DATA PROCESSING REPORT FOR MALAWI MAIN CWIQ SURVEY
By Mwangala Marilyn Bonna
(CWIQ DP Consultant)

(August – November, 2002)

The National Statistics Office (NSO) conducted a national CWIQ survey from 21st August to 21st November 2002.

The Data Processing Consultant arrived at NSO in Zomba, Malawi on 21st August, 2002 and immediately proceeded with the task of modifying the generic CWIQ questionnaire for Malawi. This also involved incorporating the Malawi poverty model into the questionnaire. The final questionnaire was ready and activated on 2nd Sept. The changes to the generic questionnaire for Malawi are listed in Appendix 1.

A draft questionnaire was used for supervisor training which was carried out in Liwonde on 29th August to 1st Sept. While the supervisor training was going on, the data processing team was being trained on how to use the scanners and the various steps in processing the questionnaires. After this initial training, the DP team was able to administer the data processing CWIQ "Printing and shading exercises" to the supervisors and scanned them to produce their error rates. Other training of DP staff carried on concurrently with that of the supervisors. Hands-on training was emphasized such that each member had to do some processing on their own.

Enumerator training was from Tues 3rd Sept to 9th Sept. and data processing training continued during this period.

Field work started on 10th September, 2002 with 11,600 households expected to be enumerated. Because during the training, pens were found to produce less errors, it was agreed that they should be used in the field instead of pencils.

9812 households were successfully interviewed and captured in data entry.

Staff & Equipment

5 DP staff were trained in CWIQ general processing – including:

   i.   How to process questionnaires when they arrive at the data processing station,

   ii.   Scanning and Verification of questionnaires (Data entry),

   iii. Dealing with persistent missing pages,

   iv. Committing of Tele-form data to Access Database,

   v.   Validating and correcting data.

All the five were able to handle these processes such that all could work independently. This was done deliberately so that the progress of the survey did not hinge on specific persons. Of these, two were programmers who were further trained in handling problems that arose in the course of this processing. The training they had received in Microsoft Access was helpful in enhancing their understanding of the CWIQ system. The programmers also received training on how to produce the standard tables and summaries and how to do minor modifications. They were also shown how to weight the data.

Two Fujitsu 4097D scanners were used with two Compaq PCs. Windows 98 was loaded on both because of lack of Tele-form drivers for more recent versions of Windows. Two NSO laptops were also used for the Microsoft Access database committing. Two HP Laserjet 4000 Series PCL printers used. A third printer used on one of the Compaq PCs to quicken process. Therefore there were effectively 3 scanning stations.

Teleform v.7.1 was used for the processing with Microsoft Access 2000.

With this equipment and staff, an average of 25 clusters were processed a day up to validation, though this picked up tremendously in the last days.

The questionnaires used in the survey were printed, correlated and stapled in the DP office.

Poverty Predictors

These were provided by Todd Benson of IFIP together with Jim Otto of World Bank. The details are presented in Appendix 2.

Processing Schedule

Questionnaires began arriving in the data processing office on 20th September and data entry started immediately. Processing proceeded as follows:

> 20th September: Started arriving.
>
> By 15th October: 1700 (94 EAs) households received and processed;
>
> By 22nd October: 3000 (152 EAs) Households
>
> By 28th October: 3579 (170 EAs)
>
> By 14th November: 9,889 households

About half of the questionnaires arrived in the last week and a half. The last questionnaire arrived on Thursday, 14th November, 2002. The final validations were completed on 21st November and unweighted tables produced on the following morning.

Some Issues on Processing

Section I: Children Under 5

Children who were not measured had their heights and weights entered as "9999" and "999" respectively. This is to be differentiated from those whose details did not appear at all. This affects somewhat Table 5.1.

Furthermore, the standard heights and weights limits were largely ignored because there were a lot of exceptions.

The variables added for the poverty model were related to some already in the questionnaire and they were not reconciled because the differed in a lot of cases. These included the following pairs:

- F12l Radio ownership and H10 for Number of radios owned
- G5 (main source of fuel for cooking) and H3 (whether cook over collected firewood) – not very related.
- H8 (number of members receiving a salary) with section E on employment of members.

Section L & M

L1 and L2 were to be related but also many differences were found. They were not compared to each other.

Call-backs for the AIDS module meant that the end time of the interview (A9) was affected and a lot of code "0000"s were recorded.

Recommendations for Next Survey

One of the aims of CWIQ is to carry out the processing of questionnaires concurrently with the field work. However, this requires that questionnaires arrive back in the data processing office fairly uniformly. Most of the questionnaires arrived in the last days of this survey because they were not "bubbled yet" or were not collected. Perhaps in the next survey, emphasis could be put on this to ensure that the processing is done uniformly and not concentrated at the end. Supervisors could make time to "bubble" rather than concentrate on collecting data with a view of bubbling at the end.

The data processing team did a good job in printing the questionnaires but if possible, because of the nature of the CWIQ questionnaire and its sensitivity to how the printing is done, it is suggested that a commercial printer do the printing of the questionnaire. This will make for uniform printing and stapling, and reduce the amount of errors that arise because of unrecognized page sizes and reference marks. This is more so the more the

number of households that have to be processed. In this survey, a lot of "Non-forms" arising from enumerators using pages of questionnaires given during the training and not recognizable during the processing, were found.

The data processing staff need to build on what they have learnt from this survey. In order to be able the next survey, they need to take time out to study and understand the CWIQ system of programs and modules to understand how it works. The training in Microsoft Access received by the programmers provided a good basis to do this.

Visits

Mr. Muhammad Lugongo, a DP lecturer from EASTC, arrived on 10th Tuesday in Zomba. He was given an overview of the data processing preparations that go into changing a questionnaire for a CWIQ survey in Tele-form and Access. He departed on 22nd Sept.

An Ethiopian delegation visited NSO from 19th to 26th October. During the visit, they were shown the workings of CWIQ.

APPENDIX 1: Data Processing Technical Report

1.   CHANGE QUESTIONNAIRE WITH TELEFORM DESIGNER

- Questionnaire changed in Tele-form 7.1 Designer. Final version activated 1st September, 2002 as **MLW2002**. Has 12 pages, and sections up to section M.
- Update Variable names, group into sections appropriately, remembering to remove hyphens in sections
- Field List, Tab-order, Primary Auto Export Configuration for committing to Microsoft Access.

2 (a).   LIST OF CHANGES TO QUESTIONNAIRE (In addition to those done during the Pilot survey)
- Variables deleted and added, labels changed

PAGE 1 *(Section A)*
         Inserted text "National Statistics Office, ZOMBA, MALAWI"
Q1   Label Changed to Village/Place
A1   Note that Districts will be used instead of Region. There are 31 districts. This will be entered in ClusterMaster Table for processing.

PAGE 2 *(Section B)*
B6   An issue raised about whether this question should even be asked to children under 12. It was suggested to put a comment to guide enumerators but finally decided to leave it as it was because a response is required there for everybody. Perhaps note this for future CWIQ survey.

PAGE 3 *(Section C)*
C4   Question changed to include "academic".
C7   Label changed to Religious Institutions from Islam/Church
C9   Range change by adding "Lack of Food" (1-10) as option 9 before other. Therefore addition of C9j

PAGE 4 *(Section D)*
D8   Labels and Range (1-9) changes for types of medical facilities
D10   Label changes

PAGE 5 *(Section E)*
E6   Range change by adding option 1 "Mlimi" (1-5); Label change to "Family Business Worker" which is same as "Unpaid family worker".

PAGE 6 *(Section F)*
         Added decimal points for F3, F6
F1   Label change to remove nomadic
F6   For future changes, suggested that this question be re-examined. Acres versus hectares. For now, left as is.

F8   Question changed to leave only "head of cattle" instead of all big livestock.
F10   Small livestock broken down to include Sheep, Goats, and Pigs with each category provided with 3 spaces for responses. (It was suggested that this change did not add to the quality of information especially considering that separating them versus leaving them together doesn't affect how the household's economic situation is evaluated. It would in fact bring a complication in deciding which exact situation in terms of ownership means a household has improved or worsened. However, NSO still wanted them in even if not they are not in the standard tables. They can do their own analysis in SPSS.)
F12   Added "Lantern" as option 1
      Added "Telephone" as option 2
      Label changes Electric Iron changed to just "Iron" and to "Modern stove"
F14   Range change by combining Seldom with Sometimes, (1-4)
F18   "How many chickens, turkeys, ducks or guinea fowls are currently owned by household?" Looking at standard tables for this section, groupings were for big livestock and small livestock and both or none. (Similar and related to F10, see comment above about tables).

PAGE 7 (Section G)
G1   Label changed to remove "Mud"
G2   Label changes
G3   Label changes
G4a   Range change (1-7). Rename it to (a), Label changes
G4b   Insert question on toilet ownership; Range (1-3)
G5   Range change by adding "Solar Energy"
G6   Label 1 change to remove Kerosene
G7g   Added new question "Telephone Facility"

(Section H – Poverty Predictors)
      [Derived from the Integrated Household Survey (IHS), Variables with significant coefficients and Co-relates of Poverty]

H1   Does the household grow tobacco? (Y/N)
H2   Does the household grow hybrid maize? (Y/N)
H3   Does the household cook over collected firewood? (Y/N)
H4   How many members of this household receive a salary after an agreed period? (2)
H5   How many working radios does the household own? (2)
H6   Does any member of this household own a sofa? (Y/N)
H7   Are there any household members who ever migrated for work? (Y/N)
H8   Has the household used sugar in the past two weeks? (Y/N)
H9   How many hectares of land are cultivated by the household? (3)
H10   How many household members have been doing Ganyu the past one month? (2)

PAGE 8 (Section I)
I4   Label 1 changed to add "Clinical Officer", label 4 to include "HSAs", Label 5 to add self?

I5a   Note to accommodate codes for "Don't know" 999
I5b   Note to accommodate codes for "Don't know" 9999
I6   Label changed to "Under-five Clinic"

NEW MODULES

FOOD SECURITY
PAGE 9 (Section J) J1 – J12

		MS Access (HHData)	
Length of section	*Section J*	*Database*	
Variable name	*Position*	*Position*	*Ranges*
J1	1	1	1-4
J2	2	2	1-4
J3	3	3	1-2
J4	4	4	Y/N
J5	5-6	5-6	1-12
J6	7	7	1-6
J7	8	8	1-7
J8	9	9	1-4
J9	10	10	Y/N
J10	11-12	11-12	1-12
J11	13	13	1-6
J12	14	14	1-7

SAFETY AND SECURITY
PAGE 9 (Section K) K1-K5

		MS Access (HHData)	
Length of section	*Section K*	*Database*	
Variable name	*Position*	*Position*	*Ranges*
K1a	1		Y
K1b-j	2-10		Y
K2a	11		Y
K2b-h	12-18		Y
K3a	19		1-4
K3b	20		1-4
K4a	21		Y
K4b-e	22-25		Y
K5a	26		Y
K5b-h	27-33		Y

SOCIAL PROJECTS
PAGE 10 (Section L) L1-L5

MS Access (HHData)

Length of section *Variable name*	*Section L* *Position*	*Database* *Position*	*Ranges*
L1a	1		Y/N
L1b-i	2-9		Y/N
L2a	10		Y/N
L2b-i	11-18		Y/N
L3a	19		Y
L3b-g	20-25		Y
L4	26		Y/N/X
L5a	27		Y/N
L5b-m	28-39		Y/N

AIDS MODULE
PAGE 11 (Section M)
Household Module – (M1-M6)

Length of section *Variable name*	*Section Mh* *Position*	*MS Access (**HHData**)* *Database* *Position*	*Ranges*
M1	1		Y/N
M2a	2		Y
M2b-j	3-11		Y
M3	12		Y/N
M4	13		Y/N
M5a	14		0-9
M5b	15		0-9
M6a	16		Y
M6b-j	17-25		Y

HIV/AIDS Individual Information
PAGE 12 (Section M1)

Length of section *Variable name*	*Section M1* *Position*	*MS Access (**INData**)* *Database* *Position*	*Ranges*
M7a	1		Y
M7b-g	2-7		Y
M8a	8		Y
M8b-h	9-15		Y
M9a	16		Y
M9b-f	17-21		Y
M10	22		Y/N/X

PAGE 12 (Section M2)

		*MS Access (**INData**)*	
Length of section	*Section M2*	*Database*	
Variable name	*Position*	*Position*	*Ranges*
M11	1		Y/N/X
M12	2		Y/N/X
M13	3		Y/N
M14a	4		Y
M14b-g	5-10		Y
M15a	11		Y
M15b-h	12-18		Y

(b)   SUMMARY OF CHANGES

ADDED OR DELETED VARIABLES IN CWIQ CORE:

C9j(1);

F10a(3), F10b(3), F10c(3); F12l(1), F12m(1); F18a(3), F18b(3), F18c(3);

G4b(1), G7g(1);

H6(1), H7(2), H8(2), H9(3), H10(2); (section H ());

J1 – J12; (section J – (14))

K1a – K5h; (section K (33))

L1a – L5m; (section L (39))

M1 – M6b; (section M (25))

MNo – (2)

M7 – M10; (section M1 (22))

M11 – M15h; (section M2 (18))

CODE CHANGES IN VARIABLES

Q1;

C7;

D8, D10;

E6;

F1, F8, F12c, F14;

G1, G2, G3, G4a, G5, G6;

I4, I6.

3.   UPDATE AND PRINT SYSTEM SPECIFICATIONS

    (a) Questionnaire flowchart
       No change

    (b) *DATA DICTIONARY UPDATE*

       *(i) **Value.Doc*** changes done in HHData, INData, CHData, HHSummR
       tables labels and values

          (ii) TableDoc link with values to ValueDoc
             (Note: 30 regions instead of 10)

(c) *UPDATING & MODIFYING SYSTEM ELEMENTS IN DATABASES*

Databases & Tables

(a) Data & RawData: Change or add variable names and lengths that have been changed in Microsoft Access "Table Design mode". Length should match with TFData (Teleform Data structures), and with section lengths in CWIQTemp in HHData, INData, CHData.

(b) CWIQtemp: Match with Teleform Data structure.

(c) CWIQOrig: "

(d) ErrCodes

(e) Interv

(f) ClusterMaster

Modules (Programs)

(g) tfData: Lengths and location of sections, taking care to change YNto12, Yto1, YNXto13 appropriately- alphabetic to numeric conversion.

(h) tyHh: Variable definitions, MErrMsg, AIDS Module changes;
bstore, bline, aLine(), fLine(), gLine(), hLine(), jLine(), kLine(), lLine(), mLine();
Date of interview, maximum values for F3, F6, Section H maximums;
Validation checks for sections A, F, F, H, J, K, L, Mh

(i) tyMember: mLine(); C9j, bStore & bLine methods for sections B,C,D,E, M1, M2 & validation checks

(j) tyChild: No changes

Forms

(k) HHData1, INDataSF, CWIQtf – insert new variables for editing and printing

Reports

(l) CWIQtf

6.   *PROGRAM SPECIFICATIONS*

Modules

(m) mValidate

(n) mConstVar

(o) Summaries
(p) pSummList (Region names)
(q) CWIQ0Tables – CWIQ6Tables

APPENDIX 2: Incorporating Poverty Model into Programs (mConstVAr)

Calculation of household score using poverty predictors:

Define the following poverty predictor variables:

PP01 – Household size

> HHSize

PP02 – Household size squared

> HHSize * HHSize

PP03 – Education level of household head

> Derive from C3 for household head

>> 1 = standard 1-4
>> 2 = standard 5-8
>> 3 = form 1-2
>> 4 = form 3-4
>> 5 = university or higher

PP04 – Number of salaried HH members

> H8

PP05 – Household owns a bicycle

> If F12k = 1 then 1 else 0

PP06 – Household owns a car or motorcycle

> If F12l = 1 or F12m = 1 then 1 else 0

PP07 – Household owns a fridge

> If F12d = 1 then 1 else 0

PP08 – Total acreage owned

> If F3 is not missing then 2.47 * F3 else 0

PP09 – Household grows tobacco

> If H1 = 1 then 1 else 0

PP10 – Household owns a bed

> If F12f = 1 then 1 else 0

PP11 – Household used sugar in the past 2 weeks

    If H6 = 1 then 1 else 0

PP12 – household cooks over collected firewood

    If H3 = 1 then 1 else 0

PP13 – Number of cattle owned

    If F8 is not missing then F8 else 0

PP14 – Household grows hybrid maize

    If H2 = 1 then 1 else 0

PP15 – Age of head of household

    HAge

PP16 – Household gets lighting from electricity or gas

    If G6 = 2 or 3 then 1 else 0

PP17 – Urban center coefficient

    See below.

PP18 – District coefficient for rural households

    See below.

Compute the household score as follows:

<u>Households in urban centers (Lilongwe, Blantyre, Zomba and Mzuzu)</u>

```
Score = natural log (0.518 x PP07 - 0.306 x PP01 + 0.016 x PP02 + 0.005 x PP15

                 + 0.151 x PP03 + 0.061 x PP04 + 0.704 x PP06 + 0.280 x PP16

                 + 0.247 x PP10 + 2.347 + PP17)
```

<u>Rural households</u>

```
Score = natural log (-0.174 x PP12 - 0.283 x PP01 + 0.015 x PP02 + 0.078 x PP03

                 + 0.098 x PP04 + 0.153 x PP05 + 0.693 x PP06 + 0.591 x PP07

                 + 0.152 x PP11 + 0.029 x PP08 + 0.105 x PP09 + 0.263 x PP10

                 + 0.013 x PP13 + 0.076 x PP14 + 2.826 + PP18)
```

Urban center coefficient

```
Blantyre City       -0.037
Zomba Municipality  -0.347
Mzuzu City          -0.205
Otherwise            0.000
```

District coefficient for rural households

```
Nsanje               0.173
Chikwawa             0.161
Mwanza               0.025
Blantyre Rural       0.208
Zomba Rural          0.168
Thyolo              -0.296
Mulanje             -0.004
Phalombe            -0.306
Machinga             0.059
Mangochi            -0.035
Chiradzulu          -0.048
Ntcheu              -0.303
Dedza                0.030
Salima               0.141
Mchinji             -0.052
Kasungu              0.181
Dowa                 0.138
Nkhotakota           0.039
Mzimba              -0.077
Nkhata-Bay           0.123
Rumphi              -0.024
Karonga              0.253
Chitipa             -0.346
Otherwise            0.000
```

APPENDIX 3:
For DP's Interest / Errors:

Duplex printing on 4097D in Black and White mode results in black images on one side. 256 Grayscale had to be used.

When validating, not all the data printed to save paper and also because found that they were not looking at that information as much as the questionnaire. Therefore printed only household member data with all the errors.

If A8 is Null, validation will stop. Change A8 to 1 - 4.

Date:

Submitted By:

Signed: _____