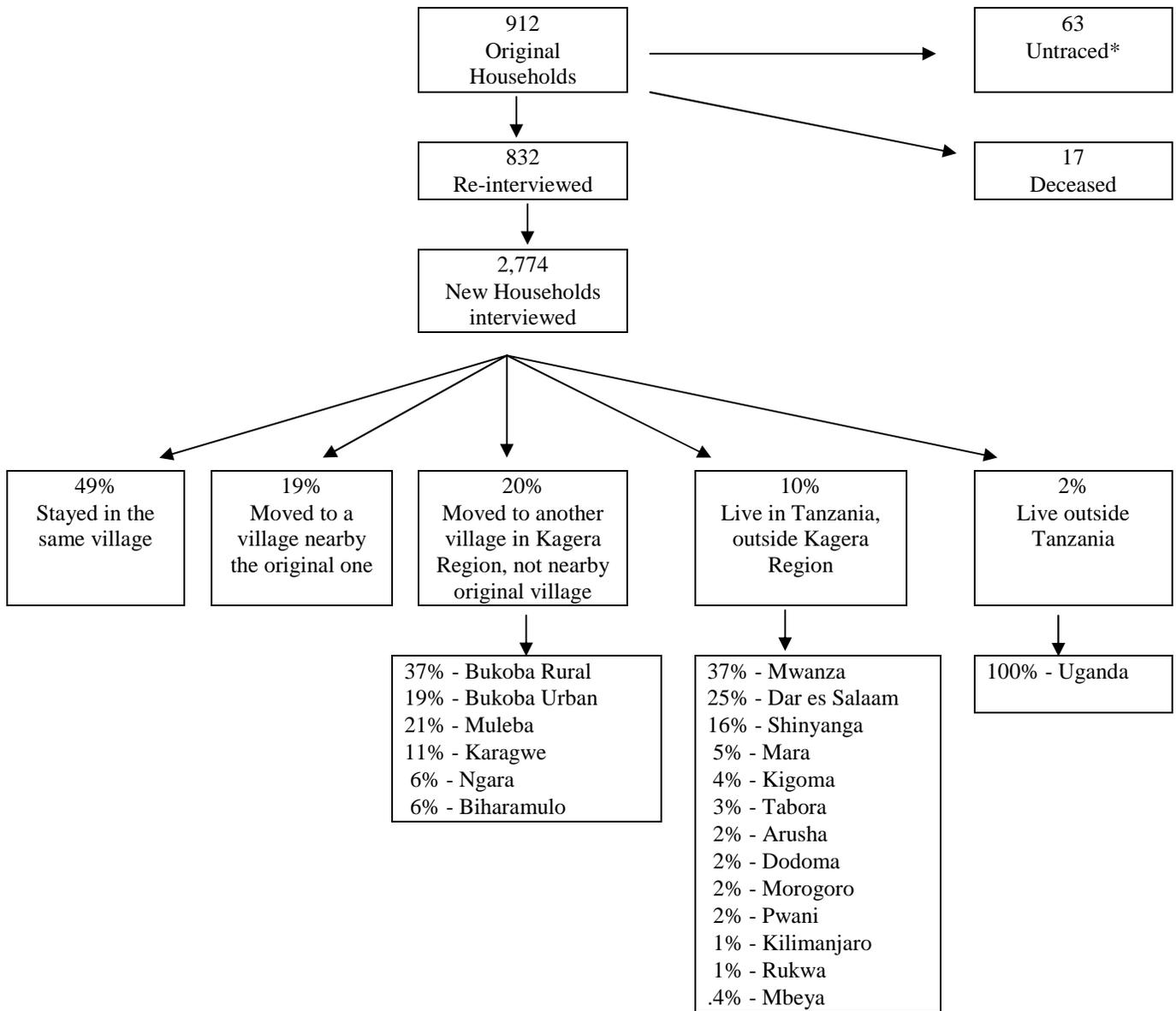
*Notes:* “Re-interviewed” means that at least one member of the baseline household was re-interviewed in the KHDS 2004. “Deceased” means that all previous household members are reported to be dead. “Untraced” means that no previous household member was re-interviewed. Of the 915 original sample households, 3 were single-person households in which the respondent died before the end of the KHDS 1991-1994 rounds, leaving 912 “surviving” households.

Excluding households in which all previous members are deceased (17 households and 27 people), the field team managed to recontact 93 percent of the baseline households. Not all 912 households received four interviews. Not surprisingly, households that were in the baseline survey for all four rounds had the highest probability of being reinterviewed. Of these 746 households, 96 percent were reinterviewed.

Figure 1 shows some preliminary statistics on the relocation of households. Because people have moved out of their original household, the new sample in KHDS 2004 consists of over 2,700 households from the baseline 832, which were recontacted. Much of the success in recontacting respondents was due to the effort to track people who had moved out of the baseline villages. One-half of all households interviewed were tracking cases, meaning they did not reside in the baseline communities. Of those households tracked, only 38 percent were located nearby the baseline community. Overall, 32 percent of all households were not located near the baseline communities. While tracking is costly, it is an important exercise because migration and dissolution of households are often hypothesized to be important responses to hardship. Excluding these households in the sample raises obvious concerns regarding the selectivity of attrition. In particular, out-migration from the village, dissolving of households, and even marriage, may be responses to adult mortality. At the same time, tracking will provide a unique

opportunity to study these coping mechanisms: who uses them, what is the effect, do they get people out of poverty or do they themselves constitute a poverty trap.

**Figure 1: Re-interviewing Respondents after 10+ years**



*Notes:* “Re-interviewed” means that at least one member of the baseline household was reinterviewed in the KHDS 2004. “Deceased” means that all previous household members are reported to be dead. “Untraced” means that no previous household member was reinterviewed. \* The locations of the sample of untraced individuals were reported by informants as: Kagera (57%), Dar es Salaam (8%), Mwanza (12%), other region (10%), other country (6%) and unknown (7%).

Turning to recontact rates of the sample of 6,204 respondents, Table 8 shows the status of the respondents by age group (based on their age at first interview in the 1991-1994 rounds). Re-interview rates are monotonically decreasing with age, although the reasons (deceased or not located) vary by age group. The older respondents were much more likely to be located if living, which is consistent with higher migration rates among the young adults in the sample. Among the youngest respondents, over three-quarter were successfully re-interviewed. Excluding people who died, 82 percent of all respondents were re-interviewed. Without tracking, re-interview rates of surviving respondents would have fallen from 82 percent to 52 percent. Non-local migration is not trivial; restricting the tracking to nearby villages would have resulted in 63 percent recontact of survivors. Migration proved to be an important factor in determining whether someone was recontacted. Respondents who were untraced were much more likely to be residing outside Kagera (52 percent) compare to their counterparts who were re-interviewed (9 percent).

KHDS 2004 tracked international migrants for Uganda only. Although the location of those in other countries was known, they were not traced. For those respondents who were not reinterviewed, the KHDS 2004 gives some information about their interactions with the reinterviewed respondents. Survey modules on the frequency of contact with all previous household members inform on the cash, in-kind and labor interactions between former household members (Section 18A).

**Table 8: KHDS Individuals by Age**

Age at baseline 1991-1994	Re-interviewed	Deceased	Untraced	Re-interview rate among survivors
<10 years	1,606 (79%)	122 (6%)	317 (15%)	83%
10-19 years	1,408 (73%)	97 (5%)	413 (22%)	77%
20-39 years	828 (66%)	234 (19%)	189 (15%)	81%
40-59 years	436 (74%)	119 (20%)	34 (6%)	93%
60+ years	163 (41%)	228 (57%)	10 (2%)	94%
Overall	4,441 (72%)	800 (13%)	963 (16%)	82%

*Notes:* Sample of individuals interviewed in KHDS 1991-1994 and alive at last interview. Age categories are based on age at first interview. “Re-interviewed” means that at least one member of the baseline household was re-interviewed in the KHDS 2004. “Deceased” means that all previous household members are reported to be dead. “Untraced” means that no previous household member was re-interviewed.

**Table 9: KHDS Re-interview Rates by Location**

	Number	Location	%
Baseline sample	6,204		
Re-interviewed	4,441		
		Same community	63
		Nearby village	14
		Elsewhere in Kagera	15
		Other region	7
		Uganda <sup>a</sup>	1
Untraced	963		
		Kagera	57
		Dar es Salaam	8
		Mwanza	12
		Other region	10
		Other country <sup>b</sup>	6
		Don't know	7
Deceased	800		

*Notes:* Location for untraced respondents is reported by other household members from the baseline survey who were successfully located, interviewed, and able to provide location information on the respondent. In some cases, this information comes from other relatives or neighbors residing in the baseline communities.

a. The KHDS 2004 tracked international migrants for Uganda only. b. Countries to which the 53 untraced respondents had moved are: Uganda (24), Rwanda (16), Norway (3), Burundi (2), Kenya (2) Sweden (2), Botswana (1), Egypt (1), England (1), and Germany (1).

### **3.4 Community Survey**

The community questionnaire was administered in all KHDS baseline communities. There are 49 unique communities; as noted above, the sample has 51 enumeration areas but 2 pairs are in the same community (areas 44 and 45; areas 46 and 47).

In 2004, the community questionnaire was administered in the same manner as in 1991-1994. The respondents for this questionnaire are people who are well informed about the activities, events and infrastructure of the community being surveyed. The group of respondents consists of the following people: chairman of education committee, secretary of development committee, one person from the community leadership, someone familiar with the health problems of the community, and someone familiar with agricultural and livestock practices of the community

There were two other questionnaires included in the 1991-1994 survey which were dropped in the 2004 survey. These were the health facility questionnaire (administered in all four waves of KHDS 91-94) and the traditional healer questionnaire (administered only in wave 3 of KHDS 91-94).

### ***3.5 Primary School Survey***

The school questionnaire was completed for every primary school in the enumeration area, both public and private. In 2004, the school questionnaire was administered in the same manner as in 1991-1994. The number of schools per enumeration area ranged from one to three schools per enumeration area. A total of 72 school questionnaires were administered in 49 baseline communities.

### ***3.6 Price Survey***

Price questionnaires were completed for markets and shops in every enumeration area. In 2004, the price questionnaire was administered in the same manner as in 1991-1994. Where possible two questionnaires were completed per enumeration area. In most enumeration areas one questionnaire was done in shops and one in markets, although some enumeration areas have only one questionnaire and one enumeration area has three questionnaires. A total of 90 price questionnaires were administered, 47 from markets and 43 from shops.

## **4 Organization of Field Work**

The project headquarters of KHDS 2004 were at the EDI (Economic Development Initiatives) offices in Bukoba Town. Here the human resources and finances of the project were managed, legal and contractual matters were taken care of, transport arrangements for field teams were made, the stock of field equipment was managed, future work was planned and all other activities necessary for the successful completion of the project were implemented.

Details on recruitment, pre-fieldwork tracking, main field work, and tracking are given below. In addition to these activities, the questionnaire itself was piloted by supervisors in non-sampled households in Kibeta and Kitendaguro areas in Bukoba District prior to training.

### ***4.1 Recruitment***

Field staff recruitment started in spring 2003 with field supervisors. Four supervisors were recruited. After training, supervisors were involved in developing the survey instruments, planning field work, piloting the questionnaire, and preparing interviewer training manuals and materials. In November 2003, 36 interviewers were recruited; they were trained for three weeks. Interviewers were trained on the household questionnaire, mortality questionnaire, anthropometrics and basic communication methods with respondents. The training included actual household interviews in one rural area for all trainees. After training, an assessment of each interviewer on and off the field was conducted; the best 28 interviewers were retained. The field teams consisted of 4 teams of 7 interviewers and 1 supervisor. The final field team included one supervisor and 5 interviewers had worked on the KHDS 91-94.

## ***4.2 Pre-Fieldwork Tracking***

In order to facilitate the field work and prepare for tracking of movers, the KHDS 2004 had a pre-fieldwork tracking phase. Field management, supervisors and three interviewers did the pre-field work tracking in October 2003. The team visited all 51 baseline communities with rosters from the 1991-1994 survey to complete a Household Tracking Form. This form identified the status and location of all panel respondents (all previous household members). When possible, panel respondents still residing in baseline communities were contacted in order to collect information on the status (alive/deceased) and location of all surviving panel respondents with whom the respondent resided in KHDS 91-94. When none of the panel respondents could be located in the baseline community, this information was collected by an informant (either a neighbor, relative or village leader).

For respondents who had moved out of the baseline community, tracking information was collected on the Individual Tracking Form. The form contained information on the name, age, and sex of the person tracked. It also included area of residence, which was divided into country, region, district, ward, village and sub-village. In addition, their marital status, name of spouse, contact details, professional details, hang out places, other names used and physical characteristics were noted. Furthermore, information on potential informants was collected on the Informant Tracking Form, in the event that tracking information appeared unreliable or insufficiently detailed to allow for the tracking of the panel respondent.

Upon completion of the pre-field work tracking, data collected consisted of:

- Household Tracking Form which recorded the status of all panel respondents. If a panel respondent was reported to be alive, basic information on current location (same community, nearby community, district, region etc...) was recorded
- Individual Tracking Form for panel respondents who did not reside in or nearby the baseline community
- Informant Tracking Form for potential informants in the event that information on the individual tracking form proved insufficient to locate panel respondents who lived far from baseline communities.

The data collected were entered in the headquarters in Bukoba. These data were used to estimate the total number of households expected to be interviewed in KHDS 2004, considering migration and splitting of households. These data also allowed for careful planning of the main and tracking phases of the field work.

## ***4.3 Main Field Work***

The main field work started in January 2004. It consists of field team visits to the 51 baseline communities, as well as tracking of panel respondents who had moved to villages nearby the baseline communities. It excluded the tracking of panel respondents who had moved far either within the region or outside the region.

The field teams were divided into four groups of seven. One supervisor led each group. For the first two enumeration areas, teams were paired up in order to ensure data quality and consistency, as well as identify any over-sights in field procedures. For the remaining field work, each enumeration area was assigned one field team.

When the survey team arrived in a community, the supervisor met with the village chairman to introduce the team, and, if necessary, explain about the survey in more detail. The supervisor also collaborated with the village chairman to find accommodation and guides who are familiar with household locations in the village. They also compiled a list of respondents for the community questionnaire and planned actual dates for administering the questionnaire. Households in the village were grouped according to sub-villages. Appointments for administering the household questionnaire were then made with the household members. The interviewers completed the household questionnaire in two to three separate interviews depending on the size of the household and the number of sections that applied to the household.

Fieldwork supervision was done in several stages to ensure high quality data collection. Supervision included revisits to households and direct observation during interviews by field supervisors and management team. During the revisits, the supervisor re-administered some of the sections in the household questionnaire, took anthropometric measurements and crosschecked clinic cards for children under the age six to verify the validity of the data. This also enabled collection of missing household data and anthropometric measurements for household members who had not been available during the initial interview.

Questionnaire checks were done in four stages:

- First, interviewers checked their own questionnaire after every household interview.
- Second, interviewers exchanged questionnaires among themselves for further checking.
- Third, the supervisor checked the questionnaires.
- Finally, field management occasionally checked the questionnaires.

For the field work, each team was given pre-printed 1991-1994 Household Rosters and 1991-1994 Children Living Elsewhere Rosters. They used these to check and confirm identities, relations and identification codes of respondents, which link them to the 1991-1994 survey (including completion of Section 1 question 10 in the household questionnaire, Section 2, and the Network Roster Card ).

Supervisors reported to the main office once a week to give an update of the work. They also reported any problems or queries that arose which enabled the field management to develop addendums. Addendums to the field manual were produced and distributed in the first few months of the fieldwork to clarify some aspects of the fieldwork and questionnaires. The field teams returned to the main office after completion in every enumeration area.

#### ***4.4 Tracking***

The tracking phase started in June 2004 and ended in August 2004. During this phase, the teams were sub-divided into smaller teams of about three people. The size of the teams changed

according to the location and number of panel respondents who needed to be tracked. A team leader led each team, while a supervisor was responsible for monitoring several teams located close to each other.

During the tracking phase, field staff tracked panel respondents who migrated to areas far away from their baseline 1991-1994 dwelling. Panel respondents who had migrated to nearby villages were visited during the main fieldwork.

In some cases, when the field team arrived at the location on the Individual Tracking Form, the panel respondent had re-located. In this case, a second (or third, etc..) Individual Tracking Form would be completed and entered at operation headquarters.

## **5 Data**

The following section provides information on how the data are organized, what unique identifiers can be used to link data across sections of the questionnaire, and across different questionnaire types. It also provides information on data processing and cleaning, as well as linking records across surveys rounds.

### ***5.1 Data Processing***

Data entry was done at the main office in Bukoba, concurrent with the main fieldwork. The data entry team consisted of seven data entry operators and one data entry supervisor. Data was entered in CsPro then transformed to Stata format. Questionnaires were entered and verified after each entry. Although internal consistency checks were performed in CsPro, in addition to more elaborate checks for inconsistency and outliers were done in Stata.

All responses obtained from individual, household, and community level interviews were recorded in questionnaires. In cases where the respondent did not know the answer, the interviewers recorded “DK” (Don’t know) in the questionnaires. Data entry were trained to input this as nine (9) which represents missing information in the datasets. In cases where nine was an eligible code, the highest value for the number of digits was entered. For example, DK’s for questions with up to two eligible digit codes were entered as 99; 999 was entered for DKs for questions with eligible three digit codes (assuming 999 was not otherwise an eligible response).

For the mortality questionnaire, in some cases, multiple informants were interviewed. The data were consolidated such that each baseline household has one mortality questionnaire in the data files (with, perhaps, multiple deceased therein).

### ***5.2 Data File Structures***

The data are contained in many separate Stata data files. Table 10 gives an overview of the household data files, including the level of observation for each record in the data file and the

unique variables that identify each record. Most of the data are in a data file named such that one can identify the section to which the data pertain. An exception is “hh.dta” which is a file that pulls household-level data from across several sections. Table 10 gives an overview of the price data. As the school and community data sets contain only one data set each, they are not presented in tables here.

In addition to containing the data covered in the household questionnaire, the household data include a tracking data set which contains information on tracking efforts, including the various zones to which individuals were tracked and the final outcome of tracking (that is, a completed interview or otherwise).

**Table 10: KHDS 2004 Household Data**

Section	Topic	Data file	Level of Observation	Identification Variables				
				#1	#2	#3	#4	#5
00	Survey Information	sec00	Q'aire section	hhid2	sectnum2			
1	HH Roster & HH Roster Card	sec1	Individual	hhid2	id2			
1	Network Roster Card	sec18a	Network member	hhid2	netwkid2			
2	Previous Children Residing Elsewhere	sec2	Individual	hhid2	cluster	hh	cid	
2	Previous Children Residing Elsewhere	hh	Household	hhid2				
3	Main Activities	sec13	Business	hhid2	entprid2			
3	Main Activities	hh	Household	hhid2				
4	Information on Parents	sec4	Individual	hhid2	id2			
5	Education	sec5	Individual	hhid2	id2			
6	Health	sec6	Individual	hhid2	id2			
7	Labor	sec7	Individual	hhid2	id2			
8	Individ Exp (2 week and annual)	sec8	Individual	hhid2	id2			
9	Migration	sec9	Individual	hhid2	id2	cluster	hh	id
9	Migration	hh	Household	hhid2				
10	Shocks in the Last 10 Years	sec10	Individual	hhid2	id2			
10	Shocks in the Last 10 Years	hh	Household	hhid2				
11	Farming	hh	Household	hhid2				
11A	Land	sec11a	Plot of land	hhid2	shambaid2			
11B	Crops	sec11b	Crop	hhid2	cropid2			
11C	Farm Inputs	sec11c	Input	hhid2	inputid2			
11C	Farm Inputs	hh	Household	hhid2				
11D	Sales of Products From Home Grown Crops	hh	Household	hhid2				
11D	Sales of Products From Home Grown Crops	sec11d	Processed crop	hhid2	crprodid2			
11E	Hand Tools	sec11e	Hand tool	hhid2	toolid2			
11F	Farm Equipment	sec11f	Equipment	hhid2	fequipid2			
12	Animals	hh	Household	hhid2				
12A	Animals	sec12a	Livestock	hhid2	lvstckid2			
12B	Sale of Animal Products	hh	Household	hhid2				
12B	Sale of Animal Products	sec12b	Animal product	hhid2	amprdid2			
12C	Livestock Expenditures	sec12c	Livestock exp item	hhid2	lvsinpid2			
13	Non-Farm Enterprise	sec13	Business	hhid2	entprid2			
14A	Type and Ownership of Dwelling	sec14a	Dwelling	hhid2	bldngid2			
14A	Type and Ownership of Dwelling	hh	Household	hhid2				
14B	Housing Expenditures	hh	Household	hhid2				
15A	Durable Goods	sec15a	Durable good	hhid2	dlinenum2			
15A	Durable Goods	hh	Household	hhid2				

**Table 10: KHDS 2004 Household Data**

Section	Topic	Data file	Level of Observation	Identification Variables				
				#1	#2	#3	#4	#5
15B	Household Annual Expenditures	sec15b	Expenditure item	hhid2	hhexpid2			
15C	Household Two-Week Expenditures	sec15c	Expenditure item	hhid2	hhexpid2wk			
15D	Inheritance or Bride Price	hh	Household	hhid2				
15D	Inheritance or Bride Price	sec15d1	Inheritance	hhid2	inherid2			
15D	Inheritance or Bride Price	sec15d2	Bride price	hhid2	bridepid2			
16A	Seasons of the Past 12 Months	hh	Household	hhid2				
16B	Food Consumption of Home Production	sec16b	Food item	hhid2	typfood2			
16C1	Food Expenditures, Seasonal Foods	sec16c1	Food item	hhid2	typsfood2			
16C2	Food Expenditures, Non-Seasonal Foods	sec16c2	Food item	hhid2	typnsfid2			
16C2	Food Expenditures, Non-Seasonal Foods	hh	Household	hhid2				
17A	Informal organizations	hh	Household	hhid2				
17A	Informal organizations	sec17a	Informal organizations	hhid2	groupid2			
17B	Ability to Cope	hh	Household	hhid2				
17C	Assistance from Outside Organizations	hh	Household	hhid2				
17C	Assistance from Outside Organizations	sec17c	Assistance organization	hhid2	assistid2			
18A	Network Members Interactions	sec18a	Network member	hhid2	netwkid2			
18A	Network Members Interactions	hh	Household	hhid2				
18B	Gifts/Loans Received from Others	sec18b	Other person	hhid2	giftfid2			
18B	Gifts/Loans Received from Others	hh	Household	hhid2				
18C	Gifts/Loans Given to Others	sec18c	Other person	hhid2	giftoid2			
18C	Gifts/Loans Given to Others	hh	Household	hhid2				
Separate Forms	Tracking Information	tracking		cluster	hh	id		
Separate Form	Anthropometry	anthro	Individual	hhid2	id2			
Separate Form	Anthropometry	hh	Household	hhid2				
Separate Form	Mortality of Previous Household Members	mortality1	1991-1994 Household	cluster	hh			
Separate Form	Mortality of Previous Household Members	mortality2	Deceased person	cluster	hh	id		

**Table 11: Price Data**

Topic	Data file	Level of Observation	Identification Variables		
			#1	#2	#3
Cover	prcvrcls	Market	cluster	mearplc	
Food	pr1__fdp	Food item	cluster	mearplc	foodid
Pharmacy	pr2__ph1	Phara. item	cluster	mearplc	pharmid
Nonfood	pr3__nf1	Nonfood item	cluster	mearplc	nfoodid
Charcoal	pr3__nf2	Market	cluster	mearplc	

### 5.3 Linking 2004 Data Sets

The community, price and primary school data are *only* relevant for households located in the vicinity. That is, these questionnaires were only administered in the original 51 enumeration areas (which are 49 unique communities). Households that are located in or near the baseline community can be identified by the question *si2c* on the first page of the household questionnaire. The enumeration area number for these households is the first two digits of the six-digit household identification number. For example: HHID 150105 has *si2c*=1, meaning that the household resides in the original sample community 15.

In a strict sense the community, price and primary school data can only be used for people living in the same village. Some households reside nearby, although not in the same community. Many of the variables collected at community level may be valid for people tracked nearby the original enumeration areas (variable *si2c* in *hh.dta* equal to 2). One can, in theory, link them to their baseline community data, although it is not necessarily the best community data to describe the community of that household, since some of these nearby communities were actually several kilometers away and in another village entirely.

### 5.4 Linking Individuals Over Time

In the KHDS 1991-1994 survey household identification was based on two-digit enumeration area number (*cluster*) and two-digit household number within the enumeration area (*hh*). Individuals in the household were assigned a person-ID number (equivalent to their roster line number) (*id*). Since a very small number of people during the baseline survey moved out of one panel household and into another, in order to uniquely identify people, each person is also assigned a 6-digit panel respondent (*pid91\_94*) which is almost always the combined of *cluster+hh +id*. *pid91\_94* uniquely identifies every person ever interviewed in the KHDS (be it one of the first four rounds of 1991-1994 or 2004).

Households in the 2004 survey were assigned 6-digit identification numbers (*hhid2*). Household identification numbers in 2004 were designed to allow the user to easily link back to the 1991-

1994 community and household. The first four digits of the 2004 household identification are the same as *cluster* and *hh* from 1991-1994. The last two digits number the 2004 household such that households with the same origin household are not given the same 6-digit identification code. In rare cases, two panel respondents from two different baseline (1991-1994) households now reside together. Thus, in these rare cases, *hhid2* may not refer to the KHDS 91-94 household for each panel respondent in that household.

Individuals can be linked back to their 1991-1994 data through data from Section 1 question 10 in the household questionnaire. The respondent’s identification number from the household roster of their baseline household is recorded in this question. The roster ID of a person in KHDS 2004 (*id2*) does not correspond with their roster ID in the KHDS 91-94 (*id*). Data from Section 1 question 10 must be used in order to link panel respondents to their KHDS 91-94 data. Section 1 question 10 appears as four variables in the data set, corresponding to the variables described above: *cluster*, *hh*, *id*, and *pid91\_94*.

For example, fictional household 571701 in the 2004 survey has 4 household members. Three of these household members were respondents in at least one of the KHDS 1991-1994 survey rounds.

<i>hhid2</i>	<i>id2</i>	Name	Section 1 Q9	Section 1 Q10			
				<i>cluster</i>	<i>hh</i>	<i>id</i>	<i>pid91_94</i>
571701	01	Johanna	1	57	17	04	571704
571701	02	James	1	57	17	05	571705
571701	03	Anna	2				
571701	04	Radhia	1	57	17	07	571707

Note that in a very small number of cases, a panel respondent may have moved into another panel households. In this case, Section 1 question 10 *cluster* won’t match the cluster imbedded in the household ID in KHDS 2004. For example, consider the fictional household where the third person is from a different KHDS 91-94 household than other members:

<i>hhid2</i>	<i>id2</i>	Name	Section 1 Q9	Section 1 Q10			
				<i>cluster</i>	<i>hh</i>	<i>id</i>	<i>pid91_94</i>
631001	01	Johnson	1	63	10	02	631002
631001	02	Mariana	1	63	10	03	631003
631001	03	Godlike	1	63	03	05	630305

In very few cases, a panel respondent could reside in two households at the same time. These are cases where two observations in Section 1 have the same *pid91\_94*. Variables in *sec1.dta* (including *s1q10\_oth* and *s1q10\_plgm*) explain the reasons why occurred, including:

- Section 1 question 8 is no: the person was listed on the roster by the household head but doesn't qualify as a household member by the stated criterion.
- The person moved to another sample household during the field work and qualifies as a household member in both households.
- The person was reported as the household head in one household (which automatically qualifies the person as a household member), although is actually residing in another location.
- The person is polygamous and maintains two separate households.

## Appendix A: List of KHDS 2004 Data Sets

**Table 12: KHDS Data Sets**

<b>Date File</b>	<b>Questionnaire</b>	<b>Description</b>
hh.dta	Household	Household-level information compiled across multiple sections
sec00.dta	Household	Survey information
sec1.dta	Household	Roster
sec2.dta	Household	Children residing elsewhere
sec4.dta	Household	Information on parents
sec5.dta	Household	Education
sec6.dta	Household	Health
sec7.dta	Household	Activities and Non-Labor Income
sec8.dta	Household	Individual expenditures
sec9.dta	Household	Migration
sec10.dta	Household	Shocks
sec11a.dta	Household	Land
sec11b.dta	Household	Crops
sec11c.dta	Household	Farm inputs
sec11d.dta	Household	Sales of products from home grown crops
sec11e.dta	Household	Hand tools
sec11f.dta	Household	Farm equipment
sec12a.dta	Household	Animals
sec12b.dta	Household	Sales of animal products
sec12c.dta	Household	Livestock expenditures
sec13.dta	Household	Non-farm self employment
sec14a.dta	Household	Types and ownership of dwelling
sec15a.dta	Household	Durable goods
sec15b.dta	Household	Household annual expenditures
sec15c.dta	Household	Household two-week expenditures
sec15d1.dta	Household	Inheritance
sec15d2.dta	Household	Bride price
sec16b.dta	Household	Food consumption of home production
sec16c1.dta	Household	Food expenditures, seasonal foods
sec16c2.dta	Household	Food expenditures, non-seasonal foods
sec17a.dta	Household	Informal organizations
sec17c.dta	Household	Receipt of assistance from outside

**Table 12: KHDS Data Sets**

<b>Date File</b>	<b>Questionnaire</b>	<b>Description</b>
sec18a.dta	Household	Interactions with network members
sec18b.dta	Household	Gifts and loans received from others
sec18c.dta	Household	Gifts and loans given to others
anthro.dta	Household	Anthropometrics
tracking.dta	Household	Tracking status of individuals not re-interviewed (not traced) in 2004 and reported to be alive
mortality1.dta	Mortality	Cover page
mortality2.dta	Mortality	Mortality data for deceased panel respondents
community.dta	Community	Community data
prcvrcls.dta	Community	Price data: Cover page
pr1__fdp.dta	Community	Price data: Food Items
pr2__ph1.dta	Community	Price data: Pharmaceuticals
pr3__nf1.dta	Community	Price data: Non-food items
pr3__nf2.dta	Community	Price data: Charcoal
primary.dta	Community	Primary school

## **Appendix B: List of Related Documents**

The following KHDS 2004 documents can be obtained from the World Bank, Living Standards Measurements Study, Development Research Group (DECRG). They can be downloaded from the LSMS website: <http://www.worldbank.org/lsms/>

### ***Questionnaires***

- Household Questionnaire (English and Swahili versions)
- Community Questionnaire
- Primary School Questionnaire
- Price Questionnaire

### ***Training Manuals***

- Interviewer Manual
- Supervisor Manual
- Community, Price and Primary School Questionnaires Manuals
- Cumulative Addendum to the Field Manual
- Tracking Addendum Manual

### ***Other Forms***

- Household Tracking Form
- Individual Tracking Form
- Informant Tracking Form
- Supervisor Direct Observation Form
- Supervisor Reinterview Form
- Supervisor Verification Form

## **Appendix C: Comparison of Key Variables with other Kagera Data Sets**

### ***Introduction***

As described in Section 3 of this document, the KHDS 2004 sample is non-random for two reasons: first, it is based on the KHDS 91-94 sample which was not a random sample, and second, there is some sample attrition. Further, households surveyed as part of KHDS 2004 were those that contained KHDS 91-94 respondents; therefore, it includes all households formed from the baseline sample of households.

These very specific conditions of the KHDS 2004 sample have implications for the extent to which KHDS 2004 is representative of the Kagera population. This appendix compares the KHDS 2004 data to that of the Kagera Rural CWIQ, a cross-sectional household welfare survey that was completed a month before the start of KHDS 2004 field work (November - December 2003). This survey collects data on many household and individual characteristics also covered in the KHDS 2004.

The Kagera Rural CWIQ Survey was funded by the Netherlands Government through its District Rural Development Programme in Kagera Region. The purpose of the survey was to conduct a study of poverty, welfare, as well as accessibility and quality of services at regional and district levels. The World Bank Core Welfare Indicators Questionnaire instrument was used as the survey tool. Information collected from the random sample of 2,250 households informs on basic trends in household demographics, education, health, child nutrition, employment, as well as utilization of and satisfaction with social services. Data analysis and report writing were carried out by EDI and released in the Kagera Rural CWIQ Report in April 2004 (*Kagera Rural CWIQ: Baseline Survey on Poverty Welfare and Services in the Kagera Rural Districts*; Tanzania-Netherlands Development Co-operation: The District Rural Development Programme; April 2004. Available from the EDI website: [www.edi-africa.com/research/cwiq.html](http://www.edi-africa.com/research/cwiq.html))

While both KHDS 2004 and CWIQ surveys cover some same basic indicators, such as adult literacy rates, school enrolment rates, rates of morbidity and others, it is important to be aware of some of the key differences between the two surveys:

- The CWIQ survey was administered to a random sample of households, while KHDS 2004 used a non-random sample, as discussed above.
- The CWIQ survey was only administered in rural districts of Kagera Region. The KHDS 2004 sample included households from all the districts in the region as well as regions beyond Kagera to which some panel respondents had migrated.
- The CWIQ questionnaire is substantially shorter than KHDS 2004 questionnaire. It took an average of 30 minutes to complete one CWIQ questionnaire, compared to between 4 and 7 hours for a KHDS 2004 questionnaire (often conducted in 2 or 3 visits).
- The whole of the CWIQ interview was conducted with one respondent – the head of household or the most knowledgeable person in the household. In contrast, the respondents in KHDS 2004 differed depending on the content of the section. For

instance, while in the CWIQ, the household head / most knowledgeable person provided individual information on every member of the household, in KHDS 2004, to the extent possible, each individual in the household provided information about him/herself.

Discussion in this section will focus on the similarities and differences between the summary statistics for the CWIQ and the Kagera Rural sub-sample of KHDS 2004. In sum, statistics for the KHDS 2004 sub-sample of households located in Kagera Rural tend to be similar to those for the CWIQ. As the KHDS 2004 sub-sample is broadened to include all households in the Kagera Region more differences can be observed; these differences increase further with the inclusion of the whole sample including households in KHDS 2004 residing outside Kagera.

### ***Distribution of Households by District***

Table 13 shows that proportions of households located in Bukoba Rural, Karagwe and Biharamulo districts differ substantially between the CWIQ and KHDS 2004 surveys. While more than two fifths (42 percent) of households in the Kagera Rural sub-sample of KHDS 2004 are located in Bukoba Rural district, this is the case for only about a quarter of households in the CWIQ. In contrast, proportions of CWIQ households located in Karagwe and Biharamulo districts exceed those of Kagera Rural KHDS 2004 households by just under 10 percentage points. This is probably in large part the direct result of the first step in the KHDS 91-94 stratification strategy (See Section 3).

**Table 13: Distribution of Households by District**

	CWIQ Kagera Rural <sup>2</sup>	KHDS		
		Kagera Rural <sup>1</sup>	Kagera	All <sup>3</sup>
Bukoba Urban	NA	NA	20.0	17.7
Bukoba Rural	24.1	42.4	32.9	29.2
Karagwe	23.7	16.8	13.4	11.9
Muleba	21.1	19.4	15.8	14.0
Biharamulo	17.9	8.9	7.3	6.4
Ngara	13.1	12.8	10.5	9.3

*Notes:* 1) This data excludes households located in Bukoba Urban District. 2) While KHDS 2004 results are based on unweighted data, CWIQ data is weighted to be representative of Kagera Region, excluding Bukoba Urban District. 3) Column does not sum to 100 as some households reside outside of Kagera region.

### ***Basic Household Characteristics***

Table 14 presents statistics on basic household characteristics such as household size and characteristics of the household head. As can be seen, the gender distribution of household heads is almost the same according to the findings of the two surveys; it was found in both KHDS 2004 and CWIQ that roughly 80 percent of households in Kagera Rural are headed by men. Similarly, there is almost no difference in mean age of the household heads. In contrast, mean household

size, distribution of household heads by marital status and the rate of literacy among the household heads are different across the two surveys.

On average, KHDS 2004 households are significantly smaller than CWIQ households. Further, a significantly higher proportion of household heads were married at the time of the survey according to the results of the CWIQ, compared to those of KHDS 2004, at 77 and 72 percent respectively. The CWIQ reported a lower rate of separation among household heads than those of KHDS 2004, a higher proportion of household heads were divorced at the time of the survey according to the CWIQ. This difference, however, may be due to the ambiguity of the definition used. Finally, the literacy rate among household heads is significantly lower in the CWIQ than those of KHDS 2004, at 69 and 77 percent respectively.

**Table 14: Basic Household Characteristics**

	CWIQ Kagera Rural <sup>2</sup>	KHDS		
		Kagera Rural <sup>1</sup>	Kagera	All
Household size	5.2	4.8	4.6	4.6
Female headed households	18.2	21.0	21.6	21.5
Mean age of household head	43.1	42.3	42.2	41.4
Marital status of household head				
Not married	4.0	6.1	7.5	8.9
Married	77.4	71.9	70.0	69.4
Divorced	3.2	0.8	1.0	1.1
Separated	3.1	5.7	5.9	6.1
Widowed	12.2	15.3	15.4	14.4
Literacy of the household head	68.5	77.1	79.4	80.8

*Notes:* 1) This data excludes households located in Bukoba Urban District. 2) While KHDS 2004 results are based on unweighted data, CWIQ data is weighted to be representative of Kagera Region, excluding Bukoba Urban District.

### ***Land and Livestock Ownership***

The results of KHDS 2004 and CWIQ surveys show identical trends in land ownership (Table 15). According to both surveys, on average households in Kagera Rural own 3 acres of land (this average includes landless households). Further, households in this area use an average of about half an acre of land that is not owned by the household.

There are some significant differences in livestock ownership trends. While proportions of households holding medium and large livestock according to CWIQ and KHDS 2004 are almost equal, mean amounts of both large and medium livestock held are significantly higher according to the CWIQ than KHDS 2004. For instance, as can be seen in Table 15, while the CWIQ reports an average holding of almost 8 large livestock among owners of large livestock, according to KHDS 2004 this number is just over 5. Similarly, the results of the CWIQ show that owners of medium livestock hold an average of nearly 2 more medium livestock than reported in KHDS

2004. It should be noted, however, that the structure of the questions informing on livestock holdings is quite different in the CWIQ questionnaire than that used for KHDS 2004; whereas KHDS 2004 lists each type of livestock, CWIQ divides livestock owned into only 2 categories.

In addition to average livestock holdings, distribution of households by change in livestock holding is also different according to the findings of CWIQ and KHDS 2004 surveys. Again it should be noted than the structure of the questions used to obtain this information differs between the two surveys. Nevertheless, the proportions of households citing an increase in the number of large and medium livestock held over the year preceding the survey are roughly 10 percentage points higher according to the CWIQ than KHDS 2004. Proportions of households reporting no change or a decrease in numbers of large and medium livestock held are between 4 and 6 percentage points lower according to the CWIQ than KHDS 2004.

**Table 15: Household Land and Livestock Holdings**

	CWIQ Kagera Rural <sup>2</sup>	KHDS		
		Kagera Rural <sup>1</sup>	Kagera	All
<b>Land Ownership (acres)</b>				
Land owned	3.0	3.2	3.1	3.1
Land used, but not owned	0.4	0.5	0.5	0.5
<b>Livestock Ownership</b>				
% of households holding large livestock	12.8	11.0	10.6	10.0
Mean amount of large livestock held (among owners of large livestock)	7.7	5.2	5.2	5.3
% of households holding medium livestock	43.0	41.6	36.7	33.8
Mean amount of medium size livestock held (among owners of medium livestock)	5.9	4.4	4.5	4.5
<b>Amount of large cattle owned the year of the survey compared to the year preceding the survey (%)</b>				
More now	40.4	29.2	29.7	30.7
Same now	30.1	35.2	32.8	33.7
Less now	29.5	35.6	37.5	35.6
<b>Amount of medium cattle owned the year of the survey compared to the year preceding the survey (%)</b>				
More now	42.4	33.7	32.7	32.5
Same now	23.3	26.7	27.7	28.2
Less now	34.2	39.6	39.6	39.3

*Notes:* 1) This data excludes households located in Bukoba Urban District. 2) While KHDS 2004 results are based on unweighted data, CWIQ data is weighted to be representative of Kagera Region, excluding Bukoba Urban District.

### ***Household Dwelling and Amenities***

Table 16 presents the findings of the CWIQ and KHDS 2004 surveys regarding characteristics of dwellings and selected household amenities. The results of the two surveys show little differences in proportions of households owned by a member of the household and mean number of bedrooms in dwellings. Similarly, no substantial differences are observable between distributions of households by building materials used for the roof, walls and floor of the dwellings.

There are, however, some significant differences in the findings of the CWIQ and KHDS 2004 on the distribution of households in Kagera Rural by source of drinking water, type of toilet and source of light.

According to KHDS 2004, more than three fifths of the households in the area use drinking water from rivers, lakes or ponds. According to the CWIQ, this proportion constitutes roughly a half (48 percent) of the households. In contrast, while in KHDS 2004 only a quarter of the households were found to use water from unprotected wells or rainwater, according to the CWIQ 37 percent of households are in this category. The results of both surveys indicate that less than 2 percent of the households in Kagera Rural use water piped directly into their dwelling or compound.

Further, results of KHDS 2004 show that nearly twice as high a proportion of households in Kagera Rural have no toilets than reported by the CWIQ, at 8 and 4 percent respectively. In contrast, the proportion of households with a pit latrine is 4 percentage points higher according to the CWIQ than the KHDS; although this difference is not substantial, it is statistically significant.

Finally, while kerosene, paraffin and gas were found to be used as the main source of lighting by the great majority of households in Kagera Rural, these were slightly more widespread according to the results of the CWIQ than KHDS 2004. In contrast, the proportion of households using electricity as a source of lighting is significantly higher according to KHDS 2004 than the CWIQ, at 4 and 2 percent respectively.

### ***Other Household Characteristics***

The results of both surveys show almost identical trends in household asset ownership and food consumption trends (Table 17). The largest difference observable is in proportions of households found to possess watches and jewelry, which is slightly higher according to the CWIQ than KHDS 2004. However, even this difference, although statistically significant, is only 5 percentage points.

**Table 16: Household Dwelling and Amenities**

	CWIQ Kagera Rural <sup>2</sup>	KHDS		
		Kagera Rural <sup>1</sup>	Kagera	All
<b>Dwelling</b>				
% of households who own their dwellings	92.6	89.5	85.2	80.7
Number of bedrooms in the household	2.5	2.3	2.3	2.3
<b>Main roofing material (%)</b>				
Grass/Thatch	40.4	37.0	32.1	30.3
Iron	57.4	61.0	65.4	67.2
Other (Mud, Wood/Planks, Roofing Tiles, Asbestos)	2.3	2.0	2.3	2.3
<b>Main construction material of outside walls (%)</b>				
Stone/Burnt Bricks/Cement	16.5	16.7	6.7	7.9
Iron Sheets	0.1	0.4	0.6	0.6
Mud/Mud bricks	77.3	79.3	72.0	67.8
Wood/Bamboo	5.0	0.7	0.7	0.7
Other	1.1	2.9	2.8	2.5
<b>Main flooring material</b>				
Concrete	14.5	14.7	21.1	26.0
Mud	85.2	84.7	77.8	72.8
Other	0.2	0.6	1.0	1.2
<b>Source of drinking water</b>				
Piped into dwelling/compound	1.1	1.2	3.0	4.3
Neighbor's house	3.0	0.4	1.6	2.2
Well/Rain water	37.3	25.7	22.1	21.6
Public outdoor tap or borehole	10.4	8.8	7.9	8.4
River, lake, pond	47.9	61.6	61.3	56.5
Other (Vendor/truck)	0.2	2.1	3.7	6.8
<b>Type of toilet</b>				
Pit latrine	94.1	90.0	89.1	88.2
Flush	1.5	1.6	3.5	4.9
Other	0.1	0.1	0.1	0.1
None	4.3	8.2	7.2	6.8
<b>Type of fuel used for cooking</b>				
Charcoal	9.7	7.0	12.7	17.2
Electricity	0.0	0.3	0.3	0.7
Firewood	90.3	91.1	85.0	79.1
Other (gas, Kerosine/oil, Biogas)	0.0	1.3	1.6	2.3
<b>Main source of lighting</b>				
Electricity – Mains/Generator	1.5	4.0	8.7	12.4
Kerosine/parrafin/oil/gas	97.0	93.1	88.8	85.1
Other (candles, battery/solar, none)	1.4	2.9	2.5	2.6

*Notes:* 1) This data excludes households located in Bukoba Urban District. 2) While KHDS 2004 results are based on unweighted data, CWIQ data is weighted to be representative of Kagera Region, excluding Bukoba Urban District.

**Table 17: Other Household Characteristics**

	CWIQ Kagera Rural <sup>2</sup>	KHDS		
		Kagera Rural <sup>1</sup>	Kagera	All
<b>Asset Ownership (%)</b>				
Cars/other vehicles	0.9	0.9	1.8	1.9
Motorbikes	2.2	1.9	1.9	2.1
Video equipment / television	1.1	1.6	3.4	5.7
Bicycles	41.8	39.2	37.9	36.9
Radio / Cassette / Record / CD players	54.5	56.7	59.7	61.4
Telephone (mobile / landline)	5.0	4.3	6.8	10.3
Watches / Jewelry	46.8	42.4	47.3	49.5
Iron	19.6	17.3	21.1	24.8
% of households with a bank account	7.8	8.8	11.3	14.0
<b>Food Consumption</b>				
Number of meals consumed per day	2.2	2.1	2.1	2.1
Number of times meat is consumed per week	0.6	0.6	0.7	0.8

Notes: 1) This data excludes households located in Bukoba Urban District. 2) While KHDS 2004 results are based on unweighted data, CWIQ data is weighted to be representative of Kagera Region, excluding Bukoba Urban District.

### ***Individual Characteristics***

Table 18 shows the disaggregation of KHDS 2004 and CWIQ data by individual characteristics. The distributions of the population by gender, mean age, and disability status were found to be almost identical; none of the differences between proportions of individuals in these categories exceed 1 percentage point. Distribution of the population by marital status is also, as can be seen, very similar according to the CWIQ and KHDS 2004. A slight discrepancy is, however, noticeable in rates of divorce and separation. According to the CWIQ, separation is almost as widespread as divorce; these describe the marital status of, respectively, 2 and 3 percent of the adult population. According to KHDS 2004, separation is much more widespread than divorce, at 5 and 1 percent respectively. As mentioned previously, these differences could, in part, reflect the ambiguity of the definition.

Education trends reported by the CWIQ and KHDS 2004 surveys differ more significantly. Overall, levels of literacy and schooling rates are consistently and significantly higher according to the results of KHDS 2004 than the CWIQ. The most substantial difference between the two surveys is between proportions of men and women over the age 14 who had ever attended school. These differ by roughly 10 percentage points. Similarly, proportions of children between the ages of 6 and 14 who had attended school at some point are significantly higher according to KHDS 2004 than the CWIQ, as are those of literate men and women; differences in these rates are, however, smaller. Finally, disaggregation of individuals over the age of 19 by highest grade completed also shows some differences. While according to the CWIQ the proportion of adults who had terminated formal education at completion of primary school was higher than that found

in KHDS 2004, at 73 and 68 percent respectively, all subsequent categories contain a slightly higher proportion of adults according to KHDS 2004.

**Table 18: Individual Characteristics**

	CWIQ Kagera Rural <sup>2</sup>	KHDS			
		Kagera Rural <sup>1</sup>	Kagera	panel	all
% Female	50.7	51.7	51.6	51.9	51.5
Age	20.5	21.2	21.4	30.8	21.3
Marital Status (%) <sup>3</sup>					
Never married	26.9	26.2	27.8	31.9	28.8
Married	61.6	59.8	58.1	52.7	57.6
Divorced	2.7	0.9	1.0	0.8	0.9
Separated	2.0	4.6	4.7	5.5	4.7
Widowed	6.7	8.3	8.3	8.9	7.8
Literacy (%) <sup>3</sup>					
Males	76.2	83.0	85.2	85.3	86.2
Females	63.4	69.2	72.4	72.9	74.1
Any schooling (%)					
Boys 6-14 years	74.5	79.2	80.7	95.1	81.4
Girls 6-14 years	73.1	76.6	78.7	93.4	78.9
Males 15+ years	76.9	86.8	88.4	88.7	89.2
Females 15+ years	64.2	73.1	75.6	76.9	77.1
Highest grade (among individuals who had gone to school and are age 20+)					
Some primary	20.2	22.9	21.3	22.0	19.9
Complete primary	72.5	67.7	67.3	66.1	66.2
Some lower secondary (Form 1-4)	1.4	2.5	2.8	3.4	3.3
Complete lower secondary (Form 1-4)	5.5	6.1	7.5	7.2	9.1
Some upper secondary (Form 5-6)	0.0	0.1	0.1	0.2	0.2
Complete upper secondary (Form 5-6)	0.4	0.7	1.0	1.2	1.5
Disabled	2.1	2.3	2.4	3.2	2.3

Notes: 1) This data excludes households located in Bukoba Urban District. 2) While KHDS 2004 results are based on unweighted data, CWIQ data is weighted to be representative of Kagera Region, excluding Bukoba Urban District. 3) Martial status and literacy are computed for individuals 15+ years.