

AGRODEP
Household survey data course
Dakar, 8-10 October 2012

Data Quality



HarvestChoice
BETTER CHOICES, BETTER LIVES



LSMS
Living Standards Measurement Study

Levels of quality control

- Range checking
- Simple consistency checks
- Inter-record checks

Range checking

- Age should be a number less than 100
- Gender should be coded either “1” for *male* or “2” for *female*
- No numbers in the *name* field
- etc.

Simple consistency checks

- Age and birth date should be consistent with the date of the interview
- Head of the household should be 18 years or older
- A doctor should have completed university studies
- etc.

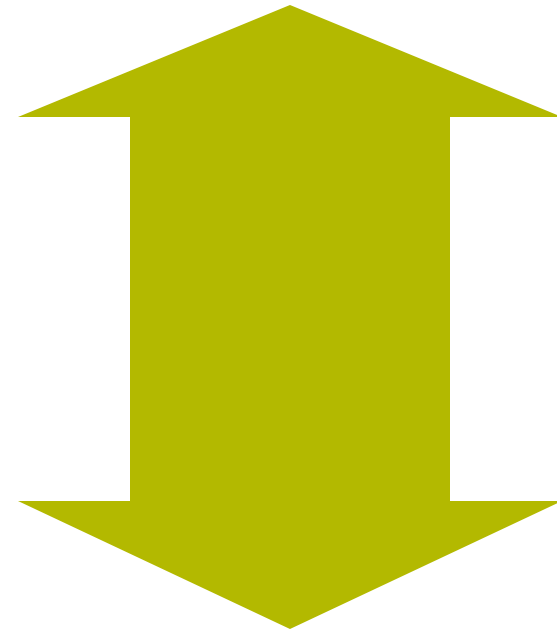
Inter-record checks

- Sub-totals
- Food consumption
- Checks with reference tables (anthropometrics)
- Cash balance
- Item-specific unit prices
- etc.

Levels of quality control

- Ranges
- Simple consistency
- Inter-record

Easier to conceive and program



Harder to conceive and program

Levels of quality control

- Ranges
- Simple consistency
- Inter-record

Errors likely to be due to miscoding or data entry



Errors likely to be due to interviewing

Why concurrent data entry?

- Quality control
- Turnaround time
- General improvement of field procedures
- Eliminates encoding as a separate task

Data entry is an **integral**
part of the survey, **not** an
afterthought

What is needed to develop a survey data entry program?

1. Integrated skills / tasks
2. Development time
3. Data entry software

1. Integrated skills

- Questionnaire and data entry program evolve synergistically (example: subtotals)
- Data manager needs to be a part of the core team from the beginning
- Integration is even more important in CAPI surveys

1. Integrated tasks

- Data entry program development
 - screen design
 - range checks
 - consistency checks
 - reference data
- Data entry manual
- Training
- Supervision during field work
- Compilation and documentation

2. Development time

- First version to be completed before the field test
- Working version needs to be ready for training
- Debugging needed during first weeks of data collection

3. Data entry software

- Traditional data entry (DE) software designed for transfer from paper questionnaire to computer
 - Benefits of relying on CAPI better realized working with software packages designed for interactive interviewing
- CAPI software packages make up a small market, with varying degrees of cost effectiveness & types of strengths
 - Key players: *Blaise*, *CASES*, *CSProX*, *MMIC* & *Surveybe*
- LSMS-commissioned comparative assessment of software programs for the development of CAPI applications (available on www.worldbank.org/lsms-isa)

Why Contemplate Transition to CAPI?

- Enhanced tools for in-field & remote management of mobile teams
 - Headquarters & Team Leaders: Assigning work, tracking progress, immediate & comprehensive feedback
- Expected gains in timeliness of data availability
 - Data entry, checking & exportation in one application
- Expected gains in data quality
 - Accommodation of [non-linear/integrated questionnaires](#)
 - [Automated routing](#) reduces the incidence of missing data
 - [Data checking](#), reporting & revision facilities during the interview
 - Range & consistency checks, flags for missing fields
 - [Improvements](#) in quantification of nonstandard units
 - Instructions on questions, note taking [facilities](#)

Non-linear Navigation

T Household number **1021000402**

CHRISTINE_MALIA_ALUMA

Roster



- 1. Start
- 2. Roster
- 3. Member Details
- 4. Housing & Energy
- 5. Income & Assets
- 6. Consumption
- 7. Misc.
- 8. Agri
- 9. Finish

We would like to make a complete list of Household members in the last 12 months including guests who slept here last night' and those that left permanently

Add new HH member

Update HH Location

Visit Number: !##:

Check

| Wave1 | Name | Gender | Residential Status | Age | Moved | Died | Demographics |
|-------------------------------------|-------------------|----------|--------------------------|-----|--------------------------|--------------------------|---------------------|
| <input checked="" type="checkbox"/> | 1 Mugalu Mohammed | Male [1] | Usual member present [1] | 3 | <input type="checkbox"/> | <input type="checkbox"/> | ▶ a |
| <input checked="" type="checkbox"/> | 2 Gadaffi Ismail | Male [1] | Usual member present [1] | 25 | <input type="checkbox"/> | <input type="checkbox"/> | ▶ a |
| <input checked="" type="checkbox"/> | 3 Kimbowa Martin | Male [1] | Usual member present [1] | 19 | <input type="checkbox"/> | <input type="checkbox"/> | ▶ a |

Tracking information

Automated Routing

T3ba Education Details

Household ID 1021000402

Check



Education 1 | Education 2 | Education 3

| | | |
|----|--|-----------------------------|
| 1 | IS Gadaffi ANSWERING FOR HIMSELF OR HERSELF? | Yes [1] |
| 2. | WHAT IS THE NAME OF THE PERSON RESPONDING FOR Gadaffi? | |
| 3. | Can Gadaffi read and write with understanding in any language? | nable to read and write [1] |
| 4. | Has Gadaffi ever attended any formal school? | Never attended [1] |
| 5. | Why has Gadaffi not attended school? . | |
| 6. | What was the highest grade/dass that Gadaffi completed? | |
| 7. | What was the main reason that Gadaffi left school? | |
| 8. | What grade/dass was Gadaffi attending in the last complet | |

Consistency Checks

T3ca Health Details Household ID 1021000402 Check

Health 1 Health 2

1 IS Mugalu ANSWERING FOR HIMSELF OR HERSELF? No [2]

2. WHAT IS THE NAME OF THE PERSON RESPONDING FOR Mugalu? Kimbowa Martin

3. During the past 30 days, did Mugalu suffer from any illness or injury? Yes [1]

4. For how many days did Mugalu suffer due to illness or injury during the past 30 days? 4

5. For how many days did Mugalu have to stop doing Mugalu's usual activities due to illness o... 12

6. Can you describe the symptoms that Mugalu primarily suffered due to Diarrhoea (acute) [1]


7. Can you describe the symptoms that Mugalu primarily suffered due to

8. Was anyone consulted (e.g. a doctor, nurse, pharmacist or traditional healer) for the majo... Yes [1]

T3c Member Health Household ID: 1021000402 Check

| 1. Wave 1 | 2. Name | 3. Age | 4. Health Info |
|-----------|-----------------|--------|----------------|
| 1 | Mugalu Mohammed | 3 | a. Health |
| 2 | Gadaffi Ismail | 25 | a. Health |
| 3 | Kimbowa Martin | 19 | a. Health |

Consistency Checks (Cont'd)

T Household number **1021000402** CHRISTINE_MALIA_ALUMA **Roster** 

1. Start 2. Roster 3. Member Details 4. Housing & Energy 5. Income & Assets 6. Consumption 7. Misc. 8. Agri 9. Finish

COMPLETE validation check

Update Visit Result ► V

General comments **Question specific comments**

Use of Media for Better Quantification


T6AE Food consumption item **105 Sweet potatoes (fresh)**

Household ID **1021000402**

Over the last **7** days

If none, enter 0

1_1 How many days in the past 7 days was Sweet potatoes (fresh) consumed?

2. Unit of quantity 

| | Quantity | Value |
|--|----------------------|----------------------|
| 3. Amount of Sweet potatoes (fresh) consumed in the HOUSEHOLD over the last 7 days which came from PURCHASE... | <input type="text"/> | <input type="text"/> |
| 4. Amount of Sweet potatoes (fresh) consumed AWAY FROM HOME over the last 7 days which came from PURCHASES... | <input type="text"/> | <input type="text"/> |
| 5. Amount of Sweet potatoes (fresh) consumed over the last 7 days which came from HOME PRODUCTION? | <input type="text"/> | <input type="text"/> |
| 6. Amount of Sweet potatoes (fresh) consumed over the last 7 days which came from GIFTS/IN-KIND sources? | <input type="text"/> | <input type="text"/> |
| Totals | 0.0 | 0.0 |

Use of Media for Better Quantification

◀ ▶ **Nomi Tin(500g)**
(FU-000120.jpg : 5 of 8)

use unit

+

-



Managing Expectations

- Data quality control principles in CAPI set-up no different than surveys based on paper-and-pencil interviewing (PAPI) with computer-assisted field-based DE (CAFE)
- CAPI tools useful as much as enumerators & field supervisors take advantage of available facilities & act on inconsistencies
- Relative impact of CAPI on data quality: Open empirical question
 - Limited evidence on improved data quality with respect to a well-supervised survey based on PAPI with CAFE
 - Fafchamps, M., McKenzie, D., Quinn, S., and Woodruff, C. (2010). Using PDA consistency checks to increase the precision of profits and sales measurement in panels. CSAE Working Paper Series No. 2010-19.
 - Caeyers, B., Chalmers, N., and De Weerd, J. (2012). “Improving consumption measurement and other survey data through CAPI: Evidence from a randomized experiment.” *Journal of Development Economics*, 98, pp. 19–33.

Cost Implications

- CAPI generates (minimal) savings in printing costs & data entry
 - Savings increase with the complexity & frequency of survey
- Significant up-front costs in hardware procurement
 - More cost-effective if machines are used in other survey operations
- Transition into CAPI also driven
 - by field work structure -size of the enumerator corps may be prohibitively large-
 - by demand for increased data quality & availability

Summing up

- CAPI is the option of the future, ensuring high data quality
- It is used successfully by some statistical agencies for simple surveys (LFS and CPI price collection) and by international institutions (WB)
- Recent experiments have shown that
 - Technology is already available (Lightweight notebooks and software development platform – both Windows based)
 - Can be cost-effective
 - No negative serious externalities
- We still need to solve (partially):
 - Interviewer training
 - Development of supervision procedures adapted to the new technology (voice recording, use of GPS', etc.)
 - Mistrust