

LESSONS FROM TIA 96 TO IMPROVE FUTURE TIA ACTIVITIES

Paul Strasberg
Michigan State University
March 31, 1998

The Department of Statistics (DEST) and the Ministry of Agriculture in Fisheries have gained much valuable experience concerning household-level data collection over the last five years, during which three national agricultural surveys of smallholders have been carried out. Notwithstanding these achievements, there are a number of specific suggestions offered below to improve the quality of implementation of future TIA surveys.

SECTION ONE: QUESTIONNAIRE FORMAT AND DESIGN

1. Key variables and household identification

a. General Problem:

Data in any agricultural survey is typically collected at multiple levels (e.g., household, plot and product). Frequently, in data verification and analysis, it is necessary to match the cases of each household found in different files. Key variables identifying each case, where each case has a unique set of key variables are required.

b. Problem from 1996 and Suggestion:

The following is an example of how the unique identifying variables for the households in two villages was implemented in data entry.

There were four identifying variables:

PROV = province
DIST = district
VIL = village
AF = af

prov	dist	ald	af
1	3	1	467
1	3	1	549
1	3	1	693
1	3	1	759
1	3	1	807
1	3	1	906

1	3	1	961
1	3	1	979
1	3	2	255
1	3	2	277
1	3	2	557
1	3	2	787
1	3	2	935
1	3	2	1301
1	3	2	1474
1	3	2	1834

An improved approach would be to number each of 1, 2, 3, 4, 5, 6, 7, 8 for each eight households in each village.

Another problem concerns the numbering of villages. Villages were not numbered consecutively in all instances. Consider the following example from the current data set:

prov	dist	ald
3	12	3
3	12	6
3	12	7
3	12	9
3	12	10
3	12	11
3	12	12
3	12	13

It is recommended that all villages be numbered consecutive 1, 2, 3, 4, 5, 6, 7, 8. Likewise, the same procedures should be used to number districts. The numbering system for provinces should be maintained, as it is already done with consecutive numbers.

c. How will these changes improve TIA?

The 1996 survey generated 16 pages for each of 3,889 households interviewed, or 62,224 pages of survey data. In the process of data entry and processing, it is easier for DEST staff to accurately process these documents if they are well-organized. The suggested approach should improve organization at each step of the process.

2. Variable names

a. General Problem:

Once collected, data on survey questionnaires is entered by a data entry staff member. The process around data entry is crucial, and the design of the underlying questionnaire can improve the process. Experience from the 1994/96 MAP/MSU FSP Nampula/Cabo Delgado Socio-Economic Survey should guide how we conceptualize the naming of variables for future TIAs. It is recommended that a method which is as systematic as possible be developed, as this will ensure easy communication between team members and analysts about particular data issues throughout the research and report production process. Once developed, this method should be explicitly written up in a document. It is then recommended that such a system be taught to DEST staff.

b. Problem from 1996 and suggestion:

Let us analyze the variable naming systems practiced in TIA 1996. Questions at the household-level, such as

A.1: A familia e originaria desta zone?

A.2: Desde quando a familia vive aqui? and

A.17 Quanto tempo leva por semana a buscar lenha?

were all found on one page, and all started with A. Each had a dot after the "A."

Other questions at the household level were:

D0: Quantas pessoas compoem este agregado?

It is recommended that for all variables at a particular level, such as the set of above examples at the "household" level, the following system be adopted.

AF1	_____	Quantas pessoas compoem este agregado?
AF2	_____	Desde quando a familia vive aqui?
AF3	_____	
AF4	_____	
AF5	_____	
AF6...	_____	

c. How will these changes improve TIA?

TIA collected data at 11 levels in 1996. It would be best if for each data level, a system is used that connects data level, data file name and variable names. Assuming the same data levels would be collected in the next TIA, a workable system would be the following:

NOME	DESCRIÇÃO	VARIÁVEIS CHAVE	<u>VARIABLE NAMES</u>
AF		PROV DIST ALD AF	<u>AF1, AF2, AF3...</u>
MEM		PROV DIST ALD AF MEM	<u>MEM1, MEM2, MEM3...</u>
MEIO		PROV DIST ALD AF MEIO	<u>MEIO1, MEIO2, MEIO3...</u>
CULT		PROV DIST ALD AF CULTURA	<u>CULT1, CULT2, CULT3...</u>
FRUTA		PROV DIST ALD AF FRUTA	<u>FRUTA1, FRUTA2, FRUTA3</u>
PROD		PROV DIST ALD AF CULTURA	<u>PROD1, PROD2, PROD3...</u>
HORT		PROV DIST ALD AF HORTIC	<u>HORT1, HORT2, HORT3, ...</u>
PEC		PROV DIST ALD AF ANIMAL	<u>PEC1, PEC2, PEC3...</u>
MACH		PROV DIST ALD AF MACH	<u>MACH1, MACH2..</u>
TF		PROV DIST ALD AF MEM	<u>TF1, TF2, TF3...</u>
TE		PROV DIST ALD AF X2	<u>TE1, TE2, TE3...</u>

The result of such a plan would likely result in all individuals involved in the research process becoming familiar in a practical sense with this systematized method of naming variables.

3. Value Labels / Categorical Coding

a. General Problem:

Data collection, processing and analysis are facilitated by a set of value labels which has a well-developed set of logic behind it. As a general principle, it is preferred to have the code on the questionnaire for a particular item, if the appropriate response as designed by the respondent, be entered in exactly the same form as which it appears on the questionnaire. **There are three distinct problems types which will be discussed below.**

b. Problem from 1996 and suggestion:

PROBLEM TYPE 1:

Let us take the example from the MEM - level table and variable D4. The possible codes were listed as

“1.Proprio”
“2.Conjuge”...

There are two problems with this approach: a) there should be more space between the code and the text; and b) there was an unnecessary dot.

In such an instance, it is recommended that the “.” be eliminated based on the principle that it will not be entered and represents an unnecessary source of confusion for some survey enumerators. Further, it is recommended that more space be allocated between the actual number of the code and the text defining the code.

PROBLEM TYPE 2:

Let us take the example of crop code. Crops were coded as:

- 1 milho
- 2 arroz
- 3 mandioca
- 4 feijao nhemba
- 5 feijao manteiga
- 6 mapira
- 7 batata doce
- 8 gergelim
- 9 mexoeira
- 21 castanha de caju
- 22 coco
- 23 algodao
- 24 girassol
- 25 amendoim
- 26 cana de acucar
- 27 mafurra
- 28 cafe
- 29 tabaco
- 30 sisal
- 31 soja

- 32 cha
- 33 gengibre
- 34 copra
- 56 outros feijoes
- 61 bananeira
- 62 mangueira
- 63 laranjeira
- 64 limoeiro
- 65 toranjeira
- 66 ananazeiro
- 67 abacateira
- 68 pereira
- 69 videira
- 70 papaeira
- 71 pessegueira
- 72 litcheiro
- 73 macaniqueira
- 74 macieira
- 75 tangerineira
- 76 goiabeira
- 77 ateira
- 78 outro

A key problem with this approach is that products were not consistently grouped by type (food, cash, fruit, vegetable), though this apparently was the intention. That is, products 1, 2, 3, 4, 5, 6, 7, 9, 25 and 56 were “food crops,” while products 8, 21, 22, 23, 24, and 25-34 were “cash crops.” A suggested improvement would be to do the following:

- a) List all product types about which data will be collected in the questionnaire. It is recommended that this list of product types be done in such a way that “estado” of the product be eliminated as a variable. Below is an example of one possible scheme:

Cultura	Estado	Code
Culturas Alimentares: Grains, Pulses, Oilseeds, Tubers		
milho	<i>grao</i>	11
	<i>em espiga</i>	12
arroz	<i>sem casca</i>	21
	<i>com casca</i>	22
mandioca	<i>seca</i>	31
	<i>fresca</i>	32
feijao manteiga	<i>seco sem casca</i>	41
	<i>seco com casca</i>	42
feijao nhemba	<i>seco sem casca</i>	51

	<i>seco com casca</i>	52
feijao outro	<i>seco sem casca</i>	61
	<i>seco com casca</i>	62
mapira	<i>grao</i>	71
	<i>seco com espiga</i>	72
mexoeira	<i>grao</i>	81
	<i>seco com espiga</i>	82
batata doce	<i>fresco</i>	91
girassol	<i>seco</i>	101
gergelim	<i>seco</i>	111
<i>Vegetables and Fruits</i>		
alface	???	201
couve	???	202
cebola	???	203
tomate	???	204
pepino	???	205
abobora	???	206
melancia	???	207
quiabo	???	208
repolho	???	209
beringela	???	210
pimenta	???	211
batata reno	???	212
alho	???	213
cenoura	???	214
inhame	???	215
ervilha	???	216

	<i>Frutas</i>	
banana	<i>cacheo</i>	301
manga	???	302
laranja	???	303
limao	???	304
toranja	???	305
ananas	???	306
abacate	???	307
pera	???	308
vida	???	309
papaia	???	310
peche	???	311
litchi	???	312
maca	???	313
maciera	???	314
tangerina	???	315
guava	???	316
ateira	???	317
	<i>Culturas de Rendimento</i>	
castanha de caju	<i>raw nut</i>	411
	<i>processed nut</i>	412
algodao	<i>caroco</i>	413
cana de acucar	???	414
tabaco	<i>seco</i>	415
coco	<i>unidades</i>	416

PROBLEM TYPE 3: Values for YES/NO Questions. Values were coded in TIA 1996 as:

1 = yes
2 = no

It is recommended that this be changed to

1 = yes

0 = no

because it allows easier computation within a statistical software package of the percent that said yes or no to a particular question. Essentially, the advantage is that such coding facilitates the computation of statistics. For example, running MEANS of a 1/0 Yes/No variable automatically yields the percent yes; otherwise it would be necessary to recode 2 to 0.

c. How will these changes improve TIA?

It is best to have a logical system throughout a data collection system which is readily apparent to enumerators, data entry staff and analysts alike. The above system is one such attempt. The principle to keep in mind is that there should be significant investment on the part of DEST to identify, *ex ante*, all products and forms such that there is much less possibility for ambiguity *ex post*.

4. Space for enumerator to write answers to questions

a. General Problem:

It is necessary for ample space to be provided on survey instruments for enumerators to write answers to each question in a clear manner. It is also helpful to have no printed material on a questionnaire **between** two consecutive characters.

For example, on the TIA 1996 questionnaire, the space for writing crop code was written as:



b. Problem from 1996 and suggestion:

The space allowed on the 1996 instrument for many questions was too small to permit easily legible reading of values recorded by enumerators. It is recommended that space for each answer be expanded from the current format. Further, it is recommended that the “box” approach for recording each digit be abandoned. It is recommended that two approaches be used, depending on the type of question. Where data is at the household level, use a format like:

AF22 _____ Esta familia e originaria de esta aldeia?

1	Sim
0	Nao

Or, if data is collected in a table, simply provide ample space within a rectangle for the characters to be entered.

c. How will these changes improve TIA?

There will be an easier translation of data from data written on the questionnaires to the data entry program. There were a number of data entry errors which resulted from the format used in 1996 which could likely be avoided by using the approach recommended.

5. Format of member-level table

a.General Problem:

It is good methodological practice to have data on questionnaires appear as like as possible to data in an SPSS file. There were a number of cases in TIA 1996 where the tables were transposed in comparison to what might be an optimal approach. We use the member-level file as an example.

b. Problem from 1996 and suggestion:

The member level table was formatted such that each member's information was written from top to bottom on the instrument for a range of questions. However, data for each member was entered such that a particular member's characteristics are horizontal within the SPSS data base. It would be easier to reorganize this part of the questionnaire using the table format on the following page (though particular questions may vary).

It is also recommended that member numbering begin at 1, 2, 3...as was done in 1996. However, it would be useful to eliminate the "0" preceding each such number (through 9) to eliminate the probability of introduction of another error type in data entry.

c. How will these changes improve TIA?

A data analyst will be able to use the resulting electronic file in combination with original documents in an easier manner than under the previous format.

Nome	No.	Relação ao chefe 1 chefe 2 esposa/o 3 filha/o 4 pai/mãe 5 outra fam. 6 outro (esp)	Idade	Sexo 1 m 2 f	Nível de escolaridade (entrar o último ano completado, OU) 0 analfabeto 12 após décima primeira 98 nenhuma escola mas sabe ler e escrever	Esta pessoa está actualmente a frequentar escola? 0 não 1 sim	Estado civil 1 monógama/o 2 polígama/o 3 solteira/o 4 viúva/o 5 divorciada/o 6 esposa de emigrante (esposo fora mais de 6 meses)	Esta pessoa é nativa desta aldeia? 0 não 1 sim	Desde a colheita do ano passado (Junho) até agora, esta pessoa tem feito trabalho fora da machamba da família? 0 não 1 sim
NOME	MEM	MEM1	MEM2	MEM3	MEM4	MEM5	MEM6	MEM7	MEM8
	1								
	2								
	3								
	4								
	5								
	6								
	7								
	8								
	9								
	10								

7. Format of meios table

a. General Problem:

Chris Wolf, in his article “**Computer Analysis of Survey Data — File Organization for Multi-Level Data**”(1990) highlights the fundamental importance of data level and the advantages of organizing all data collected at a particular level within the same data file within the context of a set of data files representing an entire survey data set.

b. Problem from 1996 and suggestion:

Let us take the example of the MEIOs table from the 1996 design. The original questionnaire was designed such that five different files would be developed from data which was essentially at the same level : PROV - DIST - ALD - AF - MEIO

It is recommended that all data at a particular level be organized into one table. See the following example, and compare it to the TIA 1996 questionnaire, and compare the simplicity of the suggested approach.

	Proprio	Emprestado	Alugado	Custo total (meticaís)
MEIO	MEIO1	MEIO2	MEIO3	MEIO4
1 Enxada				
2 Catana...				

To the extent that more variables are required at the meio level, simply add additional columns, labeling them MEIO5, 6, ...

c. How will these changes improve TIA?

This will enable data entry staff, and analysts to clearly and unambiguously define the meaning of a case in each data file, and have it closely tied to the original questionnaire.

8. Recording of quantities of Mozambican currency

a. General Problem:

Currency values reported in a survey instrument should always be recorded in the same units to reduce errors on the part of enumerators and data entry staff.

b. Problem from 1996 and suggestion

Two approaches were used in the 1996 TIA: meticais and contos. Of course, there are 1,000 meticais in a “conto,” a slang expression which derives from the colonial era.

Many enumerator errors occurred such that meticais were incorrectly written as contos and vice versa. Correcting this problem delayed data analysis by a significant amount of time. It is recommended that one consistent approach is used throughout future TIAs, and that this approach be the metical. This is because precision in measurement is important, and a meticais value is more likely to represent accurately the amount of money in question.

c. How will these changes improve TIA?

There will be less data entry errors than in earlier TIAs in terms of prices being off by three orders of magnitude.

9. Order of questionnaire components

a. General Problem:

It is useful from both the perspective of the enumerator and respondent in any survey to have a questionnaire organized into parts, where each part reflects a particular theme.

b. Problem from 1996 and suggestion:

Questions about specific themes were, in some instances, not found in the same section of the instrument. It is suggested that future TIA questionnaires be organized in conceptual themes which help both enumerator and respondent answer questions in as logical a framework as possible. A specific example concerns land. Consider the areas where farmers were asked about their land holdings:

page 3: list each field and the member responsible

page 3: total of other land not in machambas

page 5: list each fruit tree type and area or pomar

page 9: information on the first cultivated field

loose pages: for each additional field and for measuring fields

It is recommended that a method be developed which will ask about all types of land in one section which is exhaustive and mutually exclusive. See below for further comments on this issue.

c. How will these changes improve TIA?

Logical ordering of questionnaire parts will improve overall data quality, allowing the farmer an improved opportunity to accurately reflect his/her reality as queried in the instrument.

SECTION TWO: PRODUCT CONVERSION FACTORS TO OBTAIN DRIED EDIBLE PORTION

a. General Problem:

Farmers report the production, sales and consumption of a variety of products in a range of forms and units. A statistical report must report all production, sales and consumption using a common numeraire — dried, edible portion for human consumption.

b. Problem from 1996 and suggestion:

TIA 1996 permitted any combination of

1. Smallholders responses indicate that 39 different products are cultivated across the country. Fifteen different unit types are used to refer to a series of essentially volumetric measures. Finally, 11 different product forms are used to refer to the 39 products. This presents a very complex combination of conversion factors needed to convert farmer reported information to information useful for analysis.
2. It is recommended that DEST, in collaboration with MAP/MSU FSP personnel, develop a list of all products included in a future TIA, and the set of units and forms which would be acceptable for each.
3. It is recommended that DEST, in collaboration with MAP/MSU FSP conduct measurement at the farm level of each of these products to develop a new and improved set of conversion factors for use in TIA and by other MAP researchers.
4. The following represents a proposal concerning how this process might proceed.
 - a. Develop a list of all products in a given region produced by greater than five percent of the population. Determine the set of logical product-estado combinations that are possible. For example, could manioc be on an espiga? No. Manioc can only be referred to as dried (without the skin) and fresh (with the skin). For this purpose, let us divide the country's ten provinces into three regions: North, Center and South.
 - b. Use standard scientific methods to determine the density and ratio of edible portion to weight for each product and form combination within each region. See the Table in Section One, Point 3. above for a proposed list of each product. Note that in that formulation, "estado" has been eliminated as a variable, another key modification recommended. In other words, eliminate estado by increasing the number of products such that, for example, maize grain and maize on the cob are distinct products.

MAP/MSU FSP has significant experience in developing conversion factors of this type across the country. It is recommended that DEST and FSP collaborate on this activity.

c. It is recommended that the particular methods for developing these key conversion factors be done, to the extent possible, in collaboration with the Instituto Nacional de Investigacao Agraria.

c. How will this improve TIA?

Estimates of production, commercialization and other utilization variables will be more precise than estimates from the 1996 TIA, as we will have greater confidence in the conversion factors required to convert each observation to kilograms of dried, edible portion.¹

**SECTION THREE:
LAND AREA MEASUREMENT, USE AND MUTUAL EXCLUSIVITY**

General Problem:

Farmers have land in annual crop cultivation, tree crop cultivation, fallow and pasture. TIA seeks to collect information on each type of land owned and/or used by sample households. It is important that this data be collected in an exhaustive and mutually exclusive manner, such that total farm size (including land from all categories) can be computed from the data base.

Problem from 1996 and Suggestions:

TIA 1996 asked farmers about land area owned at three different levels; had no adjustments been made to negate this possibility, areas computed in Table 6 would have had significant double-counting of particular parcels.

The three levels at which land area data were collected include:

- a. Household level (questions AF9, AF10H, AF10M, AF11A, AF11B)
- b. Fruit tree level (question AA7)
- c. Machamba level (questions M1131, M1132, M1133C, M1133L)

Let us consider an example of how these three types of questions may have resulted in double counting. Suppose that Alberto is a farmer in Niassa Province who has 3 machambas of 0.5, 0.75 and 0.25 ha, respectively, which he cultivated in 1995/96. Suppose that Alberto has 3 mango trees on the first of these three fields. Further, Alberto has 20 mango trees located 2 kms from his house on a 1 ha piece of land, one-half of which is fallow.

¹ See the table in Section 1, Point 3 and compare it to the list of codes immediately preceding that to consider this.

So, current TIA methods would have recorded for Alberto:

1.5 ha cultivated
1 ha in pomar
0.5 ha fallow
= 3 ha owned

or, would TIA methods have reported

1.5, 0.5 and 0.5=2.5 ha owned?

It is important to recognize that the source of ambiguity derives from the fact that the methods used to elicit information about all land areas were not exhaustive and mutually exclusive. Let us examine TIA 1996 performance on each type of land area. Based on the following discussion, we will develop a strategy for revising TIA methods. This discussion is divided into two parts: 1) Measurement of area cultivated and area allocated to each crop; and 2) Measurement of land in fallow, with tree crops or in pasture.

1. Measurement of area cultivated and area allocated to each crop

The 1996 TIA planned to measure the cultivated fields for three randomly selected households in each village. Note that each household may have zero, one or more fields, according to the TIA definition of household.

Field measurement is a labor-intensive activity, given the long distances which enumerators and interviewees must travel to arrive and measure each field, as well as the irregular shape of many machambas which further complicates the issue.

Did TIA actually measure the number of fields intended? How much did it cost?

For a variety of reasons, TIA actually succeeded in measuring approximately two-thirds of the fields intended to be measured. Note that for nearly all measured fields, farmers were also asked, prior to measurement, to declare the area in each field.

Assuming that measurement was correctly implemented, and given that it is much more costly from DEST's perspective to measure fields rather than simply rely on farmer estimates, we may ask, is it worthwhile to measure fields in a TIA? Is it redundant — that is, can the farmer give the same information as the field estimates at a much lower cost? Or, do measurements improve accuracy.

To gain perspective on this, consider that, nationwide, the Pearson correlation coefficient between area declared and area measured is 0.581. This would suggest a significant amount of error, on average, in declared area.

A more disaggregated analysis of this question, at the provincial level, however, demonstrates wide variability in terms of the ability of farmers to declare their field size accurately across provinces. Table 1 demonstrates that some provinces had a relatively high correlation (e.g., Cabo Delgado at 0.804 and Manica at 0.691) while, at the other extreme, there was no statistical correlation of measured and declared area in Zambezia. It is recommended that DEST note these results in enumerator training and pre-survey activities in areas of particularly poor correlation, and determine the reason for some of the low correlations in particular locations. Understanding the source of past errors may improve data quality in the future.

Table 1. Pearson Correlation Coefficient of Declared Area Cultivated and Measured Area Cultivated, TIA 1996

Province	Correlation Coefficient
Niassa	0.361**
Cabo Delgado	0.804**
Nampula	0.597**
Zambezia	0.085
Tete	0.518**
Manica	0.691**
Sofala	0.426**
Inhambane	0.578**
Gaza	0.595**
Maputo	0.315**
National	0.581**

** Statistically significant at 0.10 level or lower.

It is recommended that DEST investigate possible reasons for the differential performance of this relationship in the three provinces where the correlation was below 0.5.

2. Measurement of land in fallow, with tree crops or in pasture

Land resources are a key determinant of smallholder welfare in Mozambique. An additional hectare of land owned by a household, regardless of its current use, was shown to be associated with a ten percent increase in per capita income across each of the models from the 1994/96 MAP/MSU FSP Nampula/Cabo Delgado Smallholder Survey. Thus, knowledge about land is important...

TIA 1996 asked smallholders about the following types of land over which they hold tenure:

- a) **land cultivated** this year with one or more crops whose harvest was anticipated during the current agricultural season;
- b) **fallowed land** defined as a land area where the household either has already cultivated or plans to cultivate some crop, annual or perennial, but has chosen during this campanha to allow the land to produce no crop of value;
- c) **land in tree crops** defined as that land whose principal economic and/or social purpose is to have trees which either currently do, once did, and/or are anticipated to produce some product of value;
- d) **land in pasture** defined as land principally dedicated to providing nutrients for livestock for some commercial, consumption (or food security) purpose
- e) **other** defined as land

It is recommended that future design of data collection ensure mutual exclusivity and exhaustiveness in identifying area in different uses held by each interviewed household. It is recommended that DEST and MAP/MSU FSP personnel meet to discuss methods used by MAP/MSU FSP in the 1994/96 Nampula/Cabo Delgado Socio-Economic Study designed to achieve these objectives.

The following is table represents one method that could be used to gather key information about field size and area planted to each crop, as well as gathering key analogous information on fruit trees and fruit tree area.

M1 Area total _____

M2 Area semeada _____ [Area em pousio = Area total menos Area semeada]

M3 Colheu da toda area semeada _____ 1=sim 0=nao

M4 Favor indicar quaisquer novidades sobre esta machamba que afectaram a **produção** _____

Cultura	Area desta cultura este ano que semeou	Consociação ou monocultura 1=Monocultura 2=Consociação 3=A volta da cultura mais importante
C1	C2	C3
1a cultura		
2a cultura		
3a cultura		

	Arvores Total	Arvores Produtivos	Arvores Novos	Area destas Arvores
Tipo de arvore	F1	F2	F3	F4
Cajueiro				
Coqueiro				
Mangueira				

SECTION FOUR: LEVEL OF SALES DATA

General Problem:

Implementation of TIA occurs, by necessity, once in a given agricultural year. Farmers may sell production at any point from time of harvest until production is depleted. TIA seeks to gain insights into smallholder commercialization of own production by asking farmers about their sales in two parts: sales already made, and planned sales. **If gaining insight into farmer marketing behavior is a priority of MAP, it is recommended that the date of interview be as far after harvest as is feasible.** This will enable the data collected to represent as much of commercialized production as is possible.

Problems from 1996 and Suggestions:

It is desirable to have data on marketing behavior by season, given the fluctuations in demand and prices received by farmers. Given that the TIA 1996 was implemented in September / October 1996, it is clear that farmers could only provide a part of their marketing behavior information. This is because such farmers only provided definitive answers about what they had done since harvest (April/May/June) until the present.

A way of improving on this design would be to collect marketing data at the transaction level. If Joao sold 50 kgs of maize overall in the first 5-6 months of marketing, we would like to know if these 50 kgs were sold simultaneously or in several smaller sales. We would also like to know the price received.

It is suggested that sales data be collected at the transaction level. For each transaction, price per kg and / or preco total should be gathered, as well as month of transaction.

It is suggested that sales planned with the current year's harvest be maintained as a question in a similar fashion to 1996 TIA. This question should be at the product level, rather than transaction level, given that its purpose is simply to obtain an estimate of planned behavior.

How will this improve TIA?

TIA will be better able to inform policy-makers about who sells agricultural products, value received, timing of sales and quantities sold by adopting a more disaggregated approach. Simultaneously, this approach will allow farmers to explain each transaction to the enumerator, and if there were many, may allow him/her to remember more clearly the details of each transaction.

SECTION FIVE: ADDING SMALLHOLDER FOOD PURCHASES

From a food security policy perspective, one of the most important pieces of information about smallholders is whether they are net buyers or net sellers of food products. This requires data on food sales and food purchases. TIA 1996 collected sales, but did not collect purchases. This is an important omission.

It is suggested that an approach be developed for this purpose which could gather the best possible data at the appropriate level while not unduly complicating or lengthening the data collection process. It is suggested that DEST and MAP/MSU FSP staff collaborate on this issue.

SECTION SIX: INTERNAL CONSISTENCY CHECKS

A majority of the time and human resources invested in data cleaning following data entry experienced in the TIA 1996 survey can be avoided. Avoiding such labor-intensive activities should be a high priority for DEST. What are the key steps necessary for this to occur? The following represents a non-exhaustive list of ideas in this regard:

- 1) Improve enumerator training and ensure that each sees his/her work as a study of a farm-family, where the parts of the survey must reflect a whole picture of the household, such that responses must be internally consistent. It is recommended that a higher educational level be required for enumerators than has been required in the past, and that training and testing of enumerators be made more rigorous.
- 2) Questionnaires should be checked in the village on the date of interview by a well-trained supervisor. Where errors are identified, the enumerator should return to the interviewee for clarification.
- 3) Ranges and Rules in data entry design. It is recommended that full advantage be made of the ranges and rules features of the data entry package used for future TIAs. In this way, implausible values for particular variables, if entered, will be rejected, and the data entry staff will be given the opportunity to correct such errors.

SECTION SEVEN: AGRICULTURAL INPUTS

1. TIA 1996 collected data on fertilizer, manure, pesticides and seed in a manner which may not have yielded an accurate depiction of use of these inputs across the country. To remedy this, the TIA questionnaire should be redesigned such that questions are asked at the household:input level. This is primarily due to poor questionnaire design in this area. **Consider an alternate design presented below.**
2. It is commonly assumed that smallholder use of modern inputs is low. To the extent that this represents the conventional wisdom of TIA staff and enumerators, it is likely that little care will be used to probe these questions with each household. It is important that, as part of training and reciclagem, enumerators be told not to assume any particular response as expected concerning this issue from smallholders; rather, the enumerator should ask the question in as objective a manner as possible.

Insumo	Fez? 0 nao 1 sim	Quem decidiu sobre quando aplicar? 1 Enquadrador 2 Campones sozinho 3 Vizinho sugeriu 4 Outro (esp.)	Como aplicou? 1 Ulva da companhia 2 Ulva propria 3 Outro (esp.)	Como obteve? 1 Companhia 2 Lojista 3 Vizinho 4 Guardou 5 SEMOC 6 Outro	Quanto aplicou?	
					Qt	Unidade
INSUMO1	INSUMO2	INSUMO3	INSUMO4	INSUMO6	INSUMO7	INSUMO8
1 Semente						
2 Insecticida 1a aplicacao						
3 Herbicida						
4 Fertilizante						

**SECTION EIGHT:
AGRICULTURAL LABOR SALES AND NON-FAMILY LABOR HIRED**

General problem:

Labor sales by smallholders in Mozambique represent between 10 and 20 percent of household income, on average, according to Rui Benfica's M. Sc. Thesis. It is important that TIA gather information on off-farm labor sales.

Problems from 1996 and Suggestions:

TIA 96 gathered data at a rather ambiguous level (See the original questionnaire). An alternative is presented on the following page.

The suggested format is at the level of MEMBER-ACTIVITY. This allows a household to report, for each member who worked off-farm, the time spent and value earned for each type of work. Note that in the 1994/96 Nampula / Cabo Delgado Smallholder Survey, it was common to find individuals who worked in different types of activities off-farm during different seasons. For example, an individual may work from September to November on a neighbor's field and earn 100,000 Mts per month. Then, he may work at a sugar processing plant from March to June and earn 120,000 Mts per month. It is important to design a questionnaire that captures these various activities for each household member.

TRABALHO FORA DAS MACHAMBAS PROPRIAS

INSTRUÇÕES

Pergunta o entrevistado sobre *cada tipo* de trabalho alistado na tabela.

- Gostariamos de fazer algumas perguntas sobre *cada tipo de trabalho* que cada pessoa fez ...

Nome	No.	Tipo de trabalho <u>AGRICOLA</u> 1 Machamba da companhia 2 Machamba dum privado 3 Machamba dum vizinho <u>NAO AGRICOLA</u> 4 Fabrica da companhia 5 Função pública 6 Professor 7 Outro	Cultura (se for trabalho agricola, qual cultura)	Em que <u>mes</u> comecou este trabalho?	Por quanto tempo em total fez este trabalho?		Onde se fez o trabalho? (Distrito)	Pagamento			
					Quant	Unidade 1 dias 2 semanas 3 meses		Total de Meticais	Se recebeu géneros ...		
									Qt	Unidade	Tipo de Genero
NOME	MEM	TF1	TF2	TF3	TF4	TF5	TF6	TF7	TF8	TF9	TF10

SECTION NINE: MICRO AND SMALL ENTERPRISES (MSEs)

General Problem:

In the rural Mozambican context, off-farm income can be classified into two major types: labor sales and small scale enterprises. It is important that TIA collect information on this type of income, estimated by Benfica (1998) to account for between 6 and 21 percent of net household income in four Central and Northern Provinces.

Problems from 1996 and Suggestions:

A large section on MSEs was incorporated into the design of TIA 1996 in selected provinces. For other provinces, such data were not collected. To the extent that analysts choose to compute net household income for areas where this activity was not undertaken, these estimates will be downwardly biased. It is suggested that data be collected from on MSE activities in future TIA surveys.

A suggested format for collecting such information, at the HOUSEHOLD-ACTIVITY level is suggested on the following page.

A frequent problem with data collection related to MSEs is a confusion between gross and net income amounts. It is of crucial importance to the design of this part of a future TIA instrument that this issue be considered, and that training of enumerators focus on this. For example, a farmer may report that she sold capulanas in the village and had income of 1,000,000 Mts. Upon further questioning, this farmer would report that the cost of the capulanas, including transport to the market of purchase, totalled 700,000 Mts. The information that TIA should seek to collect is simply the net profit, in this case, 300,000 Mts.

Actividade	Quem é o principal responsável para o trabalho?	Pagamento			
		Total de Meticais	Se recebeu géneros ...		
			Quantidade	Unidade	Tipo de Genero
ACT	CP1	CP3	CP4	CP5	CP6
1=Compra/venda de grãos básicos					
2=Compra/venda de outros productos					
3=Compra/venda de roupa					
4=Artesanato (esp) _____					
5=Produzir bebida de _____					
6=Carpintaria (esp) _____					
7=Transporte					
8=Venda de comida preparada					
9=Outra actividade _____					

SECTION TEN: SAMPLING

General Problem:

TIA is conceptualized as a national survey of smallholder agriculture. It is important for TIA to be representative of smallholders, and thus broad coverage is necessary. Determination of how many households, villages, districts and agro-ecological zones should be covered are important questions if TIA is to provide a good picture of agriculture on a national level.

Problems from 1996 and Suggestions:

The documents which describe sampling procedures for TIA 1996 are, in a certain sense, internally inconsistent. That is, it was stated that sampled districts within each province were chosen “population proportionate to size.” However, given the existence of multiple agro-ecological zones within each province, an effort was made to include all such province-zone combinations. The number of such combinations in the country was stated as 28. From a purely statistical perspective, these procedures appear to be at odds with each other.

Key Questions:

At what level does DEST intend to be representative with a reasonable confidence interval?

The following represent a few ideas which should be considered in the design of future sampling activities.

1. If the survey is to be representative of the country, and if the key cluster group is to be **district**, DEST must be prepared to accept any combination of districts which result. For example, let us suppose that the next TIA selects 60 districts, with each district being selected population proportionate to size. Let us further assume that “replacement” is allowed, such that each district may be selected multiple times, with the probability of this occurrence being related to its relative population.
2. Let us further assume that, in selection of districts, **no districts from a province with a small population (e.g., Niassa)** are selected. This is a real possibility. From a statistical perspective, such sampling would be nationally representative. However, it is critical that DEST and MAP policy makers consider that the disadvantages of such an approach may outweigh the advantages. . That is, the entire survey would be conducted, yet we would not have any information on an entire province (e.g., Niassa).
3. An intermediate solution, and perhaps what was actually done in 1996, would be to **essentially do 10 different sampling procedures (one for each province)**. In this way, we could ensure that each province would be surveyed, and that information about agriculture would be perceived of as being statistically representative at the province level.

4. Let us assume that essentially ten samples are drawn, one for each province. In the past there has been an effort to have the number of sampled households in each province be proportionate to the relative populations of each province. But, from a statistical purpose, is this necessary? For example, if there are 200,000 households in province 1 and 300,000 households in province 2, from a statistical perspective, the sample sizes can be practically the same to yield an equivalent confidence interval. This may not be true for clustered sampling (choosing districts, villages and then households). It is recommended that these issues be considered thoroughly, and that the advice of a statistical sampling expert be brought to bear on the issue.

**SECTION ELEVEN:
DATA ENTRY PROCEDURES**

Ranges and rules...