Biometric review of fish loss assessment methods in West Africa.

1. Introduction

This review will consider the PRA, questionnaire survey and load tracking approaches used in Cote d'Ivoire, Ghana and Nigeria. Detailed spreadsheet information is available for Cote d'Ivoire and this forms the basis for a review of the quantitative loss estimates. More qualitative reviews of the Ghanaian and Nigerian studies are also undertaken.

Based on these examples some recommendations are made for the analysis of fish loss data, which may be incorporated into this or future projects. This review assumes that rigid definitions for loss have been agreed as in [1].

2. General comments on Cote d'Ivoire survey data

The survey data consists of three spreadsheets for the fisherman (sample size = 84), the processors (sample size = 114) and the fishmongers (sample size = 80). The data is in the form of a 'flat-file' in an excel spreadsheet. There are a number of points that should be considered for data entry of this kind.

- Column headings should be very clear and relate directly to the question numbers of the questionnaire. Using suitable coding if the column headings are too long, but include the key in a separate sheet in the same file.

- Be very careful to differentiate between missing values and zeros.

- Use one spreadsheet for the entries obtained directly from the questionnaire. Use a separate spreadsheet within the same file to include derived columns, such as price.

- Use the prescribed loss formulas for derived columns.

- Consider entering the data twice and then comparing the spreadsheets to look for discrepancies, this will improve data integrity.

- Use the excel graphical facilities to produce scatter plots of columns, which will give an indication of possible outliers.

- For international compatibility use the decimal notation 0.126 not 0,126 and don't use commas to represent 1000 separators.

- Create an audit trial for the data, i.e. make a master copy and record who has been sent copies. Any edits to the master copy should be strictly recorded and should be undertaken by a nominated person.

- Produce an archive copy of each spreadsheet accompanied by adequate documentation. Somebody unconnected with the project should be able to examine the archived spreadsheet and understand how losses were calculated.

2.1 Average weight and price calculations

The calculation of the average of a column of numbers is generally straightforward, but the number of missing values must be considered. Consider the excel column labelled A shown in Figure 3, where records 2, 8 and 12 are missing.

A
12.3
13.5
17.8
12.7
21.4
20.7
19.3
12.8
15.6
Sum(a1:a12)/12 = 12.175

Figure 3 Incorrect mean calculation

The denominator in the mean calculation is incorrect, as there are only 9 values. The correct formula should be **Sum(a1:a12)/9**, which gives a value of 16.233.

2.2 Average price calculations

The method used to calculate the average price per kg should be consistently applied throughout the analysis. At present two definitions are used:

$$\text{average price per kg} = \frac{\text{Total price}}{\text{Total weight}} \qquad \text{definition 1}$$

$$\text{average price per kg} = \frac{1}{\text{number of subjects}} \sum_{i=1}^{\text{number of subjects}} \text{price per kg for subject i} \qquad \text{definition 2}$$

Consider the three prices and weights shown in Figure 4.

Weight in kg	Price of load	Price per kg
10	5000	500
1	300	300
4	2400	600

Figure 4 Examples of price and weight figures for three subjects.

Using definition 1 the average price per kg is 513.3 and for definition 2 the average price per kg is 466.7. There is clearly a large difference and it is both important to be consistent and to pick the correct averaging method.

Definition 2 can be shown to be a weighted form of definition 1, where the weights for each subject are calculated as $\frac{15}{3*\text{nunber of weights}}$ and:

$$\text{Average price per kg definition 1} = \frac{5000+300+2400}{15} = \frac{5000}{15}+\frac{300}{15}+\frac{2400}{15}$$

$$\text{Average price per kg definition 2} = \left(\frac{1}{2}\right)*\frac{5000}{15}+(5)*\frac{300}{15}+\left(\frac{5}{4}\right)*\frac{2400}{15}.$$

Definition 2 clearly gives higher weight to those subjects with the fewest observations and for this reason definition 1 should be preferred.

The actual formula implemented in the spreadsheet for definition 1 is

$\text{average price per kg} = \frac{\text{average price}}{\text{average weight}}$. In the majority of cases this will give the correct answer, but to protect against problems with missing data the formula using totals as given in definition 1 should be used.

When using definition 1 missing values must be accounted for otherwise the average price will not be representative. Consider the amount of fish and its price in the two columns of Figure 5, where blank cells represent missing values.

Weight (kg)	Cost (CFA)
25	7500
30	12000
15	
32	11200
	10000

Figure 5 Example of price and weight columns with missing values.

A naïve application of definition 1 would be

$\text{average price per kg} = \frac{\text{Total price}}{\text{Total weight}} = \frac{40700}{102} = 399\,\text{CFA/kg}$. Although very easy to calculate in excel this has not taken into account of the missing values. Any rows with missing values should be dropped from the calculation and the average price calculation gives, $\text{average price per kg} = \frac{30700}{87} = 353\,\text{CFA/kg}$.

2.3 Accuracy of estimation

Much of the information from the survey comes in the form of point estimates for loss, weight etc. Without some qualification of their accuracy these estimates are not useful. Consider a histogram in Figure 6 for the weight (kg) caught by each fisherman.

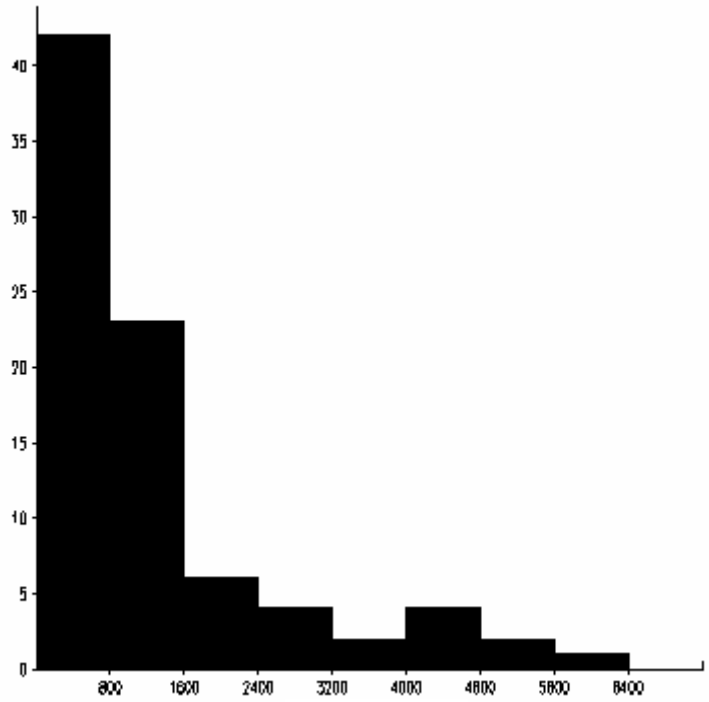

Figure 6 Histogram of weight (kg) of fish landed

The distribution is highly skewed and a conventional 95% confidence interval based on normality assumptions would not be valid. Further work would be required to establish a meaningful non-central confidence interval. It is not even clear whether the mean is a good summary statistic for the catch weight. Often the median is a better summary statistic for this type of data and the median is 800kg, compared to a mean of 1275kg.

A useful statistic for quantifying the variability is the coefficient of variation (c.v.), which is defined as (standard deviation of the data)/(the mean of the data). For the data in each of the surveys reviewed here the c.v.'s are very large and whatever definitions are used confidence intervals for the point estimates are likely to be wide.

The distribution for such quantities as the weight of poor quality fish landed by the fishermen are even more difficult to summarise as they are highly influenced by the large amount of zeros.

2.4 Weight measurement

The survey is highly dependent on the accurate determination of weight and since this is so important a detailed description of the techniques should be given.

3. Fisherman survey in Cote d'Ivoire

This section addresses the estimates obtained from the fisherman survey. Note that the fishermen number 11 has a negative weight for the weight of good quality fish landed (-198kg). This is due to the derived method used for calculating the good quality column. There is also a problem with the price calculation for the poor quality fish (see section 3.2).

3.1 Categorical variables within the Fisherman survey

The questionnaire is very heavily dependent on converting local measurement units into kg and this could be a major source of error.

The database contains two columns of qualitative variables, type of engine and place. There is considerable imbalance within the levels of these qualitative factors as shown in Table 7.

Type of engine	Nos. in category	Place	Nos. in category
1	24	1	27
2	55	2	40
3	5	3	10
		4	5
		5	0
		6	2

Table 7 Distribution within category

If these are important stratification variables then the number of samples are too low in some levels. The cross-tabulation of engine/place counts is shown in Table 8.

place	1	2	3	4	6
Engine 1	1	13	4	1	2
Engine 2	23	21	4	3	*
Engine 3	2	3	*	*	*

Table 8 Cross-tabulation of counts

Comparisons between places or engines will be difficult to interpret due to confounding effects.

There is also an uneven distribution of times for which the gear is in the water (1 @ 14 hours, 81 @ 24 hours and 2 @ 48 hours).

There is no information in the spreadsheet about which suburb the fisherman work from. The load tracking information appears to be stratified in this manner, making it difficult to perform any comparisons.

If one of the objectives of the study were to examine the effect of place, engine or any other categorical variable then it would be advisable to stratify the population before doing random sampling. This would help to reduce the imbalance between different categorical variables.

3.2 Estimates from the fisherman survey

The calculation of the value of poor quality fish is vital for the calculation of loss percentages, but one of the values for poor quality value is suspicious. Table 9 gives the fisherman number, the weight in kg, the price of the load and the price per kg, for those fisherman who have poor quality fish.

Fisherman	Weight (kg)	Price (CFA)	Cost per kg
11	4422	670000	151.5
12	2640	33000	12.5
18	330	225000	681.8
21	66	5000	75.8
26	154	16000	103.9
40	12	0	*
47	726	3500	4.8

* it is not clear whether the cost should actually be zero or missing

Table 9 Poor quality costs and weights

The shaded price is higher than any of those prices received for good quality fish, which seems unreasonable. Dropping this record (and the record for fisherman 40) from the spreadsheet, the mean calculation then gives an average cost of poor quality fish as 91 CFA/kg. This is considerably different to the value of 114 quoted in the report and will of course change the loss calculations, which have already been demonstrated as incorrect for another reason (wrong loss formula [3]).

The effective sample size for loss estimation in poor quality fish is very small, consequently the accuracy of any estimates must be questionable (see c.v. in Table 10).

Table 10 gives the recalculated estimates for the values in Table 2 of [2], and where relevant the coefficient of variation. Estimates that show a large difference from the original estimates are shaded.

Quantity	Estimate	c.v.
Average quantity of fish caught by canoe	1275 kg	103%
Average quantity lost at sea	8 kg	162%
Average quantity lost during landing	3 kg	150%
Average quantity of good quality fish landed	1109 kg	102%
Average quantity of poor quality fish landed	99 kg	567%
Average quantity used for subsistence	72 kg	138%
Average price per kg of good quality fish	154 CFA/kg	
Average price per kg of poor quality fish	91 CFA/kg	
Total quality loss in CFA	1694 + 6237 = 7931 CFA	
% quality loss	4%	

Table 10 Estimation of fishermen losses.

Note it is not entirely clear when the subsistence sample was taken. The average quantity of fish caught does not equal the sum of the averages of good quality + poor quality + subsistence since the columns have different missing values. One method for removing this inconsistency would be to remove all records with any missing values.

4. Processor survey in Cote d'Ivoire

Again the survey is dependent on converting the local measurements into weights. The column headings could be coded more clearly. The column AH is a derived column for the weight of good quality fish after smoking, the formula has a large

number of elements and should be explained particularly since it can give negative weights (see processor 81). Processor 13 has a zero for price, presumably since there is a 2550 kg load of fish, this is actually missing and not zero.

4.1 Estimates from the processor survey.

The price per kg estimates have been calculated using both definitions 1 and 2 and this will clearly give inconsistencies. The recalculated prices given in Table 11 have been calculated using definition 1 and shaded values represent discrepancies with the results of Table 3 from [2]

Quantity	Estimate	c.v.
Average quantity of fish processed	339 kg	115%
Average price per kg at purchase	207 CFA/kg	
Average quantity declared poor quality	22 kg	186%
Average quantity of fish lost	0.15 kg	182%
Average quantity of fresh fish lost at processing	0.23kg + 0.76kg	508% / 342%
Average quantity of smoked fish lost at processing	0.4 kg	300%
Average quantity of smoked fish lost at selling	0.8kg	718%
Average weight of good quality fish after smoking	125 kg	derived
Average weight of poor quality fish after smoking	10 kg	150%
Average selling price of good quality	653 CFA/kg	
Average selling price of poor quality	332 CFA/kg	
Financial loss for fresh fish loss	339 CFA	
Financial loss for smoked fish loss	843 CFA	
Total quality loss in CFA	??	

Table 11 Estimation of processor losses

The definition of quality loss needs to be refined before an estimate can be calculated.

A processor quality loss in fresh fish has been estimated in terms of kg, but there are no corresponding price estimates. It is therefore not possible to calculate the CFA loss for this quality loss.

It is also not clear whether the quality loss in fresh fish causes the quality loss in smoked fish. If this were the case then adding both fresh and smoked quality losses would be equivalent to counting the same losses twice.

5. Trader survey in Cote d'Ivoire

The data integrity for this survey is better than the previous surveys and the averages have been calculated with more care, regarding the correct treatment of missing values and zeros. However all the price calculations have been performed using definition 2.

5.1 *Estimates from the trader survey.*

Table 12 gives the recalculation for the estimates in Table 4 of [2], using definition 1 for the mean price estimates, where possible. Shaded values again highlight those estimates, which show large differences.

Quantity	Estimate	c.v.
Average quantity of fish purchased	344 kg	157%
Average price per kg of good quality	656 CFA/kg	
Average price per kg of poor quality	609 CFA/kg	
Average price per kg of mixed quality	600 CFA/kg	
Average quantity declared good quality	322 kg	171%
Average quantity declared poor quality	22 kg	derived
Selling price of good quality*	741 CFA/kg	
Selling price of poor quality*	606 CFA/kg	
Quality losses due to storage	13 kg	330%
Total losses	?	

* note definition 1 could not be used to calculate these averages as the price per kg was entered directly.

Table 12 Estimation of trader losses.

Again the quality losses are difficult to quantify. For example, quality losses due to storage + quality loss at purchase = 13 + 22 = 35kg, but to justify this addition the losses must be independent.

6. Summary for Cote d'Ivoire data

The main points raised by this survey are:

- Importance of data validation and integrity
- Accuracy of weight measurement
- Rigid definitions of loss should be followed
- Refine definition of quality loss
- Confidence intervals for estimators need to be derived
- Transport losses have not been assessed

The variation in a lot of the estimates is extremely high and it is important to know whether this is actual variation within the various stages or an artefact of the data collection methods used for the survey. If it is the former then it is implying that questionnaire surveys are never going to get accurate estimates due to the inherent variability. It might be possible to control some of this variation by using stratification and using at least 20 subjects in each level. This may be prohibitive in terms of resources and time.

The integrity of all three surveys could be improved if the inconsistencies between stages in the form of fish prices could be explained. For example the fishermen sell the fish at an average of 154 CFA/kg, but at the next stage the processors buy it at an average of 207CFA/kg.

Although it might not be very accurate the survey has shown that the largest losses occur due to quality losses, so it is important to quantify this loss in a meaningful and

accurate manner. It is not clear, whether the survey has asked the correct questions to achieve this and perhaps this is something that load tracking can do.

7. Load tracking for Cote d'Ivoire

Without electronic versions of the load tracking information and a far more detailed account it is not possible to discuss the accuracy of the figures. This section will concentrate on a simple qualitative comparison of the losses. Tables 13 and 14 give the economic losses for the questionnaire and the load tracking.

	Supply	Processing	Transport and trade
Load tracking	1.2	0.7	11
Questionnaire	7	10.5	8.6

Table 13 Summary for 15 June to 15 October

Other than for transport and trade there is little agreement between the two sets of figures.

	Supply	Processing	Transport and trade
Load tracking	13	8	12
Questionnaire	5	13	8

Table 14 Summary for 20 October 1997 to 20 February 1998

Again the results for transport and trade have the closet agreement, but it is clear there is a large seasonal effect and averaging over season is not likely to be representative.

8. PRA for Cote d'Ivoire

The use of PRA results for the estimation of loss is unlikely to be accurate, as the sampling cannot be assumed to be random. The inference from any analysis of the PRA data would be very weak. However PRA is useful for the identification of the most important areas of loss and the refinement of questionnaire and load tracking surveys.

Some simple qualitative comparison between the PRA and the survey techniques is possible, but is unlikely any statistical comparison is justified.

9. Nigeria questionnaire

The description of the questionnaire at Lake Chad fisheries is good and seems to follow all recommended guidelines. There is no data to analyse, but there are some summary tables of loss data. However is not clear how they were calculated, how they should be interpreted or compared to the Cote d'Ivoire data for example. Despite the good description of the survey there are a number of problems.

Fisherman survey

- Tilapia and clarias have been mixed
- Layout of question 8 and 9 is poor
- Weight and daily price have not been included – hence only information is in fish numbers.

Processor survey

- Poor survey layout
- Very little qualitative information

Transport survey

- Again poor layout
- Poorly filled out, with little information about unit

Marketing losses

- Poor layout
- Little or no quantitative information about units

The questionnaires are so poor that it is difficult to justify entering the data and attempting an analysis.

10. Ghanaian survey

The two surveys are short with 10 and 14 questions and rely on recording catches and losses in terms of units. This is acceptable if the units used by the respondents have the same volume. Although depending on the size of the units this type of recording is not going to give an accurate assessment of loss.

Fisherfolk survey:

The losses assessed are fish thrown before landing and those lost after landing. Two quality categories are considered and the amount of fish used for home consumption is recorded. Not sure if question 4 is relevant or easy to quantify. Losses were so small that only one basket was of low quality and the estimates of loss are effectively based on 1 sample. The small losses were a result of the season when the experiment was performed, but this type of survey is not going to estimate small losses accurately.

The sample size was small (24) due to over 50% of the fisherfolk not having gone fishing within the last 14 days.

Fish processors survey:

Again a short questionnaire, but the questions don't seem to be in a logical order. Only a very small number of baskets (3) showed deterioration between landing and smoking. It is not clear where exactly this loss occurred along the chain from fisherfolk to processors – again definition of loss needs to be clarified.

After smoking the 575 baskets of good quality fish a further 7 were then regarded as being low quality smoked fish. Altogether11 were finally classified as low quality, with the extra 4 being attributed to losses that occurred during 'packaging, loading, transportation and unloading at market'. These losses have then been counted as processor economic losses, which is not convincing. Also it is not indicated what happened to the 3 baskets that were already classified as poor when fresh.

Load tracking:

Rounding is a problem here – all scores have been rounded to a mean of 2, this is deceptive and does not mean that everything was good quality. At least one decimal place should have been used.

In this form load tracking has not added much information to the analysis. It is not clear where the value of 6.2 comes from, in the traders' quality loss cell of the table on page 25 in [4].

As described in the discussion at the end of the document [4] the very low losses at the time of the survey limit its effectiveness. Even if there were more losses it could benefit from a more accurate weight/volume determination. The load tracking in its current implementation is not going to add much quantitative information to the analysis.

11. Further work

The Cote d'Ivoire survey work demonstrates that the data analysis for this type of survey is quite involved and requires the application of a number of formulas. Missing values and zeros also complicate the problem and even with a good manual giving explicit instructions it might be difficult to get consistent and accurate results.

One solution to this is to use Access both for data entry and loss calculation. The advantages of this are:

- Data entry screens will improve data validation
- Complicated formulas can be easily applied
- Confidence intervals can be immediately calculated
- Queries that can be used to provide further useful cross-tabulations

Although this will give a more sophisticated implementation of the survey and load tracking techniques, it can be improved by using the Internet. Access 2000 (the latest version of Access) can be used to write databases that can be accessed via a web page. The NRI server would contain master copies of the Access forms, queries and functions, but these would be available to anybody who could access the web page. This is an ideal implementation for an international project and all data would be immediately saved in the master database on the NRI server. Any changes or upgrades could immediately be disseminated to all international users. A bespoke Java program could also be written to perform the role of the Access database. This might simplify software licensing problems and reduce the amount of space required on the server.

A disadvantage is that a programming task of this kind, will inevitably take some time to develop. Data from additional surveys would be required to try and determine whether the accuracy of the questionnaire justifies this type of investment.

Theoretical work should also continue on trying to determine the most appropriate accuracy indicators.

It might transpire that all three methods are required with a gradual increase in accuracy from PRA, questionnaire survey through to load tracking. Load tracking could be used to measure losses that are difficult to obtain estimate using the questionnaire approach. The questionnaire divides the stages into a number of discrete actions, which might not be a realistic model of the physical situation. Load tracking has the advantage that it does not have to adhere to these 'artificial' stages.

There does seem to be some scope for the publication of papers. Presuming that the accuracy is not so poor to render the estimates meaningless the results and inferences from well managed surveys could be publishable. The use of load tracking to compliment/extend the traditional questionnaire approach would also be interesting. Also any work involving the posting of a database on a web site would be highly publishable. This is also the sort of thing that would make an ideal third year/MSc project. (i.e. cheap labour!)

The use of databases over the Web seems to be a growing area and ISNAR have given some funding to the Zambian ministry of agriculture to produce a web page from which their national agricultural archives can be accessed.

12. References

[1] Post- Harvest fish losses FAO report – Frans Teutscher

[2] Tools and methodology developed for the evaluation of post harvest losses in the artisanal fisheries of West Africa - Paul Anoh

[3] Review of country reports and work – Ansen Ward

[4] Field testing of methodologies for evaluating post harvest losses of artisanal fisheries in West Africa – a case study of Kormantse fishing community, Central Region , Ghana – Nicholas Ntiamoah.