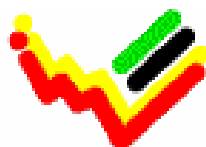# Mission Report

from a short-term mission on

Final data processing and quality control IAF 2002/03

2 September to 10 October 2003

TA for the 'Bridging Support Program to Strengthen the Institutional Capacity of the National Statistics, Mozambique

James Otto

James Otto
6092 9th Place N
Arlington, VA 22205-1608
USA

JOtto@WorldBank.org
bwJim@Compuserve.com
202-473-5169 World Bank
703-536-4763 Home

Table of contents

List of abbreviations

CO	Scanstat Coordination Office in Statistics Denmark
Danida	Danish International Development Assistance
DKK	Danish Kroner
DSt	Statistics Denmark
EUR	European Euro
IAF	Inquérito aos agregados familiares sobre orcamento familiar, Household budget survey
IDS	Inquérito Demográfico e de Saude Demographic and Health Survey
INE	Instituto Nacional de Estatística, Mozambique
INE-P	Instituto Nacional de Estatística, Portugal
MZM	Mozambique Meticais
NOK	Norwegian Kroner
QUIBB	Questionário dos Indicadores Básicos de Bem-Estar Core Welfare Indicators Questionnaire
Scanstat	Consortium between Statistics Denmark, Statistics Norway and Statistics Sweden
SCB	Statistics Sweden
SEK	Swedish Kronor
SSB	Statistics Norway
USD	US Dollars
VBA	Visual Basic for Applications
ZAR	South African Rand

1   EXECUTIVE SUMMARY

The main reason for the mission was to complete the data processing of the 2002-3 Integrated Household Survey (IAF2002-3). This was a one year survey covering 8,700 households from July, 2002 to June, 2003. The survey used a household questionnaire adapted from the QUIBB1 survey in 2000-1 together with daily and monthly expenditure and income questionnaires to estimate household expenditure and income classified by geographic and socio-economic variables. The estimates from this survey were to be compared with the results of the 1996-7 Integrated Household Survey (IAF96-7) to measure changes in the economic situation over the last six years.

The survey data processing system, developed in July and August 2002 and extended in February and March 2003 was completed by adjusting the validation, imputation, aggregation and tabulation functions to produce survey results required for publication of the principal survey reports. Cross-tabulations of the general household data, expenditure and income data were prepared in early October for official publication before the end of December, 2003. Sampling errors for the main survey estimates and a time-series of core welfare indicators (2000-1 and 2002-3) were also prepared. The basic structure and contents of the survey database were documented and presented to potential data users in INE and Ministry of Finance.

Training of INE data processing and data analysis staff was limited because the National Demographic and Health Survey (IDS) was being launched at the same time as the mission. It is recommended that future data processing technical assistance, particularly for CWIQ (QUIBB) surveys should be done in a technical workshop conducted before the pilot survey. The objective of the workshop would be for the INE data processing staff to design and implement the survey data processing system with the <u>assistance</u> of a survey data processing consultant. An outline of such a workshop for the next national QUIBB survey is included as an appendix (4.1) to this report.

The data from the IAF2002-3 survey is a very valuable resource for INE and can be used for more than the preparation of the main survey report. INE staff should be trained so that they can analyze the survey data without external assistance.

2 INTRODUCTION

The IAF 2002/03 integrated household survey was a 12 month survey conducted from July, 2002 to June, 2003 using a sample that covered 8700 households. The survey was designed to produce estimates of household expenditure, income and social characteristics at the national, provincial and rural/urban levels.

The survey used four questionnaires

(1) The general household characteristics questionnaire (Carateristicas Gerais do Aggregado Familiar - QCG), based on the questionnaire used for the 2000-1 QUIBB survey recorded:

Household information
Housing characteristics and amenities
Poverty predictors
Victimization

Household member information
Demographic characteristics
Education
Health
Employment
Victimization

(2) The daily household expenses questionnaire (Despesas Diarias - QDD) recorded daily observations of

Household purchases
Household consumption of own production
Gifts in kind received by the household

(3) The monthly household expenses and income questionnaire (Despesas e Receitas - QDR) recorded:

Inventory of durable goods owned by the household
Number and value of durable goods purchased in the last 12 months
Education expenditure in the last 12 months
Household purchases in the last 30 days
Income and revenue in the last 30 days by household member
Transfers received and paid in the last 30 days by household member

(4) In rural areas the community questionnaire (Questionario Comunitario - QC) recorded:

General characteristics of the community
Market prices for selected products

Development of the data processing system for the survey began in April 2002. The system used for the 2000-1 QUIBB survey was adapted to process the general household questionnaire (QCG). This system uses the TELEform software package to capture data from questionnaires using optical scanning technology. The captured data are transferred to a Microsoft Access database

and all subsequent processing is done with programs written in the Visual Basic for Applications (VBA) programming language.

The data entry system for the expenditure and income questionnaires was designed and implemented in July and August 2002. This system was developed using Microsoft Access to allow easy integration with the data from the household questionnaire. The system used very limited data validation procedures: (1) the questionnaire data were entered twice to insure the accuracy and completeness of the entered data; (2) then the entered data were tested for basic consistency based on the logic defined by the questionnaire. Correction procedures were limited to verifying that the data entered in the database were the same as the data recorded in the questionnaire. Development of more extensive validation, imputation, aggregation and tabulation procedures was deferred until a sufficient amount of data was available to define and test these procedures.

By February, 2003 the data from the first 13 periods of the survey (covering the six months from July to December, 2002) had been entered and were available for use to develop the rest of the data processing system. The team of consultants (David Megill, survey sampling specialist; Erwin Triebkorn, survey statistician; and James Otto, survey data processing specialist) worked together with the INE staff to define the methods and procedures needed to complete the data processing system for the survey.

The system was then extended to include the integrated validation of the data from all questionnaires, the imputation of values for observations with incomplete data, the aggregation of data to household and higher levels, the preparation of the standard QUIBB tables, provisional income and expenditure tables and the calculation of sampling errors for all principal survey estimates. This system was used to process the data from the first six months of this survey. The results produced by the system were used to prepare a provisional report for the first six months of the survey. A brief description of the systems functions follows.

<u>Preparation for income and expenditure validation</u>

Much of the validation of the daily expenditure data uses the unit price of observed purchases, consumption of own production and gifts received in kind. This validation is done using a table that summarizes all price information and defines a range of valid unit prices by product. The price table is constructed from two sources: (1) observed purchases recorded in the daily expenditure questionnaire (DD) and (2) local market prices recorded in the community questionnaire. This information is classified by province, urban/rural, period of the survey and product.

A similar approach is used for the price or value of monthly and annual expenditure in the monthly expenditure and income questionnaire. Tables are constructed with the valid ranges of unit prices by product code for durable goods and monthly purchases and the valid range of total value (cost) for monthly purchases with value only and annual education expenses. There are fewer observations for these expenditure items, so the prices/values are classified only by urban/rural and product.

Income and expenditure data validation

Daily expenditure observations with both quantity and total are validated by calculating the unit price (value / quantity) of the observation and comparing it with the range of unit prices defined for the product classified by province, urban/rural and period of the survey. Observations with out of range values are printed, compared to the data recorded in the questionnaire and corrected where necessary.

The unit price or total value of monthly/annual expenditure observations are checked against the range defined for the product classified by urban/rural for the type of observation (durables goods, education expenditure, monthly purchase). Observations with out of range values are printed, compared to the data recorded in the questionnaire and corrected where necessary.

The validation of unit prices/values of the monthly/annual expenditure data was combined with the general consistency checks of all income and expenditure observations and the consistency checks between the general household questionnaire and the income and expenditure questionnaires.

The income and expenditure observations were checked to insure that all codes (product, master unit, place of purchase) were valid and that the observation contained all pertinent data and only pertinent data. These were essentially the same checks done in the post data entry validation of the income and expenditure data.

Consistency checks between questionnaires included (1) rent expenditure with housing tenure; (2) expenditure on water and source of water; (3) salary income and employment status; and (4) education expenditure and children in school. Finally, households with no daily purchases and no consumption of own produce for the entire seven day period were identified.

Corrections were made for all errors where the data printed in the validation error listing did not correspond to data recorded in the questionnaires. Cases where suspicious data was recorded in the questionnaire were to be referred to the validation supervisor or higher authority to determine whether and how to correct the error condition.

Imputation and preparation for aggregation

Imputation is necessary when there is missing data such as the value of own consumption, receipts in kind or annual purchases of durable goods. To calculate the value of such observations, the unit price is imputed using the same price data that was used in the unit price validation described above. The logic of the imputation is to find a price for the product that is a close as possible in time and location to the observation with missing information. Once a suitable price is found the value of the observation is imputed by multiplying the observed quantity by the imputed price. For certain products such as firewood there are observations with neither quantity nor price. Imputation for these "indefinite" observations is done by using the mean value of observations with only value, but no quantity for these products. A similar approach is used for households with no reported payment of rent. For these cases the mean value of rent paid by households in the same

province and urban/rural area is used to impute the rent paid by the household.

Purchases for a few items were recorded in both the daily and the monthly questionnaires. To avoid double counting the daily observations recorded before the date of the monthly questionnaire were identified and subsequently excluded from the aggregation procedures.

Data for monthly receipts and transfers are recorded by person and type of receipt/transfer. This data is converted to the same format as the expenditure data with a six digit product code identifying the type of receipt/transfer and the total monthly value summed for all members of the household.

Aggregation of data to household level

Data are aggregated by product and type of observation (daily purchase, own consumption, monthly purchase, etc.) for each household. The quantity, value, and for daily observations the number of days supply purchased are accumulated for the household. The daily total expenditure is calculated by type of observation and summed to total daily expenditure for the product for the household.

To balance income and expenditure the value of certain observations are duplicated with special product codes in the aggregated household data. The total value of all own produce consumed by the household is generated as part of household income (product code 519001). Transfers received in kind from organizations and family members are also included in income, while receipts in kind of food, lodging, transport and other types are included as expenditure.

The data aggregated by household and six digit product classification is then aggregated by household and the first two digits of the product code (COICOP Division).

The aggregated household data is used to define the following household classification variables:

    Total value of household consumption (expenditure)
    The household expenditure quintile based on weighted population
    Per capita household consumption
    Number of adult equivalents in the household

Preparation of survey results

The survey results consist of data summaries, indicators, cross-tabulations and sampling errors. The data summaries show detailed results by question and province. Summaries are prepared for the household data, household member data and income and expenditure data aggregated at the division (two digit) level. Core welfare indicators show values of standard QUIBB indicators and national, urban/rural and provincial level with the margin of error (95% confidence interval) and values for the rural and urban poor (lowest quintile). Income and expenditure indicators show the mean household expenditure and income estimates at the division level by at

national, urban/rural, geographic zone (North, Center, and South), province levels and margin of error of the national estimates.

Cross tabulations from the household questionnaire (QUIBB) are organized into sections: (1) Response rates and quintile distributions; (2) Household characteristics; (3) Education; (4) Health; (5) Poverty predictors; and (6) Employment.

The income and expenditure cross tabulations show (1) estimates of mean household expenditure, expenditure patterns, per capita expenditure and expenditure per adult equivalent at the division level classified by household characteristics; (2) estimates of mean household expenditure at the national, urban/rural geographic zone and expenditure quintile level by the three digit product code (COICOP group) level modified for compatibility with 1996 IAF classifications.

The sampling errors show estimates of the core welfare indicators and the two digit income and expenditure aggregates at national, urban/rural and provincial level with the standard error, 95% confidence interval, number of observations and an estimate of the design effect.

3   ACTIVITIES DURING THE MISSION

The principal reason for the mission as defined in the terms of reference was to assist INE to complete the processing of the data collected during the 12 months of the IAF 2002/03 survey. The data processing system completed during the six week mission in February and March, 2003 was to be modified based on the experience gained preparing the preliminary report and extended to process all questionnaires and prepare all requested tabulations. In particular, the following tasks needed to be accomplished:

(1) Integration of the community questionnaire data with all other data
(2) Preparation of tables from the victimization module
(3) Facilitate comparison with the 1996/97 IAF survey.

The work was to be done in collaboration with the IAF data processing supervisor, Eugenio Matavel in order to familiarize him with the details of the data processing system and to increase his ability to process the IAF and QUIBB data independently. Unfortunately, close collaboration was not possible because INE was launching the national demographic and health survey (IDS) at the same time that I arrived to start my mission. In fact the data processing advisor for the demographic and health survey arrived for a three week mission the week after my arrival. The exploratory nature (described below) of much of the work of this mission and the pressure to produce results quickly combined to reduce the possibilities for on the job systems training.

The work plan adopted for the mission was to systematically evaluate each phase of the data processing system to correct any shortcomings noted during the processing of the first six months data and the subsequent processing of the remaining data and to fine tune the system to produce the results needed for the survey report as quickly as possible.

Income and expenditure validation

Data entry and basic validation of the survey data was completed by the end of July, 2003, one month after the end of data collection. Several problems with the validation system had been noted. First, the validation of unit prices and values produced a large number of possible exceptions many of which were subsequently accepted. The number of these errors and the time needed to verify and where necessary correct them reduced the attention paid to other possibly more serious errors. Secondly, there was no way other that the double data entry to verify that all data in questionnaire had been entered. Finally, the detection of certain types of extreme values was only possible by examining the distributions of the data aggregated from all the households surveyed.

The first problem was resolved by modifying the validation system to identify exceptions that were most likely actual errors. This reduced the number of validation errors that were identified and insured that every one was closely examined. The error checking and correction was done by the data entry and validation supervisors to insure the quality of the work.

The second and third problems were addressed by "exploratory" validation. This involves preparing the distribution of aggregated household variables, identifying households with extreme values (both low and high) of these variables and then printing all the questionnaire data for these households. The listing of each household's data is checked line by line against the questionnaires to detect errors or identify the reason for the extreme value. Some of the variables used for the exploratory validation were: total household expenditure, total household income, the total value of daily purchases, and consumption of own production, receipts in kind and monthly expenditure. This type of validation is highly interactive and requires a great deal of judgement to determine the best way to resolve any anomalies.

In addition to identifying extreme values in the data, the exploratory validation also discovered some suspicious results that needed to be dealt with. Among these were (1) very large and small values for respondent estimated rents; (2) several households in one enumeration area in Cabo Delgado with very high daily and monthly expenditure and only 1 or 2 household members; (3) very large values (and quantities) of a single day's consumption of own production, particularly maize meal; (4) very large quantities of gifts received in kind, particularly rice. For durable goods purchases in the 12 months preceding the survey, we discovered a limited number of questionnaires where the number or value were not recorded in the questionnaire. All of these cases were examined individually before we decided what action to take. In some cases, the data recorded in the questionnaires were accepted, but most cases were resolved by systematic correction (imputation) of values based on data from other households. A description of the specific methods used follows.

Systematic correction/imputation and aggregation

Except for durable goods, it was possible to apply systematic corrections to the aggregated data while leaving the questionnaire (or raw data) untouched. For the purchases of durable goods, corrections were made directly in the durable goods (BD) table for missing number of items purchased in the last

12 months (13 cases) or the value of items purchased in the last 12 months (49 cases). The mean purchase value of the item calculated from those observations with both number and value was used for correction.

The household aggregation procedures were modified to include systematic correction procedures for extreme and missing values of rent paid by the household and extreme values of gifts in kind for rice. These corrections were made to the aggregated household data leaving the observed questionnaire data unchanged.

Rent paid by the household was recorded in the questionnaire in two ways: actual rent paid (374 households) or estimated rent value of the dwelling (7,710 households). There was no rent information for 616 households. To impute values for households with no rent information and to correct extreme values of estimated rents detected during validation, the distribution of estimated rents was summarized by province, urban/rural and the housing, water/sanitation and energy characteristics of the household. The median value of estimated rent was used to correct households with no rent (616 households) or estimated rents outside the range defined by the 20th percentile and the 90th percentile (1,297 households).

Detailed examination of receipts in kind identified observations with quantities that could not be consumed by the household in one day. Households with values of rice received/person greater than 1,000 meticais were systematically changed to 1,000 times the number of household members. There were 24 such cases.

The household aggregation procedure was changed to include the calculation of use value of the durable goods owned by the household as an alternative to the total value of purchases of durable goods in the 12 months preceding the survey. The use value is a function of the number of items purchased in the 12 months before the survey, the number of items owned for more than 12 months, the median value of all new purchases and an estimate of the depreciation rate for the item. Both the use value and the 12 month total expenditure on durable goods are stored in the aggregated data. This allows analysts to choose the value appropriate to the type of analysis.

The only other change to the household aggregation procedures was the duplication of the value of estimated (product code 042000) and imputed (product code 242201) rent paid by the household as income (product code 518001) to balance the expenditure value.

Data organization and management

The survey databases were modified by (1) creating a new database for the community questionnaire data; and (2) adding variables needed for analysis. The community questionnaire data was entered using the CS Pro survey processing package. CS Pro data is stored in ASCII text format; these data were stored in a Microsoft Access database (IAFQC.mdb) to be compatible with the data from all other questionnaires. There is data for 520 out of 523 rural enumeration areas; one enumeration area was not surveyed and two questionnaires were lost. The community questionnaire data can be linked to other survey data using the enumeration area code.

During the analysis of the data from the first six months of the survey two questions in the victimization module (H7f, H7n) of the household questionnaire were found to have a large number of observations categorized as other.   It was decided to check the questionnaires to determine if interviewers had specified the actual responses in the questionnaire.   This had been done in a sufficient number of cases and it was decided to code these responses and record them in new variables to allow more detailed analysis of this information.   The manually recoded other responses were entered into SPSS data files.   These files were converted to Microsoft Access and linked with the other data from the household questionnaire for further analysis.

Several variables were added to the HhData table to simplify analysis by the period when the data were captured.   These variables are: IntYear – the calendar year of interview (02/03); IntMonth – the calendar month of interview; SurvMonth – Month of survey (01-12); and SurvQuart – the quarter of survey (1-4).   Two additional household weights were calculated for national accounts purposes.   These are CNWeight2002 – the household weight for households interviewed in 2002 and CNWeight2003 – the household weight for households interviewed in 2003.   These weights are used to compute estimates of the aggregated total value of expenditure for calendar year 2002 and calendar year 2003.

Two tables were created to show total expenditure and mean household expenditure classified by the six digit COICOP product code.   These tables (ProdAggr and ProdAggrHh) are aggregated at the national, urban/rural, zone (North, Center, South), national 2002 and national 2003 levels.

Preparation of survey results

The only problem with the tabulations produced with the data from the first six months of the survey concerned the method used to calculate mean per capita household expenditure. This was done by averaging the per capita expenditure calculated for each household (total household expenditure / number of household members). An alternative definition was proposed using total expenditure for all households divided by the total population (i.e. total number of household members). We decided that both definitions were legitimate depending on the type of analysis intended. Two versions of the main tables were produced: tables produced by averaging household results are called Micro tables and tables produced by totaling the household results and subsequently dividing are called Macro tables. The decision on which tables were to be published and how to explain them was deferred.

Micro tables:

D.1 - Media des despesas diarias dos agregados familiares
   - Average household daily expenditure
D.2 - Estrutura percentual da despesa diaria dos agregados familiares (MICRO)
   - Percentage distribution of daily household expenditure [expenditure patterns]
D.3 - Media des despesas diarias per capita dos agregados familiares (MICRO)
   - Average household daily expenditure per capita
D.4 - Media des despesas diarias per adulto equivalente dos agregados familiares (MICRO)
   - Average household daily expenditure per adult equivalent

Macro tables:

DM.1 - Total des despesas diarias dos agregados familiares (em milhãoes de Meticais)
   - Total daily household expenditure (in millions of Meticais)
DM.2 - Estrutura percentual da despesa diaria dos agregados familiares (MACRO)
   - Percentage distribution of total daily household expenditure [expenditure patterns]
DM.3 - Media des despesas diarias per capita dos agregados familiares (MACRO)
   - Average per capita daily expenditure

To facilitate comparison with the results published from the 1996-97 IAF survey, tables were added to show expenditure in the seven categories published for that survey. Both Micro and Macro versions of the tables were prepared using the 1996-97 categories. For the same reason, the basic Micro and Macro tables were produced using 1996-97 definition of urban/rural.

Estimates of mean household income at the national, urban/rural geographic zone and expenditure quintile level by the three digit product code were added to complement similar tables prepared for household expenditure.

For the first six months all tables showed estimates of daily expenditure and income. Tables were added showing estimates of monthly expenditure and income (by multiplying all the cells in the daily tables by 30.4).

A simplified time series showing common core welfare indicators from the QUIBB1 (2000-1) and IAF2002/3 (2002-3) were prepared to facilitate comparison of the results of the two surveys.

Documentation of the survey database contents and structure was prepared to familiarize the INE staff with the data from the survey. The documentation was distributed to all INE staff interested in understanding and using the data. A general presentation of the documentation was made to all interested

parties during the third week of September. The documentation was produced as an Adobe pdf (portable document format) file and can be viewed interactively. The main survey results were presented just before the end of the mission. A CD with the survey documentation and results was prepared and distributed on the last day of the mission.

There was not sufficient time during the mission to complete all the outstanding work. Certain tasks scheduled for the last week of the mission were not done as there were last minute changes to the aggregation and tabulation procedures up to the end of the mission.

The following is a list of important tasks that were not done:

Tables for the community questionnaire

Tables for the individual level data in victimization module

Calculation and application of price deflators in order to prepare spatially adjusted estimates (as an alternative to "nominal" estimates).

Limited Access/Excel training for users of IAF2002/3 data.

4   RECOMMENDATIONS

INE has developed its capacity to collect and process survey data efficiently. However, the capacity to design and develop survey data processing systems and to analyse and publish survey results needs to be considerably strengthened. Because of heavy pressure for fast survey results, technical assistance for these activities has been biased towards producing results rather than building capacity.

Several of the recent surveys conducted by INE, such as the CWIQ and DHS surveys, are complete survey packages that include pre-defined tabulation plans and data processing systems. Typically, the packages are modified to INE's specifications by technical assistance personnel who install the systems and train the INE data processing team to use them. This affords little opportunity for INE staff to increase their ability to design questionnaires, tabulation plans and data processing systems.

The CWIQ project has tried several approaches to enable INE to use the CWIQ technology independently, but none of these have been completely successful. Training courses have been organized during technical assistance missions to introduce the INE data processing staff to the CWIQ development tools: TELEform, Microsoft Access/Office and Visual Basic for Applications. The problem has always been that the pressure of the day-to-day work commitments of trainer and trainees has reduced the time and attention devoted to the training course. Further, two of the data processing staff attended a CWIQ workshop at the Eastern Africa Statistical Training Center (EASTC) in May, 2003. The workshop was originally designed to train data processing personnel to use CWIQ data processing methods independently. Unfortunately, the workshop objectives were changed at the last minute to include survey and questionnaire design and there was not sufficient time to cover the data processing material completely.

Lack of time and personnel are the principal constraints to developing INE's data processing and data analysis capacity. Sufficient time must be available to include technical training as part of survey preparation (development?). Ideally, such training should be done [as part of the survey design prior to] before the pilot survey. The principal objective of the training would be for INE personnel to design and implement the survey data processing system as well as the survey questionnaires, and tabulation plan. Technical assistance personnel should be responsible only for the training and supervision of the INE data processing staff. The actual work must be done by INE personnel. Because system development will be done as a training exercise, it will take longer to develop and sufficient time must be allocated to insure the system is ready before the start of the pilot survey.

The system will be used to completely process and evaluate the pilot survey data. There must be enough time between the pilot and main survey to make and test necessary changes to the questionnaire, tabulation plan, data processing system and manuals. Furthermore, there should be enough time between surveys to insure that essential personnel are available to complete the analysis of the survey data, prepare the survey reports and to archive survey data, results and other information for dissemination.

Appendix 4.2 is an outline for an advanced data processing workshop that could be conducted as part of the preparations for the next CWIQ survey. This could be done in collaboration with INE Angola which is also planning a CWIQ survey in 2004. Funds may be available from the World Bank if needed.

INE should use the data available from the IAF 2002/03 survey to enhance their capacity for data analysis. Once the main survey reports are published, INE staff should prepare a methodological analysis of the survey methods and data. Such a report would be useful for INE and other national statistical offices preparing future income and expenditure surveys. Preparation of the report would be an ideal opportunity for INE staff to learn more about the IAF data and to use different software programs to analyse the data. This should be done as a specific project with both training and practical work components. The training should cover at least Microsoft Access, SPSS and possibly STATA. Most of the training could be accomplished by INE staff assisted by long term technical assistance personnel from INE and the poverty monitoring unit of the Ministry of Planning and Finance. I would be happy to assist with this work via e-mail.

5 APPENDIX 1. Persons met

5.1 Collaborators

João Dias Loureiro	INE President
Manual da Costa Gaspar	INE Vice President, Social & Demographic Statistics
Arão Balate	INE director of surveys and censuses
Fatima Zacarias	INE director of demographic and vital statistics
Jon Teigland	Scanstat resident advisor
Anastácia Honwana	INE data processing manager
Eugenio Matavel	IAF2002 data processing supervisor
Cristovão Muahio	INE director of survey methodology
Carlos Creva Singano	INE survey sampling specialist
Cassiano Chipembe	IAF2002 survey team
Elisio Mazive	IAF2002 survey team
Pedro Duce	IAF2002 survey team
Gustavo Pioris	IAF2002 survey team
Erwin Treibkorn	Statistical consultant
David Megill	Sampling consultant
Channing Arndt	Poverty monitoring unit, MOF
Ken Simler	IFPRI poverty analyst

6 APPENDIX 2. Programme for the Mission

The first three weeks of the mission were devoted to developing and implementing the final data validation. Starting at the end of the third week regular meetings were held with survey staff, technical advisors and data users to discuss the methods to be used for systematic correction, imputation and aggregation of the data. The tabulation plan was also discussed during these meetings.

I also had several meetings with Dr. Loureiro, the President of INE concerning the core welfare indicators time series that was prepared from the indicators common to the 2000-1 and 2002-3 QUIBB questionnaires.

7   APPENDIX 3. Terms of Reference

TERMS OF REFERENCE
Within the Scandinavian Support Program

For a 6-week mission in September/October 2003
on
Final data processing and quality control IAF 2002/03

Background

INE is in the process of finalizing the second national household budget survey (IAF 2002/03). The main objective with this survey is to establish an updated and improved base for evaluating living conditions and poverty in Mozambique, and especially the changes that takes place. Other important objectives are to collect necessary data for updating the national accounts and adjust price indexes.

This survey covers a full year (July 2002- June 2003). To reduce production time, the data processing has been organized as a continuous process. After the first 6 months of data were collected a more thorough quality check of errors and consistency was done with assistance mainly in the shape of gap filling from 2 short-term experts who visited INE in March/April 2003. A preliminary report on some important findings were also prepared and published as a working paper. Limited time made it, however, necessary to postpone:
- quality control of one important part of IAF 2002/03; data collected by the community questionnaire.
- Comparisons with the IAF 1996/97 and QUIBB-surveys.

After June 2003 when the collection of new IAF-data is finished, a final quality control is needed. It is also a need to organize the data both from IAF 1996/97 and 2002/03 so comparisons can be made as efficiently as possible to document the data processing process and to continue to systematically develop INEs competence

The short-term mission presented in this document is one of two planned missions in September/October 2003 that will help INE with the final quality control and planning of the main report with results from IAF 2002/03 and changes from 1996/97.

Main reasons for the mission

During the data processing mission in March/April 2003 the staff of INE got an introduction to the quality control of the new IAF-data. Two members of INE´s staff have later on also participated at a workshop in Tanzania where they were trained in how data can be organized for producing standard tables, based on preliminary table plans for the community data and data on victimization. Despite all these efforts, INE does not have sufficient capacity to fulfill the data processing activities on its own, but need more gap filling support. In particularly INE lacks sufficient competence in control of data consistency and more advanced data processing... Further training will also be carried out during the mission to avoid future need of gap filling missions. It is also a clear advantage to have expertise available that can advice INE to:

- reduce time used on data processing,
- integrate data from the four different questionnaires used in IAF-2002/03 and the 1996/97-survey,
- prepare necessary data-files for analysis, and
- contribute to the documentation of the IAF data-files.

Benefactors of the mission

The mission will benefit the decision makers and other users that depend on rapid and reliable statistics on level of living and poverty in Mozambique by reducing production time and improving data quality. INE-staff working with data processing will improve their qualifications in the area. It will improve INEs competence in documentation and processing of data for large surveys.

Objectives of the mission

- Enhance INEs capacity in data processing and documentation to avoid situation of gap filling
- Contribute to the production of the final report from IAF 2002/03……..

Activities by the Consultant

- Carry out in house training activities according to a programme that INE has prepared and been agreed with the Consultant
- Document the data processing procedures in close cooperation with the counterpart to enable the counterpart to fulfill the documentation after the departure of the consultant (if need arises)
- Check the quality controls that have been done by INEs staff on data from the final 6 months of IAF 2002/03 and secure consistency with the first 6 months.
- Do a quality control of errors and consistency on the data from the community survey.
- Develop specifications for the integrated validation of the QUIBB questionnaire and the income and expenditure questionnaires
- Develop specifications for the aggregation of the income and expenditure data. This includes methods for the imputation of values for all transactions without values, and calculation of spatial and temporal weights to allow annual national estimates of household income and expenditure.
- Develop programs to implement these specifications.
- Document the data processing process, data-bases and files
- Organize IAF-data in an efficient way for analysis by standard statistical packages used at INE, including necessary data preparations for comparisons of IAF 1996/97 and IAF 2002/03.

Expected results

- Sufficient capacity at INE to be able to fulfill the data processing work (including documentation)
- Instructions to assist INE to fulfill the work
- Recommendations on future support (if any)

- Data that has high quality and are easy to use as base for the main report from IAF 2002/03 and later analytical work at INE and at other institutions

- INE-staff is more qualified for data control and producing.

Agenda for the mission

Tasks to be done by INE to facilitate the mission

- Analyze INEs need for in-house training that can be carried out by the consultant and propose a programme for this.
- Have finished the data-punching processing before the consultant, including the community data
- Supply good working conditions for the consultant

Consultant, Counterpart and LTA

Consultant:
James Otto, The World Bank

Main counterparts :

Arao Balate and Elisio Mazive

LTA:
Jon Teigland

Timing of the mission

6 weeks September/October 2003

Report

The consultant will prepare a draft report to be discussed with INE before leaving Maputo. He will submit a final draft to INE for final comments within one week of the end of the mission. Statistics Denmark as Lead Party will print the final version within 3 weeks of the end of the mission. The structure of the report should be according to Danida-format. The Counterpart has to ensure that the final printed report has at least a summary in Portuguese if the main report is in English – or vice versa.

These Terms of Reference were prepared by

Day   /   /
....

Approved by/in the name of the President of INE

Day   /   /   ....

8   APPENDIX 4.1

Proposal for an advanced CWIQ data processing workshop

Participants:

Statisticians and survey data processing staff from INE Mozambique
Statisticians and survey data processing staff from INE Angola
Long term technical assistance advisors from INE Mozambique
Long term technical assistance advisors from INE Angola
World Bank CWIQ consultant(s)

Objectives:

Capacity building:

1. Increase capacity of INE Mozambique to independently use CWIQ methodology and technology.

2. Increase capacity of INE Mozambique to assist other organizations to use CWIQ methodology and technology.

3. Develop capacity of INE Angola to use CWIQ methodology and technology.

Survey preparation:

1. Design questionnaires for pilot and national CWIQ surveys in 2004.

2. Define validation rules and tabulation plans for these surveys.

3. Develop, test and implement data processing systems for these surveys.

   Data entry: conversion to Access and transfer to CWIQ database
   Data validation and correction
   Preparation of survey results

   Data summaries
   Standard indicators
   Cross-tabulations
   Sampling errors

4. Develop survey manuals and training materials

   Interviewers manual
   Supervisors manual
   Training exercises
   Data processing manual