Assessing economic and welfare values of fish in the Lower Mekong Basin

Project funded by ACIAR

Welfare Component

ROLE AND VALUE OF FISH IN THE WELFARE OF RURAL COMMUNITIES IN CAMBODIA

DEVELOPMENT OF THE SURVEY METHODOLOGY AND PROTOCOL DESCRIPTION

Joshua NASIELSKI[1]

Rudolph WITT[2]

Gareth JOHNSTONE[1], **Olivier JOFFRE**[1], **Eric BARAN**[1]

[1] WorldFish Center, Phnom Penh, Cambodia.

[2] Center for World Food Studies, Amsterdam, The Netherlands

July 2012

Abstract

Freshwater capture fisheries in the Lower Mekong Basin are an important source of food, income, jobs and livelihood opportunities for Cambodians (e.g. 2 million people in Cambodia alone). However there has never been a solid estimate of the total economic value of inland fisheries. As a consequence the importance of these fisheries remains poorly recognized by institutions and governments and in development plans, which hampers rural development. Furthermore the role of fish resources in promoting household welfare, as well as its place in the livelihood strategies of Cambodian households, has never been quantified.

The welfare valuation component is part of the "Valuation of Fisheries in Cambodia" project, funded by the Australian Center for International Agricultural Research (ACIAR). By assessing welfare values of fishery resources, this project hopes to increase the prominence of fishery resources in broader agriculture and rural development strategies and programs within Cambodia. The welfare valuation component is made up of two integrated research methods: a large fish-focused household welfare survey along with integrated focus group discussions.

This report details the development of the household welfare survey questionnaire. It provides an in-depth treatment of the rationale behind the questionnaire design, sources and materials used in its creation and modifications made after field testing and expert input.

Keywords:
Cambodia; welfare; socioeconomics; household survey

Table of contents

1. INTRODUCTION

1.1 PURPOSE OF THE WELFARE VALUATION COMPONENT

The WorldFish Center supporter by ACIAR launched a project in 2012 called "Valuation of Fisheries in Cambodia".

The overall objective of the project is to quantify the multiple values of fish resources and convey information to national decision-makers and development agencies for sustainable and improved rural livelihoods. The objectives of the project are to:

i. assess the economic value of capture fisheries in Cambodia;
ii. assess the welfare value of fish for rural populations in Cambodia and identify strategies that maximize this value;
iii. establish a coordinated monitoring of fish resources through a network of universities;
iv. improve national statistics about fisheries resources;
v. inform a large range of stakeholders about the actual role of fisheries in national economy and livelihoods.

The Welfare Valuation Component is designed to address objective ii. Our approach integrates two research methods: a household welfare survey and village-level focus group discussions.

1.2 PURPOSE OF THIS REPORT

This report describes the methodology and protocol of the household welfare survey and makes explicit for readers the rationale behind its development. Notably, it explains the major methodological flaws of previous fish-focused welfare studies done in Cambodia and how understanding these flaws allowed the welfare valuation team to move forward with an understanding of what to avoid. It follows the development of the baseline welfare survey protocol from the selection of a population of interest up to the selection of villages (the primary sampling unit). This report will be of use for readers of future reports derived from the "Valuation of Fisheries in Cambodia" project who are looking for a detailed explanation of the survey protocol. It will also be of use to researchers who plan on developing a fish-focused household welfare survey of their own.

2 REVIEW OF SURVEY METHODS IN FISH-FOCUSED WELFARE STUDIES

2.1 INTRODUCTION

It is not surprising that so much data exists on the subject of Cambodia's inland fisheries, given the large amount of people who rely on these fisheries both directly and indirectly for food security, income and other measures of household welfare. A wide variety of household surveys, welfare assessments and socio-economic studies have been conducted with the aim of quantifying, in some way, this dependency. These studies have given us the only quantitative information we currently have about the importance of Cambodia's inland fisheries. They have provided the evidence needed to create and reinforce persuasive arguments for the benefits of freshwater fisheries to Cambodian household welfare.

What remains surprising is the lack of attention these studies, for the most part, have paid to the way in which they have collected their data. That is to say, the little care taken in developing a sound study methodology is surprising. Surveys, more so than other types of studies require sound methods in order to be credible and statistically valid. If the methodology is flawed, or worse, simply not discussed in publication, then they will largely fail to persuade sophisticated readers of the validity of the results, regardless of what they are. Without a published discussion of a study's methodology, we are forced to trust the biases and discretion of the study authors, people who are typically strangers to us and with agendas of their own. Reading previous fish-focused welfare studies done in the Mekong basin (including Cambodia), it becomes apparent that most methodologies are unclear, grossly oversimplified, statistically invalid or simply absent from written reports.

This report does not aim to dismiss the findings of previous fish-focused welfare studies, nor does it wish to criticize their findings. These studies have made large contributions to the state-of-knowledge of Cambodia's inland fisheries. Given Cambodia's political instability during the time several of these studies were done and the general inexistence of necessary background data (e.g. census information) until relatively recently, the shortcomings of these studies are less surprising. However this report does not pander to them.

This report argues that previous fish-focused welfare studies done in Cambodia and around the Mekong basin have several methodological flaws that can and should be avoided by future studies. The first part of this report discusses the methodological flaws of previous surveys and is based on an intensive review of 13 fish-focused welfare studies done in Cambodia, Laos and Vietnam[1]. In general terms, we explore why and how did these flaws damage the credibility of the study results. Specific details of each of the reviewed studies are mostly left to the appendices. Again this report does not aim to criticize specific studies but argues that their methodologies can be improved upon for the future.

The second part of this report presents the methodology developed by the welfare valuation team for the household survey component of the project. Understanding the drawbacks associated with previous survey methods allowed us to move forward with an understanding of what to avoid. This section explains our methodological approach as well as the reasoning behind it. The methodology was designed to be statistically valid by avoiding sampling and non-sampling errors and to avoid survey design errors caused by inadequate treatment of fishing

[1] As well as similar studies conducted by WorldFish in Africa

dependence. Survey design is often couched in statistics because, obviously, they are closely related. This report uses a bare minimum of statistics and statistical terms and concepts to ensure that this report remains comprehensible to a wide variety of readers. However some statistics cannot be avoided entirely.

It is hoped that this report will help future studies in the development of their own methodologies, particularly fish-focused welfare studies done in Cambodia

2.2 LITERATURE REVIEW

As a first step in developing the protocol for the household welfare survey, the welfare valuation group undertook a review of previous fishery-focused welfare surveys conducted in the region. The purpose of the review was to inform the methodology of the welfare valuation survey. Thirteen surveys in total were reviewed: eight were done in Cambodia, two in Vietnam, one in Lao PDR and one in Nigeria/Cameroon. Eleven of the surveys were done at a household or individual level, two were based on participatory interviews with the entire village and one focused on commercial fishing lot operators. Because the household welfare survey is the main component of the welfare valuation component, the team was interested primarily in the protocols used in other household welfare surveys. The following surveys were reviewed by the welfare valuation team:

- Chiwaula, L.S., Witt, R., and H. Waibel. 2010. An Asset-Based Approach to Vulnerability: The Case of Small-Scale Fishing Areas in Cameroon and Nigeria. Journal of Development Studies. 47; 2; 338-353.
- Dubeau, P., Ouch Poeu., and Sjorslev, J. No Date Specified. Estimating Fish and Aquatic Animal Productivity/Yield per Area. Report for the Project: “Assessment of Mekong Fisheries Component (AFMC)”. Mekong River Comission, No Location of Publication Specified. 21 pp.
- Gregory, R., Guttman, H., and T. Kekputherith. 1996. Poor in all but Fish A study of the collection of ricefield foods from three villages in Svay Theap District, Svay Rieng. Working paper No 5. AIT Aquaculture Outreach, Cambodia. 21 pp.
- Hortle K.G., Troeung R., and S. Lieng. 2008. Yield and value of the wild fishery of rice fields in Battambang Province, near the Tonle Sap Lake, Cambodia. Technical Paper No 18. Mekong River Commission, Vientiane, Lao PDR. 62 pp.
- Kaing Khim., Thay Somony., Nao Thuok., and M, Ahmed. 2003. Understanding Livelihoods Dependent on Fisheries. Report for the Project: “Understanding Livelihoods Dependent on Inland Fisheries in Bangladesh and Southeast Asia”. Department for International Development. 42 pp. < *http://www.fmsp.org.uk/Documents/r8118/r8118_22.pdf*>
- Keskinen, M. 2003. The Great Diversity of Livelihoods – Socio-economic survey of the Tonle Sap Lake. WUP-FIN Socio-economic Studies on Tonle Sap 8 . Report for the project: “Tonle Sap Modelling Project (WUP-FIN)”. MRCS/WUP-FIN, Phnom Penh, Cambodia. 121 pp.
- Lem, A., and N. Nghia. 2004. “Economic Modelling and Fish Consumption.” *Fish Marketing and Credit in Viet Nam*. Ed. N. Nghia. Vol. 468. Rome: Food and Agricultural Organization. 123-34. Print. FAO Fisheries Technical Paper
- M, Ahmed., Hap Navy., Ly Vuthy., and Tiongco M. 1998.Socio-Economic Assesment of Freshwater Capture Fisheries of Cambodia, Report On a Household Survey. Report for the Project: “Management of Freshwater Capture Fisheries of Cambodia”. Mekong River Commission, Phnom Penh, Cambodia. 186 pp.

- Rab, M.A., H. Navy, M. Ahmed, K. Seng and K. Viner. 2006. Socioeconomics and Values of Resources in Great Lake-Tonle Sap and Mekong-Bassac area; Results from a sample survey in Kampong Chhnang, Siem Reap and Kandal Provinces, Cambodia. WorldFish Center Discussion Series No. 4. WorldFish Center, Penang, Malaysia. 98pp.
- Sjorslev, J.G. (Ed.). 2001. An Giang Fisheries Survey. Report for the "Assessment of Mekong Fisheries Component of the Mekong River Commission Fisheries Program". Mekong River Commission, Vientiane, Lao PDR. 120 pp.
- Sjorslev, J.G. (Ed.). 2000. Luangprabang Fisheries Survey. Report for the "Assessment of Mekong Fisheries Component of the Mekong River Commission Fisheries Program" Mekong River Commission, Vientiane, Lao PDR. 75 pp.
- So Nam., Eng Tong., Souen Norng., and K. Hortle. 2005.Use of freshwater low value fish for aquaculture development in the Cambodia's Mekong basin. Proceedings of the "Regional Workshop of Low Value and "Trash Fish" in the Asia-Pacific Region". 7-9 June 2005. Hanoi, Vietnam. 25 pp.
- Troeung Roth, Aun Sinath, Lieng Sopha, Deap Loeung and N. Zalinge. 2002. A comparison of fish yields and species composition between one fishing lot in Battambang province and two fishing lots in Prey Veng province. Proceedings of the "5th Technical Symposium on Mekong Fisheries".11-12 December 2002. Phnom Penh, Cambodia. 7 pp.

The next section explores the typical methodology used in these surveys and highlights exceptions when a study deviates from the norm. To organize this section I use the survey design flow chart developed by the United Nation's and presented in their *Household Sample Surveys in Developing and Transitional Countries* handbook (UN, 2005). This section focuses on the initial steps of survey design: how previous surveys defined their target population, developed a sampling frame, created stratifications, designed their sample (e.g. sample size determination), clustered households, and finally selected households for survey inclusion. It should be noted that the two largest surveys (by sample size), by Ahmed et al (1998) and Rab et al (2006), figure prominently in this review, since they are considered the definitive studies that focus on the contribution of fish to household welfare in Cambodia.

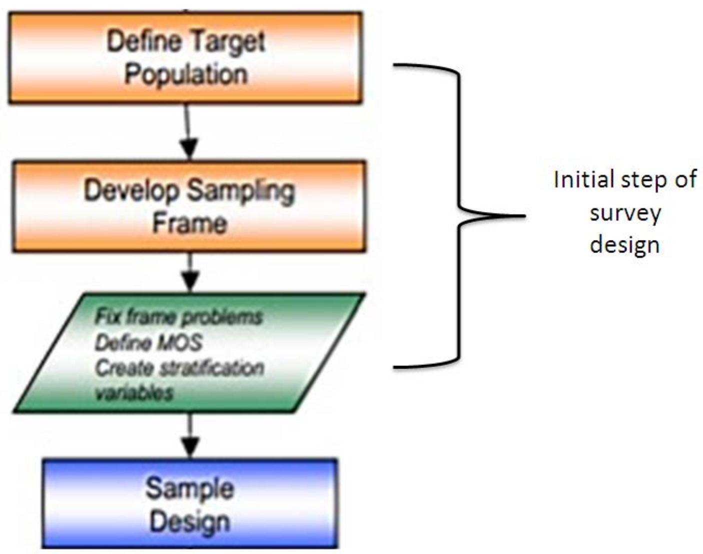


Figure 1: The Initial Steps of Survey Design

2.2.1 Selecting a population of interest

Almost all surveys start out by defining their population of interest: the population of people they wish to study. A well defined population of interest allows for a more focused approach to questionnaire design, can lower survey cost and allows for a population count and the corollary determination of the minimum sample size required for statistically valid inferences (United Nations, 2005). Without exception the reviewed studies have based their population of interest on where they live. That is, geography. Surveys define their population as those living in specific provinces, basins, etc. The reasons for selection of a particular geographic area varies: geographical areas were selected based on logistic considerations such as ease of access (e.g. Sjorslev 2000), their access to fishery resources (as determined by expert interviews) (e.g. Ahmed 1998) or in order to achieve some sort of representativeness (e.g. Rab selects Kandal province in Cambodia for survey inclusion to represent the fishery in the Mekong-Bassac basin).

Gregory, Hortle and Dubeau define their population geographically, but use a different sample technique that bears mention here. Both studies define their population of interest as all those who use a specific fishery resource regardless of where they live. For example Dubeau selected an area of about 40 Km^2 and selected fishers for a survey based on whether they fish in the area, not on the location of their home. Hortle visited 9 rain-fed rice fields once a week and interviewed 20-50% of the fishers that were present at the time.

There are several issues researchers should be made aware of when selecting their population of interest, whether based on geographic regions or not. First, if the study is meant to be inferential, the results cannot be generalized beyond the population of interest that constitutes the sample frame. Ahmed selects provinces, districts and communes in Cambodia based on their dependence on fisheries. Only areas classified as fishery-dependent are included in the survey. That is, the population of interest is the people living in fishing-dependent areas of Cambodia. Generalizing any of Ahmed's results to other populations (such as the population not living in fishing dependent communes) cannot be done.

A second issue is selecting geographic areas based on incorrect notions of "representativeness". Most studies that select sites based on some sort of representativeness assume that their results can be generalized to other areas that are "represented" by the target population. For example, in Hap's survey: "the authors selected three study villages to represent the three major ecosystems of Cambodian fisheries". The study goes on to generalize their results as representative of all of Cambodia. No matter how representative these villages are, it is misleading to claim that the results can be applied to the rest of Cambodia (at least without a wide margin of error). Authors must be aware that representativeness is not a substitute for statistical robustness. In addition, claims of "representativeness" should be substantiated and explained, not assumed as in most studies. Sjorslev (2000), in a household welfare survey done in Luang Prabang province in Lao PDR, writes: "Luang Prabang Province was chosen as the overall sample frame, partly for logistic reasons. The province is easily accessible from Vientiane; it has a relatively good local infrastructure and is reasonably representative of the northern Lao environment." Here, the report is potentially contradicting itself. If Luang Prabang province has *relatively* good local infrastructure and is easily accessible from Vientiane, then isn't it possible that the province differs sufficiently from other provinces in northern Lao PDR such that it isn't actually reasonably representative of these provinces? It stands to reason that better infrastructure and access to Vientiane would give fishers better access to markets and hence, better prices for their catch, than the rest of northern Lao PDR. Of course, Sjorslev's assertion may in fact be correct, but the report does nothing to substantiate the claim.

A final issue that authors should be aware of is the danger of selecting geographic areas based on biological or ecological criteria for socio-economic studies such as household welfare surveys. Generally, authors assume that biological factors (e.g. agro-ecological zones) dictate the livelihood strategies of the population and thus stratify their sample based on geographic location and then select provinces within those stratifications as "representative" of the geographic area. For example, Kaing identifies three main ecosystems in Cambodia (mainstream of Mekong, Tonle Sap Lake and the downstream Mekong) and selects one province within each major ecosystem to study. Notice these ecosystems refer to fishery ecosystems. Like most studies they implicitly assume that all provinces within these groups are similar enough that one province can be considered "representative" of the rest.

Especially with respect to socio-economic surveys, this may not always be the case. Other considerations, such as the seasonality and access to other alternative livelihoods (e.g. rice farming in the wet season), may modulate fishery dependence and the seasonality of this dependence to a larger extent than biological parameters. In other words, household surveys may not need to group populations into biological or ecological based zones. And even if they should, none of the reviewed surveys provide a rationale for this decision.

2.2.2 Developing a sample frame

Once a population of interest was determined, surveys typically compile a sampling frame: a list of units (communes, village, household) that cover the entire population of interest. Proper construction of the sampling frame minimizes sampling bias. In most of the reviewed studies, the construction of the sampling frame is not mentioned. This is unsurprising for the smaller case studies. Ahmed, lacking detailed information such as a census and citing the lack of resources and poor organizational strength of provincial authorities charged with keeping fishery statistics, consulted experts in order to construct the sample frame. Sjorslev (2001) makes no mention of the sample frame construction, nor does Rab. Recent studies in Cambodia, with large robust datasets such as the 2008 census and the Seila commune database, should be able to construct proper sampling frames without relying on experts or researcher discretion.

2.2.3 Clustering

Most of the studies are household surveys that utilize a two-stage clustered sample survey design. Surveys intend to target households but because of time and cost considerations households aren't randomly selected. Rather, households are first clustered, typically in Cambodia at the village or commune level. These clusters, known as primary sampling units, are first selected and then households within the primary sampling units are chosen for the survey. This cuts down on travel time and costs. Households in villages are found close together making surveys cheaper and quicker to implement. Ahmed, for example, first randomly selects communes and then randomly samples households within the commune. Without clustering, surveys will spend more time and money than is necessary, such as when enumerators must travel several days to get to a village in order to interview a single household.

2.2.4 Stratification variables

After clustering, primary sampling units and/or household are typically stratified into non-overlapping groups. . Stratification is commonly used to decrease the variability of estimates and is usually done by stratifying the sample using criteria that is highly correlated with the variables of interest. (Deaton, 1997)

Rab stratifies villages (the primary sampling unit) into fishing villages, farming villages and fishing-cum-farming villages and selects an equal number of villages from each of the three strata. Rab also stratifies households according to wealth (low, medium, high) and selects an equal number of households from each of those strata. Almost all other studies rely on a similar stratification procedure. Lem and Nghia's study of household fish consumption in Vietnam groups households into urban, rural and suburban strata.

One of the major problems with most of the previous surveys is their reliance on subjective criteria for stratification of the sample frame. The literature review has found that previous studies rely on anonymous and unverifiable experts, result in overlapping strata and are by their nature non-replicable and hence unusable in future follow-up studies.

Subjective stratifications typically rely on local knowledge collected by experts and informants who are dispersed geographically. Without coordination subjective stratifications can overlap when attempting to compare one to another. For example Rab asks village chiefs to stratify households in their village into three wealth categories (low, middle, high). At a village level there is no overlap because each house is either low wealth, middle wealth or high wealth. But if absolute wealth varies across villages, and if village chiefs vary in their assessment of low, middle and high wealth, then strata will overlap. A chief from a wealthy village might classify one household as poor and a chief from a poor village may classify that same household as wealthy. Strata are not held constant and comparisons across strata are not possible at any level other than village level.

Studies that rely exclusively on expert opinion for sample stratification force readers to make a leap of faith: readers must assume that experts are non-biased and are actually experts, that information and observations by key informants are accurate, etc. When Ahmed stratifies their sample frame into fishing and non-fishing dependent districts and communes, the authors use information collected during: "1) meetings with province/district level fishery officials 2) interviews with Key Potential Informers 3) personal observations by project experts during site visits." Chiwaula uses experts exclusively to select the 13 villages in the survey sample. Regardless of the credibility of the experts and the information they provide, the opaque nature of these kinds of stratifications erodes the credibility of the study itself since strata can easily be manipulated to achieve a particular result.

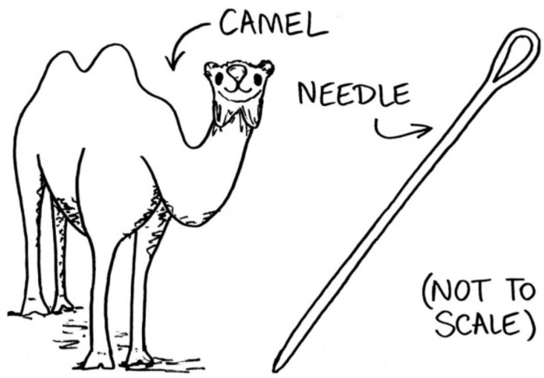


Figure 2: Without an objective scale or benchmark, subjective stratification criteria often lead to mix-ups[2]

Subjective stratifications cannot be duplicated and hence cannot be repeated in later studies. Keskinin developed an objective stratification of village fishing dependence along the Tonle Sap lake based on sea-level elevation. Villages located between 0-6 meters above sea level were placed in one stratum while villages located between 6 and 10 meters were placed in another. Compared to a subjective stratification such as "near sea level" or "slightly above sea level", Keskinin's stratification is repeatable and transparent.

Another point should be mentioned, both Ahmed and Rab stratify their population into fishing/non-fishing groups (or farmer/fisher or other binary stratifications). Ahmed classified provinces, districts, communes and households into either fishing or non-fishing strata and Rab used three strata to classify villages: fishing, farming and fishing-cum-farming.

These stratifications oversimplify reality to such a degree that the distinctions between the two break down. For example, in the Ahmed study 33% of households categorized as fishing dependent own a plough and 19.5% did not purchase rice in the previous year. This indicates fishing dependent households do quite a bit of farming, nearly 1/5th of them do not even need to purchase rice. According to Rab, the average fish catch in farming villages was 557 Kg per household and 29% of households in farming villages fish year round. The latest socio-economic survey from the National institute of Statistics (Cambodia, 2009) shows that about 56% of surveyed households engage in fishing activity but only about 1.8% of families in Cambodia indicate that they are dependent on fish, according to the 2010 SEILA database. Gregory et al (1996) notes that many of the self-identified rice farmers surveyed actually earn a greater proportion of their income from fishing and the collection of other aquatic animals.

The point is that classifying a population into fishing/non-fishing strata ignores the large variation in fishing dependence across households regardless of what they consider their main occupation. The sustainable livelihoods approach argues that rather than categorizing households based on what they tell enumerators, fishing should be viewed as part of a set of diversified livelihoods which include farming, hunting, petty trade, etc, and that the intensity of livelihood practice varies seasonally (Bene and Friend, 2011; Bene et al, 2009; Allison and Ellis, 2001). The sustainable livelihood approach, an approach which has been gaining increasing acceptance by development practitioners and researchers, offers a more accurate framework for studying fishery dependence than those available to previous researchers. The welfare valuation team recognized very early on that stratifying our sample into fishing/non-fishing

[2] Image taken from: www. http://borderlessnewsandviews.com under a creative commons lisence

groups would be unhelpful and we adopted a sustainable livelihood approach from the outset. The sustainable livelihood approach is discussed in much greater detail in the Questionnaire Report.

2.2.5 Conclusions from the literature review

Previous fishery-focused household and village-level surveys have generated important findings and have resulted in numerous peer-reviewed reports. The point of the literature review was not to assess these findings but to evaluate their methods, in order to inform the development of our own survey methodology.

From the literature review, we can draw several conclusions which are helpful not only for our own research project, but for future studies as well. The findings of this literature review apply mainly for fishery-focused household surveys. Developing a rigorous methodology will involve statistics, but subject-specific knowledge should play a role. Based on our review, the following lessons can be drawn:

- Dependence on fish (nutrition, livelihood) is controlled not only by fish biology or fish ecology but by a host of other factors such as availability and size of markets, access to other livelihoods, etc. It should not be assumed that populations living within geographic areas defined by biological or ecological criteria (e.g. agro-ecological zones or climate areas) have a homogenous fish dependency and thus villages sampled within these areas should not necessarily be considered “representative”.
- For larger studies, subjective stratifications should be avoided because of problems associated with overlapping strata, the over-reliance on experts in determining stratification criteria and their non-replicable nature
- Stratification criteria that rely on expert opinion may have groupings that overlap and generally undermine the credibility of the study
- Stratifications based on fishing dependence should avoid binary fishing/non-fishing and similar classification strategies since they confuse more than they clarify

3 THE WELFARE BASELINE SURVEY PROTOCOL

3.1 INTRODUCTION

The literature review provided a base on which to move forward with the development of our own survey protocol and methodology. This section provides readers with an in-depth, almost step-by-step report of the process behind the development of the welfare baseline survey protocol. This report covers the process from the welfare evaluation teams initial data gathering gathering data up to the selection of villages (primary sampling units) for survey inclusion.

3.2 GATHERING DATA

As a first step, the welfare valuation team decided to explore what national-level statistics were available from Cambodia's National Institute of Statistics (NIS). We wanted to avoid the problems associated with over-reliance on experts and incomplete sample frames. Knowing what information was available to base our survey design decisions on was the first step in this process. The NIS is responsible for implementing the national censuses as well as some other datasets with nationwide village-level coverage. We believed this was a good place to start looking for data.

We first reviewed the questionnaires used in NIS surveys, to know which questions interested us and wanted retrieved for our own analysis. After a quick review, two questionnaires became immediately important for us:

The 2008 National Census: Cambodia's 2008 census surveyed every household in Cambodia for basic demographic information. This data was aggregated at the village level. We were interested in the following census questions:

- Commune Classification[3]: Urban or Rural
- Total Number of Households
- Total Number of Persons
- Total Number of Person's whose Primary Occupation is Fishing
- Total Number of Person's whose Secondary Occupation is Fishing

The 2010 SEILA Commune Database[4]: This database grew out of the Cambodian government's experimental decentralization reforms (Anderson, 2004). Cambodia set up SEILA as an experiment in giving more decision-making power to communes. Because data was needed to help communes with their decision-making, the NIS created the SEILA database (Anderson, 2004). The village leader in virtually every village is tasked with answering a 243 question survey every year. Although the SEILA experiment seems to have run its course, the data is still collected annually and 2010 was the latest year that data was available for. We were interested in the following questions:

- Total Number of Families
- Total Number of Families whose Primary Occupation is Fishing
- Number of Female Headed Households
- Total Number of People Aged 18-60 with Uncertain/Irregular Jobs

[3] If a commune is classified as urban, all villages within it are also classified as urban, and vice versa.
[4] The name is somewhat of a misnomer because the data is actually available at the village level as well.

- Number of Families Who Have Rice Land Less Than one Hectare
- Number of Families Who do not Own Any Rice Land
- Number of families with fish cage culture
- Distance from Village centre to the Nearest Market
- Number of Families Living in a House Located on Public Land
- Number of Family Living in Thatched Roof
- Number of Families with Row Boats Used for Fishing
- Number of Families with Motor Boats Used for Fishing

The welfare valuation team visited the NIS office and met one of our contacts, Mr. Sam Saroeurn, the Deputy Director of Statistical Standards and Analysis at the NIS. We indicated the questions we wanted retrieved, met with more NIS staff, paid some money and less than one week later all the information we needed was given to us in the form of a Microsoft Excel spreadsheet. Combining both datasets into a single one proved to be troublesome; village name spelling (in English) remains to be standardized (as do the village ID codes). For several thousand villages there was no way to automatically sync the data and about a week was spent manually combining the datasets. We considered the NIS 2008 Census dataset authoritative and always deferred to it when discrepancies in village name and village ID code arose.

It is important to note that our dataset remains imperfect: Our dataset is missing information from 16 communes and 700 villages. In the context of the Cambodian total: 14,073 villages and 1621 communes, these numbers are relatively small (Cambodia, 2008). If these omissions are random, which they appear to be (villages are missing all over the country and not only in specific regions), then these missing villages will not affect the representativeness of the dataset[5].

3.3 SELECTING A POPULATION OF INTEREST

To be precise, our population of interest is the non-transient Cambodian population residing in rural villages in the following eleven provinces[6]. The National Institute of Statistics of Cambodia classifies every village within the country as either rural, urban or a “special settlement”. This provided a useful way for us to remove all urban villages and special settlements from our sample frame.

The welfare valuation team believes that the nature of fish dependence varies sufficiently between urban and rural areas that they should be considered separate populations. This is not to say the intensity of fishing dependence is different. About 13% of the total numbers of primary occupation fishers in Cambodia are located in urban areas (mostly in Koh Kong Province). However, these numbers mask the difference in the nature of livelihood strategies employed by urban and rural households. In the welfare valuation team’s experience, rural households typically rely on natural resource-based livelihoods while urban households have a different set of livelihood opportunities that generally relies less on natural resources. Household consumption and asset ownership composition also changes across the urban-rural divide; for example urban households tend to have running water and electricity while rural

[5] When contacted the MIS about these omissions, we were led in circles and our contacts wouldn’t give us further help. We decided to leave the dataset as it is in the hopes that someone will correct it in the future.

[6] Siem Reap, Pursat, Kampong Chhnang, Battambang, Kampong Thom, Svay Rieng, Takeo, Kandal, Prey Veng, Kratie and Stung Treng

households do not. Urban households also have greater access to public services, such as garbage collection and medical care. In sum, we believe that the difference between urban and rural households is sufficiently large that they should be considered distinct populations. This has less to do with differences in the intensity of fishing dependence (which our data indicates is surprisingly high in urban areas) and more to do with how fishing fits into their livelihood portfolio and its impact on household welfare.

Because we wish to focus on inland fisheries, we do not sample in the coastal provinces of Kep, Koh Rong, Kampot or Preah Sihanouk. Nor do we sample in the provinces of Ratanakiri, Mondukiri, Banteay Meanchey, Pailin, Preah Vihear, Kampong Speu, Kampong Cham and Oddar Meanchey. Although fishing dependent villages exist in these provinces, our survey focuses on three floodplains (Tonle Sap, Low Land, Mekong) and includes only provinces within those floodplains[7].

3.4 DEVELOPING A SAMPLE FRAME AND CLUSTERING UNITS

We decided to use villages as the primary sampling unit of the welfare baseline survey, for reasons that will be explained in section 3.4. Our sample frame is based off of the 2008 National Census dataset. All villages included in our baseline survey were randomly selected from this frame.

The eleven provinces selected for the survey were placed into one of three agro-ecological zones:
Tonle Sap Floodplain: Siem Reap, Pursat, Kampong Chhnang, Battambang, Kampong Thom
Low Land Floodplain: Takeo, Kandal, Prey Veng, Svay Rieng
Mekong Floodplain: Kratie, Stung Treng

Figure 3: Outline of Provinces Included in Sample Frame (in blue)

[7] Kampong Cham was not included in our sample because as a province it straddles the Mekong, Tonle Sap and Low Land floodplains. All other provinces included in the sample belong to a single floodplain meaning that our survey could assume that the villages within them all belonged to that same floodplain. Kampong Cham is the exception and the villages within it had the possibility of falling into any one of the three floodplains, which would have made our initial analysis more difficult.

The division of the sample frame into agro-ecological zones is not to be mistaken for a stratification. Each zone is treated as an independent agro-ecological system, for which a sample is to be drawn. In a sense, we break up the welfare baseline survey into three smaller independent ones.

3.4 Stratification variables

Given the problems associated with the stratification variables used in previous fish-focused household surveys, the welfare valuation team decided to create our own. Because one of our primary research questions is about the *relative* contribution of fish to household welfare, we required a stratification variable that would allow us to stratify our sample frame according to fish dependence, and then to sample villages along this spectrum of fish dependence. Of course, the villages would be broken down into agro-ecological zones before selection.

Taking into account the sustainable livelihood approach, which stresses that fish dependency is better modelled on a spectrum rather than binary fisher/non-fisher categories, we created the fish dependency indicator, a variable designed to model village-level fish dependence, using fishing activity as a proxy variable for fish dependence. The fish dependency indicator is a composite variable of 6 variables gathered from 3 datasets:

- **The 2008 National Census**
    - Total Number of Persons
    - Total Number of Person's whose Primary Occupation is Fishing
    - Total Number of Person's whose Secondary Occupation is Fishing

- **The 2010 SEILA Commune Database**
    - Total Number of Families whose Primary Occupation is Fishing
    - Number of Families with Row Boats Used for Fishing
    - Number of Families with Motor Boats Used for Fishing

- **The 2006 Ministry of Planning Village Poverty Score:**
    - The 2006 Ministry of Planning is itself a composite indicator of village-level poverty. It uses a variety of variables such as the number of televisions in the village, literacy rates and the proportion of children in school to create a poverty score for each village in Cambodia. The score is standardized with a mean of 100, meaning a village with a score of 100 is as poor as the average village in Cambodia and a village with a score of 101 is slightly poorer than the average village in Cambodia.

3.4.1 Designing the fish dependency indicator – fishing activity as a proxy for fish dependence[8]

The first step in creating the fishing dependency score was to generate a variable to estimate the amount of fishing activity occurring in the village. Fishing as a livelihood activity, that is fish production and consumption, is the essential economic feature of fishing dependence. Although fishing dependence can be characterized in a number of ways (based on concepts of, for example, food security, access to fishery resources, risk mitigation, reduction in household vulnerability), fishing activity can serve as a proxy variable for these other measures of fish dependence. For example, villages with many active fishers are likely to support fish-dependent livelihoods not captured in our dataset (e.g. fish sellers, fish processors) and to have a higher proportion of fish in their diet than villages with a lower number of fishers.

For each village we compare the number of persons who identified their primary occupation as fishing (National Census) to the number of families who identified their primary occupation as fishing (2010 SEILA database). Because people who identify as fishers (primary occupation) are likely to belong to families that also identify as fishing families, we did not want to double count these two groups. Thus, we take whichever number is greater and use it in our calculation.

We then add to this the number of persons who identify as fishers (secondary occupation) taken from the National Census. This gives us the total number of persons in the village who identify as engaging in fishing activity. We did not assign additional weight for primary occupation fishers or less weight to secondary occupation fishers. Livelihood strategies in rural Cambodian households change seasonally and in response to shocks; occupational designations such as primary and secondary are less accurate when put into this context. Rather than static, non-changing occupations, households engage in a highly variable household occupational portfolio (e.g. Béné and Friend, 2010)[9]. According to Béné and Friend (2010): "The majority of rural people do not classify themselves as 'fishers' per se, but rather as 'going fishing". The National Census was completed in March 2008; it is likely that if the census took place during a different month, respondents would have identified different primary and secondary occupations. Because the distinction between primary and secondary occupations is quite porous, we do not assign any weights in our fish dependency indicator.

As previously mentioned (and explored in greater detail in the Questionnaire Report), fishing activity is seasonally and opportunistically pursued. In other words, fishing activity is highly variable and likely modulated through the demands of other livelihood activities (e.g. farming). Thus, it is likely that a one-time cross sectional survey will miss many people who engage in significant fishing activities but were not fishing during the survey period. To take into account these "unaccounted" fishers, we use the number of fishing boats in the village[10]. If a household owns a fishing boat, they likely engage in fishing activities (or rent it out to people who do), although perhaps only part of the time.

Because the number of fishing boats in a village is not as variable through time as the number of self-identified fishers are (e.g. the number of fishing boats in a village does not change

[8] A more formal treatment of the creation of the fishing dependency indicator is found in Appendix A.

[9] A more in-depth discussion of the sustainable livelihoods approach as an alternative to over simplistic binary fishing/non-fishing occupations, see the Questionnaire Report

[10] The SEILA 2010 questionnaire makes the distinction between boats used for fishing and boats used for other activities such as transport.

seasonally), we use this variable as a proxy for "unaccounted seasonal/opportunistic" fishers. We take the number of fishing boats (motor boats + row boat), subtract the number of previously identified fishers (who we assume already use one fishing boat each) and if there are any fishing boats that remain, we incorporate them into the fishing dependency indicator as "unaccounted" fishers.

Figure 4: We assume that if there are more fishing boats in a village than self-identified fishers, there are still fishers unaccounted for in the village

After totalling the number of fishers in the village, we divide this number by the village population. This figure is essentially the probability of a random villager being a fisher.

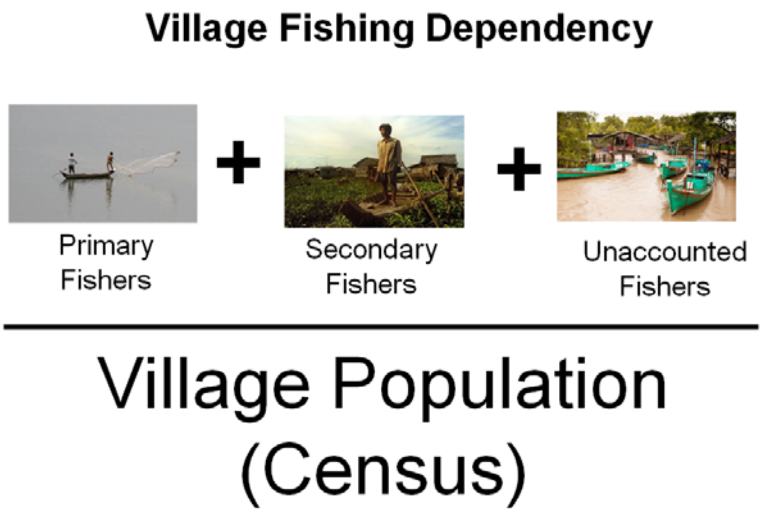


Figure 5: Calculating the Probability of a Random Villager being a Fisher

3.4.2 Designing the fish dependency indicator – incorporating poverty into fish dependence[11]

We then assign a bonus or deduction to this fish dependence score based on the wealth level of the village. The law of diminishing marginal returns, a well accepted tenant in classical economics, dictates that the welfare benefit of one additional dollar of income derived from fisheries (or an additional calorie derived from fish) diminishes as the recipient becomes wealthier. Given the same number of fishers in a village (which is our proxy variable for fishery dependence), a poorer villager is in reality more dependent on fish since they earn a larger share of their income (or calories) from aquatic resources. We wanted this reality to be reflected in our stratification variable. Thus, if a village is 5% poorer than the average Cambodian village, their fishing dependence score is given a 5% bonus. If a village is 3% wealthier than the average Cambodian village, their fishing dependence score is given a 3% penalty. In this way, we incorporate the dimension of poverty into our fishing dependence variable. A more formal treatment of the creation of the fishing dependency indicator is found in Appendix A.

3.4.3 Stratifying the sample frame and selecting villages

Because our survey aims to study the *relative* contribution of fisheries to household welfare in diversified farming systems, we aim to study households along the entire spectrum of fishery dependence, including household with none or little participation in fishing activities. Once each village in the selected provinces was given a fishery dependence score, we stratified our sample frame into five categories. Villages with a fish dependency score of 0 (about 50% of the sample frame) were placed in one group and the rest (villages with a fish dependency score greater than 0) were placed into quartiles.

Given the maximum feasible household sample size *n=720*, which has been determined by the available budget and the estimated survey costs, we calculated the number of households to be included in the sample for each study site, given the relative distribution of households among floodplains, i.e. floodplains with bigger population size receive a higher weight (Table 1).

Table 1: Distribution of the population and the sample among study areas

	Number of households	Share	n(floodplain)
Tonle Sap	568,471	0.3462	250
Low Land	682,809	0.4159	300
Mekong	390,561	0.2379	170
Total	1,641,841	1.0000	720

The distribution of the sample by floodplain is not to be mistaken for stratification. Each study site is treated as an independent agro-ecological system, for which a sample is to be drawn. Nonetheless, in order to attain comparable representativeness of the sample for all three study sites, the allocation of *n* households is done proportional to size.

[11] A more formal treatment of the creation of the fishing dependency indicator is found in Appendix A.

Due to the size of the geographical area covered by each of the study floodplains, a simple random sampling would not have been feasible, given the logistical and budget constraints of the study. We use a two-stage stratified sampling design to ensure the feasibility of the survey without compromising our sample's validity for statistical inference.

The population is geographically stratified into contiguous villages (our primary sampling unit). Selecting communes as primary sampling unit would possibly have facilitated the logistical organization of the survey. However, as shown in the next figure, the variation of the fish dependency indicator within the communes is high, especially for high values of fishing dependence. For this reason, our survey decided to use villages (as opposed to communes) as the primary sampling unit.

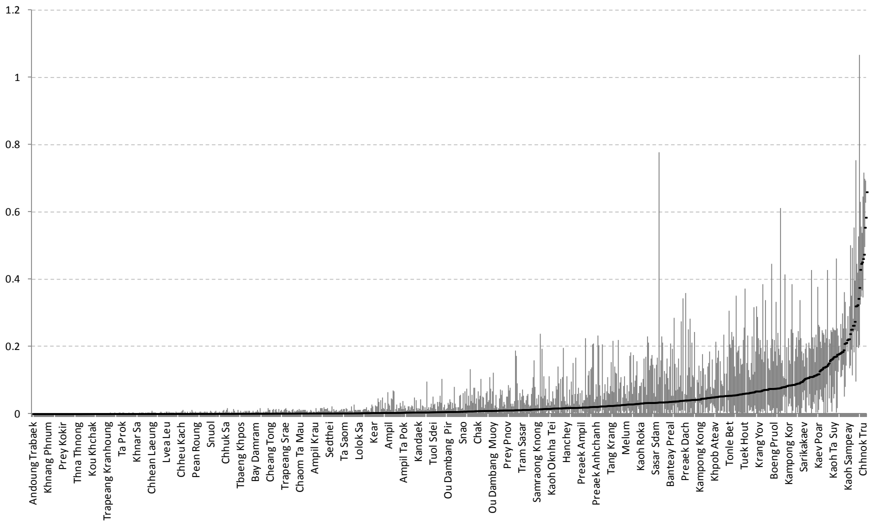


Figure 6: Distribution of Fishing Dependence by commune (mean, min and max values)

Village selection was done proportional to size, in order to account for the differences in number of villages per fishing dependence stratum and floodplain. Since the share of zero fish dependency villages in the frame is high, the weighting was done only for villages with $\delta^v > 0$, then randomly selecting two villages from each floodplain from the subset with $\delta^v = 0$.

The budget allows for approximately 720 households (approximately 37 villages) to be surveyed. The 37 villages were selected from each agro-ecological zone in proportion to the total number of villages in the agro-ecological zone. Within each agro-ecological zone two villages are selected with a fish dependency indicator score of 0 and the rest are distributed across the fish dependency quartiles in proportion to the population within those quartiles (see Table 2 and Table 3)

Table 2: Number of villages, by fish dependency quartile and agro-ecological zone

	0	1	2	3	4	Total
Tonle Sap	1,290	450	420	382	524	3,066
Low Land	2,122	439	435	381	261	3,638
Mekong	810	206	240	332	309	1,897
Total	4,222	1,095	1,095	1,095	1,094	8,601

Table 3: Number of villages to be sampled, by fish dependency quartile and agro-ecological zone

Quartile	**0**	**1**	**2**	**3**	**4**	**total**
Tonle Sap	2	3	3	2	3	13
Low Land	2	4	4	4	2	16
Mekong	2	1	1	2	2	8
Total	**6**	**8**	**8**	**8**	**7**	**37**

3.5 SELECTING VILLAGES

Villages were selected randomly using the random number generator in Microsoft Excel. Once villages were selected, they were plotted on a map. Three villages had to be reselected at random because they were obviously remote and inaccessible (i.e. they were located in the Cardamom Mountains of western Cambodia).

CARDI and WorldFish staff then visited each selected village in advance of the enumerators. There were several reasons for doing this:

- To ensure selected villages were accessible
- To map out directions to the selected villages so that that enumerators would not get lost en route
- To meet with village leaders (and deputies) to explain to them in advance about the Welfare Valuation Project. We also gathered their contact information
- To access and make copies of the village-level household list. This list is always kept by the village leader and the village deputy; village leaders allowed us to make a copy of the list once we met with them.

3.6 SELECTING HOUSEHOLDS

We randomly selected households within selected villages. To arrive at a household sample size of approximately 720, 8.5% of households in the village were randomly selected from the village-level household list.

4 CONCLUSION

This report has provided a detailed overview of the sampling design and methodology developed by the welfare valuation team for the household welfare survey. The rationale behind our design, based off of a review of previous fish-focused welfare surveys, was explained. We also explored the development of the fish dependency indicator, a key component of the overall survey design, and how it was used in our sampling protocol. We provided an in-depth explanation of our sampling procedure and the practical steps we took to ensure our villages were randomly selected but also accessible for enumerators.

The analysis detailed above provides a transparent account of the methodology put in place for the field surveys of the current project. The results will be, for the first time, statistically comparable, with known degrees of confidence. This detailed methodological report should also be useful to other researchers willing to develop their own fish-focused household surveys.

This methodological analysis also produced, for the first time in Cambodia, a map detailing fish dependency nationwide at the commune level. This result in itself paves the way for targeted development interventions.

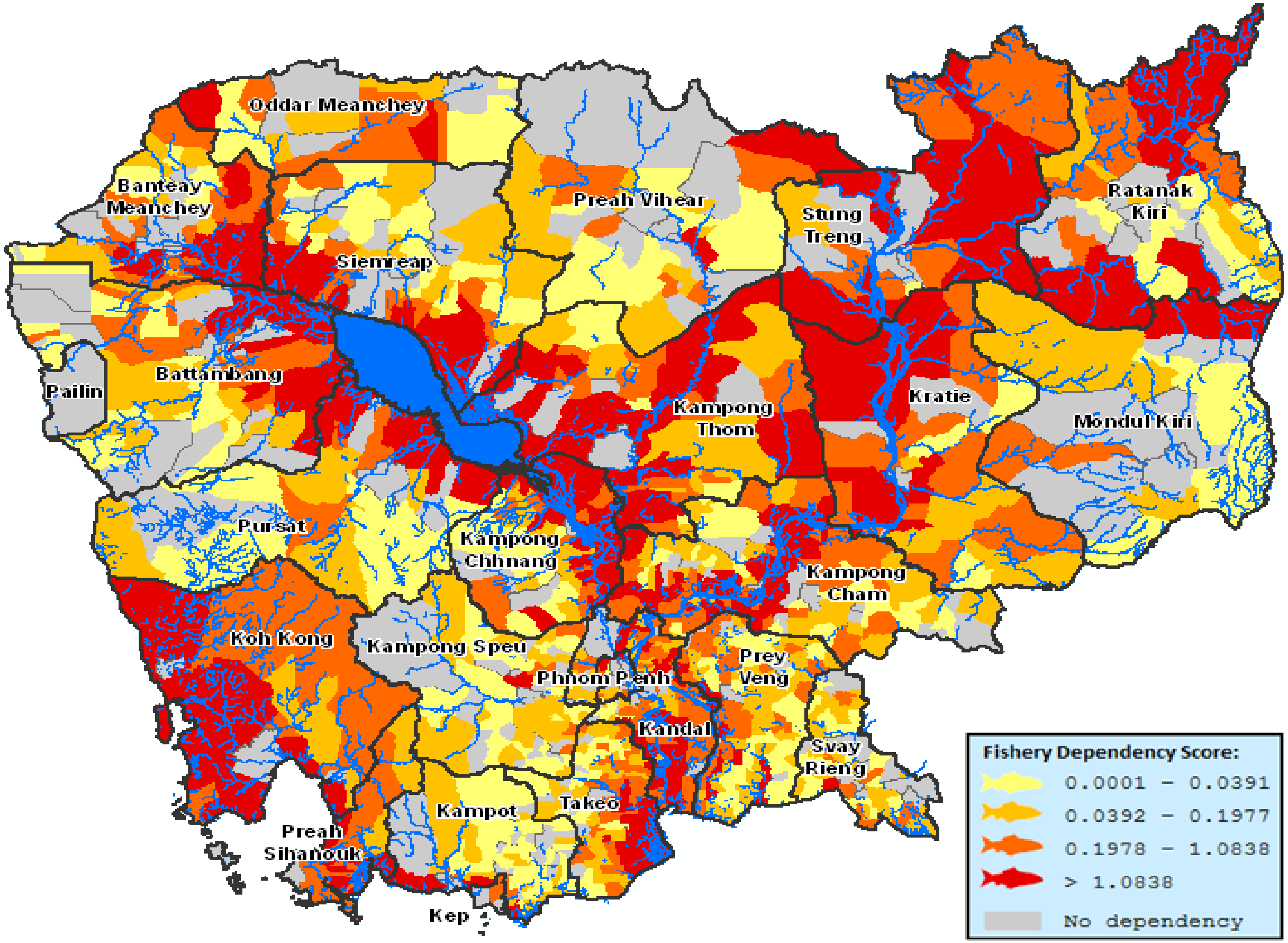


Figure 6: Fishing Dependency by commune in Cambodia

5 BIBLIOGRAPHY

Andersen, H. 2004. Cambodia's Seila Program: A Decentralized Approach to Rural Development and Poverty Reduction. Proceedings of the International Symposium "Scaling Up Poverty Reduction: A Global Learning Process and Conference". May 25-27 2004, Shanghai. World Bank.

Cambodia. 2008. National Institute of Statistics. Provisional Population Totals. Phnom Penh: n.p., Print.

Cambodia. 2009. National Institute of Statistics. Cambodia Socio-Economic Survey Tables (2009).Web. <http://www.nis.gov.kh/index.php/social-statistics/cses/cses-tables>

Chiwaula, L.S., Witt, R., and H. Waibel. 2010. An Asset-Based Approach to Vulnerability: The Case of Small-Scale Fishing Areas in Cameroon and Nigeria. Journal of Development Studies. 47; 2; 338-353.

Deaton, A. 1997. The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. The Johns Hopkins University Press, Baltimore. 479 pp.

Dubeau, P., Ouch Poeu., and Sjorslev, J. No Date Specified. Estimating Fish and Aquatic Animal Productivity/Yield per Area. Report for the Project: "Assessment of Mekong Fisheries Component (AFMC)". Mekong River Comission, No Location of Publication Specified. 21 pp.

Gregory, R., Guttman, H., and T. Kekputherith. 1996. Poor in all but Fish A study of the collection of ricefield foods from three villages in Svay Theap District, Svay Rieng. Working paper No 5. AIT Aquaculture Outreach, Cambodia. 21 pp.

Hortle K.G., Troeung R., and S. Lieng. 2008. Yield and value of the wild fishery of rice fields in Battambang Province, near the Tonle Sap Lake, Cambodia. Technical Paper No 18. Mekong River Commission, Vientiane, Lao PDR. 62 pp.

Kaing Khim., Thay Somony., Nao Thuok., and M, Ahmed. 2003. Understanding Livelihoods Dependent on Fisheries. Report for the Project: "Understanding Livelihoods Dependent on Inland Fisheries in Bangladesh and Southeast Asia". Department for International Development. 42 pp. < *http://www.fmsp.org.uk/Documents/r8118/r8118_22.pdf*>

Keskinen, M. 2003. The Great Diversity of Livelihoods – Socio-economic survey of the Tonle Sap Lake. WUP-FIN Socio-economic Studies on Tonle Sap 8 . Report for the project: "Tonle Sap Modelling Project (WUP-FIN)". MRCS/WUP-FIN, Phnom Penh, Cambodia. 121 pp.

Lem, A., and N. Nghia. 2004. "Economic Modelling and Fish Consumption." Fish Marketing and Credit in Viet Nam. Ed. N. Nghia. Vol. 468. Rome: Food and Agricultural Organization. 123-34. Print. FAO Fisheries Technical Paper

M, Ahmed., Hap Navy., Ly Vuthy., and Tiongco M. 1998.Socio-Economic Assesment of Freshwater Capture Fisheries of Cambodia, Report On a Household Survey. Report for the Project: "Management of Freshwater Capture Fisheries of Cambodia". Mekong River Commission, Phnom Penh, Cambodia. 186 pp.

Rab, M.A., H. Navy, M. Ahmed, K. Seng and K. Viner. 2006. Socioeconomics and Values of Resources in Great Lake-Tonle Sap and Mekong-Bassac area; Results from a sample survey in Kampong Chhnang, Siem Reap and Kandal Provinces, Cambodia. WorldFish Center Discussion Series No. 4. WorldFish Center, Penang, Malaysia. 98pp.

Sjorslev, J.G. (Ed.). 2000. Luangprabang Fisheries Survey. Report for the “Assessment of Mekong Fisheries Component of the Mekong River Commission Fisheries Program” Mekong River Commission, Vientiane, Lao PDR. 75 pp.

Sjorslev, J.G. (Ed.). 2001. An Giang Fisheries Survey. Report for the “Assessment of Mekong Fisheries Component of the Mekong River Commission Fisheries Program”. Mekong River Commission, Vientiane, Lao PDR. 120 pp.

So Nam., Eng Tong., Souen Norng., and K. Hortle. 2005.Use of freshwater low value fish for aquaculture development in the Cambodia’s Mekong basin. Proceedings of the “Regional Workshop of Low Value and “Trash Fish” in the Asia-Pacific Region”. 7-9 June 2005. Hanoi, Vietnam. 25 pp.

Troeung Roth, Aun Sinath, Lieng Sopha, Deap Loeung and N. Zalinge. 2002. A comparison of fish yields and species composition between one fishing lot in Battambang province and two fishing lots in Prey Veng province. Proceedings of the “5th Technical Symposium on Mekong Fisheries”.11-12 December 2002. Phnom Penh, Cambodia. 7 pp.

United Nations. Department of Economic and Social Affairs. Statistics Division. Household Sample Surveys in Developing and Transition Countries. New York: United Nations Publications, 2005. Web.

6 APPENDIX

Sampling protocol for the "Assessing economic and welfare values of fish in the Lower Mekong Basin" project – Cambodia case study.
Rudolf Witt, June 2012

1. Measuring dependence on fishery

The major objective of the welfare evaluation study is to estimate the contribution of fisheries to the economic well-being of the non-transient Cambodian population residing in rural villages. The units of analysis are defined as households that belong to the above defined population of interest. As such, the essential characteristic of the study population is its dependence on fisheries.
Although dependence on fisheries is defined and measured in a number of ways (based on concepts of, for example, food security, nutritional security, risk mitigation, reduction of vulnerability, or monetary well-being of the household), the essential economic feature of dependence on fisheries is the level of production and consumption of fish products by the household.
For the design of the welfare evaluation study we use a composite indicator of fisheries dependency, which combines information on fish production and consumption. The indicator is defined on village (or community) level, and is calculated as follows:
Let $y_{fi} \geq 0$ denote the output from fish production of individual i, and let $F = \{i : y_{fi} > 0\}$ be the set of all individuals whose output from fish production is strictly greater than zero. Then, we define the level of fishery dependency of village v in terms of fish production as the relative frequency of individuals belonging to the set F :

$$\tau_p^v = \frac{\sum_{i^v \in F} 1}{\sum_{i^v} 1}.$$

τ^v can be interpreted as the probability that an individual i^v in village v is dependent on fishing, i.e. $\tau^v = p\left(i^v \in F\right)$. It is obvious that $0 \leq \tau^v \leq 1$ such that fishery dependence increases as τ^v increases.
In order to account for the consumption side of fishery dependence, we adjust τ^v by a village-level weight ω^v with $E\left[\omega^v\right] = 1$, which represents the poverty level of the village. Fish production is arguably of greater value for poorer households, such that communities with higher levels of poverty should receive a higher weight, while fishery dependence is abated for villages with a higher average well-being.
Hence, the final fishery dependence indicator is calculated as:

$$\delta^v = \tau^v \omega^v.$$

1. Use of a priori information on fishery dependence for sampling design

To arrive at the distribution $f\left(\delta^v\right)$, we use three datasets, the National Institute of Statistics 2008 census, the Ministry of Planning's 2006 Poverty Score and the 2010 Seila[12] database.

[12] The SEILA commune database is part of the wider SEILA program of poverty-alleviation organized by the Government of Cambodia and the United Nations Development Program. The database is managed by the National Institute of Statistics, the Ministry of Planning as well as Provincial Department's of Planning Statistics. Since 2002, socio-economic information is provided annually from all communes in Cambodia. The data is collected at a village level by village leaders who fill out the questionnaire on behalf of the village.

Since all data sets are covering the whole population of Cambodia, the parameters calculated from these data can be taken as the true parameters of the population (provided that measurement errors or other possible sources of bias such as interviewer effects, question wording etc. can be neglected).

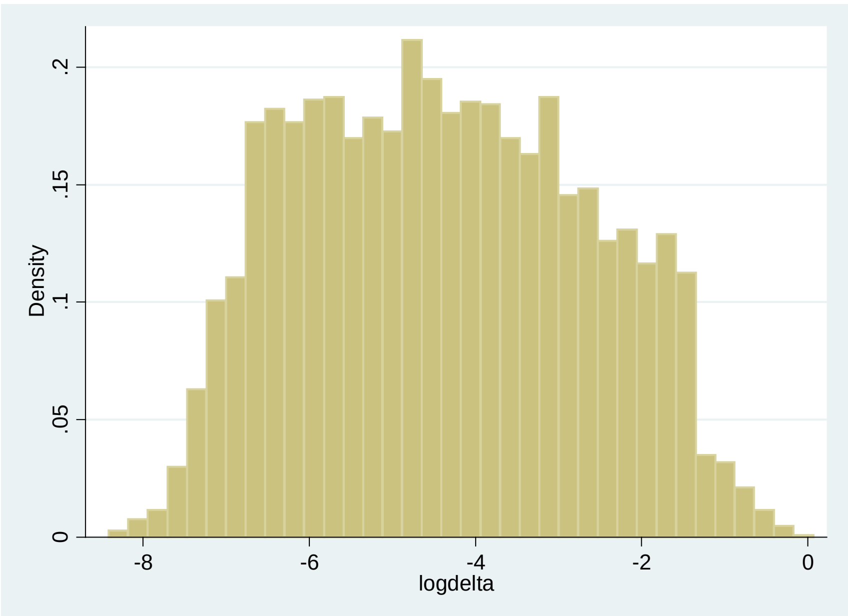
τ^v has been calculated using the number of people who consider fishing as their primary or secondary occupation (2008 census data). This information has been triangulated by the number of families who depend mainly on fishing, and the number of fisher boats, from the 2010 Seila database. The triangulation of information on fishing activities from the census has been done in order to account for the difficulties in obtaining an objective view on the involvement of households or individuals in fishing activities. This is to say that it is difficult to classify households as either fishers or non-fishers solely based on the question: "What is your main/secondary occupation?". We must look at fishing as part of a set of diversified livelihoods which include farming, hunting, petty trade, etc, and that the intensity of livelihood practice varies seasonally. For this reason, the number of individuals that categorized themselves as fishers (primary occupation) has been corrected upwards, if the number of fish-dependent families (Seila data) was greater. Also, the number of fisher boats per village was used to check whether the number of fishers (primary and secondary occupation) should be corrected upwards. The final number of fishers per village was estimated as follows:

$n_{i^v} = \max\left[\max\left(n_{pf}^{census}, n_{pf}^{seila}\right) + n_{sf}^{census}, n_{boat}^{seila}\right],\ i^v \in F$,

where the subscript *pf* stands for primary fisher and *sf* denotes secondary fisher.

The 2006 poverty score data gave us a scoring of poverty for every village ω^v. The score is standardized with a mean of 1. So a village with a score of 1 is as poor as the average village in Cambodia. A village with a score of 1.01 is slightly poorer than the average village, etc. Note that the weighting is not assigning a preference, nor is it introducing a bias with respect to the final selection of households in the sample, since the sampling is done randomly on the whole continuum of the fish dependency indicator. Hence, no specific preference is given to poor or highly fish-dependent villages in terms of the sample. The weighted fish dependency indicator is simply a means to possibly capture different levels of fish dependency (from zero fish dependency to very high fish dependency), while keeping the sample size at a manageable level.

The following figure shows the density of *ln(delta)*:



Density distribution of ln(delta) (i.e. excluding villages with delta=0)

The village census for the selected study sites (Tonle Sap, Low Land and Mekong floodplains) with *N=8,601* is then used to create four quantiles, conditional on $\delta^v > 0$, which are used as strata in the subsequent sampling design. The following table shows the boundaries, means, standard deviation and number of households in each quartile as well as for the category $\delta^v = 0$:

Distribution parameters of fish dependency indicator (delta) and number of units, by floodplain and stratum

Quartile		0	1	2	3	4
Upper bounds		0	0.0030118	0.0116265	0.0472074	1.066526
Tonle Sap	mean(delta)	0	0.001608	0.006476	0.025261	0.17164
	sd(delta)	0	0.000699	0.002486	0.010755	0.137814
	n(hh)	204,340	115,758	82,818	70,409	95,146
	n(village)	1,290	450	420	382	524
	n(commune)	271	213	214	169	159
Low Land	mean(delta)	0	0.001625	0.006375	0.023828	0.135463
	sd(delta)	0	0.000677	0.002495	0.009773	0.087223
	n(hh)	305,458	101,476	96,687	109,846	69,342
	n(village)	2,122	439	435	381	261
	n(commune)	289	229	221	187	119
Mekong	mean(delta)	0	0.001496	0.006683	0.026476	0.117663
	sd(delta)	0	0.00074	0.00247	0.009966	0.063253
	n(hh)	134,296	64,173	53,876	78,326	59,890
	n(village)	810	206	240	332	309
	n(commune)	137	105	122	127	116
Total	mean(delta)	0	0.001594	0.006481	0.025131	0.147763
	sd(delta)	0	0.000699	0.002487	0.01023	0.112209
	n(hh)	644,094	281,407	233,381	258,581	224,378
	n(village)	4222	1095	1095	1095	1094
	n(commune)	697	547	557	483	394

Determining total sample size per study site

Given the maximum feasible sample size *n=720*, which has been determined by the available budget and the estimated survey costs, we calculated the number of households to be included in the sample for each study site, given the relative distribution of households among floodplains, i.e. floodplains with bigger population size receive a higher weight (next Table).

Distribution of the population and the sample among study areas

	Number of households	Share	n(floodplain)
Tonle Sap	568,471	0.3462	250
Low Land	682,809	0.4159	300
Mekong	390,561	0.2379	170
Total	1,641,841	1.0000	720

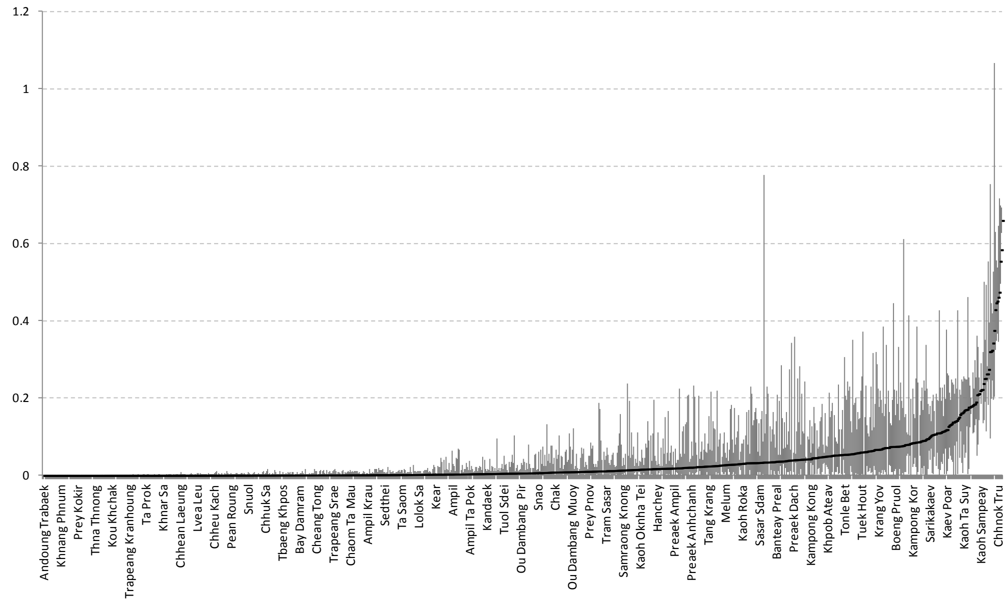
The distribution of the sample by floodplain is not to be mistaken for a stratification. Each study site is treated as an independent socio-ecological system, for which a sample is to be drawn.

Nonetheless, in order to attain comparable representativeness of the sample for all three study sites, the allocation of *n* is done proportional to size, with $\frac{n_{floodplain}}{N_{floodplain}} = 0.0438\%$.

Considerations concerning two-stage random sampling

Due to the size of the geographical area covered by each of the study floodplains, a simple random sampling would not be feasible, given the logistical and budget constraints of the study. Hence, a *two-stage stratified sampling* is proposed, which ensures the feasibility of the survey without compromising the sample's validity for statistical inference.

The population is geographically stratified into contiguous communes consisting of 2 or more villages. The total number of communes in the three floodplains is 901. The primary sampling unit could be either village or commune. Selecting communes as primary sampling unit would possibly facilitate the logistical organization of the survey. However, as shown in the next figure, the variation of the fish dependency indicator within the communes is high, especially for high values of delta.



Distribution of delta by commune (mean, min and max values)

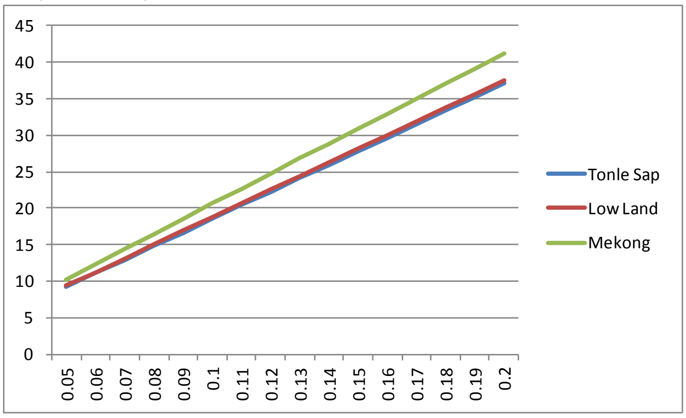
As such, it is advisable to select the villages as the primary sampling unit, and conduct the stratification of the sampling frame on village level. The total number of villages in the sampling frame is 8,601. The following table shows the number of villages in each stratum.

Number of villages, by fish dependency quartile and floodplain

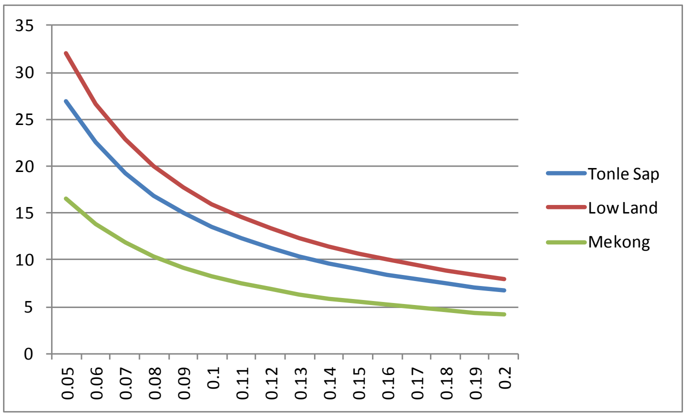
	0	1	2	3	4	Total
Tonle Sap	1,290	450	420	382	524	3,066
Low Land	2,122	439	435	381	261	3,638
Mekong	810	206	240	332	309	1,897
Total	4,222	1,095	1,095	1,095	1,094	8,601

In the first step, a certain number m of villages has to be selected within each floodplain and each fish dependency quantile. The selection is done proportional to size, in order to account for the differences in number of communes per stratum and floodplain. Since the share of zero fish dependency villages in the frame is high, the weighting should be done only for villages with $\delta^v > 0$, then randomly selecting a limited number of villages from the subset with $\delta^v = 0$. In a second step, households are randomly selected within the selected villages.

The number of villages to be selected is inversely proportional to the share of households per village that will be selected in the second step. Hence, a first decision has to be made concerning the sampling frequency per village. The next figures depict (1) the average number of households that will be selected, and (2) the number of villages to be selected, depending on the sampling frequency (x-axis). For example, for a sampling frequency of 10% per village, 13, 16 and 8 villages will have to be selected in the Tonle Sap, Low Land, and Mekong floodplains, respectively.



(1)



(2)

This decision needs to take into account that a higher number of villages will most probably increase the variable survey costs per household due to increased travelling between villages.

For a given average sample size of households per commune, the exact allocation of communes to each stratum is then done by use of stratum weights (number of communes per stratum) in order to achieve a SPS sample, or self-weighted sample. In order to arrive at a self-weighted random sample of communes per floodplain and fish dependency stratum, we calculate the frequency distribution of communes (without the "zero quantile"):

Share of villages, by fish dependency quartile and floodplain

Quartile	1	2	3	4
Tonle Sap	0.253	0.236	0.215	0.295
Low Land	0.290	0.287	0.251	0.172
Mekong	0.190	0.221	0.305	0.284

This frequency distribution is then used to calculate the number of villages to be sampled in each stratum. As such, a sampling frequency of 10% would result in the total number of 37 villages to be sampled. The next table shows the stratum-specific allocation of villages.

Number of villages to be sampled, by fish dependency quartile and floodplain, for $\frac{n_v}{N_v} = 0.1$

For $\delta^v = 0$ (the "control group") it is suggested to select two villages per floodplain.

Quartile	**0**	**1**	**2**	**3**	**4**	**total**
Tonle Sap	2	3	3	2	3	13
Low Land	2	4	4	4	2	16
Mekong	2	1	1	2	2	8
Total	**6**	**8**	**8**	**8**	**7**	**37**

A complete household list from each selected commune needs to be obtained from village elders during the focus group discussions, which will then serve as the village-level sampling frame for the final selection of households. The number of households per village will depend on the total number of households in each selected village. The resulting sample will be self-weighted (except for the control group, which will have to be weighted due to the fixed *m* in the selection procedure)

The randomness of selection on commune and household level can be assured by using a pseudo-random number generator which produces (almost) random numbers from the $[0,1]$ interval, which are then used to select the sampling units (e.g. in STATA).