



A household data collection system with a continuous multipurpose Household Budget Survey Module (INCAF/IOF)

**Report from a second short term mission to the National Statistical
Institute of Mozambique, Maputo Mozambique**

09 - 24 April 2015

within the framework of the

***AGREEMENT ON CONSULTING ON
INSTITUTIONAL CAPACITY BUILDING,
ECONOMIC STATISTICS AND RELATED AREAS***

between
INE and Scanstat

Lars Lundgren



INSTITUTO NACIONAL DE ESTATÍSTICA

Ref: Contract DARH/2008 /004
October, 2014

Lars Lundgren

larsalundgren@hotmail.com

Contents

SUMMARY	3
THE CONSULTANCY	4
RECOMMENDATIONS FOR INCAF/IOF 2014/15.....	4
Data quality.....	4
Missing information	5
Sample weights.....	6
A modernized dissemination strategy.....	6
Well-being	9
Next mission	9
IOF continues	14
Annex 1 Terms of Reference	
Annex 2 Preliminary workshop program	
Annex 3 How to use a three-quarter survey – draft letter	
Annex 4 How to use a three-quarter survey – A separate PowerPoint presentation	

Summary

INE has modernised its household survey system by introducing a continuous multi-purpose survey (INCAF). The current INCAF has a household budget module IOF (Inquérito ao Orcamento Familiar), with an extended sample. The two first quarters are surveyed and data can be analysed, to present some results.

The third quarter has not been surveyed, awaiting the Parliament's budget confirmation. Some districts in Zambezi as well as a few other district was missed during the second quarter, due to flooding.

Part of the mission was used to explain how the survey still will be useful, especially as it was confirmed that the fourth quarter can be surveyed. Income and semi-durable expenditure for the third quarter will also be collected during the last visit to the households.

Most of the mission was dedicated to training in analysis, mainly of labour force statistics. The national unemployment definition was particularly discussed.

Issues	Findings and recommendations
Data quality	The general quality is so far adequate in relation to the complexity of the survey. Expenditure data were still entered and edited and labour force data for the second quarter were still edited.
Missing quarter/s	<p>The budget for IoF has been released, so the fourth quarter can be surveyed. There is not much use of another quarter outside the planned year. It is better to use available time and resources to restart the continuous INCAF as soon as possible (maybe 2016-01-01).</p> <p>The missing third quarter can partly be surveyed by extended questions in the fourth quarter. It is also possible to use structural data from the previous IOF to impute data.</p>
Missing districts	The missing data due to flooding for the second quarter cannot be recovered. Two options are available: To exclude those districts in the analysis or to impute values on them. The use of the data should determine the solution. For poverty estimates it may be best to impute data to avoid the impact of the flooding.
Weights	Weights have been calculated for each of the first two quarters and for the panel, i.e. for household participating in both quarters. Preliminary analysis shows small differences between full quarterly estimates and estimates based only on the panel. If this still is the case after the 4 th quarter, it is recommended to only use the panel households and weights. The database is already complex and adding several different weight datasets can confuse, with risk for mismanagement.
Reporting strategy	With the new technics INE should leave the traditional report system with production oriented releases (with repeating quarterly, annual and survey reports) and adopt a more user oriented strategy where data are released when needed and in the form that is needed. Quarterly only a few indicators (e.g. employment rates and worked hours) should quickly be released (similar to the Consumer Price Index). Deeper analysis (e.g. of the labour force) should be released annually. Specific modules should be released separately as soon as possible.

IOF tabulation and dissemination	With the data gaps due to flooding and lack of funds, it is recommended to wait with further release of data till the fourth quarter is available and then report previous quarters in the annual report.
Employment, unemployment and under-employment	A preliminary report has a very high unemployment rate, based on the national definition. This definition can be interpreted in different ways. The used ILO concept for questioning means that employment and unemployment are asked for a sampled week. Activities outside this period should not be moved into the period. A relaxed definition of unemployed can be used, where people 15+ not working but available for work can be seen as unemployed even if they didn't actively seek for job. Any other groups according to the national definition should be presented as under-employed, not unemployed.
Well-being	The previous mission recommended a new module monitoring well-being (subjective welfare). With the extended questions for semi-durables and income (to partly cover the third quarter) the response burden is already increased. It is therefore recommended to postpone the module to a later INCAF.
Next mission	The next mission will check the quality of the complete database and give recommendations for calibration for the missing third quarter and missing districts in the second quarter. Labour force data and expenditure data will be analysed more in depth as part of continued training.

The consultancy

The aim of the consultancy was to secure that the work goes along international standards. This and a coming mission is a continuation of previous missions. The main expected activities are:

1. Support the INE in the verification of the database to identify any potential problems that could hide within the data;
2. Support the INE in making tabulations plans both quarterly as well as yearly,
3. Support the INE with the procedures of data aggregation for the first two quarters;
4. Make recommendations for the coming quarters related to the problems encountered in the actual data base;
5. A second line of activities is to review and propose a solution of a possible integration between INCAF and the annual agricultural survey IAI.
6. The consultant will work closely with the Continuous Multi-propose Survey team at INE

The first mission focused on activities 1-2 and 4. Some recommendations were given on activity 3 and 5. This mission focused more on activities 3 and 4, particularly how to manage the missing third quarter and the flooded areas in the second quarter. Full ToR are found in annex 1.

A workshop (7 days x 2 hours) with 4-6 participants was discussing the analysis of labour force statistics and links to welfare statistics. The preliminary program is in annex 2. More time was in fact devoted to the unemployment definitions.

Recommendations for INCAF/IOF 2014/15

Data quality

The general quality is adequate in relation to the complexity of the survey.

The use of paper questionnaires for the expenditures has delayed the results. The central data entry for the second quarter was still going on. The editing/cleaning of the data will probably take more time than usual, as it is in fact three data sources: one on paper, one in the tablets and one from central registration. The editing of labour force data for the second quarter was not yet finalised. Hopefully will the missing third quarter give time to catch-up.

Missing information

The fourth quarter of the survey will be surveyed as funds are released. It has been questioned how useful the IOF will be with only three of four quarters.

First it should be clarified that statistics are about making sense of more or less partial information. Few statistics are based on full information. E.g. sampling means that the information is limited to a few of all households and the consumption is limited to a short period in time. Many countries are doing this type of survey without covering the full calendar year. The daily consumption will now be covered for three weeks, which is much better than the previous IOF with only one week of daily consumption. This is particularly important when classifying each household as poor or rich.

As IOF is a sample in both space and time, the current IOF will have a more than three times bigger sample compared to the previous IOF. With the use of tablets, the quality of the data will also be better.

To compensate for the missing third quarter it is proposed to extend the questioning of semi-durables from past three to past six month. In general it is more difficult to remember a the longer period. It is not only the length of the period, but also the start of the period. In our case, however, we can and will refer to the last visit.

Similarly it is believed that income can be remembered for a longer time, often asked for the past 12 months. It is proposed to collect income in the fourth quarter for the past month and the calendar month four month ago. The tablets are programmed to ask for the specific month in the 3rd quarter.

Non-durables like food cannot be remembered for long and should not be asked for more than one-two weeks back. This consumption varies less over the year, different from semi-durables, for which a lot (e.g. toys) are bought in December. A quick analysis of the previous IOF reveals that the same period as our third quarter has one quarter of the total consumption (24%) and one quarter (25%) of the full year food consumption. Besides the more even daily consumption it can be explained by the fact that our missing period is between the harvests. Households tend to consume less before the harvest and more after the harvest, when prices are lower and the household can't store all food for the future.

Table 1 Consumption by month in IOF 2008/2009

Month	Households	Food	Alc, tob	Cloth	Hous e	Furn ish	Health	Transp	Comm unic	Recre ation	Educ	Hotel, rest	Misc	Total
8	333,112	2828	13	160	554	82	3	119	59	6	2	33	38	3897
9	192,040	3944	123	367	998	182	15	229	209	28	1	59	67	6221
10	297,984	3889	102	248	727	138	14	152	109	12	1	56	61	5511
11	327,745	2779	35	165	562	91	13	103	39	7	0	19	38	3851
12	298,467	4513	137	286	912	143	12	194	211	19	0	131	67	6627
1	299,065	3677	59	272	886	122	11	167	109	13	2	59	67	5443
2	272,966	3344	33	111	586	67	6	81	39	20	2	18	29	4335
3	274,289	3547	18	109	921	94	18	184	92	33	1	53	72	5142
4	232,562	3723	45	155	694	127	22	132	88	28	2	46	65	5128

5	298,023	3255	34	138	577	80	8	91	65	6	1	36	36	4328
6	210,636	3451	53	197	887	126	9	210	139	19	2	100	84	5277
7	233,425	3071	30	186	682	100	7	142	180	8	0	63	45	4514
Total	3,270,314	42022	681	2395	8987	1351	138	1805	1339	199	15	673	668	60272
2-4		10613	95	375	2202	288	46	397	219	81	5	117	165	14605
Year %		25%	14%	16%	25%	21%	34%	22%	16%	41%	35%	17%	25%	24%
1-12		70%	1%	4%	15%	2%	0%	3%	2%	0%	0%	1%	1%	100%
2-4		73%	1%	3%	15%	2%	0%	3%	1%	1%	0%	1%	1%	100%

First it can be noted that the data collection in IOF 2008/2009 varied a lot from month to month (col. 2). The data should be calibrated so each week in a month and each month will have the same representation in the final statistics. Besides the higher consumption related to Christmas and New Year, Ramadan can have an impact. Two Ramadan were included in the previous IOF. It is also worth noting that consumption deviating much from 25% in the missing period have very small impact on total consumption (last row). The higher food share (73%) in the missing period is not because of higher food consumption, but because of lower other consumption.

It will also be possible to use structural data from the previous IOF to impute data, taking into consideration different regional patterns.

The missing data due to flooding for the second quarter cannot be recovered. Two options are available: To exclude those districts in the analysis or to impute values on them. The use of the data should determine the solution. For poverty estimates it may be best to impute data to avoid the impact of the flooding. Previous IOFs have been blamed for not showing typical years.

On initiative from SIDA a seminar was conducted for the donors and the Ministry for Planning and Development to explain the possibilities to utilise the partial survey. SIDA also asked for a letter to circulate to other stakeholders. A draft letter is found in annex 3 and the PowerPoint presentation at the seminar in annex 4.

It is not recommended to continue with another quarter outside the planned year, as it would mirror the first quarter of IOF. Then it is better to restart the continuous INCAF. The new INCAF should be simplified but still measuring full consumption.

Sample weights

Sample weights have been calculated for each of the first two quarters and for the panel, limited to households participating both quarters.

The main reason for fewer households in the second quarter is flooding in some districts in Zambezi and a few in other regions.

Preliminary analysis shows small differences between full quarterly estimates and estimates based only on the panel. If this still is the case for the full year, it is recommended to only use the panel households and weights. The database is already complex and adding several different weight datasets can confuse, with risk for mismanagement.

Table 2 Employment to population ratio based on full data for each quarter compared to the panel data

Dataset	1 st Quarter	2 nd Quarter
Full data each quarter	78%	82%
Panel data	79%	82%

A modernized dissemination strategy

With the new technics INE should leave the traditional report system with production oriented releases (with repeating quarterly, annual and survey reports) and adopt a more user oriented strategy where data are released when needed and in the form that is needed. Quarterly only a few indicators (e.g. employment rates and worked hours) should quickly be released (similar to the Consumer Price Index), comparing with previous quarter and the same quarter one year ago. Deeper analysis (e.g. of the labour force) should be released annually. Specific modules should be released separately as soon as possible. Quarterly employment index could be:

Indicator	This quarter	Previous quarter	Same quarter previous year
Employed with salary			
Employed without salary			
Unemployed (available)			
Under-employed			
Worked hours with salary			
Other worked hours			

Tabulation

In the previous report it was recommended that production tables should be managed in a systematic way. The best would be to manage all data in a relational database, such as SQL or its daughter MS Access, where all data can be linked without repeating redundancies. Weighting and other manipulations of the data can either be done before storing them (as now in SPSS) or more safe within the database program. The production tables in e.g. Excel should automatically be updated if some data are updated.

Analysis of IOF 2014/15

With the data gaps due to flooding and lack of funds, it is recommended to wait with further release of data till the fourth quarter is available and then report previous quarters in the annual report based on panel data.

Analysis of employment, unemployment and underemployment

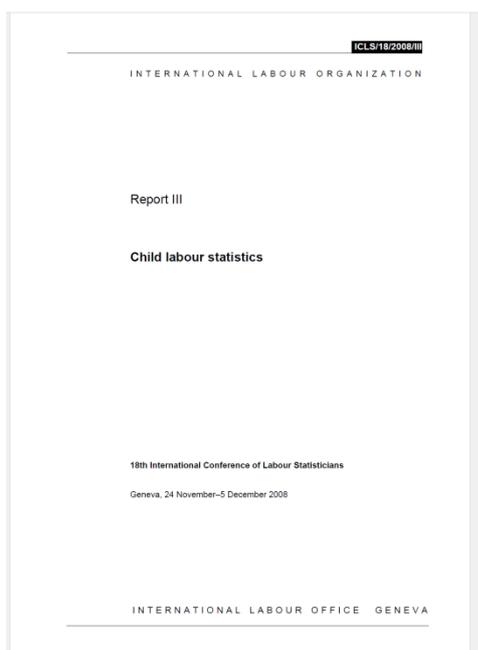
A preliminary report has a very high unemployment rate, based on the national definition. This definition can be interpreted in different ways. The used ILO concept for questioning means that employment and un-employment are asked for a sampled week. Activities outside this period should not be moved into the period in the same way as a neighbour cannot be included in the sample. A relaxed definition of unemployed can be used, where people 15+ not working but available for work can be seen as unemployed even if they didn't actively seek for job. Any other groups according to the national definition should be presented as under-employed. A necessary criteria for both unemployment and underemployment is that the person want to work more. Preliminary results from the first quarter reveals that the unemployment rate is 3% according to the ILO definition, 7% according to the relaxed definition and 13% according to the national definition (which should be called under-employment). If the "want more work" criteria is removed, the national unemployment (under-employment) rate is doubled.



The Labour Force Data Analysis by the African Development Bank is a very good manual for labour force surveys and was used as reference material in the workshop/training and was left as a soft copy (unfortunately only available in English)

Child Labour

“Child labour” is not just about employment and unemployment by younger ages. “Child labour” is something bad. Bad means that work is hazardous or hindering going to school or working too many hours. Therefore criteria have to be used interpreting international recommendations and national laws. The latest MICS by UNICEF made up such criteria.



The ILO Guidelines for Child labour statistics is also a good manual and was used as reference material and left as soft copy.

Format

Some basic rules for reports were also discussed at the workshop, e.g. not to use decimals or graphs with cut off scales. When using %, it should always be mentioned what it is percentage of. Important definitions like unemployment rate should be clarified in the text or a nearby fact box, not in an appendix.



The User Friendly Presentation by Statistics Norway and Paris 21 is an up to date guideline for modern statistical reporting (The Portuguese version was left as soft copy.

The guide can also be found at <http://www.paris21.org/knowledgebase/> and <http://www.ssb.no/english/int/>).

Well-being

IOF will be extended to cover even the third quarter for all semi-durable goods and incomes. It is therefore advisable not to add any extra module. Training of the interviewers must start almost at once, and the proposed module is far from ready.

Next mission

The next mission will check the quality of the complete database and give recommendations for calibration for the missing third quarter and missing districts in the second quarter. Labour force data and expenditure data will be analysed more in depth as part of continued training.

The mission should not start until all data are cleaned and weights for quarter four and the panel has been calculated.

TERMS OF REFERENCE

for 3 short-term missions on

INCAF/IOF the Household Budget module of the Continuous Multi-purpose Survey

23 November – 20 December 2014,
23 February – 20 March 2015 and
12 October – 6 November 2015

within the

AGREEMENT ON CONSULTING IN INSTITUTIONAL CAPACITY BUILDING,
ECONOMIC STATISTICS AND RELATED AREAS
between INE and Scanstat.

Consultant: Lars Lundgren through Statistics Sweden

Counterparts: Arão Balate and the INCAF team at INE

Background

The National Statistics Institute of Mozambique (INE) has re-designed the household budget survey Inquérito sobre o Orçamento Familiar (IOF) to function as a module of the continuous multi-purpose survey (INCAF) which covers the seasonality in household income and expenditures over a period of 12 months. The International Sampling Consultant will provide technical assistance and training to INE on the INCAF/IOF sampling and estimation procedures. This consulting assignment will take place in three separate missions planned for November 2014, February 2015 and September 2015 and the consultant will coordinate the work with other Scanstat consultants working in the same field.

Main reasons for the mission

The INCAF/IOF has been going on since July 2015 and results are obtained each quarter. There is a need for oversight, assistance and further advice on how to best go forward.

Objective

To secure that the work goes along international standards.

Activities

This is a continuation of previous missions and the Consultant is responsible for the following activities during each of the missions:

7. Support the INE in the verification of the database to identify any potential problems that could hide within the data;
8. Support the INE in making tabulations plans both quarterly, semi-annual as well as yearly
9. Support the INE with the procedures of data aggregation by quarter, semester and year, given the methodology of a panel survey;
10. Make recommendations for the coming quarters related to the problems encountered in the actual data base.

11. A second line of activities is to review and propose a solution of a possible integration between INCAF and the annual agricultural survey IAI.
12. The consultant will work closely with the Continuous Multi-propose Survey team at INE

Expected outputs

The consultant will prepare a written documentation of the above activities;

Beneficiaries of the mission

The mission will benefit INE and the whole National Statistical System of Mozambique including users of statistical information.

Tasks to be done by INE to facilitate the mission

- Elaborate ToR for the missions
- Prepare and supply the consultant with necessary documents and information, like the new visual identity of INE
- Supply good communication conditions for the consultant.

Source of Funding

Project: MPD-2008-0006 – Inquérito Sobre Orçamento Familiar – IOF
 PAAO14 – 1.2.2 Inquérito Contínuo aos Agregados Familiares
 PAAO15 – 1.2.2 Inquérito Contínuo aos Agregados Familiares

Timing of the mission

Three missions on four weeks each, as written above. Number of workdays for each mission: 18 days in Maputo (6 days x 3 weeks) and a further 6 days of remote work.

Place

The premises of INE in Maputo with possible allocations to the provinces.

Language

English.

Report

The consultant will prepare a short final report to be discussed with INE before ending assignment. Statistics Denmark as Lead Party will publish the final version on www.dst.dk/mozambique within 3+ weeks of the end of the mission. The structure of the report should be according to Scanstat format.

Approved by Arão Balate, INE/DCI

Day / /

Confirmed by Leia Gimo Macamo, INE/DARH and Contract Manager for the INE – Scanstat Contract

Day / /

Day	English	Portuguese
1	Update quarter 1	Atualização 1 trimestre
2	Child labour	Trabalho infantil
3	Quarter 2 and trends using panel data	Trimestre 2 e tendências usando dados em painel
4	Extended analysis	Análise Alargada
5	Missing quarter/s	Faltando trimestre / s
6	Welfare	Bem-estar
7	Multidimensionality	Dade Multidimensional
8	Future INCAF	Futuro INCAF

1 Make more sense of quarter 1/Faça mais sentido do trimestre 1

Topic	English	Portuguese
A	Modernising - From paper to web - From producer quality (reliability) to user quality (relevance) - From producer view to user view	Modernizar - Do papel para o web - De qualidade produtor (confiabilidade) para a qualidade do usuário (relevância) - Do ponto de vista produtor de vista do usuário
B	Explanations	Explicações
C	% of ?	% De?
D	Graphs	Graphs
E	15-64 or 15+	15-64 ou 15+
F	New data and weights	Novos dados e pesos
G	QP01: 9	QP01: 9
H	Unemployment?	Desemprego?

2 Make more sense of child labour/ Faça mais sentido do trabalho infantil

Topic	English	Portuguese
A	Why child labour?	Por que o trabalho infantil?
B	Laws and definitions	Leis e definições
C	Child labour	Trabalho infantil
D	Labour among children	Trabalho entre as crianças
E	New data and weights	Novos dados e pesos

3 Make sense of trends/Faça sentido de tendências

Topic	English	Portuguese
A	Quarterly, semi-annual and annual	Trimestral, semestral e anual
B	Cross-sectional or panel or both	Transversal ou painel ou ambos
C	Flooding	Inundação
D	General and current	Geral e atual
E	Index	Index Índice
F	Panel data for quarter 1 and 2	Dados em painel para quarto 1 e 2

4 Make more sense of data/Faça mais senso de dados

Topic	English	Portuguese
A	Long-term development	Desenvolvimento a longo prazo
B	Utilisation of other statistics	Utilização de outras estatísticas
C	Population and labour force	População e força de trabalho
D	Worked hours	Horas trabalhadas
E	Employment and education	Emprego e educação
F	Employment and welfare	Emprego e bem-estar

5 Filling the gaps/Pree tonchendo as lacunas

Topic	English	Portuguese
A	Q1+Q2+Q4/Q1+Q2	T1 + T2 + T4 /T1 + T2
B	Sample size	O tamanho da amostra
C	Bias?	Viés?
D	Adjustments	Ajustes
E	Revised Q4: income	Q4 Revisado: renda

6 Looking for the money/Olhando para o dinheiro

Topic	English	Portuguese
A	Money-metrics	Dinheiro-metrics
B	Engel's law	Engel's lei
C	Income	Renda
D	Durables	Durables
E	Scale of equivalence	Escala de equivalência
F	Poverty line	Linha da pobreza
G	Gini	Gini

7 Multidimensional welfare/Multidimensional bem-estar

Topic	English	Portuguese
A	Welfare more than money	Bem-estar mais do que dinheiro
B	Well-being	Satisfaction
C	Housing and welfare	Ben-estar e CONDIÇÕES DA HABITAÇÃO
D	Welfare and labour	Bem-estar e de trabalho
E	Welfare and education	Bem-estar e educação
F	Welfare and health	Bem-estar e saúde
G	Welfare and tourism	Bem-estar e turismo
H	Resilience and welfare	Bem-estar e CALAMIDADES NATURAIS
I	Subjective welfare	INDICADORES DE POBREZA

8 Future INCAF/Futuro INCAF

Topic	English	Portuguese
A	Simplifications	Simplificações
B	Core labour indicators	Indicadores fundamentais do trabalho
C	Hierarchical questions	Questões hierárquicas
D	No quantities	Sem quantidades
E	1 day auto-consumption	1 dia auto-consumpt.
F	Smaller sample	Menor amostra
G	Module program	Programa Módulo

IOF continues

Funds are released, so the fourth quarter of IOF can be surveyed. This is a very positive news, but the question arise: What to do with a three quarters survey? Is it useful at all?

Absolutely! It may not give all details as planned, but the data are enough for main purposes of the survey.

First it should be clarified that statistics are about making sense of more or less partial information. Few statistics are based on full information. E.g. sampling means that the information is limited to a few of all households and the consumption is limited to a short period in time.

Many countries are doing this type of survey without covering the full calendar year. The daily consumption will now be covered for three weeks, which is much better than the previous IOF with only one week of daily consumption. This is particularly important when classifying each household as poor or rich.

As IOF is sample in both space and time, the current IOF will have a more than three time bigger sample compared to the previous IOF. With the use of tablets, the quality of the data will also be better.

To compensate for the missing third quarter the questioning of semi-durables will be extended from past three to past six month. In general it is more difficult to remember, the longer is the period. It is not only the length of the period, but also the start of the period. In our case we can and will refer to the last visit.

Similarly it is believed that income can be remembered for a longer time, often asked for the past 12 months. It is proposed to collect income in the fourth quarter for the past month and the calendar month four month ago. The tablets are programmed to ask for the specific month.

Non-durables like food cannot be remembered for long and should not be asked for more than one-two weeks back. This consumption varies less over the year, different from semi-durables, for which a lot (e.g. toys) are bought in December. A quick analysis of the previous IOF reveals that the same period as our third quarter has one quarter (24%) of the total consumption and one quarter (25%) of the full year food consumption. Besides the more even daily consumption it can be explained by the fact that our missing period is between the harvests. Households tend to consume less before the harvest and more after the harvest, when prices are lower and the household can't store all food for the future.

Table 1 Consumption in the third quarter in IOF 2008/2009

	Food	Alc , to b	Clot h	Hous e	Fur nis h	Trans Healt p	Com muni c	Recr eatio n	Educ	Hotel , rest	Mis c	Total	
3 rd quarter			16		21	34						25	
share	25%	14%	%	25%	%	%	22%	16%	41%	35%	17%	%	24%
				15								100	
Full year	70%	1%	4%	%	2%	0%	3%	2%	0%	0%	1%	1%	%
				15								100	
3 rd quarter	73%	1%	3%	%	2%	0%	3%	1%	1%	0%	1%	1%	%

It is worth noting that consumption deviating much from 25% in the missing period have very small impact on total consumption (last row). The higher food share (73%) in the missing period is not because of higher food consumption, but because of lower other consumption.

It will also be possible to use structural data from the previous IOF to impute data, taking into consideration different regional patterns.

A permanent survey like INCAF will make such interruptions less sensitive.

It is also worth mentioning that this type of surveys are much more partial in developed countries, due to very low response rates (mostly below 50 %). The IOF will have more than 80% response rate.