Kenya Trip Report

Prepared by: Thomas Ondra **Date Prepared: February 14, 2001**

Site Visited: Nairobi, Kenya **Submitted to: Joanne Jeffers**
USAID/G/PHN/POP/PE

Period Covered: January 21 to February 11, 2001

Trip Purpose and Summary

At the request of USAID/Kenya and the Kenyan Central Bureau of Statistics (CBS), Thomas Ondra traveled to Nairobi, Kenya, to assist the CBS in the development of computer programs for the 2000 Post-Enumeration Survey (PES) and in the review the computer programs used to run the edit specifications and the tabulations for the 1999 Population and Housing Census (PHC).

Upon arrival, Ondra and Armando Levinson, Mathematical Statistician of the U.S. Census Bureau met with David Nalo, CBS Director; J. Kekovole, Census Project Manager; F. Munene, CBS Deputy Director; P. W. Nyongesa, Chief of Data Processing Department; and Fred Otieno, Demographer, to discuss the status of census and PES operations and the details of the work.

The CBS conducted the 1999 Population and Housing Census (PHC) in August 1999, with August 24-25, 1999, as the reference night and the PES in February 2000. They released the census report during the first week of Ondra's visit. Ondra worked with Levinson and prepared computer programs to reformat and analyze the PES data, based on specifications provided by Levinson.

Trip Details

Post-Enumeration Survey (PES)

The CBS took a sample of 529 enumeration areas (EAs) for the PES. At the time of Ondra's visit, they had entered data for 508 EA's. They plan to complete data entry for the remaining EA's by the end of February 2001.

The International Programs Center – Population Division – US Census Bureau – Department of Commerce -
Washington, DC 20233-8860
Phone: (301) 457-1444 – Fax: (301) 457-3033 – E-mail: ipc@census.gov
www.census.gov

G. Sierra - Zorita

In a previous visit, Levinson recommended a format for the PES questionnaire. This form listed the matched census data on a line immediately following the data collected in the PES. (See Kenya Trip Report, December 1999, Levinson, U.S. Census Bureau.) The questionnaire used did not follow this recommendation and instead placed matched census data on a separate page from the corresponding PES data. Because of this, CBS had to manually write matched census line numbers on the PES page. This introduced the possibility of transcription errors and significantly increased the chance of data entry errors.

The data file, as entered, placed the data from the outmover records at the end of PES records thus combining PES data with outmover data. For processing, outmover records and PES records needed to be considered separately. Ondra worked with Sammy G. Kinyanjui, Senior Statistical Officer, Census Data Processing, and wrote IMPS data dictionaries and CONCOR programs to reformat data file from two record types (PES combined with outmover data, census data) to three record types (PES data, outmover data, census data).

During Ondra's visit, the PES data had not yet been edited and problems still exist in the data such as census records matched both to PES records and outmover records (resulting in double counting) and unmatched census records. CBS has prepared some very limited edit specifications for the PES. These specifications need to be refined and the PES edits need to be completed prior to finalizing the PES analysis. Ondra and Kinyanjui prepared a CONCOR edit program to list questionnaires containing these problems. The program is designed so that additional edits can easily be inserted when specifications are complete.

A special file needed to be created for input to CENVAR. Ondra worked with Levinson and Kinyanjui and prepared CONCOR programs to reformat the PES data, based on specifications provided in Levinsons's trip report. (See Kenya Trip Report, February 2001, Levinson, U.S. Census Bureau.) Ondra made some minor additions to the reformatted file. The specifications for the reformatted file can be found in attachment B. A flowchart showing the flow of the programs and data to create this file can be found in attachment C.

The weights to be applied to the data are in Excel files. The Excel file can be used to produce a text file (*.prn) in a format that can be used as a CONCOR lookup file. The specifications for this file are in the PESMatch.CN CONCOR program. This program reads the weights from the Excel text file (*.prn) and adds them to the data file. Ondra demonstrated this procedure to Kinyanjui.

Levinson's trip report specified three series of tables for the analysis to be prepared using the IMPS CENTS module: the "I" series (initial tables) containing 9 tables, the "A" series (Estimates of Coverage Measures) containing 5 tables, and the "B" series (Content Error Estimates) containing 5 tables. Ondra and Kinyanjui developed the CENTS format file and the CENTS TAB programs to produce these tables.

The final tables are to produce weighted data. For purposes of program development and testing, the weight variable (WGT) in the CENTS programs has been assigned a value of one (1) (unweighted). After the CENTS programs have been tested, the weight variable should be assigned a value for the weight field in the data file.

Census Edits and Tabulations

Ondra reviewed the census edit programs. He noted that the CONCOR programs did not make use of the frequency reports. Frequency reports show the allocation of values of imputed data. By reviewing these reports the user can determine if the imputations are introducing bias into the data. Ondra did add frequency reports to the CONCOR reports for some key variables (age, sex, relationship). The reports did not indicate any significant bias being introduced by the edits.

Analysis for sources of error is important to minimize non-sampling errors for future Censuses and Surveys. This census marks the first time that optical character recognition (OCR) was used for a Kenya census. Ondra recommends that a sample of the questionnaires be taken to compare with the raw data file to determine the rate of data misreads. The edit statistics indicate that the majority of data errors are blanks in the data. This is commonly caused by OCR misreads. See attachment D to for a report on one province listing a summary count of blanks where data should be.

CBS released the first volume of census tabulations during Ondra's visit.

Recommendations

CBS has not yet completed the data edits for the PES. Edit specifications and the corresponding CONCOR programs need to be completed for the analysis of the PES. When the CONCOR programs are run, CBS staff should carefully review the edit statistic reports.

The output from the CENTS tabulation programs should be carefully reviewed for accuracy and consistency.

Analysis of the sources of error is important for future work. CBS should sample check, comparing questionnaires with the data files to determine error sources such as enumeration errors, transcription errors, or scanning errors.

CBS should carefully document and review its procedures for future censuses and surveys.

Attachments
cc:

T. Takona	USAID/Kenya
D. Vogel	USAID/Kenya
D. Nalo	CBS Director
F. Munene	CBS Deputy Director
J. Kekovole	Census Project Manager
F. Otieno	CBS Demographer
P. W. Nyongesa	Chief of Data Processing Department
G. Sierra-Zorita	USAID/AF/E, Desk Officer
J. M. Anderson	USAID/AF/E, Director
J. Jeffers	USAID/G/PHN/POP/PE

ATTACHMENT A

Persons Contacted

USAID/Nairobi

Mailing Address: Box 30261
Nairobi, Kenya

Street Address: ICIPE Complex
Kasarani
Nairobi, Kenya

Telephone: (254-2)-862-400; 862-402
Fax: (254-2)-862-949

Timothy Takona, Project Manager, Health, Population, and Nutrition Officer
E-mail: ttakona@usaid.gov ([redacted])
Michael Strong, mstrong@usaid.gov

Central Bureau of Statistics (CBS)

Street Address: Herufi House (between Central Bank and Ministry of Finance)
Nairobi, Kenya

Mailing Address: P. O. Box 30266
Nairobi, Kenya

Telephone: (254-2)-333-970
Fax: (254-2)-333-030
E-Mail: Not available

Census Operations Nyayo House (Near the InterContinental Hotel)

David Nalo, CBS Director (Phone: 340-929)
Francis M. Munene, Deputy Director (cellular phone 072-751-737)
Dr. Kekovole, Census Project Manager
(Office 17th floor, Room 1720; phone: 22-43-85; 21-29-55; fax: 21-61-34)
P. W. Nyongesa, Chief of Data Processing Department
Sammy G. Kinyanjui, Senior Statistical Officer, Census Data Processing
Fred Otieno, Demographer (Fredrick_otieno@hotmail.com)
Sammy L. Oyombe, Economist Statistician I, National Sample Survey and Evaluation Program (NASSEP), Division of Field Administration, NASSEP, and Cartography
Collins Opiyo, Demographer (Collins_opiyo@hotmail.com)
Henry Kuria, Assistant Statistical Officer
Rosemary Kong'ani, Statistical Officer
Omondi Omenya, Statistical Officer I

United Nations Fund for Population Activities (UNFPA)

Mailing Address:	P. O. Box 8714 Addis Ababa, Ethiopia
Telephone:	251-1-51-12-88, 51-31-34
Fax:	251-1-51-71-33
Telex:	21240

Jean Marc Hie, Regional Advisor in Data Processing (jm_hie@hotmail.com) (hie@unfpa.org)
Jason Ondora Osembe, Regional Advisor in Survey Methods Sampling (jonsembe@hotmail.com) (onsembe@unfpa.org)

ATTACHMENT B

PES Data Dictionary Specifications (CENVAR)

Variables names written in *italics* are listed in parentheses below; the variables names are the names used in the corresponding IMPS data dictionaries and programs. Variable names written in *italics* with no other variable name in the cell are variable added to the original specification for special use in the computer processing of the data.

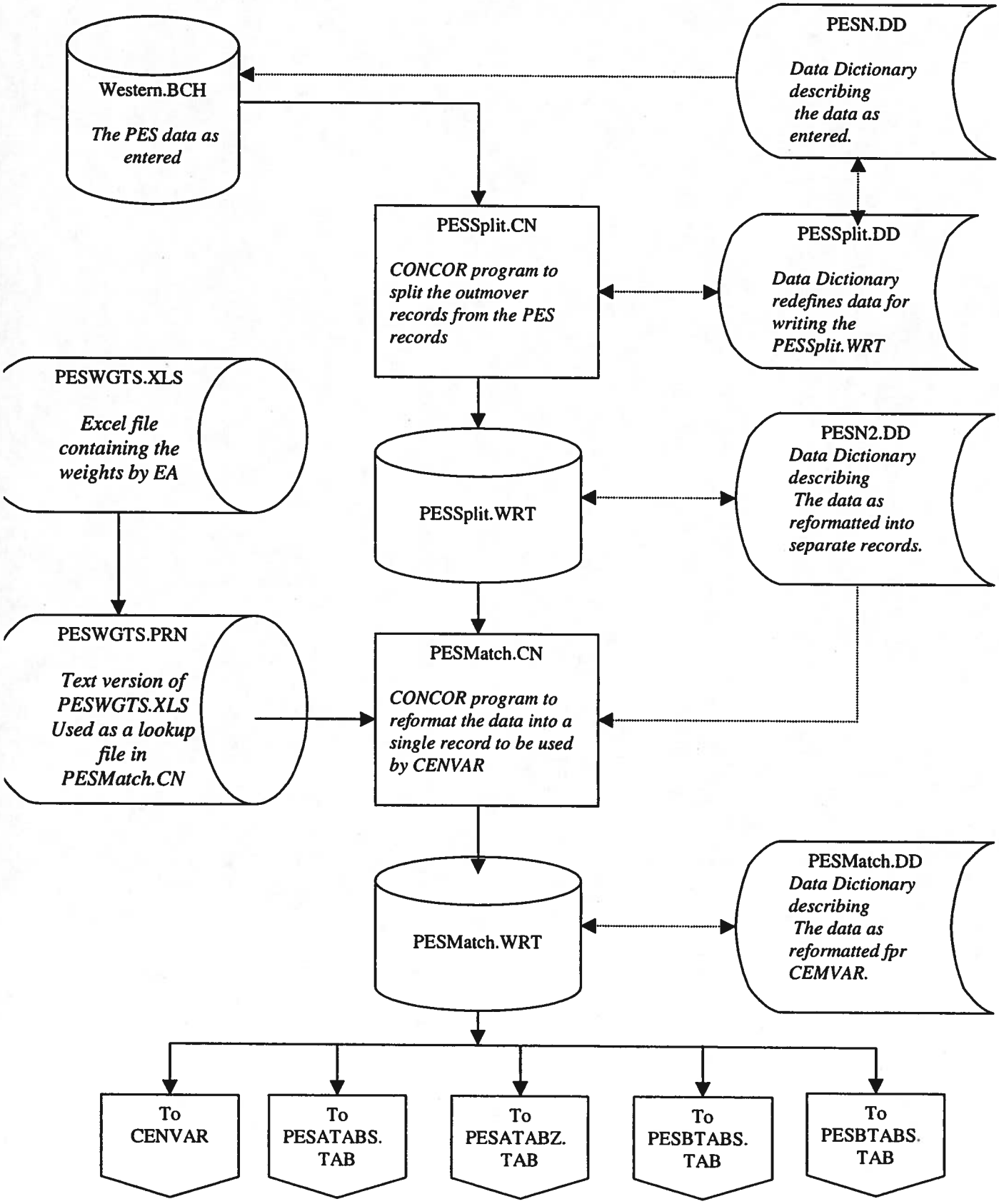
Variable Name	Data Type	Position	Length	Value and Value Name
STRATUM	N	1	1	1 = Nairobi 2 = Central 3 = Coast 4 = Eastern 5 = Northeastern 6 = Nyanza 7 = Rift Valley 8 = Western
EA *(CLUSTER)*	N	2-3	2	*1 to n within stratum*
WEIGHT	N	4-10	7 (4 dec.)	
RELATIONSHIP-P-SAMPLE *(P10-RELATIONSHIP)*	N	11	1	1 = Head 2 = Spouse 3 = Son 4 = Daughter 5 = Brother/Sister 6 = Father/Mother 7 = Other relative 8 = Non-relative 9 = NS/DK
RELATIONSHIP-E-SAMPLE *(E10-RELATIONSHIP)*	N	12	1	1 = Head 2 = Spouse 3 = Son 4 = Daughter 5 = Brother/Sister 6 = Father/Mother 7 = Other relative 8 = Non-relative 9 = NS/DK
SEX-P-SAMPLE *(P11-SEX)*	N	13	1	1 = Male 2 = Female
SEX-E-SAMPLE *(E11-SEX)*	N	14	1	1 = Male 2 = Female
AGE-P-SAMPLE *(P12-AGE)*	N	15-16	2	1 = 0 to 4 years 2 = 5 to 9 years 3 = 10 to 19 years 4 = 20 to 29 years 5 = 30 to 39 years 6 = 40 to 49 years 7 = 50 to 59 years

				8 = 60 years and above
AGE-E-SAMPLE *(E12-AGE)*	N	17-18	2	1 = 0 to 4 years 2 = 5 to 9 years 3 = 10 to 19 years 4 = 20 to 29 years 5 = 30 to 39 years 6 = 40 to 49 years 7 = 50 to 59 years 8 = 60 years and above
MARITAL-STATUS-P-SAMPLE *(P15-MARITAL-STAT)*	N	19	1	1 = Never married 2 = Monogamous 3 = Polygamous 4 = Widow/er 5 = Divorced 6 = Separated 7 = NS/DK
MARITAL-STATUS-E-SAMPLE *(E15-MARITAL-STAT)*	N	20	1	1 = Never married 2 = Monogamous 3 = Polygamous 4 = Widow/er 5 = Divorced 6 = Separated 7 = NS/DK
RELIGION-P-SAMPLE *(P14-RELIGION)*	N	21	1	1 = Catholic 2 = Protestant 3 = Other Christian 4 = Muslim 5 = Traditionalist 6 = Other 7 = No Religion 8 = NS/DK
RELIGION-E-SAMPLE *(E14-RELIGION)*	N	22	1	1 = Catholic 2 = Protestant 3 = Other Christian 4 = Muslim 5 = Traditionalist 6 = Other 7 = No Religion 8 = NS/DK
NON-MOVER *(X01-NON-MOVER)*	N	23	1	1=Non-Mover 0=Otherwise
OUT-MOVER *(X02-OUT-MOVER)*	N	24	1	1=Out-Mover 0=Otherwise
IN-MOVER *(X03-IN-MOVER)*	N	25	1	1=In-Mover 0=Otherwise
MATCHED-NM *(X04-MATCHED-NM)*	N	26	1	1=Matched-NM 0=Otherwise
MATCHED-OM *(X05-MATCHED-OM)*	N	27	1	1=Matched-OM 0=Otherwise
MATCHED-IM *(X06-MATCHED-IM)*	N	28-31	4 (3 dec.)	0000:1000

(Obtain the match rate for outmovers by sex and age group, which is expressed as a percentage, and apply that rate divided by 100 to a person who is an inmover.)				
TOTAL MATCHED=(M-NM) + (M-IM) *(X07-TOTAL-MATCHD)*	N	32-35	4 (3 dec.)	1000:2000
ERRONEOUSLY INCLUDED *(X08-ERR-INCLUDED)*	N	36	1	1=Erron. Included 0=Otherwise
CORRECTLY ENUMERATED, MISSED IN PES *(X09-CE-MISS-PES)*	N	37	1	1=Corr. Enumerated 0=Otherwise
CENSUS POPULATION = (M-NM) + (M-OM) + (ERR INC)+ (CORR. ENUM.) *(X10-CENSUS-POP)*	N	38	1	1 0
PES POPULATION = Non-Mover + In-Mover *(X11-PES-POP)*	N	39	1	1 0
OMISSIONS = PES-POP - (M-NM) - (M-IM) = PES-POP - (TOTAL-MATCHED) *(X12-OMMISSIONS)*	N	40-43	4 (3 dec.)	0000:1000
X13-CENMATCH-FLG *This field was added so programs using this file can determine if the record contains PES data, Outmover data, and/or census data.*	N	44	1	0=Census record never matched to PES or outmover record. 1=PES record matched to a census record 2=PES record with no matched census record 3=Outmover record matched to a census record 4=Outmover record with no matched census record 5=PES record and outmover record matched to same census record
X15-SEQNO *This is a sequential number assigned to each person in the cluster to identify a single person as a questionnaire. It is used to eliminate the "questionnaire processing in parts" warning messages.*	N	45-50	6	*1 to n within a cluster*

ATTACHMENT C

PES Data Reformatting Flow Chart



ATTACHMENT D

Blanks verses other type of invalids – Edit Statistics

Page 1 27/02/01 11:56:59 SUMMARY REPORT -- Dictionary: URBAN

Input: ..\CENDATA\5.TXT Program: CHEKEDIT

Count	Percent	Message
12,013	1.45	P10-RELATIONSHIP is BLANK
15	0.00	P10-RELATIONSHIP is invalid but not BLANK
2,978	0.36	P11-SEX is BLANK
1,029	0.12	P11-SEX is invalid but not BLANK
4,140	0.50	P12-AGE is BLANK
4,883	0.59	P12-AGE is invalid but not BLANK
19,408	2.34	P14-RELIGION is BLANK
2	0.00	P14-RELIGION is invalid but not BLANK
19,504	2.35	P15-MARITALSTAT is BLANK
-	-	P15-MARITALSTAT is invalid but not BLANK

QUESTIONNAIRE AND RECORD SUMMARY

Record Type	Input Quests	Input Records	Invalid Records	Output Quests	Output Records	Out - In Change	Write Records
POPULATI		829,898			-	-	
HOUSE		147,915			-	-	
TRAVELER		-			-	-	
VAGRANTS		-			-	-	
HOTTELS		-			-	-	
Total	147,638	977,813	-	-	-	-	-