EICV2 Mission: Data review and preparation for analysis

January 2006

Draft report

Patrick Ward, Oxford Policy Management

February 2006

Table of Contents

Executive Summary

OPM is providing technical support to the Enquete Integrale sur les Conditions de Vie des Menages (EICV Survey), a household living standards survey, a part of a larger project.This mission was undertaken to review the first two cycles of data and to assist in the development of syntax in preparation for the analysis of the full first quarter of data.

The mission

The consultant was in country from Friday 20th January to Thursday 2nd of February 2006. In addition to the two regular OPM advisers (Oumar Sarr and Geoffrey Greenwell), Mary Strode (project manager) and Andy McKay (income/expenditure poverty specialist) were present during the mission. The review work was divided between me and Andy McKay in line with the respective terms of reference. AM focussed on income, consumption and agriculture, while I focussed on other areas.

The main activities undertaken by the consultant were:

- To review selected variables from some of the relevant sections of the questionnaires (sections 1, 2, 3, 5, 8, 10) for the two cycles of data. Some of the relevant parts of the EICV1 questionnaire and some parts of the EICV1 reports and data were also reviewed.

- To develop some of the syntax to derive estimates for the key indicators for the first quarter of the data.

- To review a selection of analysis syntax files developed by Oumar Sarr (OS) and provided feedback and comments on it.

- To briefly reviewed the data entry and data processing systems.

- Together with other team members, to define and agree a set of working procedures to ensure that the data cleaning and analysis is effectively coordinated, quality controlled, and properly documented.

Circumstances meant that it was not possible to undertake external stakeholder consultation during the mission. However other team members were planning to do that after my departure.

Given the current circumstances of the Institute, no staff members were available to work with the consultant during the mission. It is hoped that this will be resolved during future missions once the re-staffing of the Institute is complete.

Data quality

While only two cycles of (unweighted) data were examined the findings are nevertheless reasonably informative. Overall, the quality of the data seems to be high. Sample losses are modest and there are few variables with significant problems with missing values.

Data entry procedures are rigorous and there are does not appear to be general concern about inconsistencies. It is important that the data entry staff habitually use a ‘user missing’ code to replace inconsistent data points that cannot be reliably inferred from other data in the questionnaire. There should be no imputation procedures at that point.

The analysis of errors by interviewer and consistency checks for part B were being strengthened during the mission. It was also agreed to track sample completeness and an indicator of

consumption reporting frequency on a monthly basis, in addition to those indicators that are currently being tracked.

Some concerns about particular variables were identified during the mission and are listed in section 3.1. They include concerns on age reporting and in the section on health. Household re-interviews by supervisors should be strengthened through the provision of a standard follow-up questionnaire.

Outputs and workplans

The team developed proposals for the key outputs and associated workplans during the mission, which need to be finalised in consultations with the NSIR and users. As they stand, they propose two main outputs for an external audience: preliminary estimates from the first quarter and a final 'poverty highlights report'.

The first is aimed at meeting the immediate needs of the Cabinet in the PRS drafting process. It would provide information on a few selected indicators. Estimates would be highly provisional and should be marked as such. It would not include consumption or income poverty measures. For the selected indicators, it would aim to assess trends since EICV1, although caution would be required in their interpretation.

The second would provide information on trends in the main indicators between EICV1 and EICV2, including income poverty. The estimates contained in it would be publicly available and final. It would contain appropriate information on methods and on sampling errors.

In addition, the project should support the production of the draft final tables for a full EICV2 report. However, it will not be possible to produce the final report itself within the project lifetime. Neither is a full poverty profile likely to be possible.

The PRS process requires estimates before the end of 2006, while the fieldwork only finishes in October. As a result there is very little time to complete the cleaning and analysis for the full data in line with the PRS timetable. 'Dummy runs' , producing the tables with data sets of either two or three quarters of data, are proposed as a means of ironing out any problems in advance of the final estimates. Broad workplans are provided for these activities. The team also agreed on a set of conventions on directories and files to ensure consistency and adequate documentation.

A number of other documents are proposed, to ensure that the survey is adequately documented and can be repeated more easily in future rounds.

Issues in the analysis

In addition to the need for consultation with users about outputs, a number of general issues in the analysis were discussed amongst the team members. Foremost amongst these is the importance of assessing comparability between the two surveys when estimating trends. EICV1 tables should be re-run to ensure comparability. In addition, it will be necessary to re-tabulate the EICV1 estimates using the new regional and urban-rural classifications, so that comparable disaggregated information may be presented for the two surveys. Some issues of comparability for particular variables were also identified for the first quarter estimates and are outlined in Section 3.3.

1. Introduction

OPM is providing support to the new Rwanda National Institute of Statistics (NSIR) under a DFID-financed project. The NSIR is a newly created independent institute, replacing the previous Department of Statistics in MINECOFIN. There is some disruption of normal activities as a result of this ongoing transition.

OPM is supporting the process of institutional change and is providing technical support to the Enquete Integrale sur les Conditions de Vie des Menages (EICV Survey) and some other areas. The EICV is a Household Living Conditions Survey which was last conducted in 2000-02. This mission was undertaken to review the early data and to assist in the development of syntax in preparation for the analysis of the first quarter of data.

Background to the survey

The sample size is 6,900 households, of which 1,620 will be in urban areas. The sample is divided into 10 cycles spread over a 12 month period. In urban areas each household is visited every 3 days over a 33 day period, in order collect consumption data. In rural areas households are visited every other day over a 16 day period. The sample matches as closely as possible the EICV1, except that the sampling frame changed following the Population Census becoming available and the definitions of urban and rural were changed. In the EICV1 urban and rural households were interviewed during different years, and the timing of interviewing was not spread over agricultural/climatic zones in an optimal manner to minimize seasonal effects. Rwanda has three growing seasons, and these vary depending on geography, despite the country being small in size. The EICV2 sample evened up the spatial/temporal distribution.

Despite the upheaval in the organization due to its ongoing transformation to an independent institute, the EICV2 started fieldwork on 10th October 2005, and data entry commenced the following month. The fieldwork will run for a full calendar year and finish in October 2006.

The EICV team is supported by two OPM advisers, Oumar Sarr (Resident Survey Expert) and Geoffrey Greenwell (Data Management Expert). Both advisers worked on EICV1 and as the subject matter, question form and methodology is similar, the local team is well supported by experienced professionals

The new Rwanda PRSP is under preparation and the date of its release has been delayed in order to incorporate results from the EICV2. In order to provide the PRSP team with some early information, it is possible to provide preliminary results on a quarterly basis. The final results are required in December, and this is a high priority for Government and donors alike.

The mission

Data for the first two cycles of fieldwork were available at the time of the mission. The objective of this mission was to prepare the ground for the local team to provide some preliminary first quarter estimates once three cycles of data became available.

The specific objectives of the mission were for the consultant to:

1. Consult with key stakeholders on their needs from the EICV, both immediate and in the longer term.

2. Familiarize himself with the analysis of EICV1

3. Liaise with the EICV team and resident expert to propose a tabulation plan for survey analysis. The consultant should focus on the PRSP social indicators, and develop an

analysis plan to provide stakeholders with results relating to these indicators. Analysis of agricultural activities and consumption aggregates will be undertaken by other consultants.

4. Review the available data and comment on its quality and completeness. The results of this review should be communicated to the team, and possible remedies for fieldwork and data management weaknesses explored.

5. Liaise with the team to prepare programmes and data treatment operations for the analysis proposed to meet users' data requirements, in particular the preliminary quarterly output.

6. Carry out on the job training with the team in preparation for the quarterly results releases.

7. Propose a work programme for the analysis of the survey results, between now and the end of the project in December 2006.

8. Review the resources which will be required to carry out the analysis, including local experts and other expert input.

9. Prepare a report of the findings within four weeks of the end of the mission.

Full terms of reference are attached in Annex 1.

2. Activities undertaken

The consultant was in country from Friday 20th January to Thursday 2nd of February 2006.[1] In addition to the two regular OPM advisers (Oumar Sarr and Geoffrey Greenwell), Mary Strode (project manager) and Andy McKay (income/expenditure poverty specialist) were present during the mission. For the first week, two DFID representatives (Kim Bradford Smith and Roger Edmunds) were also present in country to review the progress of the project.

The review work was divided between myself and Andy McKay (AM) in line with our respective terms of reference: AM focussed on income, consumption and agriculture, while PW focussed on other areas. There were some areas of overlap. Both worked closely with the other team members. Given the current circumstances of the Institute, no staff members were available to work with the consultant during the mission. It is hoped that this will be resolved during future missions once the re-staffing of the Institute is complete.

The main activities undertaken by the consultant were as follows:

- Reviewed a selection of variables from some of the main relevant sections of the questionnaires (section 1, roster; section 2, education; section 3, health; section 5, housing; and small parts of section 8 (agriculture) and section 10 (credit, goods and savings) for two cycles of data (1,380 households). Relevant parts of the EICV1 questionnaire and some parts of the EICV1 reports and data were reviewed, although there was not time to review all of it during the mission.

- In parallel with this, developed some of the syntax that can be used to derive estimates for the key indicators that are expected to be calculated in the analysis of the first quarter of data.

[1] The day of arrival and one subsequent day were spent completing other work and will not be billed to the project.

- Reviewed a selection of analysis syntax files developed by Oumar Sarr (OS) and provided feedback and comments on it.

- Briefly reviewed the data entry and data processing systems.

- Together with other team members, defined and agreed a set of working procedures to ensure that the data cleaning and analysis is effectively coordinated, quality controlled, and properly documented.

It was agreed that the best way to develop and peer review syntax for the first quarter estimates and subsequently is for OS to draft it and for other consultants to review and comment on it. This might be done by directly reviewing the syntax or by producing a selection of the same estimates in parallel and comparing results. In either case, OS will remain responsible for the final analysis syntax files.

The mission coincided with a planned PRS workshop that was then cancelled, and the unfortunate sickness and travel to South Africa of the DG. This meant that it was not possible to undertake any external stakeholder consultation during the mission. However, Mary Strode and Andy McKay were able to do that after my departure. Some of the other team members also undertook a visit to the field.

The consultant also participated in one meeting with the DFID review team.

3. Key findings and recommendations

3.1 Data quality

While only two cycles of (unweighted) data were examined the findings are nevertheless reasonably informative. Overall, the quality of the data seems to be high.

Data entry procedures are rigorous, with 100% double entry and reconciliation of differences. Data entry is in CS-Pro and extensive range and consistency checks are run during entry and afterwards on the complete files. Procedures to deal with inconsistencies are generally clear and where necessary the field teams are contacted to correct problems in a questionnaire, although this is currently only possible for part A. As a result, there are no general concerns about inconsistencies. It is important that the data entry staff habitually use a ‘user missing’ code to replace inconsistent data points that cannot be reliably inferred from other data in the questionnaire. There should be no imputation procedures at that point in the data management process.

During the mission, Geoff Greeenwell developed a programme to examine errors by interviewer and by team, which should assist with controlling and improving quality in areas where there are concerns. It was also agreed to track sample completeness and an indicator of consumption reporting frequency on a monthly basis, in addition to those indicators that are currently being tracked. Consistency checks for part B should also be completed and it may be that this shows the necessity of some field revisits for part B data. It would also be useful to ensure that the household identifier (‘key’) is included in all data files on export to SPSS.

Figure 1: Age reporting in single years

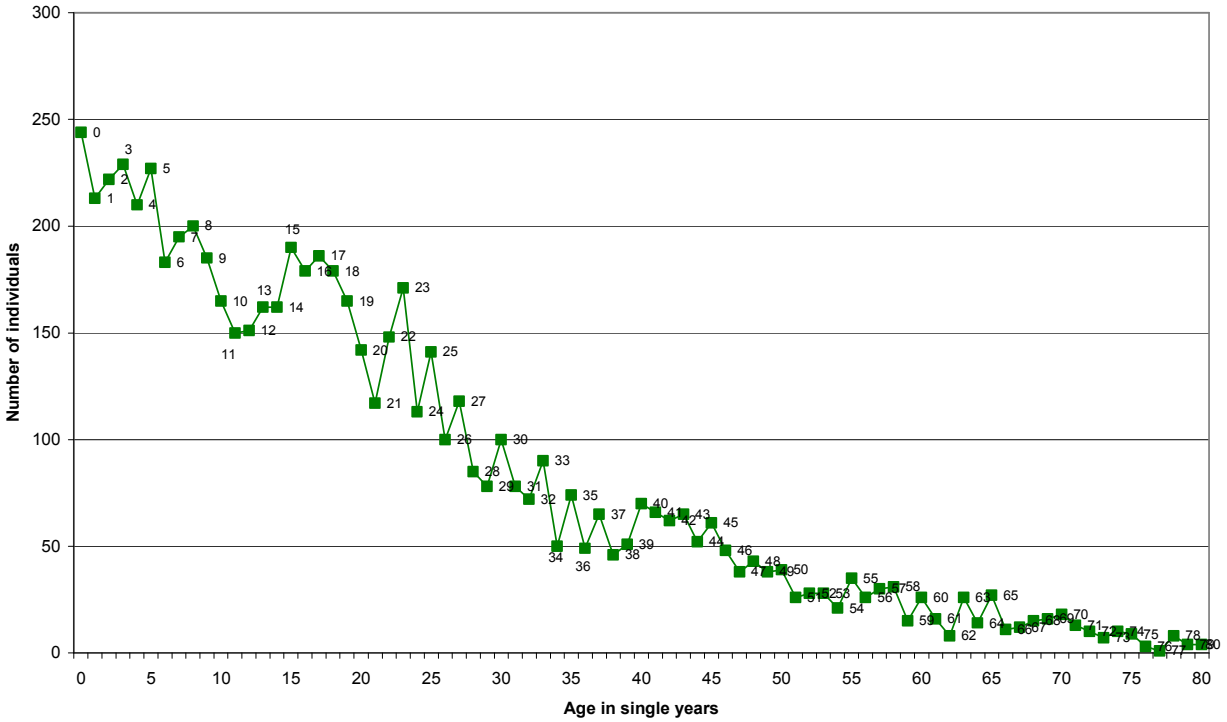
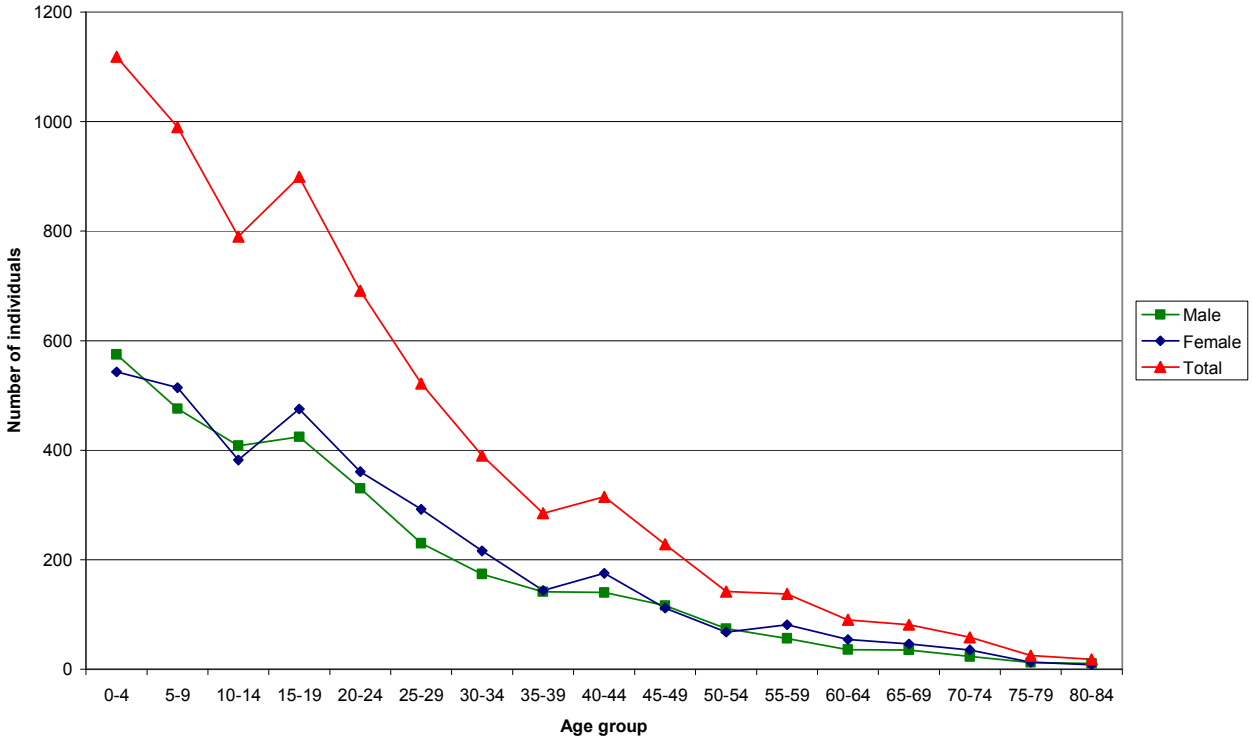


Figure 2: Age reporting by sex in five-year age groups



Sample losses and refusals are reasonable for a survey of this type - some 95% of the original sample has been interviewed and outright refusals are rare. It would be useful to investigate the reasons specified for households that were not interviewed and where the reason was coded as 'other' (1.2% of the sample). No households are missing entire sections of the questionnaire and none of the variables examined had significant proportions of missing values. One or two variables have a high proportion of 'other' responses, which need to be followed up (eg reasons for not using services in section 5); this is generally something that it is useful to monitor.

The age and sex structure is reasonably consistent with the census. However there seems to be some patterns that are not found in the census. There are: some age heaping at age 40-44 and (strangely) at particular ages in the twenties; an apparent tendency for girls aged 10-14 to report themselves as 15-19 (although a some of the lack of 10-14 year olds is probably genuine and is also shown by the Census data); and a lack of six year olds and an excess of five year olds that might reflect interviewers (consciously or not) shifting children back to avoid having to complete an education module for them (see Figures 1 and 2). While some inaccuracies are inevitable the survey should try to ensure that interviewers record age more precisely where possible and should act on the apparent age 5/6 shifting.

Education, Section 2: the data examined seems to be of reasonable quality and no obvious problems were identified. It is necessary to infer current class from class completed, which is not ideal, but enrolment rates calculated using this approach appear broadly consistent with other data sources (Rwanda in Statistics).[2]

There are some apparent concerns in the health data (section 3), which ought to be followed up:

1. Person consulted (s3aq13): health assistants are mentioned surprisingly rarely (<1%). It should be checked whether respondents and interviewers able to distinguish, doctors, nurses and other types of health workers effectively.

2. There seem to be extremely high levels of reported use of antenatal care services (s3bq7) – some 95% of women report using it, compared with the last DHS reporting around 20%.

3. Part C: Many five year olds do not have information recorded in this section, despite the instructions on the questionnaire. It might be decided to collect it only for children age four and under, but whichever group is included should be covered consistently. The skip in Q3 (if no vaccination card skip to Q5) seems to be inconsistent with the instructions in Q4 (if don't have a card, get information from the respondent). From the data it looks as though the skip is being followed. This does need to be done consistently, but as it stands it will give us only 'card only' vaccination rates, which are not so useful in places where many children are vaccinated but don't have cards (in the survey data this is true of about one third of children). This will reduce the value of the immunisation data. There are also quite a lot of 'not applicable' codes in the data from the vaccination cards (Q4), which should be clarified.

In Section 5, reported time to reach each specified facility/service is very heaped at 30 minutes. This is not unusual and might be difficult to improve but its impact on the analysis, where 30 minutes is often used as a cutoff, should be considered.

Data on the ownership of consumer durables (section 11 B) seems to be generally consistent with EICV1, with the exception of ownership of tables and chairs, both of which appear to show erratic trends and/or unlikely urban/rural differentials which suggest that they are being reported/recorded differently in the two surveys.

[2] Although the secondary gross enrolment rate given in that document appears to be wrong since it is lower than the net enrolment rate, which is impossible.

Other team members reported that fieldwork was being done well. It may be useful to provide the 'controlleurs' (field team leaders) with a short standardised questionnaire of key variables for re-interviews, since there is currently no clear guidance on what they should collect and they should not be repeating the entire interview. Submitting these questionnaires would also provide evidence that the re-interviews have been undertaken. It is essential that transport needed for the supervisors (central staff) to visit the field be provided. OPM consultants should also continue to visit the field. The use and fate of the completed diaries should be clarified. It is unfortunate that the agriculture section of EICV2 remains so long despite the fact that the light agricultural survey is being undertaken in the same households and it had been recommended, as a result, to cut it back substantially.

3.2 Proposed outputs and workplan

The team discussed and developed proposals for outputs and workplans during the mission. However, discussions with the NSIR and consultations with users about the key outputs were still to be completed by the other team members after I departed. The proposals as they stood at the end of the mission are as follows.

There will be five main analytical outputs from EICV2 supported by the project. These are shown in Table 1 below.

Table 1: Main outputs from EICV2 supported by the project

Output	Based on data from:	Target audience	By when
A. Analytical outputs:			
1. Provisional trends in selected welfare measures	3 cycles	The Cabinet, for the PRS process	March / early April 2006
2. Poverty highlights with half of the data	5 cycles	Internal to NSIR	End July 2006
3. Draft final tables for main report with three-quarters of the data	7 cycles	Internal to NSIR	September 2006
4. Poverty highlights report	All cycles	Public, esp. PRS drafting team	Late Dec. 2006
5. Draft final tables for main report	All cycles	Internal to NSIR	Late Dec. 2006
B. Other documentation			
6. Fieldwork report	All cycles	Internal	November 2006
7. Sampling report	3 cycles	Internal	March 2006
8. Data processing documentation	All cycles	Internal	December 2006
9. Report on lessons learnt for future surveys	All cycles	Internal	November 2006

The main outputs with an external audience are the first and fourth. The first is aimed at meeting the immediate needs of the Cabinet in the PRS drafting process. Estimates would be highly provisional and should be marked as such. A list of proposed indicators is included in Annex 2. It would not include consumption or income poverty measures. For the selected indicators, it would aim to assess trends since EICV1, although caution would be required in their interpretation.

The other main external output would be some sort of 'poverty highlights report', which would provide information on trends in the main indicators between EICV1 and EICV2, including income poverty. The estimates contained in it would be publicly available and final. It would contain

appropriate information on methods and on sampling errors. The PRS process requires estimates before the end of 2006, while the fieldwork only finishes in October. As a result there is very little time to complete cleaning and analysis for the full data set in order to meet the PRS timetable. It is proposed that a dummy run for these tables be undertaken in July, using half of the data, to ensure that any problems have been sorted out before the final analysis with the full data set. The results of the dummy run should remain within the NSIR, given their provisional nature.

Given the short time between the end of the fieldwork and the end of the project, it will not be possible to support the production of a full survey report under the project. However, the project should support the production of all of the tables for such a report. This would ensure they are available for a subsequent process of finalisation and publication early in 2007. For the same reasons as given above, a set of these tables should be produced with a partial data set as a dummy run. This might best be done in August/September, before the work on the full data set happens.

A number of other internal documents are proposed, to ensure that the survey is adequately documented and can be repeated more easily in future rounds. All reports will be produced initially in English.

It should be noted that a number of important outputs are not likely to be produced within the timescale of the project. These are: the complete final statistical report of EICV2; a full poverty profile from EICV2; and analysis of linked data from EICV2 and light agricultural survey. This work will need to take place in 2007. However, it may prove possible to develop draft text for a main report during the project, based on the 7 cycles of data. In this case, subsequent work would consist of editing and revising it in the light of the final tables.

Working procedures for cleaning and analysis

Working procedures for the cleaning and analysis were agreed amongst the team members. This distinguished two clear phases in the cleaning process. Any imputation will be carried out in the second phase of the cleaning and will be done using SPSS syntax so that it is properly documented and can be replicated. The stages outlined were as follows:

1. **Data entry and first cleaning**

   This consists of data entry and resolution of discrepancies from double entry:

   Data entry staff undertake double entry and resolve any differences between the two entered files. After this, data entry staff run the second consistency check programme and, if necessary, return to the questionnaires to resolve problems. They change a data point only if it is found to incorrectly entered from the questionnaire (which should be very rare) or if the correct value can be unambiguously inferred from other information on the questionnaire. Otherwise they either leave it or set the value to user missing. Some data editing will also be carried out to ensure skip consistency (eg where have user missing values and should be sysmis).

   The data from this process is exported to SPSS to give the raw data. There are no further manual edits after this stage.

2. Second cleaning

GG and OS will run the second consistency check programme. Any recoding and imputation is done in SPSS, using syntax files. Generally, missing or inconsistent values should be coded to user missing codes (not 'other' etc). Every effort should be made to deal with outliers without returning to the questionnaires again. OS will be responsible for this syntax. The resulting clean data files are written to another directory and used for the analysis.

3. Analysis

OS drafts and controls the master syntax files for analysis. AM and PW review and comment on these files; and/or run independent tables as a cross-check. All operations are written in syntax (if for any reason this is not possible, what has been done should be noted in the syntax and starred out "*! Action taken".)

4. Results and reporting

Draft results are shared for comments amongst team members and staff at the NSIR. The review process with users needs to be agreed.

It is important that files are clearly identified and annotated. It is also useful for team members to use a common directory structure. These issues were discussed and a draft set of conventions agreed. They are attached in Annex 3, but will be revised and finalised by GG and OS.

Workplan

A provisional workplan was agreed by the team. This is attached with this report as a separate Excel spreadsheet. Some of the main activities and deadlines are shown below.

1. First quarter estimates

Activity	Person responsible	By when
Agree indicators with stakeholders	All	10 Feb
Review data and syntax	All	10 Feb
Complete first cleaning of full three cycles	GG	10 Feb
Complete full cleaning of data	GG & OS	18 Feb
Compute and apply weights & calculate sampling errors	DM & GG	25 Feb
Production of preliminary estimates	OS & GG	25 Feb
Collection of information on comparable sources	OS	25 Feb
Review of preliminary estimates	AM & PW	3 March
Sharing with stakeholders	OS & GG?	11 March

Some of the syntax needed for the first quarter estimates was developed during the mission. This needs to be completed once the indicators are finalised with users. It is proposed that OS draft the complete syntax for the March estimates and that it is reviewed long distance by PW and AM.

OS will also draft the remaining syntax needed for the income-expenditure analysis and for other tables during and after March, for comments and review by AS and PW during subsequent missions.

2. 'Poverty highlights' report

Activity	Person responsible	By when
Preliminary tables and report (5 cycles of data)		
Agreement of report and table contents	All	April?
Drafting syntax for tables	OS & GG	End May
Re-run estimates for EICV1	OS & GG	End June
Complete data entry and first cleaning	OS & GG	9 June
Complete second cleaning and basic frequencies	OS & GG	7 July
Produce draft tables	All	21 July
User seminar on methods & progress	All	21 July
Production of 'dummy' report	OS	End Aug
Full analysis		
End of entry and first cleaning	OS & GG	Mid Nov
Full cleaning and basic frequencies	OS & GG	Late Nov
Produce draft tables and poverty estimates	OS & GG	Early Dec
Review and comment on tables	AM & PW	Mid Dec
Drafting of report	All	Mid Dec
Presentation of results to users	All	Mid Dec
Finalisation of report	All	Late Dec

3. Tables for the main survey report

Activity	Person responsible	By when
Preliminary tables (7 cycles of data)		
Agreement of report and table contents	All	April?
Drafting syntax for tables	OS & GG	End May
Re-run estimates for EICV1	OS & GG	End Aug
Complete data entry and first cleaning	OS & GG	Early Sept
Complete second cleaning and basic frequencies	OS & GG	Mid Sept
Produce draft tables	All	End Sept
Full analysis		
End of entry and first cleaning	OS & GG	Mid Nov
Full cleaning and basic frequencies	OS & GG	Late Nov
Produce draft tables	OS & GG	Mid Dec

It is hoped that the consultants will work closely with NSIR staff members in future. This has proved so far because of the uncertainties associated with staffing the new Institute. Once this is finalised then there should be more scope for it. The consultants' work should then include some specific capacity building and training. It is proposed that:

- Oumar Sarr provide training on the use of SPSS;
- Andy MacKay provide a short (one day) seminar on measuring income poverty

- David Megill provide a short (half-day) seminar on sampling

3.3 Analysis issues

Some of the general issues to be considered in undertaking the analysis were discussed and agreed amongst the team. These included:

- The need to agree the relationship between the contents of the proposed 'poverty highlights' report and the main survey report in consultation with users. This should include a consideration of which (or if both) of the reports should present information on trends.

- The importance of assessing comparability between the two surveys when estimating trends - EICV1 tables should be re-run to ensure comparability. It is also important to check the consistency of levels and trends with other sources (including DHS, CWIQ, Census and routine data sources).

- The need to re-tabulate the EICV1 estimates using the new regional and urban-rural classifications, so that comparable disaggregated information may be presented for the two surveys.

Comparability issues are also important in the production of the preliminary estimates in March. A number of issues came up in this mission that will need care in the analysis. They include the differences in the education module coverage between the two surveys (age 6/7 and above age 40); and differences in the health questions. The adult literacy rate in the EICV1 report also looks unlikely (rural higher than urban) and should be carefully checked. There is no household level information on time to services in EICV1, so assessing trends in these indicators will be difficult or impossible.

4. Conclusions

This mission was undertaken to review some of the key non-consumption data and to develop syntax for the first quarter estimates. Selected data was reviewed and on the whole appears to be of good quality. There are some specific areas of concern that have been identified for follow up. The syntax was developed for some of the estimates. However this needs to be completed by OS. The estimates so produced should be reviewed by PW/AM.

The timetable for the cleaning and analysis for the full survey data is extremely tight, with fieldwork ending only in October but estimates needed before the end of the year. A proposal for key outputs and a workplan was developed by the team in the light of this and is presented here. Agreements were also reached on standardising processes and file management for the second stage of cleaning and for analysis.

The discussions and planning amongst the team members during the mission was useful. However circumstances meant that I was unable to consult more widely during this mission; that important process was left to other team members after my departure. Some of the proposals outlined in this report may need to be adjusted in the light of those consultations.

Annexes

Annexe 1: Terms of Reference

Background

The OPM project to support the transformation of the Rwandan statistical office from a Department to an independent Institute has one year left to run. In September the laws were approved by Parliament setting up the new Rwanda National Institute of Statistics (NSIR). A Director-General was appointed in September 2005 and the Board was appointed in the last few weeks. All former staff members have to apply for posts in the new Institute along with external candidates. The office is therefore undergoing major transformation, and regular activities are disrupted.

The EICV Survey is a Household Living Conditions Survey which was last conducted in 2000/2001/2002. The sample size is 6,900 households, of which 1,620 will be selected in urban areas. The sample is divided into 10 cycles spread over a 12 month period. In urban areas each household is visited every 3 days over a 33 day period, in order collect consumption data. In rural areas households are visited every other day over a 16 day period. The sample selected matches as closely as possible the EICV1, except that the sampling frame changed following the results of the Population Census becoming available, at the same time definitions of urban and rural were changed. In the EICV1 urban and rural households were interviewed during different years, and the timing of interviewing was not spread over agricultural/climatic zones in an optimal manner to minimize seasonal effects. Rwanda has three growing seasons, and these vary depending on geography, despite the country being small in size. The EICV2 sample evened up the spatial/temporal distribution.

Despite the upheaval in the organisation, the EICV2 started fieldwork on 10th October 2005, and data entry commenced the following month. There were some minor logistical field problems early in the fieldwork, but the results to date are reported to be of reasonable quality, and data checks have already been carried out on the first cycles. The fieldwork will run for a full calendar year, and finish in October 2006.

The EICV team is supported by two OPM advisers, Oumar Sarr (Resident Survey Expert) and Geoffrey Greenwell (Data Management Expert) who is available in Rwanda on a half-time basis. Both advisers worked on EICV1 and as the subject matter, question form and methodology is similar, the local team is well supported by experienced professionals

The new Rwanda PRSP is under preparation and the date of its release has been delayed in order to incorporate results from the EICV2. In order to provide the PRSP team with some early results, it is proposed that preliminary results be made available on a quarterly basis. The sample design enables nationally representative estimates to be released each quarter. The final results are required in December, and this is a high priority for Government and Donors alike.

The results to be released will be limited, and their circulation will also be limited, but they are vital for PRSP preparations. The full first three months of data will not be available in time for the mission, but data for at least the first 2 months should be available for inspection and practice. The EICV local team is expected to run the first quarter's tabulations and produce the first results using the products of this mission. The consultants are expected to incorporate training in their work, and leave clear recommendations for the team to follow.

There will be an OPM/EICV team in Rwanda during this mission. The team members will comprise:

- Geoffrey Greenwell (data management and computer applications)
- Andy McKay (consumption aggregates and poverty analysis)
- Patrick Ward (analysis of other variables, especially social sector)
- David Megill (preparation of sampling and corrective weights; estimation of sampling errors)
- Oumar Sarr (resident survey expert)
- Mary Strode (project management & support; quarterly project review)

Questionnaires are available in French. The consultant will be expected to work in both the English and French languages. The report should be prepared in English.

Most of the EICV prefer to communicate in French, but can understand English and other team members can provide translation where necessary. Stakeholders tend to be proficient in both languages.

Mission Context

The purpose of the mission is to review the first months of data in preparation for the analysis of data from the first quarter.

The consultant will;

1. Consult with key stakeholders on their needs from the EICV, both immediate and in the longer term.
2. Familiarize himself with the analysis of EICV1
3. Liaise with the EICV team and resident expert to propose a tabulation plan for survey analysis. The consultant should focus on the PRSP social indicators, and develop an analysis plan to provide stakeholders with results relating to these indicators. Analysis of agricultural activities and consumption aggregates will be undertaken by other consultants.
4. Review the available data and comment on its quality and completeness. The results of this review should be communicated to the team, and possible remedies for fieldwork and data management weaknesses explored.
5. Liaise with the team to prepare programmes and data treatment operations for the analysis proposed to support users' data requirements, in particular the preliminary quarterly output.
6. Carry out on the job training with the team in preparation for the quarterly results releases.
7. To propose a work programme for the analysis of the survey results, between now and the end of the project in December 2006.
8. Review the resources which will be required to carry out the analysis, including local experts and other expert input.
9. Prepare a report of the findings within four weeks of the end of the mission.

Timing

The consultant is expected to be in country by 24 January 2006. The consultant will work up to 8 days in country and spend up to 5 days in preparation, report writing and follow-up. A further mission is anticipated later in the year, and a mission in November/December 2006 will be required.

Annex 2: Proposed preliminary estimates for the first quarter

1. Outcome Indicators

- ☑ Adult Illiteracy Rate

2 Output Indicators

- ☑ Net and gross primary and secondary enrolment ratios
- ☑ Entry rates for children age 7
- ☑ Population with access to safe water
- ☑ Population with access to health services
- ☑ Population with access to sanitation
- ☑ (A) Population with access to primary education

3 Proxy Indicators

- ☑ (A) % employed in agriculture/formal/informal sectors
- ☑ (A) Asset ownership
- ☑ (A) Livestock ownership
- ☑ (A) Agricultural inputs (fertilizers/insecticides etc.)

(A) = additional non-PRS indicator proposed

Annexe 3: Standardising conventions – draft for finalisation

There is a need to standardise: directories, syntax file labelling and notation, data file management, file names.

Data files:

Use a single one of each type, replacing older one of fewer cycles (copied somewhere) ie not named by cycle contents.

Generate a household master file with key household data in it from various sections.

Syntax files

Organisation and naming conventions: organised and named by section and/or analysis purpose.

2 sets: March estimates, main report (keep a copy of March syntax)

Structured by and annotated with details of the estimates produced by each section of the syntax – add table numbers to this documentation once the report structure finalised.

Use 'get file' and 'save file' commands.

Directories – possible structure

C:\EICV2\

Subdirectories:

\data\raw\partieA\cycle1

\data\raw\partieB\cycle1

\data\raw\partieC\cycle1

\data\clean\partieA

\data\clean\partieB

\data\clean\partieC

\programmes

\documentation\questionnaires and field docs

\documentation\reports

Consider how to structure for the phases of data and analysis;

And the same for EICV1 – at least for the data. How to organise any revisions to syntax files.