Final technical report on the development of Master Sampling Frames (MSF) for fishery and aquaculture statistics

December 2018
Final technical report on the development of Master Sampling Frames (MSF) for fishery and aquaculture statistics

Table of Contents

Acronyms and Abbreviations.................................................................................. 6
Glossary.................................................................................................................. 7
Acknowledgements.................................................................................................. 13
Preface..................................................................................................................... 14
Overview of previous reports.................................................................................. 15
Executive summary.................................................................................................. 16
1. Background and objectives................................................................................ 18
   1.1. Need for reliable frames for fisheries and aquaculture............................ 19
   1.2. Sources of frames for fishery and aquaculture surveys........................... 20
   1.3. The MSF concept..................................................................................... 21
   1.4. Sampling frames and the survey process............................................... 22
   1.5. Relevance and objectives of this research project................................... 23
2. Defining survey goals......................................................................................... 25
   2.1. Fisheries production.................................................................................. 25
   2.2. Aquaculture production.......................................................................... 27
   2.3. Social/demographics.............................................................................. 27
   2.4. Economics............................................................................................... 27
   2.5. Verification and audits............................................................................ 28
3. Dividing the population..................................................................................... 29
   3.1. Fisheries.................................................................................................... 29
   3.2. Aquaculture............................................................................................. 30
   3.3. Spatial and temporal contexts.................................................................. 31
4. Aspects of the survey process that relate to frames for fisheries and aquaculture........................................................................................................ 32
   4.1. Identification of the target population..................................................... 32
   4.2. Measurement and variables..................................................................... 34
   4.3. Data collection and contact methods...................................................... 35
   4.4. Sample design.......................................................................................... 37
   4.5. Estimation................................................................................................ 40
5. The Master Sampling Frame............................................................................ 42
   5.1. Types of frames....................................................................................... 42
   5.2. Combining frames.................................................................................... 52
   5.3. Updating the frame.................................................................................. 54
6. Steps for building an MSF for fisheries and aquaculture

6.1. Purpose of activities

6.2. Need for new activities

6.3. Identification of populations

6.4. Determine variables and measurement

6.5. Determine contact method

6.6. Outline sampling design

6.7. Construct sampling frames

6.8. Identify potential non-sampling errors

6.9. Develop estimation strategy

6.10. Develop framework for storage and retrieval of data and results

6.11. Logistics and coordination

References

ANNEX - CASE STUDY - Master sampling frames for fisheries surveys:
  desktop study using data from Burkina Faso

Annex references

Appendix
List of tables and figures

LIST OF TABLES

Table 1. Example of fishing gear categorization................................. 29
Table 2. Example of frame-updating plan............................................ 55
Table 3. Biases (multiplied by 10 000) for two parameters and three methods................................................................. 86
Table 4. Monte Carlo bias, standard deviation, and root mean squared error (multiplied by 10 000)......................................................... 87
Table 5. Estimate and 95 confidence interval for total catch from dual-frame design............................................................... 92

LIST OF FIGURES

Figure 1. Categories of aquaculture populations.................................. 30
Figure 2. Correspondence between contact methods and frames.............. 37
Figure 3. Number of villages per region for Burkina Faso. Top row drawn from the Burkina Faso 2006 population census. Bottom row drawn from the subset of fishing sites in the 2008 fishery survey that measured catch-related variables........................................ 70
Figure 4. Land cover map of Burkina Faso.............................................. 71
Figure 5. Burkina Faso's 13 administrative regions................................ 73
Figure 6. Major cities and rivers.......................................................... 73
Figure 7. Reservoirs and rivers in Burkina Faso.................................... 74
Figure 8. Permanent and non-permanent waterbodies by region (Yes = permanent, no = non-permanent; far right column is total)........ 75
Figure 9. First fishing month by region.................................................. 75
Figure 10. Last fishing month by region............................................... 76
Figure 11. Dominant species 1............................................................. 76
Figure 12. Dominant species 2............................................................. 77
Figure 13. Dominant species 3............................................................. 77
Figure 14. Conversion factors from heaps to kg (number of kg per heap), by species and time-point......................................................... 81
Figure 15. Number of household heads by sex and region (unweighted)..... 82
Figure 16. Number of non-household heads by sex and region (unweighted).............................................................................. 82
Figure 17. Sex of head of household by region (1 = male; 2 = female)........ 83
Figure 18. Histograms of 100 estimates from village sample.................... 90
Figure 19. Area frame; green and yellow represent fishing sites................. 92
## Acronyms and abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPUE</td>
<td>Catch Per Unit Effort</td>
</tr>
<tr>
<td>CIAT</td>
<td>Centro Internacional de Agricultura Tropical</td>
</tr>
<tr>
<td>CWP</td>
<td>Coordinating Working Party on Fishery Statistics</td>
</tr>
<tr>
<td>DfID</td>
<td>Department for International Development</td>
</tr>
<tr>
<td>FAO</td>
<td>Food and Agriculture Organization of the United Nations</td>
</tr>
<tr>
<td>GSARS</td>
<td>Global Strategy to improve Agricultural and Rural Statistics</td>
</tr>
<tr>
<td>ID</td>
<td>Identification Number</td>
</tr>
<tr>
<td>IBGE</td>
<td>Instituto Brasileiro de Geografia e Estatística</td>
</tr>
<tr>
<td>ISSCFG</td>
<td>International Standard Statistical Classification of Fishing Gear</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Squared Error</td>
</tr>
<tr>
<td>PPS</td>
<td>Probability Proportional to Size</td>
</tr>
<tr>
<td>PSU</td>
<td>Primary Sampling Unit</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Squared Error</td>
</tr>
<tr>
<td>SAC</td>
<td>Scientific Advisory Committee</td>
</tr>
<tr>
<td>SDS</td>
<td>Statistical Development Series</td>
</tr>
<tr>
<td>SSF</td>
<td>Small-Scale Fisheries</td>
</tr>
<tr>
<td>UNSD</td>
<td>United Nations Statistics Division</td>
</tr>
</tbody>
</table>
Glossary

Area sampling frame

An area frame is a set of land elements, which may be either points or segments of land. The sampling process may involve single or multiple stages.

Bycatch

Part of the catch of a fishing unit taken incidentally in addition to the target species, or on a different size range of the same species towards which fishing effort is directed. It may be retained for human use, or some or all of the bycatch that has no commercial value may be returned to the sea as discards, usually dead or dying. See Discards.

Catch

The quantity of fish that is retained by the fishing gear during fishing operations.

Census

A fisheries census is a survey in which the value of each variable for the survey area is obtained from the values of the variable in all reporting units, which are usually fishing households. The primary objective of fisheries censuses is to provide a detailed classification of the fisheries structure of the country. It provides estimates for each household, and therefore, aggregate data for the smallest administrative, political or statistical subdivisions of the country and for classifications of households by size or other subgroups of interest.

Cluster sampling

A sampling method having the aim of reducing frame development and data collection costs. The population is partitioned into primary units (clusters). Each primary unit is comprised of secondary units, which may be listings of farms, segments of land units or points, in the case of agricultural surveys. Clusters are land areas which are defined either in administrative terms (villages, counties, etc.), geographical terms (using natural boundaries), or by using georeferenced boundaries. A sample of clusters is selected, using any sampling method, and is surveyed in its entirety. This is referred to as a “complete cluster”. However, as indicated in Food and Agriculture Organization of the United Nations (FAO) Statistical Development Series 3 (SDS), “often, the sizes of available and identifiable clusters are both too large and variable for efficient sampling.
Subsampling of the primary clusters then becomes necessary if population listing or smaller units are not available. This leads to two-stage sampling of primary selections and of elements from them” (Kish, 1989; p. 74).

**Catch per unit effort (CPUE)**

The quantity of fish caught (in terms of number or weight) with one standard unit of fishing effort, such as the number of fish taken per 1 000 hooks per day, or weight of fish, in tonnes, taken per length of gillnets. The CPUE is often considered an index of fish biomass or abundance. Sometimes referred to as catch rate, that is, a proportional change in the CPUE, is hoped to represent the same proportional change in the target species’ true abundance. A decreasing CPUE indicates overexploitation, while an unchanging CPUE indicates sustainable harvesting. However, it is known that many factors (including economics and geographical distributions) may affect the CPUE but do not represent changes in abundance. Therefore, CPUEs are often “standardized” using a variety of statistical techniques, to remove the effect of those factors which are known not to be related to abundance.

**Discards**

The portion of the total organic material in the catch (1) to be released or returned to the sea, whether or not such fish are brought fully on board a fishing vessel; or that is (2) part of the catch, but is not retained and is returned to the sea. Discard typically consists of “non-target” species or undersized specimens. While some species (clams, sea stars, etc.) might survive the process, most fish will die. See By-catch.

**Fishing effort**

The total amount of fishing activity on the fishing grounds over a given period of time, often expressed for a specific gear type: for example, number of hours trawled per day, number of hooks set per day or number of hauls of a beach seine per day. Fishing effort would frequently be measured as the product of (a) the total time spent fishing; and (b) the amount of fishing gear of a specific type used on the fishing grounds over a given unit of time. When two or more kinds of gear are used, the respective efforts must be adjusted to a standard type before being added.
**Fishing gear**

Tool used to catch fish, such as hook and line, trawl, gill net, trap, spear. The gear is classified according to the fishing license of the vessel or the owner/operator, using the Coordinating Working Party on Fishery Statistics (CWP) International Standard Statistical Classification of Fishing Gear (ISSCFG).

**Fishing vessel**

It refers to any vessel, boat, ship, or other craft that is used for, equipped to be used for, or of a type that is normally used for the exploitation of living aquatic resources or in support of such activity. This definition may include any vessel aiding or assisting one or more vessels at sea in the performance of any activity relating to fishing, including but not limited to preparation, supply, storage, refrigeration, transportation or processing.

**Frame**

The set of source materials from which the sample is selected (UNSD, 2005). It is the basis for identifying all statistical units to be enumerated in a statistical collection.

**Homeport**

It refers to the port from which fishing units operate, irrespective of where they are registered. Boat and gear activities are sampled from homeports, in contrast to catches and species composition, prices, etc., which are sampled at **Landing sites**.

**Household**

“The concept of household is based on the arrangements made by persons, individually or in groups, for providing themselves with food or other essentials for living. A household may be either (a) a one person household, that is to say, a person who makes provision for his or her own food or other essentials for living without combining with any other person to form part of a multi-person household, or (b) a multi-person household, that is to say, a group of two or more persons living together who make common provision for food or other essentials for living. The persons in the group may pool their resources and may have a common budget; they may be related or unrelated persons, or constitute a combination of persons both related and unrelated” (FAO, 2015).
Landing site

A landing site covers a certain physical area; the infrastructure in place; the technical, financial and social services available; the activities taking place and the users deriving all or part of their livelihoods from its activities.

Logbook

A record of the fishing activity registered systematically by the fisher, including catch and its species composition, the corresponding effort, and location. In many fisheries (for instance, industrial fisheries) completion of logbooks is a compulsory requirement of a fishing license. In this report, logbooks, besides catch cards and fishermen diaries, are presented as administrative sources of information that contain information on catch, effort, and possibly socio-economic variables.

Multistage sampling

A sampling method that, for agriculture, uses large geographical areas or clusters as the first stage. The final sample frame is then developed only within the selected clusters in one or more stages of sampling. In a two-stage sampling design, the clusters are subsampled and the secondary units sampled are the reporting units. In a three-stage sampling design, sampled selected units are subsampled again. Generally, a multistage sampling design is the sub-sampling (in two or more stages) of primary sampling units (clusters).

Multiphase sampling

In this type of sampling, a large sample is selected in the first phase; from this, subsamples are selected in a second phase. If a given stratification approach is too expensive to be applied to the entire population, it can be applied only to the sample obtained in the first phase (incomplete stratification). The procedure is often used for area frames of points.

Non-sampling error

Any error that may arise in the entire survey process (from frame development to data analysis) that is systematic or random and is not related to a random error in sampling. These errors include over- or undercoverage of the sample frame, errors resulting from poorly worded questionnaires, etc.
Observer

Any certified person serving in the capacity of an observer employed by the relevant management authority, either directly or under contract. Usually embarked on large fishing vessels, on-board observers are responsible for monitoring fishing operations (areas fished, fishing effort deployed, gear characteristics, catches and species caught, discards, etc.).

Primary Sampling Unit (PSU)

See Cluster.

Probability proportional to size (PPS)

A sampling procedure whereby the probability of selection of each unit in the universe is proportional to the size of some known relevant variable (OECD, 2018).

Recreational fishing

Any fishing for which the primary motive is leisure rather than profit, the provision of food or the conduct of scientific research, and which may not involve the sale, barter, or trade of part or all of the catch.

Sample survey

The collection of data from a sample of units, rather than from all the units (as occurs in a census).

Sampling error

Any random sampling method can produce several different samples that can produce a set of statistics. The sampling error is the variability in the results that are obtained from the different samples.

Single-stage sampling

Sampling scheme in which the sample is selected directly from a list of units covered by the survey.
Small-scale fisheries (SSF)

The SSF can be broadly characterized as a dynamic and evolving sector employing labour-intensive harvesting, processing and distribution technologies to exploit marine and inland water fishery resources. The activities of this subsector, conducted full-time or part-time or only seasonally, are often targeted on supplying fish and fishery products to local and domestic markets, and for subsistence consumption.
Acknowledgements

The line of research on Master Sampling Frames for Surveys of Fisheries and Aquaculture was implemented by Emily Berg and Mark Kaiser of the Center for Survey Statistics and Methodology of Iowa State University (United States of America). This document is the final technical report of the project. The report was reviewed by Gertjan de Graaf,1 Aymen Charef,2 Jennifer Gee,2 Dramane Bako3 and the members of the Scientific Advisory Committee (SAC)4 of the Global Strategy to improve Agricultural and Rural Statistics (GSARS).

The research project was supervised by Dramane Bako3 under the overall coordination of Flavio Bolliger.5 The technical reports of the project benefited from valuable inputs from Jacques Delincé,6 Tiago Almudi,7 Gertjan de Graaf,1 Jennifer Gee2 and Aymen Charef.2 Participants8 in the Expert meeting on Master Sampling Frame for Fisheries and Aquaculture, held on 11–12 May 2017 at the Food and Agriculture Organization of the United Nations (FAO) headquarters in Rome, provided valuable comments and recommendations to improve the methodology of the research.

This publication was prepared with the support of the Trust Fund of the Global Strategy, funded by the UK’s Department for International Development (DfID) and the Bill & Melinda Gates Foundation.

---

1 Nefisco, the Netherlands.
2 Fisheries and Aquaculture Department, FAO.
3 Statistician, Technical Assistance Coordinator, GSARS.
5 Senior Statistician, former Research Coordinator, GSARS.
6 Senior Consultant, GSARS.
7 Agriculture Research Analyst at the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatistica, IBGE)
8 Emily Berg (Iowa State University – United States of America), Mark Kaiser (Iowa State University – United States of America), Pierre Morand (Institut de Recherche pour le Développement – Côte d’Ivoire), Christophe Bene (International Center for Tropical Agriculture – Centro Internacional de Agricultura Tropical [CIAT] – Colombia), Tiago Almudi (IBGE – Brazil), Sébastien Djienouassi (Ministry of Livestock, Fisheries and Animal Industries – Cameroon), Philippe Tous (independent consultant – France), Jin Kim (Statistics Korea – Republic of Korea), Jacques Delincé (GSARS – Belgium), Naman Keita (GSARS – Italy), Valérie Bizier (GSARS – Italy), Jennifer Gee (FAO, Fishery Statistics and Information Branch (FIAS)-Italy), Cristina Ribeiro (FAO, FIAS – Italy), Aymen Charef (FAO, FIAS – Italy), Marc Taconet (FAO, FIAS – Italy), Dramane Bako (GSARS – Italy), Flavio Bolliger (GSARS – Italy)
Preface

The GSARS is an effort to provide statistical offices in developing countries with guidance on how to conduct surveys related to agriculture and natural resources. The FAO has hosted several GSARS research projects that address different dimensions of survey sampling. The ultimate product of each research project is a set of guidelines for a statistical office to use when preparing for a survey of agriculture or natural resources. Published guidelines cover a variety of survey-related topics, including the role of administrative data, small area estimation, censuses, and sample designs.

One GSARS project that is closely connected to this publication covers Master Sampling Frames (MSFs) for agricultural surveys. One finding of the research on frames for agricultural surveys is that the sampling methodology recommended for agriculture does not directly apply to surveys of the fishery and aquaculture sectors. Populations of interest in fishery and aquaculture surveys are distinct from populations of interest in agricultural surveys. The nature of the population elements in the fishery and aquaculture context is very different from those pertaining to agricultural settings. The resources available for constructing sampling frames are very different for surveys of fisheries or aquaculture than for surveys of agriculture.

The inclusion of aquaculture in this final report should not be taken as an indication that prescriptions appropriate for sampling of catch fisheries are necessarily appropriate for aquaculture. Aquaculture operations may bear considerable similarity to other activities associated with traditional agriculture, or to mobile fishing operations. Sampling for aquaculture may require approaches described herein, or may require approaches stemming from agricultural surveys.

This final report presents guidance for constructing an MSF to use for surveys of fisheries or aquaculture. An MSF is a coherent structure for implementing surveys of a particular sector over time. This final report is not prescriptive; rather, it brings together issues for a statistical office to consider when developing a strategy for designing an MSF. Survey planners are advised to consider how the issues and recommendations presented in this final report relate to the objectives and context of their country.
Overview of previous reports

The process of producing this final report involved writing four intermediate technical reports. Technical report 1 (Global Strategy, 2017a) reviews literature on surveys of fisheries and aquaculture. The literature review highlights concepts with applicability to developing and emerging economies, where SSF present unique challenges. Technical report 2 (Global Strategy, 2017b) analyses the interplay between the sampling frame and the overall survey process, explaining the role of the sampling frame in each step of the survey process. Technical report 2 also identifies methodological issues associated with surveys of fisheries and aquaculture that are underdeveloped in the literature. Technical report 3 (Global Strategy, 2017e) lays the foundations for several possible analyses to conduct for the desktop-study component of this project. Several possible approaches are outlined, a preliminary data analysis is described, and available resources for conducting a complete desktop study are evaluated. Technical report 3 also reflects on feedback received at an expert meeting held in Rome in May 2017. Technical report 4, titled Master sampling frames for fisheries surveys: desktop study using data from Burkina Faso, which is annexed to this report, describes the outcome of the desktop analysis. Using data from Burkina Faso, four analyses are conducted that address different issues associated with sampling frame construction. The selected issues include updating a list of fishing sites, incorporating area frame information through a multiple-frame approach, conducting a multistage sample design, and establishing conversion factors between local and standard measurement units. These particular aspects of the frame problem are selected because they genuinely relate to issues arising in data collection programmes in Burkina Faso. These four reports culminate in the current final report, which synthesizes the material from the previous four reports.
Executive summary

Data-driven policies and management programmes for surveys of fisheries and aquaculture require reliable information. Probability sampling is a scientifically defensible tool for obtaining information about a target population. Estimators from probability samples have well-understood statistical properties. One aspect of conducting a probability sample involves forming a sampling frame. Simply put, a sampling frame is a list from which the sample can be selected.

In practice, many subdivisions of a particular sector are of interest and different policies may benefit from information on different types of variable. This is particularly true in surveys of fisheries and aquaculture. Management programmes benefit from information on production-related variables (that is, catch and effort), demographic information (that is, education level, household size), and economic data (income, costs). The fishery and aquaculture sectors are composed of diverse subpopulations, ranging from industrial marine vessels to small communities of nomadic fishers.

Because of the diversity of populations and variables of interest, a single list is likely to be inadequate as a sampling frame to support all survey goals. For example, a list of households may be well suited to measuring socio-economic information about small-scale fishers, while a list of vessels may be more appropriate to collect production-related data about industrial fisheries. Similarly, distinct administrative processes for fisheries and aquaculture may result in a need to utilize multiple sources of information.

An MSF is a collection of lists that support surveys of a particular sector over time. Four important types of frame are reviewed in this document. First, a direct enumeration of units is an explicit data collection effort that results in a listing of units of interest, along with auxiliary information. Second, administrative data are lists constructed for non-statistical purposes. Third, previous surveys or censuses may serve as the frame for a subsequent fishery or aquaculture survey. Fourth, a spatio-temporal representation of fishing areas over time can provide a frame for on-site data collection or aerial surveys. The overall MSF strategy may include one or more of these types of frames. The choice of the MSF will depend on the financial resources available, existing data sources, and the nature of the populations under study.

In addition to specifying the relevant lists, defining an MSF strategy involves specifying mechanisms for using, updating and combining relevant sources of information. Thus, the role of the MSF is intimately connected to all aspects of
the survey process. The objectives of a particular programme play an important role in guiding the choice of the MSF. In turn, the MSF has implications for which sample designs and estimators are possible.

This publication provides guidance for a statistical office to use when forming an MSF. The goals of this document are to (1) describe a process for building, maintaining and using MSFs; (2) discuss interrelations among survey objectives, core data elements, construction of sampling frames, estimation, and archival of results; and (3) offer guidance for using multiple sources of information. The statistical office will need to consider how the directives presented in this document apply to a particular situation. When deciding on an MSF approach, the statistical office should consider the ultimate uses of the collected data and the desired estimates.
Background and objectives

Is fishing important in your country? The answer is probably “yes.” The FAO 2018 edition on the State of World Fisheries and Aquaculture highlighted the critical importance of fisheries and aquaculture for the food, nutrition and employment of millions of people. Total fish production in 2016 reached an all-time high of 171 million tonnes, of which 88 percent was utilized for direct human consumption. The fishery and aquaculture sectors provide more than 20 percent of the average per capita animal protein intake for 3 billion people (more than 50 percent, in some less developed countries) and is especially critical for rural populations, which often have less diverse diets and higher rates of food insecurity. Estimates of fish consumption per capita reached a record of 20.3 kg per person in 2016. It has grown steadily in developing regions from 6 kg in 1961 to 19.3 kg in 2015. (FAO, 2018). An estimated 119 million people depend directly on fisheries for their livelihoods, of whom 96 percent live in developing countries. SSF alone employ an estimated 109 million people and contribute to 45 percent of global fish production – 48 million megatonnes (de Graaf et al., 2011).

Sustainable management of fishing resources is crucial if the fishery and aquaculture sectors are to remain viable. Effective management strategies require reliable data on diverse aspects of fisheries and aquaculture operations. As explained in de Graaf et al. (2015), “[i]n practice, fisheries assessments should always combine biological, economic, sociocultural, and compliance indicators to guide management decisions…” (de Graaf et al., 2015). Further, the fishery and aquaculture sectors are composed of diverse subpopulations, each with unique characteristics. Policies targeting fishery and aquaculture sectors require scientifically defensible estimates for a wide variety of populations and parameters.

Probability sampling is a scientifically defensible data collection method. In probability sampling, every element of the population has a chance to be included in the sample, and the probability of being included is known for every sampled element. As discussed in Pinello, Gee and Dimech (2017), using data from a probability sample can avoid biases that can arise when data are collected from non-probabilistic data collection methods, such as convenience samples. Common types of data from non-probability samples in fishery or aquaculture contexts are logbook data, vessel monitoring systems or local experts. Selecting
a probability sample typically requires a sampling frame – a representation of the target population from which a sample can be selected. Obtaining a sampling frame that contains all population elements of interest can be expensive and challenging. When lists become outdated, they can exclude eligible population elements or include ineligible elements. Combining multiple lists is often necessary to cover the entire target population. The choice of the sampling frame relates to multiple aspects of the survey process, including data collection, sample design and estimation.

1.1. Need for reliable frames for fisheries and aquaculture

Reviews of data collection programmes for fishery and aquaculture (de Graaf et al., 2011) identified the absence of a reliable and up-to-date frame as the major challenge in developing reliable data collection programmes. In a review of fishery data collected from 28 countries, de Graaf et al. (2011) conclude that the “main priority is to determine the best sample frame”. Administrative registers and censuses of fisheries are two types of resource for constructing sampling frames. Administrative procedures may furnish lists of fishing vessels or licenses. A fishery census may have been conducted to enumerate major fishing sites, fishing vessels or fishing gear. Registers and fishery censuses, however, have important limitations. Administrative processes often only cover commercial fisheries and large-scale fisheries. Fishery censuses are expensive to conduct and the resulting lists thus often become outdated.

Registers and fishing censuses can be particularly problematic as frames when SSF are of interest. SSF in developing countries is an important sector to understand, as recent studies estimate that SSF in developing countries accounts for almost 40 percent of global fish production. Collecting data for SSF in developing countries is challenging because these populations tend to be geographically dispersed, highly variable, located in regions that are difficult to reach, and isolated from government processes. As a consequence, current estimates of the contribution from SSF to global fish production are thought to be underestimates. Part of the problem is that administrative registers and censuses of vessels or fishing sites can miss important segments of SSF populations. Therefore, de Graaf et al. (2011) recommends using resources outside the fishery and aquaculture sectors to construct sampling frames for the purpose of targeting small-scale fishers. This same advice applies to nomadic fishers who travel to different fishing locations throughout a year or season. Resources for sampling small-scale or nomadic fishers may include population censuses or geographic representations of fishing areas.
1.2. Sources of frames for fishery and aquaculture surveys

For readers unfamiliar with sampling frames, this section briefly introduces the types of frame that may be of use for surveys of fisheries and aquaculture. These are discussed in detail in section 5, but refer to the terminology explained in sections 1-4. The frames are classified as follows.

- **Direct enumeration of operations** involves an explicit effort to list all operational units that compose the population of interest. In the fisheries literature, a direct enumeration is often called a “frame survey”. We prefer the term “frame census” to avoid confusion with a sample survey. In the context of fisheries, a frame census may involve enumerating all landing sites or enumerating all vessels that compose a fleet. In aquaculture surveys, a direct enumeration might involve constructing a list of all aquaculture facilities.

- **Administrative lists** can provide lists of individual people, lists of operations or lists of fishing sites. Sources of administrative lists include logbooks, licenses or registration in government programmes.

- **Other surveys or censuses** can serve as frames for subsequent surveys of fisheries and aquaculture. The most important types of survey or census considered here are previous population or agricultural censuses. Using a population census is typically preferred relative to an agricultural census for studying fishery populations, because the agricultural census can fail to cover important segments of these populations.

- **Spatio-temporal** frames consist of units that are geographic areas, time-points, or intersections of geographic regions and time-points. These types of frame typically support data collection schemes that take place at the location where fishing occurs, or within the context of aerial surveys.

Frames are often classified into two broad categories: list frames and area frames. List frames are typically thought of as physical listings of a population, such as a phone book or a business register. Area frames are geographic representations of the population, such as a map. In fact, the line between area frames and list frames is blurry, as an area frame can be viewed as a particular type of list frame. For instance, dividing an area into a collection of geographic segments results in an enumeration of segments; that is, a list frame of segments. The subdivision of frames into categories of list frames and area frames is
relatively unimportant for the purposes of this final report. In case readers are familiar with the “list frame” and “area frame” terminology, the types of frames listed above are aligned with the standard terminology. The first three types of frame listed above are typically associated with “list frames”, while the last category is typically considered an area frame. Section 5 further discusses a type of frame called an aerial frame, which is considered a special type of a spatio-temporal frame.

1.3. The MSF concept

Developing a comprehensive frame for probability surveys of fishery and aquaculture populations is complex. The fishery and aquaculture sectors consist of multiple subpopulations, ranging from commercial vessels to individual nomadic fishers. Data collection programmes may target diverse variables of interest, including catch, effort and socio-economic attributes of fishers. As a further complication, even the relevant definition of a population unit depends on the objectives of the data collection program. For studies of catch, the unit may be a fishing vessel. For socio-economic studies, the unit may be an individual fisher. Because of this complexity, a single list may be inadequate as a sampling frame. The concept of an MSF brings together the collection of lists that may be necessary to construct sampling frames for surveys of fisheries and aquaculture over time.

An MSF is a permanent institution for administering data collection programmes over time. In relatively simple cases, the MSF may be a single list that is maintained over time. In more complex situations, the MSF may be a collection of multiple lists. When developing an MSF, the analyst should consider mechanisms for updating each list and for linking multiple lists together.

As a formal definition of an MSF, the definition proposed by Carfagna (2013) is adopted here: “[a] master sampling frame is a sampling frame that provides the basis for all data collections through sample surveys and censuses in a certain sector, allowing to select samples for several different surveys or different rounds of the same survey, as opposed to building an ad-hoc sampling frame for each survey. The aims of the development of a master sampling frame are: avoiding duplication of efforts, reducing statistics discrepancies, connecting various aspects of the sector, allowing the analysis of the sampling units from the different viewpoints, and having a better understanding of the sector.” See Martinez (2013) and Tuner (2003) for further discussions of the concept of an MSF.
In summary, an MSF is a structure, or collection of structures, that cover target populations of interest, along with a set of procedures for using these structures. The structure may be a list (list frame), such as a list of vessel licenses or addresses. The structure may be a spatio-temporal description of a target population (area frame), such as the two-way classification defined by geographic locations and time-points. The choice of the MSF depends on the nature of the populations under study and the variables of interest for the survey. The MSF in turn has implications for the sample designs and estimation procedures that are possible. The MSF is therefore intimately tied to all aspects of the survey process.

1.4. Sampling frames and the survey process

The overall approach to a survey of fisheries or aquaculture (as with any survey) involves several components. In designing a survey, an approach for each component must be selected. The nature of the frame is central to many aspects of the survey process. The objectives may guide the choice of the frame. Subsequently, the choice of the frame has implications for data collection, sample design and estimation. As a prequel to future sections of this report, which will discuss implications of the frame for the survey process in more detail, the following list briefly reviews how the frame relates to four aspects of the survey process:

- **Defining objectives.** The first part of defining objectives involves establishing broad survey goals. After these are established, primary variables of interest, along with units of measurement, must be identified. Defining the target population and population elements is also a part of defining the objectives. If the population element is a fishing vessel, then a register of vessels or a census of fishing sites may provide an important contribution to the overall MSF. If the unit of interest is a household, then deriving a sampling frame from a population census may be considered an appropriate approach.

- **Designing the sample.** The sample design is driven partly by the structure of the frame. Available auxiliary information to use in design depends on the auxiliary information available in the frame. Multiple stages of selection may be needed if the ultimate population elements are grouped into “clusters” on the frame.
• **Developing and implementing a data collection protocol.** The choice of the data collection procedure may guide and be limited by the choice of the frame. Different data collection procedures can introduce different sources of non-sampling errors. Non-sampling errors are selection or measurement errors arising from factors other than random sampling. An example of measurement error is recall error – that is, when a fisher forgets the exact catch. As the choice of the frame is related to the data collection protocol, the frame has implications for the types of non-sampling errors that are likely to arise.

• **Estimating parameters of interest.** Estimation procedures should be compatible with the sample design. They may also account for non-sampling errors that arise, for example, due to overcoverage or undercoverage of the sampling frame.

### 1.5. Relevance and objectives of this research project

FAO and GSARS recently published three reports with connections to this publication: (1) the Handbook on Master Sampling Frames for Agricultural Statistics (GSARS, 2015a); (2) the Handbook for Fisheries Socio-Economic Sample Survey (Pinello, Gee and Dimech, 2017); and (3) Guidelines to Enhance Small-Scale Fisheries and Aquaculture Statistics through a Household Approach (GSARS, 2017d). Given these other publications, one may ask why a separate document on MSFs for fisheries is necessary.

The principles for constructing MSFs for agriculture do not apply readily to the fishery and aquaculture contexts. An important reason for this relates to the concept of a unit, the fundamental building block of a sampling frame. In agricultural surveys, the relevant units are composed primarily of the farm, the plot and the household. In surveys of fisheries, the possible units and interrelations between them are more numerous and more complex. Even if the focus lies only on surveys of catch or effort, many possible units exist. The sampling units may be intersections of time-points (that is, date in the month) and geographic region. The sampling unit may be a landing or an individual vessel. For socio-economic surveys, the relevant unit may be a household or a fisher. The relevant unit, and hence appropriate choice of frame, for a fishery survey can vary greatly depending on the objective of the data collection program.

In considering the existing guidelines for agricultural surveys, it is worthwhile to note that while fishery and agriculture populations overlap, fishery populations are not completely contained in agricultural populations. As
explained in (Pinello, Gee and Dimech, 2017), small-scale and the poorest fishing communities often do not have any holdings, but are, rather, landless households.” Therefore, an agricultural census is unlikely to provide a complete frame for a survey of fisheries.

The objectives of the current research are broader than the objectives of GSARS (2017d) or Pinello, Gee and Dimech (2017). GSARS (2017d) provides a comprehensive treatment of variables to consider collecting in surveys of fisheries and aquaculture. They distinguish between core and supplemental items. They also provide guidelines on how to construct questionnaires that solicit information about these variables. The guidelines on a census-based framework, however, give relatively little attention to the survey process. The current report is more statistically oriented than GSARS (2017d). Rather than focus on variables of interest, this publication focuses on the survey process and the role of the sampling frame in that process. The scope of the guidelines on socio-economic surveys of fisheries is narrower than the scope of the current work. Pinello, Gee and Dimech (2017) focuses on a relatively unique situation with two main characteristics. First, the primary variables of interest are related to socio-economic aspects of the fishery. Second, a comprehensive and up-to-date list of vessels exists and defines the relevant target population. This final technical report expands on Pinello, Gee and Dimech (2017) by considering a broader class of variables and populations. As such, a broader class of frames than simply vessel registers is considered.

This report has three main goals:

1. Describe a process for building, maintaining, and using MSFs.
2. Discuss interrelations among survey objectives, core data elements, construction of sampling frames, estimation and archival of results.
3. Provide guidance on using multiple sources of information.

The most salient section of this final technical report is section 5, which reviews types of lists that can constitute an MSF. Sections 2 to 4 lay the necessary foundations, discussing categories of variables, types of populations and the survey process. Section 6 concludes with a synthesis that proposes a possible procedure for developing a master sampling plan.
Defining survey goals

The first step in the survey process involves defining the survey objectives. Questions to consider when defining the survey goals are: “What are the needs of the data users?” and “What estimates are desired in the final report?” One may also consider gaps in existing data or data collection programmes. Estimates are often composed of indicators that synthesize multiple variables. FAO (2015) provides a thorough list of core and supplemental variables to consider collecting in surveys of fisheries and aquaculture. Pinello, Gee and Dimech (2017) focuses on variables related to socio-economic attributes of fishers. Possible survey goals are classified into four broad categories and a brief overview of each is provided.

Biological and environmental variables are omitted from the discussion below. While these variables are obviously important for understanding the health of the ecosystem, they are often collected through dedicated studies. This final technical report is geared, rather, towards surveys of production and socio-economics rather than of environment and biology.

2.1. Fisheries production

**Catch and effort** are the primary variables involved in determining the total production of a fishery. In this context, catch refers to harvest, or kept catch. Multiple measures of catch and effort exist. Three measures of catch are weight, volume and number of fish. Less standard (local) measures of catch are baskets or cartons. When a landing site does not have a weighing station, using local units might be necessary. Conversion factors may be established to convert the measurement of catch from local units to standard units (for instance, converting heaps to kilograms (kg): see case study in annex). Effort can be measured in terms of the time spent fishing, the number of trips, or the number of boats. The appropriate measure of effort often depends on the gear type. Decoupling the time spent fishing from time engaged in more general fishing-related activities, such as travel to the fishing ground or time searching for fish, can be difficult.

When deciding on a measurement strategy, obtaining an interpretable and feasible index should be the goal. It is recommended to use measures that are consistent temporally and geographically, and are relatively easy to obtain. Even if somewhat inaccurate, a consistent index may allow for more meaningful
analyses than highly accurate but expensive measurements. As discussed in section 2.5 below, verification studies can be used to assess relationships between accurate measurements and less expensive indexes.

Catch and effort are often combined to form an index called catch per unit effort (CPUE), the ratio of catch to effort. Monitoring CPUE over time is a common way to assess trends in stock size. Declines in CPUE can indicate a decrease in stock size. When interpreting CPUE as an indicator of relative stock size, it is important to recall that both catch and effort vary greatly across space. Variations in fishing patterns across space can account for observed changes in CPUE that do not actually reflect variations in stock size. Caution is particularly important when comparing CPUE estimates at small levels of geographic or temporal detail.

Breaking down measures of catch, effort or CPUE by species or gear type is advised in de Graaf et al. (2015). When sufficient data are available, detailed analyses by species or gear type can provide a more nuanced picture of the nature of the harvest or of the effect of fishing on stock size. When defining more detailed divisions of the data, it is important to recall that sample sizes become smaller and the variability of estimators increases.

**Discard and Bycatch** are additional variables involved in determining total mortality due to fishing operations, for both target and non-target species. Like CPUE, discard and bycatch can be used to assess the impacts of fishing on the general health of an ecosystem. For instance, discard or bycatch of endangered species can suggest a need to change fishing practices. Observer programmes aim to provide data related to the health of marine ecosystems. Zollett et al. (2015) and Brooke (2012) discuss discard and bycatch observer programmes in North America and Europe. Dietrich (2011) provides a training manual targeted for West African fisheries.

**Disposition** refers to how the fisher uses or sells the kept catch. Examples of possible types of disposition are export, market, commercial processing, barter or trade, direct consumption, or bait. Disposition categories will depend on the sector of the overall fisheries activity of concern.
2.2. Aquaculture production

**Growth/production.** In aquaculture, gross versus net production may be more of an issue than with fisheries.

**Disposition** issues for aquaculture are similar to those for catch fisheries; however, domestic versus export may be more important.

2.3. Social/demographics

**The demographic** characteristics of individuals involved in fisheries or aquaculture operations may be of interest. Examples of possible demographic variables include sex of head of household, level of education of fisher, household size, and age.

**Employment** has at least two components. One measures the extent to which fishing or aquaculture contributes to the total income of the individual. Some individuals involved in fishing may be part-time fishers, while others are full-time. The second identifies employment opportunities outside the fishery or aquaculture sectors that are available to the fisher. These employment opportunities may depend on the level of education of the fisher, proximity to a large city, and agricultural productivity of the area where the fisher lives.

**Cultural** aspects of interest may include attitudes toward government, traditions that may have an impact on when fish are caught, or participation in policy development. Understanding cultural attitudes can be important for developing policies that fishers are likely to follow and that officials are able to enforce.

2.4. Economics

**Remuneration or earnings** is “cash and in kind paid to employees, as a rule and at regular intervals, for time worked or work done.” Remuneration provides information on earnings of employees and profits generated by owners. Pinello, Gee and Dimech (2017) argues that remuneration is among the most important economic variables to measure in a survey. Measuring remuneration involves measuring numerous economic components, including costs and inputs.

**Costs** can be classified as fixed costs, variable costs, and investments. Pinello, Gee and Dimech (2017) provides a detailed discussion of different types of costs for fisheries and their implications for understanding remuneration.

**Investments** may include purchases of engines, gear or other equipment. The market value of the vessel (excluding the license) contributes to total
investments. Investments may be measured annually, as they change more slowly than certain other economic variables.

**Income and profits** measure the earnings of the fishers and of the owners of the fisheries. In some cases, fishers’ earnings are a share of the catch in proportion to the profit.

### 2.5. Verification and audits

A number of surveys may be conducted for the purpose of justifying or verifying the use of an alternative, easier or cheaper survey. For example, a survey based on intercept sampling of marine boats for catch may be used to verify that logbook information provides a reliable source of data for estimation of total catch. Similarly, a direct survey of cage culture operations in a given region might be used to verify the efficacy of adding questions on aquaculture to a household survey relative to production of one or more species. FAO (2017) contains an argument in support of verification surveys, noting that they may be used to simplify measurements of catch in surveys of SSF. They suggest accompanying data from logbooks with information from periodic verification surveys. In this process, “information to be collected by the fisher is kept at a minimum, but with a higher level of information which can be validated.”
Dividing the population

Any number of divisions of both catch fisheries and aquaculture have been proposed, for various uses. Divisions of the population are often seen only as potential stratification variables; however, they should also be considered in terms of the impact on data collection methods. That is, a division based on the operational units of a fishery would seem useful for designing and conducting a survey, because it is operational units that will need to be dealt with in any direct sampling efforts. The divisions suggested are certainly related to size and target species; however, they are also related to the types of sampling designs that are effective. Readers are also referred to GSARS (2017d) for definitions of types of subpopulation of interest in surveys of fisheries and aquaculture.

3.1. Fisheries

One possible division of a population of fisheries is presented in the table below. The subdivision in the table below is based on gear category. Depending on the context, the classification may be further refined to include size classes (see Pinello, Gee and Dimech, 2017).

<table>
<thead>
<tr>
<th>Type</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marine vessels</td>
<td>Moderately large to large, trip length may be multiple days, often trawls or longlines</td>
<td>Tuna, groundfish</td>
</tr>
<tr>
<td>Marine boats</td>
<td>Smaller size, shorter trips, small trawls, hook and line</td>
<td>Shrimp, snapper, crab</td>
</tr>
<tr>
<td>Marine diver</td>
<td>Scuba, spearfishing</td>
<td>Reef fish, rays</td>
</tr>
<tr>
<td>Inland boat</td>
<td>Small boat, mobile, short trips, intermittent use, mostly handheld gear or traps</td>
<td>Freshwater finfish, catfish, eels</td>
</tr>
<tr>
<td>Inland or near shore</td>
<td>Individuals, hand-held gear such as traps/hook and line/hand-set trot lines</td>
<td>Freshwater fish, inshore marine fish (seabass, redfish, etc.)</td>
</tr>
</tbody>
</table>

Source: Authors elaboration
The division of fisheries on the basis of gear type is advantageous because it avoids a need to use ambiguous classifications of fisheries, such as “artisanal” and “small-scale.” Obtaining a universally applicable definition of “artisanal” or “small-scale” fishing is notoriously difficult: rather, many such definitions exist. The classification above implicitly encompasses small-scale and nomadic fishers who may use modest tools, such as a hook and line or an inland trap.

3.2. Aquaculture

It is proposed to divide aquaculture populations into three categories: (1) semi-permanent structures; (2) movable structures; and (3) open water. Semi-permanent structures are operations that are not easily moved to other locations, such as both fresh- and seawater ponds, tanks, raceways and large cages. Some moveable structures may be moved annually and others, such as surface and subsurface lines for shellfish (such as oysters, scallops and mussels), racks and small sea cages, are moved on even shorter time frames. Open water refers to sea farming involving the seeding” of open water with no containment of organisms – such as sea cucumbers or scallops – followed by harvest, much as for capture fisheries.

Figure 1. Categories of aquaculture populations.

<table>
<thead>
<tr>
<th>Aquaculture populations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semi-permanent structures</td>
</tr>
<tr>
<td>Fresh- and seawater ponds, tanks, raceways, large cages...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Movable structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface and subsurface lines for shellfish, racks, small sea cages</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Open water</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Seeding&quot; of open water with no containment of organisms</td>
</tr>
</tbody>
</table>

*Source: Authors elaboration*
3.3. Spatial and temporal contexts

Surveys of fisheries or aquaculture involve the need to define relevant domains for both temporal and spatial dimensions. Spatial divisions may be political (region, province, county, administrative district, etc.) or ecological (ecotypes). Temporal divisions (that is, quarter, year) depend on the survey goals. Pinello, Gee and Dimech (2017) provides guidelines on temporal divisions for economic surveys. Calendar divisions may differ from reference periods of interest when the goal of the survey is monitoring over time or change detection.
Aspects of the survey process that relate to frames for fisheries and aquaculture

The choice of the frame is tied to all aspects of the survey process. In practice, it is not possible to consider which frame to use until the survey problem as a whole has been considered. The survey process is divided into five main components. For each, the following subsections provide a basic overview, discuss issues to consider in the context of fishery and aquaculture surveys, and examine the relation to the problem of developing an MSF.

4.1. Identification of the target population

The target population is the collection of elements of interest for the survey. Section 3 discusses possible subpopulations of the fishery and aquaculture sectors. The target population may be composed of a single subdivision or a collection of multiple subpopulations. The target population is the collection of all population units of interest. A distinction is drawn between a population unit and a sampling unit.

A population or observation unit is the unit on which a measurement is taken. Examples of population units include a fisher, a landing, a vessel, a time-point, or a household. In certain socio-economic studies, a population unit is a community, such as a village.

A sampling unit is the unit selected from the frame to be included in the sample. A sampling unit may be a single or a collection of population units or may be associated to one or many population units.
4.1.1. Special considerations for surveys of fisheries and aquaculture

Surveys of fisheries and aquaculture often have multiple stages. Consequently, a single survey can have multiple population units and multiple sampling units. In multistage surveys, the definition of a sampling unit is different at different survey stages. Similarly, a single survey can have multiple levels of population units. For example, certain variables may be measured at the level of a household (for example, does the household have access to clean drinking water?), while other variables are measured for individual fishers (for example, what is the highest level of education for each fisher in the household?).

Identification of population and sampling units is a process that is intimately linked to the data collection process used. Suppose it is determined that trips by marine vessels are the population units, and intercept sampling at landing sites is to be the data collection scheme. In this case, the landings may be considered the sampling units, while the marine vessels accessed through each landing would be the population elements on which the measurements are taken. Suppose it is determined that landing sites by time are primary sampling units (PSUs). Some data may be collected for individual inland boat trips, while other information is obtained for each angler aboard an individual boat trip. In this case, the survey has two levels of population elements: boat trips and individual anglers. Estimates for the population of boat trips or estimates for the population of individual anglers may be constructed within the same survey. Breidt et al. (undated) discusses an example in the context of recreational fishing surveys.

Spatial or temporal units may be considered strata or may be considered part of the definition of a population or sampling unit. If the entity sampled at any stage of the survey is a space by time intersection, then the space by time intersection is a sampling unit. If a collection of units is selected from each of a set of mutually exclusive and exhaustive space-by-time intersections, then each space-by-time intersection forms a stratum.

4.1.2. Relation to the sampling frame

The frame can be thought of as the best possible representation of the target population. The frame should contain all population elements. An ideal frame would contain no ineligible elements and could cover all eligible elements.

The types of frames available can dictate the definitions of the sampling units. For instance, the population units may be individual fishing trips. Even if a list of fishing trips is not available, a list of intersections of landing sites and timepoints could be constructed. In this case, a sampling unit may be defined as an
intersection of landing sites and time-points. Then, the population units may be accessed – the fishing trips – at each randomly selected sampling unit.

The definitions of the population elements can also have implications for the types of frames that are most desirable. If individual fishing trips are population elements, then a frame defined to be a collection of intersections of time-points and landing sites may provide a practical mechanism through which to access the population elements. If households are population elements, then using a frame constructed from a previous population census may be a good approach.

### 4.2. Measurement and variables

The goals for estimation and inference (section 2) should drive the selection of variables to measure or observe; however, a simple correspondence does not exist. A careful evaluation of options and the feasibility of obtaining different measurements is required for each specific context. Many variables for the same construct exist, and the choice will depend on the sector of the population to be observed.

#### 4.2.1. Special considerations for surveys of fisheries and aquaculture

As alluded to in section 3, there is often a trade-off between measures that provide an accurate reflection of fishing effort, and measures that provide an index of effort that can be consistently measured, even if they are not totally accurate. The goals identified for estimation clearly have considerable influence in determinations of the most appropriate quantity to measure.

#### 4.2.2. Relation to the frame

The variables of interest, as determined by the survey goals, have important implications for the choice of the frame. Common types of frames for surveys of socio-economic variables are lists of households, communities, or telephone numbers. On the other hand, spatio-temporal frames are more common for surveys of production-related variables, such as fishing catch and effort.
4.3. Data collection and contact methods

Surveys of fisheries and aquaculture employ diverse data collection schemes. The contact method (the avenue through which the data collector reaches the population element) is an important component of the overall data collection strategy. In many cases, the choice of the sampling frame is an important factor in determining the contact method.

Contact methods for surveys of fisheries may be divided into two broad categories: on-site (direct) methods and off-site (indirect) methods. In on-site surveys, the data are collected at the location where fishing occurs. Off-site methods include telephone surveys, household surveys, and data collected from logbooks.

GSARS (2017a) provides a relatively detailed account of contact methods used for surveys of fisheries and aquaculture. They discuss the distinguishing characteristics of several contact methods, provide examples of their use, and discuss implications for non-sampling errors. An overview of important contact methods, similar to GSARS (2017b), is given below.

**Onboard observers** are trained to collect data on vessels during fishing trips. Onboard observers typically collect data on bycatch and discard in industrial, marine fisheries. Onboard observers sometimes collect data related to total catch or effort; however, these are typically considered secondary to discard variables.

**Port agents and dockside samplers** intercept vessels at landing sites after completion of the fishing trip. This type of intercept sampling is applied primarily in marine environments. Intercept sampling is often judged reliable for collecting variables related to catch and effort. Disposition or first sale price can also be collected through intercept sampling. Dockside monitors can collect socio-economic variables; however, these are typically considered secondary.

**Roving creel surveys** are a type of contact method where the creel clerk contacts fishers during the fishing trip by boat or by foot. Catch rate or CPUE is the most common variable collected in roving creel surveys. In some data collection methods (that is, roving creel) the probability of selection is related to fishing effort. Estimation for such surveys requires care. Estimators that do not account for the unequal selection probabilities may be biased. On the other hand, estimators that use time spent fishing directly as the selection probability can have large variances due to small values (Robson, 1961; Hoenig et al., 1997; and Hoenig et al., 1993).
**Aerial surveys** collect instantaneous counts of fishers or boats within selected segments viewed from the air. The data may be collected from an airplane flying over the selected segment, from an aerial photograph, or from a satellite image. Aerial surveys have the advantage of providing a measure of effort that is not subject to recall bias. Visibility bias, however, can be important for aerial surveys.

**Logbooks, diaries and catch cards** are administrative sources of information that contain information on catch, effort and possibly socio-economic variables. More generally, this data collection method involves gleaning information from existing data sources. Most commonly, the data source is a type of administrative data file. Logbook data are inexpensive to collect because the information is recorded for the purpose of administration of or compliance with administrative processes. Because logbook data are typically self-reported, they can be subject to various forms of intentional or non-intentional misreporting.

**Household surveys** contact the fisher at the location where he or she lives. The individual may be contacted via post, the telephone, or an in-person interview. Primary variables collected in household surveys are related to social or demographic topics. Recall and measurement errors are sometimes a concern when trying to collect catch or effort data through household surveys. Misreporting can also occur from household surveys, particularly when the individual contacted is not the actual fisher.

Some surveys, called complemented surveys, use more than one contact method. Complemented surveys take advantage of the strengths that different contact methods have for different purposes. For instance, an aerial survey may be used to measure effort, while a landing site survey is used to measure catch and a household survey used for sociodemographical data. As another example, logbook data can be used in combination with an intercept or sampling approach through verification surveys, as discussed in section 2.5.

**4.3.1. Special considerations for surveys of fisheries and aquaculture**

Different contact methods are typically associated with different types of non-sampling errors in surveys of fisheries and aquaculture. Recall bias and misreporting are concerns when obtaining catch or effort from self-reported data. Serious measurement errors in estimators of catch or effort are widely documented when production data are collected from household surveys. Possible sources of measurement error include recall bias and correlations between fishing avidity and the probability of responding. Inaccurate data can
also arise if the individual contacted is not directly responsible for record-keeping activities. Direct contact methods are often judged to be more reliable than off-site methods for measuring production-related variables because a trained data collector (ideally) makes an objective determination (for instance, by weighting catches on landing site). To minimize the complexity of logistics and the potential for introducing non-sampling errors during on-site data collection, the questionnaires for on-site methods should be as simple as possible. This is another reason why socio-economic data are usually collected off-site rather than through on-site methods. Sometimes, contact information is collected during the on-site study and more detailed data related to socio-economic variables are collected during a follow-up survey.

4.3.2. Relation to the sampling frame

The data collection method is closely tied to the choice of frame. Certain contact methods are typically used with certain types of frames. To make the point clear, note that an aerial survey would obviously not be conducted using a frame of telephone numbers. The diagram below depicts typical associations between frames and contact methods.

**Figure 2. Correspondence between contact methods and frames.**

<table>
<thead>
<tr>
<th>Contact Method</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-board observer</td>
<td>Spatio-temporal</td>
</tr>
<tr>
<td>Dock-side monitor</td>
<td>Frame census of fishery</td>
</tr>
<tr>
<td>Roving creel</td>
<td>Administrative data</td>
</tr>
<tr>
<td>Aerial</td>
<td>Census of population</td>
</tr>
<tr>
<td>Logbook/administrative record</td>
<td>List of villages</td>
</tr>
<tr>
<td>Household</td>
<td>List of households obtained otherwise</td>
</tr>
</tbody>
</table>

*Source: Authors elaboration*

4.4. Sample design

In fishery and aquaculture surveys, as with any survey, one should aim to design the sample to achieve the desired level of precision for population parameters of interest. In other words, the objectives of the survey should guide the choice of the sample design. The sample design may include stratification, unequal
probabilities of selection, or clustering. GSARS (2017a) reviews sample design issues for surveys of fisheries. The following subsections highlight certain components that are particularly important in surveys of fisheries or aquaculture, and discuss connections between design issues and the choice of the sampling frame.

4.4.1. Special considerations for surveys of fisheries and aquaculture

Stratification – In surveys of fisheries and aquaculture, it can be tempting to define detailed subdivisions of the population based on variables such as gear type, length class and geography. Before defining a deep stratification, it is important to recall the objectives of the survey. What estimates are of interest? What estimation domains are of interest? If the objective of the survey is to produce estimates for an overall population mean or total, then strata should be as homogenous as possible relative to the study variable of interest. If estimates for subdomains are of interest, then aligning the strata with the estimation domains to some extent is sensible.

Cluster and multistage samples – Fishery populations are often arranged hierarchically, making cluster or multistage samples natural. For instance, landings are nested in landing sites, and anglers are nested in landings. Variance estimation for cluster and multistage samples ideally account for all stages of selection; however, the most important variance component typically comes from the first stage.

Multiphase samples – In a multiphase sample, variables collected in a first survey determine the sample design for a subsequent survey. Multiphase samples are useful if an inexpensive measurement is a good proxy for the variable of interest. In a multiphase survey, the selection probabilities for the second phase are random, as they depend on the outcome of the first phase survey. If a population survey is used as the frame for a subsequent fishing survey, then the ultimate design may be a two-phase design. In this case, estimators and variance estimators should be constructed to appropriately account for the two-phase sampling.

Unequal probabilities of selection – In unequal probability sampling, units assigned a relatively large size measure are selected with a higher probability. Unequal probability samples are efficient for estimating totals of a variable correlated with the size measure. Measures of fishing effort may be used to select locations or time periods with higher fishing activity with a greater probability.
Administrative data on the size of an aquaculture operation can also be used in probability-proportional-to-size (PPS) sampling.

Two types of sample designs that have been used in surveys of fisheries in developing countries are one-per-stratum designs and systematic sampling designs (see case study in the annex to this publication). These sample designs are often used in one part of an overall survey design that involves multiple stages of selection. Because these sample designs have been used for surveys of the types of population of interest for this research, these two sample designs are distinguished from one another and their important properties are emphasized. In a systematic sample, the elements in the list are ordered. The designer chooses a sampling interval, \( k \), and a random start, \( r \). The elements included in the sample, are element \( r, r+k, r+2k, r+3k, \ldots \). For a one-per-stratum design, the elements are placed into \( H \) groups, and one element is selected from each group. Systematic samples are efficient if the elements are ordered in such a way that the elements in a group of \( k \) consecutive units are as different from each other as possible. One-per-stratum designs, in contrast, are efficient if the groups of units (the strata) contain elements that are as similar to each other as possible. Both one-per-stratum and stratified sample designs have the property that it is impossible to construct a design-unbiased variance estimator. Two-per-stratum designs are close to one-per-stratum designs, and constructing design-unbiased variance estimators is possible for two-per-stratum designs. In most situations, it is advised to not use designs at the first stage of selection for which design-unbiased variance estimation is not possible. If one-per-stratum or systematic sampling is used in the second or third stage of a multistage design, a variance estimator may still be constructed for the first stage of sampling. Variance estimators for one-per-stratum or systematic samples make implicit assumptions about the structure of the population under study that cannot be verified with the sampled data.

When producing a table of estimates from a survey sample, it is recommended to include a measure of uncertainty, such as a confidence interval or a standard error. The data user can use the measure of uncertainty to gauge decisions based on the published data. Producing a variance estimate can also enable subsequent statistical analysis. In combining multiple sources of information through measurement error models, having a measure of uncertainty for the estimates from survey samples can help determine how much weight to assign to the survey estimator and may avert a need to make strong assumptions. Having a variance estimate can also be useful for small area estimation. Some procedures for constructing small area estimates use an estimator of the variance of the estimators obtained from survey samples. Constructing variance estimators for surveys of fisheries and aquaculture uses the same methodology as other types...
of surveys. Wolter (2007) gives a thorough discussion of variance estimation for survey samples. The methodology of Wolter (2007) is applicable to surveys of fisheries and aquaculture. Pollock, Jones and Brown (1994) present variance estimators using notation specific to the estimation of effort or catch from surveys of fisheries. Although Pollock, Jones and Brown (1994) focus on the case of recreational surveys, the notation applies to a survey of any type of fishery.

4.4.2. Relation to the frame

The choice of the sampling frame can both expand and limit the possible sample designs. If a frame is rich with auxiliary information, then the designer will have many options for defining strata and unequal selection probabilities. If the frame lists clusters of population elements (rather than listing the individual elements themselves), then using a cluster sample design or a multistage design may be needed.

4.5. Estimation

GSARS (2017a) reviews estimation procedures for survey samples. Three classifications of survey estimators are; (1) direct; (2) model-assisted; and (3) model-dependent. In reality, the differences between these categories are much more of a gradient than a discrete classification.

Direct estimators are functions of weighted sums of the data collected in the survey. The weights are the inverses of the selection probabilities. The most famous direct estimator is the Horvitz-Thompson estimator. The Hajek estimator is a useful estimator of means from unequal probability samples and can be more efficient than the Horvitz-Thompson estimator when the realized sample size is random. Multiple-frame estimators are direct estimators used to combine data collected from different frames.

Model-assisted estimators go one step further and modify the direct estimator based on a postulated relationship between the variable of interest and an auxiliary variable. A common model-assisted estimator is a calibration estimator, which modifies the weights to ensure that weighted sums of selected auxiliary variables are equal to known population totals (or means). Model-assisted estimators and direct estimators have the property that the estimator gets closer, on average, to the true population parameter as the sample size for the survey increases, regardless of whether any postulated models actually hold.
Model-dependent estimators are different from direct estimators and model-assisted estimators in that the properties of model dependent estimators depend more heavily on a specified model. If the model is incorrectly specified, then model-dependent estimators will be biased. Model-dependent estimators are often needed to adjust for non-sampling errors, such as measurement errors and selection errors.

Model-dependent estimators are also used for small area estimation. In small area estimation, direct estimators are judged unreliable for the domains of interest, usually because of small sample sizes. Model-based estimators “borrow information” from across areas, allowing more precise estimators for domains of interest. If the model is incorrectly specified, then the apparent decrease in variance can be artificial if model misspecification leads to an unmeasured increase in bias.

4.5.1. Special considerations for surveys of fisheries and aquaculture

Model-based procedures have multiple applications in surveys of fisheries and aquaculture. In the application of Volstad et al. (2014), data could not be collected at some fishing sites. Volstad et al. (2014) implicitly used models to adjust for this problem. For other examples, see Breidt et al. (undated), Hoenig et al. (1997), and Lopez et al. (2003).

4.5.2. Relation to the frame

Characteristics of the frame can necessitate or enable certain types of estimators. If a frame is rich with auxiliary information, then the frame can enable the use of a wide variety of model-assisted estimators. If a frame fails to cover the target population of interest, then some form of a model-based estimator should be applied to correct for undercoverage. If multiple frames are used to completely cover the target population, then multiple-frame estimators can be applied. A relatively common application of multiple-frame estimators uses information collected from an area frame that completely covers the target population to estimate the undercoverage of a list frame.
The Master Sampling Frame

As discussed in section 1, an MSF can be thought of as a permanent structure for implementing surveys of a particular sector over time. An MSF consists of a list or collection of lists along with a mechanism for using, linking and updating the lists. As previously stated, an area frame can be considered as a list frame. The lists are used to support various aspects of the survey process, including sample design, data collection and estimation, as discussed in section 4. Linking the lists may be important if an aim is to construct a single list that fully covers the population from several smaller lists, each of which partially covers the population. Because populations change over time, updating lists is important so that they remain adequate representations of the target populations. Section 5.1 reviews important resources for obtaining lists that can contribute to MSFs for surveys of fisheries and aquaculture. Section 5.2 provides guidance to address issues associated with linking lists. The updating component of the MSF is discussed in section 5.3. The focus is on aspects of frames deemed important for surveys of fisheries and aquaculture.

5.1. Types of frames

This section reviews several types of frames for surveys of fisheries or aquaculture. For each type of frame, possible resources that might be used to construct the frame are discussed. Each type of list discussed below might individually constitute an MSF, or an MSF might be a combination of several types of lists. The appropriate list or combination of lists depends on the context.

5.1.1. Frame from direct enumeration of population or sampling units

Overview. The direct enumeration approach involves conducting a census to obtain a list of operational units or sampling units that compose the target population of interest. Building a frame through the direct enumeration approach involves an explicit data collection effort to enumerate the population and sampling units of interest. The resulting list is an enumeration of all units operating in the target population, along with limited auxiliary information about the units. An example of a unit in a census for fisheries might be a vessel. Auxiliary information about the vessel might include the vessel size, the dominant gears used, or the average number of boat trips per month. The direct
enumeration approach is often called a “frame survey” in the fisheries literature. The term “survey” is often used in association with samples, or subsets of the population. The direct enumeration approach, however, involves a census of population elements, instead of a sample. To avoid confusion with survey samples, the term “frame survey” will be avoided and “direct enumeration” or “census” will be used instead.

Stamakopolous (2002; section 8) describes the direct enumeration approach in the context of fisheries surveys as “a census-based approach in which data is collected on all fishing vessels and gear (at all homeports/fishing sites), which could be potentially operating within the estimation context.” Variables gathered in the fishery census may include the extent of landing sites, times of the day or year when fishing activity occurs, gear types, employment, access to markets and demographic characteristics (Stamatopoulous, 2002). The results of a fishery census may indicate a need to exclude homeports that are no longer in operation or to reflect changes to boat or gear classifications.

For aquaculture, the direct enumeration approach is best suited to large and stable operations. For example, large coastal marine enclosures for finfish culture can be located via aerial monitoring and are not easily moved. A unit in a direct enumeration of aquaculture operations might be an aquaculture facility. In the Quarterly Agricultural Survey of the Philippines, these operational units are called “aquafarms.”

The process of enumerating the population elements may require utilizing a variety of resources. Administrative data may contain information about the fleet. Local experts can provide information about popular fishing areas and variation in fishing activities over time. Pinello, Gee and Dimech (2017) classifies sources of information for conducting a frame census into “direct” and “indirect” sources. Direct sources of information include the fleet register (or administrative data more generally), visits to landings and markets, and taking photographs of vessels in ports. Indirect sources include interviews with other fishers or with individuals involved in fishing for their occupation, logbooks data, gathering existing photos from Internet or other publications, and reviewing literature on the structure of fisheries in a particular sector.

The definition of the unit and the type of data collected depend on the context. The choice of what to collect in the frame census depends on the nature of the subdivision of the industry of interest. The issues discussed in section 4.1 and in section 3 are relevant when deciding what data to collect during the frame census. When planning the frame census, it is important to consider the ultimate
use of the census information in survey design and estimation, not only census logistics.

The definition of the target population and distinctions between population units and sampling units are important considerations. The choice of the PSU for a survey that uses the result of a direct enumeration as the frame is limited to some extent by the definition of the units identified during the census. Defining the units is not always straightforward. For instance, in a fishery census, a vessel may have more than one major gear type. In this case, the unit may be defined as the cross-classification of gear type and vessel.

For situations in which the national statistical agency has control over the census, the designer may provide input into the nature of the auxiliary information collected during the census. Collecting appropriate auxiliary information during the census can enhance the options for improving the efficiency of sample designs that use the results of the census as a sampling frame. For example, if strata related to operation characteristics are desired, then information related to gear type or vessel characteristics should be collected during the census. In this case, a census unit might be defined as the cross-classification of gear type and vessel. However, for a subsequent sample survey, the gear type to be used by a vessel at the time of the survey may not be available information at the time of sample selection. In this case, gear type should also be collected during the survey. In this example, there would be two different variables: gear type at the time of frame construction and gear type at the time of sample selection. The gear type at the time of frame construction can be used as auxiliary information in sample design or estimation, while gear type at the time of sample selection may be a domain variable or a response variable of interest in the survey. If collecting any information on gear type is difficult at the time of constructing the sampling frame, it may be preferable to define a census unit as a vessel and consider gear type to be the only variable of observation.

**Applications.** Lists obtained from direct enumerations have traditionally formed the basis of sampling frames for on-site surveys of fisheries. Production-related variables are the primary variables collected in surveys conducted from lists based on direct enumerations. In unique situations in which a census of vessels is updated routinely, a frame constructed from direct enumeration can provide the frame for a socioeconomic survey. Pinello, Gee and Dimech (2017) considers use of an updated vessel register as a sampling frame for socioeconomic surveys of fishers. The census approach is best suited to industries consisting of relatively large operations that are easy to identify and are stable over time. This is particularly important for censuses targeting aquaculture facilities. Using a frame census as a frame can be most practical in situations in
which the frame census is mandatory. For example, some countries legally require a frame census for regulatory purposes. In these situations, the frame census is more likely to be complete and up-to-date.

**Strengths.** If implementing direct enumeration is practical, direct enumerations can provide efficient sampling frames. The statistical office may have control over the process of implementing the direct enumeration. In this case, definitions of units in the frame can be aligned with definitions of units of interest for statistical purposes. Likewise, the statistical office can choose to collect auxiliary information that is expected to be useful for defining sample designs and estimators. Direct enumerations are typically well suited to populations composed of large and stable units. This is particularly true for aquaculture, where large facilities may be located through direct enumerations; however, smaller aquaculture operations may be more difficult to identify in this way.

**Weaknesses.** Expense may be the primary limiting factor associated with the census approach. Because of the high costs associated with conducting a census, intercensal periods can be lengthy and irregular. This can lead to overcoverage or undercoverage of relatively ephemeral or transient units. Information obtained in the census may become outdated. Operations may go out of business or new operations may form. To mitigate the potential for undercoverage or overcoverage when using the results of the census as a sampling frame, the direct enumeration should be conducted regularly and as frequently as possible. Augmenting the census data with other sources of information, such as administrative data, may improve coverage.

5.1.2. Administrative records

*Overview.* Administrative data are collected for purposes of administration operational programmes rather than for the purpose of producing official statistics. GSARS (2017c) follows the Administrative Data Liaison Service of the United Kingdom (http://www.adls.ac.uk/adls-resources/guidance/introduction/) and defines administrative data as “information collected primarily for administrative (not statistical) purposes by government departments and other organizations usually during the delivery of a service or for the purposes of registration, record keeping or documentation of a transaction.” Management or record-keeping are examples of processes that generate administrative data. As discussed in GSARS (2017c), administrative data have diverse uses at all stages of the survey process. One use of administrative data for producing statistics is to support the construction of sampling frames. Administrative data support sampling frames for surveys of many sectors, including fisheries and aquaculture. A single administrative list
may not cover the entire target population. This is particularly true if the administrative programme generating the list is voluntary or only applies to a subset of the target population of interest. Therefore, a single administrative list by itself is often unsuitable as a sampling frame. Use of administrative data for constructing a sampling frame often involves combining multiple lists to improve the coverage level.

An important issue to note when using administrative data for frame construction is that the units in the administrative list are at the discretion of the administrative agency. The appropriate definition of units for administrative purposes may differ from the desired definition of a unit for statistical purposes. Furthermore, different administrative processes use different definitions of units, which can create complications when trying to link multiple administrative files.

The fishery and aquaculture industries are rich with administrative data. Sources of administrative data for fisheries include registration of boats or vessels, logbooks or fishing licenses. Examples for aquaculture include aquaculture facility registration or taxes. Processor registration or taxes may be useful for either surveys of fisheries or aquaculture.

**Applications.** Frames based on administrative lists can be used to address multiple objectives. Administrative data have been used in situations in which primary target variables are related to production or commercial value. Administrative lists can support collection of data on socio-economic characteristics, particularly if the administrative list contains contact information (telephone, email address and address) for population elements. Administrative data can support household surveys if the administrative file contains contact information (that is, as above, telephone number, email address and address) for the population elements of interest. This may allow for using administrative lists as frames for socio-economic studies. Administrative lists as frames for household surveys have been used to target recreational anglers in developed countries. To our knowledge, this approach has not been attempted in a developing country, perhaps because the administrative processes necessary to support this approach are unreliable or nonexistent. If measuring value is important, processor registration or taxes can be useful sources of administrative data. Using administrative data may be most applicable in situations in which fishers must compulsorily register in the administrative programme that produces the data. When administrative programmes are not required, the resulting administrative data sources are more likely to be incomplete. Administrative data are most reliable if the administrative process underlying the data collection is consistent and completely covers the target population.
**Strengths.** The most commonly stated reason for using administrative data sources may be that the statistical office does not need to incur the cost of data collection because the data are collected naturally for the purpose of an administrative process. While the direct enumeration method may be the most expensive procedure for constructing a frame, use of administrative data may be the cheapest.

Another possible strength of administrative data is the availability of auxiliary information. Some administrative programmes result in a wealth of auxiliary information, including data on production-related variables such as catch or effort. In extreme cases, the information on the administrative file may be tabulated directly to form the ultimate statistical product. In less extreme cases, auxiliary information on an administrative file can be used in defining strata or selection probabilities, to help locate units or as covariates for imputation.

**Weaknesses.** The challenges arising from the use of administrative data for frame construction (and for statistical purposes more generally) stem from the fact that the data were originally collected for administrative purposes, not statistical ones. Interestingly, many of the strengths of administrative data, when cast in a different light, also become their weaknesses. The cost savings associated with not needing to collect certain data items often translate to different costs, associated with data management and maintenance. The availability of auxiliary information to use in design or estimation depends on what is collected on the administrative data source, and the administrative process may not produce the auxiliary information required for statistical purposes. In practice, combining multiple administrative lists is often necessary to obtain complete coverage. This can be accomplished through record linkage or multiple-frame survey designs. Linking multiple files is often complex, and the process can introduce errors. Because the lists and associated data elements are maintained for an administrative purpose, rather than a statistical purpose, definitions used by the administrative agency may be misaligned with definitions used by the statistical office. Similarly, identifying variables may be poorly maintained, and a one-to-one correspondence between identifying variables used in the administrative source and identifying variables used by the statistical office may not exist. Definitions of units (population elements labeled with identifying variables) may vary between administrative data sources, which can complicate record linkage when combining lists to construct a single sampling frame. If different administrative processes use different measurement units, construction of conversion factors may be a necessary part of the estimation process. Related practical issues involve the centralization (or lack thereof) of records from various levels of government, archival and storage media and policies, and the frequency with which official records are updated.
If variables of interest are recorded on the administrative file (for example, catch on a logbook), it may be tempting to use the administrative data directly for estimation. It is important to recall, however, that administrative data are typically self-reported and may be contaminated with intentional or non-intentional reporting errors. One common example occurring in recreational fisheries surveys is “prestige bias” resulting from intentionally inflating catch on a catch card. Even if administrative data are contaminated with measurement errors, they have the potential to provide useful auxiliary information in estimation when used in combination with variables that are accurately measured in the survey.

Coverage is an important potential source of non-sampling error when using administrative data. Administrative data may provide either overcoverage or undercoverage of fishing equipment. Selection bias resulting from undercoverage is a particular concern if participation in the administrative process is voluntary. Selection bias arising from undercoverage may be especially important when using a single administrative list directly as a sampling frame. Registers typically do not cover small boats with short lifespans.

5.1.3. Other surveys or censuses

**Overview.** An existing survey or census may be used as a frame for a subsequent survey of fisheries or aquaculture. This publication focuses on government-sponsored population or agricultural surveys or censuses. Such data collection efforts may produce frames in which the units are individual fishers, households or communities such as villages. The contact information on the list from the previous survey may be a list of telephone numbers, addresses or enumeration areas. Previous surveys or censuses are most useful for surveying individuals engaged in fishing or aquaculture if the survey or census contains a screening question that collects information on participation in activities related to fishing or aquaculture. The screening question could then be used to identify individuals engaged in fishing or aquaculture. GSARS (2017d) provides guidelines on how to write effective screening questions. Agriculture surveys or censuses have traditionally collected information on aquaculture, and a traditional approach to sampling aquaculture operations is to sample agricultural operations that contribute to aquaculture production. To improve coverage for fisheries surveys, the use of a population survey or census is preferable to using an agricultural survey or census as the basis for the sampling frame.

When subsetting agricultural or population surveys (samples rather than censuses), estimation procedures need to respect the design of the original
survey. The PSU is the PSU used in the agricultural survey or census. Variance estimators should account for the hierarchical nature of the selection process. The distinction between multiphase and multistage sampling is important for the analysis of data from surveys that use a previous survey as the sampling frame. The distinguishing feature of a multiphase sample is that variables collected in the first survey determine the design of the second level of data collection. Use of an agricultural or demographic survey to identify individuals active in aquaculture or fisheries is an example of a multiphase approach. Sarndal, Swensson and Wretman (2003) provides estimation procedures for multiphase designs.

**Applications.** Household surveys are best suited to situations in which socio-economic characteristics are of primary interest. Household surveys are typically not recommended for collecting production-related variables because of the potential for serious measurement errors. The individual contacted at the household may not accurately recall production-related information. If the individual contacted is not the primary fisher, then such individual may not have access to accurate information about production. The approach is the most natural one to apply if the population or agricultural survey or census contains a screening question related to participation in fishing or aquaculture. If estimation over time is an objective, then the procedure is best suited to situations in which the underlying population or agricultural survey or census is conducted on a relatively routine basis (for example, annual surveys). Despite the potential for non-sampling errors, obtaining information on production from household surveys is sometimes the best or only feasible approach. Because of the potential for cost savings, Bayley and Petrere (1989) advocate the use of household surveys for estimating fishing catch and effort in developing countries, despite the measurement issues discussed in section 4.3.1 above. From their viewpoint, the problems in measurement are offset by the savings in terms of frame maintenance and data collection costs. As discussed in section 4.2 above, for measuring production-related variables, using an index that is practical to collect can be more effective than trying to obtain a highly accurate but expensive measure of fishing production. The viewpoint of Bayley and Petrere (1989) is that household surveys can provide a cost-effective way to obtain a consistent index of fishing production when the cost of on-site data collection is prohibitive. Further, the household survey approach may be necessary if a frame survey is too expensive or if nationally consistent administrative processes do not exist.

**Strengths.** Household surveys are typically less expensive to conduct than surveys that require data collection at the location where fishing takes place. Because the same basic frame is used for the household survey of fishers as for
the original survey or census, the use of existing surveys as a frame for a fishery survey allows the statistical agency to share costs among multiple activities. Regardless of cost, household surveys are sometimes deemed more appropriate than on-site surveys for collecting socio-economic data. It may be feasible to create a lengthier questionnaire for household surveys than for on-site data collection methods. This lengthier questionnaire may include questions about socio-economic variables that are not practical to collect at the location where fishing occurs. When using a previous survey or census as the frame for a subsequent survey of fishers, questions on the original survey (the frame for the fishery survey) may serve as auxiliary information for the fishery survey. This auxiliary information collected in the first survey may be used to improve efficiency at the design stage or the estimation stage of the fishery survey. This auxiliary information could also be used in imputation procedures to account for nonresponse.

**Weaknesses.** The primary weaknesses of household surveys relate to the limitations for collecting production-related information, discussed in section 4.3.1 above. Obtaining accurate information related to production of fisheries through household surveys has been documented to be associated with numerous sources of measurement error. Recall bias is a widely documented problem. Response rates are often documented to be positively correlated with fishing avidity, which, if not accounted for, may lead to estimates of fishing effort with a positive bias. A one-to-one correspondence probably does not exist between a household and a fishing vessel or aquaculture facility. In some cases, individuals may participate in multiple aspects of fishing or aquaculture. Individuals who participate in fishing or aquaculture activities may have only tangential connections to record-keeping. Inaccurate data may also arise if the individual contacted is not directly responsible for record-keeping activities.

**5.1.4. Spatial, spatio-temporal and aerial frames**

**Overview.** A spatio-temporal frame represents a geographic area or time frame. The units in a spatio-temporal frame may be defined geographically, temporally, or as intersections of spatial units and time-points. Spatial units may be segments that partition a geographic area or individual locations. A location may be a coordinate defined by longitude and latitude. A location may alternatively be a waterbody, a landing, or a more general location at a fishing site. The temporal dimension of the spatio-temporal frame may be defined by boat trips, hours, days or months, for example. The target population must be specified to ensure that the spatio-temporal frame has appropriate coverage. Aerial frames may be viewed as a subset of spatial frames. Aerial frames may be used for aerial surveys of fishing effort. Typically, the spatial aspect of spatio-temporal frames
is more easily managed than their temporal aspect. The relation of landing sites, lakes, or river access points to the type of catch changes slowly relative to the relation between time and catch type, except perhaps for migratory species.

Stratification and multistage sampling are commonly used with spatio-temporal sampling frames. Stratification by both geographic and temporal dimensions can be used. Examples of geographic strata are an administrative region or a waterbody. Examples of temporal strata are months, times of day, or weekdays or weekends. Many surveys that use spatio-temporal frames involve multiple stages of selection. For instance, the first stage may be a landing site on a particular day, the second stage may be a boat trip and the third stage may be an angler.

Increasingly, satellite imagery is an important resource for constructing spatio-temporal sampling frames. Although immediate uses of satellite imagery are largely focused on regulation and monitoring for illegal fishing activities, images of fishing boats obtained from satellites have potential applicability to the construction of spatio-temporal sampling frames. Global Fishing Watch (http://globalfishingwatch.org/) uses satellite data to track commercial fishing fleets on oceans across the world. They produce an activity map (http://globalfishingwatch.org/map/) that shows the locations of fishing boats detected via satellites. They also provide processed datasets with gridded fishing activity data, vessel identification and classification lists, and information on encounters between refrigerated cargo vessels and fishing vessels. In an important application of the Global Fishing Watch data, the encounter information is used to predict instances of illegal transshipment. Information such as that available from Global Fishing Watch is also published commercially. For example, Marine Traffic (http://www.marinetraffic.com/en/ais/home/centerx:-62.2/centery:23.4/zoom:4) provides a user-friendly map of the spatial distribution of different types of fishing boats. The sources of satellite imagery that facilitate identification of fishing activities known to these authors only apply to large marine vessels. As such, satellite data seems to have limited use in constructing sampling frames for SSF.

Applications. Spatio-temporal frames typically support on-site data collection methods. As a consequence, spatio-temporal frames are best suited to estimating parameters related to production (that is, catch or effort), although limited socio-economic data can be obtained as a secondary objective. Bycatch and discard are also collected using spatio-temporal frames, typically in industrialized settings. When the landing site defines the spatial dimension, spatio-temporal frames are most appropriate for target populations with landing sites that are
clearly defined and easy to access. Typically, on-site data collection methods are used in combination with spatio-temporal frames. Measures of catch, bycatch, discards and effort are often considered more accurate than corresponding measures obtained from off-site methods because the data are collected by a trained data collector with an objective viewpoint. Socio-economic data can be collected on site or through a follow-up survey.

**Strengths.** Spatio-temporal frames can have good coverage of target populations. They can be less expensive to construct than direct enumerations and still support on-site data collection methods. They are better suited to aerial surveys than other types of frames. Spatio-temporal frames allow the designer to control for variation across space and across time during the sample design process. For instance, homogeneous geographic regions or time frames can be used as sampling strata. One can define spatio-temporal frames in such a way that the units (that is, intersections of geography and time-points) are stable over time, while the activities in the units change. Consequently, spatio-temporal frames are less sensitive to changes in the population than administrative lists or direct enumerations. Their stability over time also makes spatio-temporal frames well suited to estimating change over time.

**Weaknesses.** Potential limitations are associated with costs, respondent burden, and potential for inaccessible fishing sites. Because data collection often involves visiting the location in some fashion, data collection for surveys that use a spatio-temporal frame can be costly. Several types of non-sampling errors can arise in surveys based on spatio-temporal frames. For instance, aerial surveys may introduce visibility bias. Undercoverage may occur if data collectors are unable to access locations where fishing takes place.

### 5.2. Combining frames

As discussed in section 5.1, each type of frame has unique strengths and weaknesses. Combining multiple sources of information allows the data user to reap the benefits of each data source. When using two data sources with complementary strengths and weaknesses, one data source can adjust for the weaknesses of the other. Reasons to combine multiple data sources and methods for such combination when constructing an MSF are overviewed below.

**Reasons to combine**

- Improve coverage. The union of multiple lists often has better coverage than a single list. Two ways to combine lists are through multiple-frame survey designs and record linkage. In the multiple-frame survey approach, discussed
in further detail in section 3.1 of GSARS (2017a), separate samples are drawn from different lists and estimates are combined. Record linkage is the process of combining multiple lists to create a single frame. Record linkage is often complicated by factors such as errors in identifying variables and inconsistencies among identifying variables in different lists. Because of these issues, record linkage is a vast topic.

- Understand measurement error properties. As discussed in section 2 above, different frame types are typically associated with different data collection methods. Therefore, the choice of frame affects the types of non-sampling errors that are more or less likely to arise in the survey. Collecting the same target variable (that is, catch or price received) with more than one data collection method can provide information about potential biases due to measurement errors.

- Address diverse objectives. Section 5.1 shows that some types of frames are better suited to addressing certain types of objectives than others. Each type of frame is typically associated with a particular data collection method. Thus, not all frames are equally appropriate for addressing the same objectives. For example, some frames are better suited to estimating catch (such as the list of fishing sites), while others are preferable when estimating socio-economic characteristics (for example, the list of households). Similarly, certain frame structures are amenable to target populations with large, stable operations, while others can target small or transient population elements.

- Improve efficiency of estimators with auxiliary information. Combining multiple data sources is also useful at the estimation stage. Information in a list that could contribute to a sampling frame can be used as auxiliary information in estimation, even if the list was not used as the sampling frame. Examples include imputation and the types of model-based procedures that are commonly used to construct estimates for domains with small sample sizes (small area estimation; see Rao, 2003).

**Methods for combining**

- Multiple frame estimators. In multiple-frame designs, separate samples are selected from different frames. Multiple-frame estimators combine data collected from multiple-frame designs. GSARS (2017b) discusses multiple-frame estimators.
• Integrated surveys. GSARS (2015b) discusses integrated surveys. These are particularly applicable in situations in which a direct listing of the population elements does not exist. GSARS (2015b) describes an application in which an integrated survey approach is used to select a sample of kindergarteners from a list of primary schools because a list of kindergartens is not available.

• Record linkage. Combining lists is often more challenging than it may seem at a first glance. A single list may contain duplicates. Different lists may use different identifying variables. The identifying variables on a particular list may contain errors or be incomplete. The sheer size of a list could also create challenges. For dynamic populations, changes in the structure of the population lead to changes in the relationships between population elements. Probabilistic record linkage (Fellegi and Sunter, 1969) is a statistical procedure for combining multiple lists when exact matching is difficult or impossible. In probabilistic record linkage, pairs of observations are assigned probabilities of a match, and elements with high match probabilities may be linked. Day (1995) reviews alternative record linkage procedures for use in agricultural surveys. Berg and Li (2015) discuss record linkage procedures in the context of administrative data sources.

• Imputation and modelling. Imputation and modelling can be used to combine information from multiple sources at the estimation stage. Model-based approaches for combining estimates can operate at the level of the estimates themselves or at the unit level. One way to think of unit-level integration (Kim, Berg and Park, 2016) is to imagine creating a single list that contains all information on either list. Any missing information is imputed on the basis of estimated relationships among observed values. Imputation and modelling approaches often require a subset of variables to be observed on both lists. Imputation and modelling approaches also require assumptions – such as missing at random or existence of an instrumental variable – that justify applying relationships estimated on the basis of observed data in contexts in which some information is missing.

5.3. Updating the frame

The MSF approach should encompass a mechanism for updating the sampling frame. As populations change, lists become inaccurate representations of the populations that they are intended to capture. This subsection offers advice related to updating frames.
• Frame census. Reserve resources for conducting a periodic frame census. If the direct enumeration of operations approach is used, then maintaining an updated list of operations is crucial. Updating the entire list may be expensive. Integrating a plan to update sections of list periodically as part of the overall survey plan could be more practical. Table 2 below illustrates a frame updating plan in which one-fifth of the frame is updated every year, such that the entire frame is updated in a five-year period. This approach also spreads the workload evenly across years and can therefore be more practical to implement than an approach that updates the entire frame every five years. In some cases, reducing the sample sizes for surveys may be necessary to reserve financial resources for updating the list.

**Table 2. Example of frame-updating plan.**

<table>
<thead>
<tr>
<th>Fishing sites</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/5 of sites</td>
<td>Update</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/5 of sites</td>
<td></td>
<td>Update</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/5 of sites</td>
<td></td>
<td></td>
<td>Update</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/5 of sites</td>
<td></td>
<td></td>
<td></td>
<td>Update</td>
<td></td>
</tr>
<tr>
<td>1/5 of sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Update</td>
</tr>
</tbody>
</table>

*Source: Authors elaboration*

• Aerial frame. Reviewing timely satellite data may help to update aerial frames. Satellite images of fishing vessels may indicate changes in fishing activity. Satellite images of waterbodies may also show changes in waterbody size and reveal new waterbodies (see discussion in the case study in the annex to this publication).

• Local knowledge. As recommended in Pinello, Gee and Dimech (2017), local experts often have information on changes in fishing patterns. Locals may be aware of an increase in the use of a particular fishing site. They may have knowledge of fishing sites where fishing effort has declined, possibly as a result of declines in stocks from factors such as overfishing or pollution. While local expert information may be seriously biased if used directly to estimate population parameters, local knowledge is still informative. Local experts can provide useful information for updating various types of MSFs.

• Combining multiple sources. The recommendations for combining multiple sources of information, discussed in section 5.2, also apply to updating the frame. Combining updated administrative lists with older lists may help to keep information in the frame current. As populations change, definitions of
population elements and linkages between them may change, making the process of integrating new information not straightforward. As a result, methods such as probabilistic record linkage could be applied in the process of updating the frame.

- Using a contiguous population or agricultural census. Continuity of survey programmes is crucial for maintaining a consistent set of estimates over time. Using a population survey or census as the frame for a fishery survey is most appropriate when the underlying population survey or census is conducted routinely and at regular time intervals.
Steps for building an MSF for fisheries and aquaculture

This section proposes a procedure that could provide guidance on how to progress through the steps of designing a programme for producing estimates of fishery- and aquaculture-related quantities. The intention is for this guide to be applicable for uses ranging from the estimation of total fisheries or aquaculture production at a national scale through the design of a small-scale verification survey. The phrase “Activities” is used to describe the collection of surveys or programmes under consideration. For example, a proposed activity may consist solely of conducting a survey to estimate discard for a specific sector of a marine fishery; or a set of proposed activities might consist of designing a sampling structure for estimating total catch of inland fisheries on a national basis, with separate estimates for each of a set of geographic divisions. All of the material in the previous reports issued under this project, as well as other documents, should be consulted throughout the design procedure.

6.1. Purpose of activities

The first step is to develop a list of objectives. Questions related to this step include the following:

- Have all quantities for which estimation is desired been included?
- If the list of quantities for estimation is long, which of these quantities should the activities be designed to address?
- Is estimation desired for individual species or species groups? Which ones?
- Has the overall spatial domain of interest been identified?
- Has the overall temporal domain of interest been identified (including incorporation into the monitoring effort)?
- Have spatial or temporal subdivisions been identified either for partitioning populations or for use as strata?
• Are other divisions (such as gear types) involved in the estimation problem? Should these be considered an object of estimation or of stratification?

6.2. Need for new activities

In this step, the question to be considered is whether the objectives identified require new efforts, or whether existing programmes could be used or modified to provide some of the estimates desired. If it is possible to use or modify existing survey programmes, questions related to this step include:

• Are all of the estimation problems identified amenable to incorporation into existing programmes?

• Will existing programmes provide direct or indirect observation of the variables required (this may have to be revisited after a more detailed consideration of variables and measurement has been done, later in the process)?

• If data from existing programmes provide only indirect observation, are small-scale verification surveys desirable (and feasible)? What should the temporal frequency of such efforts be?

6.3. Identification of populations

Given the diversity of operational units involved in fisheries and aquaculture, at this stage, it is preferred to consider a problem as consisting of multiple populations, rather than subpopulations. For example, silver carp aquaculture may involve inland ponds or raceways, as well as small mobile cage cultures in inland lakes. In reality, these are two populations, because they comprise rather distinct units that bear little resemblance to each other, and may well require entirely different sampling plans to provide adequate coverage. Questions related to this step include:

• Are populations defined relative to the characteristics of the operational units?

• Are populations defined relative to geographic or temporal subdomains that require different sampling approaches? (if not, then these subdomains could be considered population divisions or strata rather than separate populations).
- Have all populations that contribute to the estimation purpose been identified?

The remaining steps should be conducted for each population separately, and then again for all populations as a whole, to ensure that the overall objectives can be addressed.

### 6.4. Determine variables and measurement

Variables and units of measurement should be considered separately for each population identified in the second step; however, they then need to be looked at as a whole. Questions related to this step include:

- Are variables the same or different over multiple populations? Are measurement scales the same or different?

- If multiple variables could be considered (for example, trips or time fishing, for effort), have their properties relative to the accurate reflection of the desired characteristic (versus stability in consistent reproduction) been considered?

- If one of the objectives involves monitoring or change detection over time, will the variables allow for consistent observation into the future?

- Are units of measurement consistent across populations, or can they be reasonably combined in one estimation method?

### 6.5. Determine contact method

At this point in the process, it is time to precisely determine the methods by which sampling units will be observed or measured, the resources required to do so, and how the variables identified are to be recorded. Questions related to this step include:

- Are contact methods indirect or direct (on-site or off-site)?

- Have appropriate media been identified for data recording?

- How much training and costs will be involved?

- Have points of responsibility been identified for implementation?
6.6. Outline sampling design

The type or types of sampling design that might be used have undoubtedly played a role in the progression to this point; however, this step requires those designs to be detailed – not only stratified sampling, for example, but also the number and identity of strata, how those will be sampled, how primary and secondary units in a multistage design will be sampled, etc. Consideration of all of the steps prior to this point in the process should indicate the preferable of one or more potential sampling designs, which may possibly differ for different populations or different variables.

- How many stages are required? Is the design multistage or multiphase?
- What types of frames are needed so that inclusion probabilities can be computed for all *sampled units*?
- Is the design specific enough for its description to allow individuals who are not involved in this point of the process to implement it?

6.7. Construct sampling frames

Any number of considerations prior to this step are influenced by the availability of sampling frames and, similarly to sampling designs, a fair amount of general information about possible frames should have been generated or used in previous steps. However, in the same way as for sampling designs, this step involves identifying particulars. Questions that are related to this step include:

- Has the use of each frame been explicitly identified and attached to it?
- Has the maintenance schedule for each frame been determined?
- Have unique identifiers for individual units (population and sampling) been used so that overlap in frames can be detected?
- Are the frames constructed compatible with the sampling design, variables and definitions of population units resulting from previous steps in the process?
6.8. Identify potential non-sampling errors

This step requires consideration and enumeration of possible sources of non-sampling error that may lead, in particular, to bias in estimators. The effects of some of these sources could be mitigated in estimation, while others could not. Questions that are related to this step include:

- Are sources of error amenable to correction over time (for example, an incomplete frame)?
- Can small-scale studies be designed to determine the severity of deleterious effects (and possibly suggest corrections)?
- Have modelling techniques been used to deal with the identified error sources in other contexts?

6.9. Develop estimation strategy

All of the previous steps may be conducted in an iterative fashion, with considerations and results relating to one step prompting the revisitation of a previous step. Once a reasonably stable outline for all of the previous steps has been attained, a strategy for estimation should be developed. This step may require consultation with external experts, and may be revised and improved without necessarily repeating the entire process. Estimation may be considered separately for each objective identified in the first step of the process, or may be considered as a whole from the outset. Either way, it is important to consider both estimation for each objective individually and the set of estimation problems as a coherent unit. Questions related to this step include:

- Are all quantities identified in the objectives addressed in estimation?
- Are quantities associated with uncertainty (or inference) produced by the strategy developed?
- Does the set of estimation procedures hold together in a unified (or coherent) approach?
6.10. Develop framework for storage and retrieval of data and results

Successful navigation of this procedure will result in a programme design that addresses the identified objectives in a scientifically principled manner. The information required to make such a programme operational, such as sampling frames and auxiliary information, as well as the results of prior applications of the programme (once the relevant actions have commenced) must be available in a format that can be retrieved and updated by those responsible for conducting the programme, and that can be accessed for use by decision-makers. This area has traditionally received little consideration by statisticians, and may well constitute an additional project in the overall GSARS. Within the context of this project, these authors suggest that consideration be given to the use of mathematical graphs for identifying the sources of information (such as frames) as graph nodes, and links between them (which may be of different types) as graph edges. For example, a frame used to sample for direct observation of catch and a different frame used for aerial survey of fishing activity could be joined by an edge that represents dependence in estimation (CPUE and effort). In contrast, one frame of registered fishing vessels and another of landing sites that might be sampled for activity could be joined by an edge that represents multiple frames for the same variable (active fishing vessels in one year). Questions that are related to this step include:

- Can the storage or retrieval system be used to rapidly locate information for implementation of sampling in the current spatial and temporal window of interest?

- Can the system be used to rapidly retrieve data for use with potential new estimation strategies, or for the production of estimates for population divisions that were not envisaged at the time the programme was initiated?

- Will the system allow for adaptation to new generations of software or hardware?
6.11. Logistics and coordination

Each stage of the procedure will have to account for cost and administrative constraints. A common issue particular to fisheries is the alienation of fisheries from other activities at the statistical office. Advocacy for the importance of surveys of fisheries and aquaculture could be important. After completing the proposed procedure, it may be found that the cost of the operation exceeds the budget. In this case, the proposed procedure could be iterated through a second time, beginning with the objectives. Specifying more modest objectives may be necessary to contain the cost of the survey within budget.
References


GSARS. 2017e. A Protocol for Desktop Analysis on Master Sampling Frames for Fisheries and Aquaculture.


ANNEX - CASE STUDY

MASTER SAMPLING FRAMES FOR FISHERIES SURVEYS: DESKTOP STUDY USING DATA FROM BURKINA FASO

Using data from Burkina Faso to provide context, this case study illustrates properties and issues associated with using different types of lists to maintain an MSF for fisheries surveys. Three types of lists are considered: a list frame of fishing sites, a list of villages from the population census, and an area frame. Throughout the report, four parameters are used: average catch, total catch, the proportion of female household heads, and the proportion of fishers whose primary occupation is agricultural. Section 1 introduces four illustrations that address different aspects of constructing an MSF for collecting fisheries data for Burkina Faso. Section 2 describes the basic properties of the survey and census data that support the illustrations. Section 3 elaborates on how the illustrations relate to issues arising in the Burkina Faso data collection programmes and describes implementation of the illustrations using the Burkina Faso data. Section 4 concludes with a summary and general recommendations.

1. Overview of approaches

1.1. Illustration 1: compromise between frame updating and sampling

This illustration focuses on the list frame. A problem arising in many estimation programmes for fisheries is that a list frame of fishing sites becomes outdated because conducting a complete update of the frame is expensive. The problem of integrating a frame update into the overall sampling plan is considered. A potential solution is to reduce the sample size for the survey to lower the cost of implementing the survey. The savings from implementing the survey are then devolved to updating the frame.

To be specific, suppose that the cost of implementing the survey protocol at a fishing site on one day is twice the cost of implementing the census protocol at a fishing site. Burkina Faso has approximately 600 fishing sites. If the number of visits to fishing sites in a survey year can be reduced by 300, the frame can be updated for all fishing sites in the year preceding the survey. One activity involved in updating the frame involves identifying new fishing sites, and
resources for identifying new fishing sites are considered in illustrations 2 and 3 below.

It is assumed that the bias of a survey from an updated frame is zero, and that the bias of a survey from a frame that has not been updated is \( b \). Recall that the mean squared error (MSE) of an estimator \( \hat{\theta} \) of a parameter \( \theta \) is variance plus squared bias; that is,

\[
MSE(\hat{\theta}) = Var(\hat{\theta}) + b^2,
\]

where \( Var(\hat{\theta}) \) is the variance and \( b^2 \) is the squared bias. The MSE of estimators from the combined updating/surveying approach is smaller than the MSE of estimators based on an outdated frame if the additional variance from selecting a smaller sample for the survey is smaller than \( b^2 \).

Three values for \( b \) are determined from the Burkina Faso data. First, a random sample of 10 percent of the fishing sites is removed from the population of fishing sites and the difference between average catch based on the remaining 90 percent and the average catch based on the full population is calculated (the average bias over repetitions is, of course, zero). Second, the 10 percent of sites with the highest average catch is removed. Third, the 10 percent of sites with the lowest average catch is removed.

To reduce the cost as desired, the survey sample size must be reduced by 300 visits to fishing sites per year. One way to accomplish this is to remove 12.5 interviews per month. A sampling plan whereby 24 fishing sites are randomly selected from each of the 13 regions is considered, and each of the 24 sites is assigned a time \( t = 1, \ldots, 24 \). Fishing site \( i \) is not sampled in month \( t \). The variance of the estimators from this procedure is compared to the values of \( b^2 \) obtained above. Because a random sample is available for each month, variance estimation for this design is possible.

### 1.2. Illustration 2: linking villages and fishing sites

This illustration focuses on a list of villages obtained, for example, from a population census. An approach related to the design of Bayley and Petrere (1983) is considered. The starting point is a sample of villages. For each sampled village, all fishing sites where individuals residing in that village fish are enumerated. This may be accomplished, for instance, through an interview with the village head, as in Bayley and Petrere (1983). For each sampled village, two further subsamples are selected: a sample of households and a sample of fishing
sites. The probability of selecting a village, a household and a fishing site are calculated. Estimates of demographic characteristics are obtained from the household survey and estimates of catch-related variables are calculated from the fishing site survey.

Obtaining measurements of a subset of variables using both contact methods may be desired if certain objectives are of interest. If estimating relationships between demographic variables and catch-related statistics is of interest (for example, average catch by primary occupation or education level), then a mechanism to link the two sets of variables is required. Collecting the same variables using different contact methods can provide insight into the measurement properties of different data collection methods.

As a prelude to the analysis of this design using the Burkina Faso data, the table below shows the number of villages per region in the census (top row) and the number of villages per region in the survey with catch-related statistics (second row). Clearly, more villages are represented in the census than in the fishing site survey. One reason for this is that the fishing site survey covered all fishing sites with at least seven fishermen, representing 80 percent of the total number of fishermen. Further, the fishing sites are concentrated in a relatively small number of villages, so that people must travel to a different village to fish. An additional potential issue that may occur more generally is failure of the fishing site survey to “cover” all villages, where “cover” includes the effects of the combination of frame undercoverage and nonresponse.

Figure 3. Number of villages per region for Burkina Faso. Top row drawn from the Burkina Faso 2006 population census. Bottom row drawn from the subset of fishing sites in the 2008 fishery survey that measured catch-related variables.

| Source: Authors using fishery data of Burkina Faso, 2006-2008 |

If undercoverage is more of a problem in the survey based on the fishing site list than the survey based on the census of villages, then the two-phase design, with villages as the first phase and landing sites as the second phase, can improve coverage. Using only the list of villages from the population census may be inefficient if the number of non-fishing households is large. In this case, integrating a sample based on a frame of fishing sites into the survey plan can improve the efficiency of the process. Dual-frame designs (Lohr, 2010) provide one avenue for accomplishing this. In illustration 3 below, it is proposed to
investigate dual-frame designs using the data for Burkina Faso. Although the investigation in illustration 3 below uses an area frame and a list frame, the same basic ideas would apply if one of the lists forming the multiple-frame survey were a list of villages from a population census.

1.3. Illustration 3: multiple-frame designs using an area frame and a list frame

A dual-frame design using an area frame and a list frame is illustrated. The list frame is represented with a subset of responses to a 2008 survey of fishing sites in Burkina Faso. A representation of an area frame is required. The resources necessary for constructing an area frame are considered. The map (.tif file) of land cover in Burkina Faso is obtained – see figure 4 below (https://eros.usgs.gov/westafrica/land-cover/land-use-land-cover-and-trends-burkina-faso). In the map in Figure 4, the blue locations approximately represent waterbodies. To construct an area frame, the locations classified as water are isolated. The water locations are considered, to ensure that all possible fishing sites are represented.

Figure 4. Land cover map of Burkina Faso

An area frame constructed from a map such as that in figure 4 provides the basis for estimating the undercoverage of the list frame of fishing sites. To illustrate this idea, it is supposed that the blue dots on the map in figure 4 cover all fishing sites and that p percent of all fishing sites are omitted from the list frame of sites. A site identification number (ID) is randomly assigned to all N water locations
on the map. An indicator representing inclusion on the site frame is randomly assigned to (1-p) percent of the water locations. Two independent samples are selected: one from the list frame and the other from the area frame. A dual-frame estimator of average or total catch is then constructed.

1.4. Illustration 4: estimating a conversion factor from a subsample

One interesting aspect of the Burkina Faso data is that sites without weighing facilities measure catch weight in non-standard units, such as “heaps”. The data collector then provides a conversion factor between the non-standard unit and kg. It is supposed that resources to measure weight using two procedures (heaps and weighing station) are available for a subsample of the sites. To construct the illustration, a model relating the measurement in terms of heaps to the measurement in terms of kg is specified. This relationship is estimated using the subsample for which both measurements are available. To create an adjusted estimator of catch, this relationship is extrapolated onto the part of the sample for which only the measurement in heaps is available.

2. Burkina Faso data

2.1. Basic geography

Before introducing the survey and census datasets, a general sense for the basic geography of Burkina Faso is provided. Burkina Faso has 13 administrative regions, shown in figure 5. Figure 6 shows the desert in the northern Sahel region, major cities and important rivers. Major rivers include the Niger, the Sirba, the Red Volta, the Black Volta and the White Volta. Importantly for the purpose of studying fisheries, figure 7 shows Burkina Faso’s multitude of reservoirs and relatively small rivers, where fishing also occurs.
Figure 5. Burkina Faso’s 13 administrative regions.


Figure 6. Major cities and rivers

Source: https://www.worldatlas.com/webimage/countrys/africa/bf.htm
2.2. Data from population census, fishing site census and fishing site survey

There are three datasets for Burkina Faso. The data are from (1) a census of fishing sites; (2) a survey of fishing sites; and (3) a population census. These three data sources are used for the illustrations in section 3. In preparation for the illustrations, these three data sources are described and output is provided in sections 2.2.1 to 2.2.3 below.

2.2.1. Census of fishing sites

The first data source is from a census of fishing sites conducted in 2007. The census of fishing sites contains data for 686 fishing sites. Ideally, the census contains information for all fishing sites in Burkina Faso. Variables measured in the census include whether the waterbody is permanent or not, dominant fishing time and dominant species.

Waterbody type by region

Figure 8 shows that the majority of the waterbodies (508 out of 686) are permanent. The Nord region is the only region in which the majority of the waterbodies are non-permanent. It may also be seen that Boucle du Mouhoun has the largest number of fishing sites.
Figure 8. Permanent and non-permanent waterbodies by region (Yes = permanent, no = non-permanent; far right column is total).

<table>
<thead>
<tr>
<th>Region</th>
<th>no</th>
<th>yes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CENTRE EST</td>
<td>5</td>
<td>23</td>
<td>28</td>
</tr>
<tr>
<td>CENTRE SUD</td>
<td>4</td>
<td>29</td>
<td>33</td>
</tr>
<tr>
<td>CENTRE</td>
<td>7</td>
<td>27</td>
<td>34</td>
</tr>
<tr>
<td>SUD OUEST</td>
<td>3</td>
<td>37</td>
<td>40</td>
</tr>
<tr>
<td>NORD</td>
<td>28</td>
<td>13</td>
<td>41</td>
</tr>
<tr>
<td>CASCADES</td>
<td>7</td>
<td>38</td>
<td>45</td>
</tr>
<tr>
<td>CENTRE NORD</td>
<td>13</td>
<td>42</td>
<td>55</td>
</tr>
<tr>
<td>HAUTS-BASSINS</td>
<td>8</td>
<td>48</td>
<td>56</td>
</tr>
<tr>
<td>PLATEAU CENTRAL</td>
<td>30</td>
<td>33</td>
<td>63</td>
</tr>
<tr>
<td>SAHEL</td>
<td>26</td>
<td>40</td>
<td>66</td>
</tr>
<tr>
<td>CENTRE OUEST</td>
<td>33</td>
<td>35</td>
<td>68</td>
</tr>
<tr>
<td>EST</td>
<td>5</td>
<td>68</td>
<td>73</td>
</tr>
<tr>
<td>BOUCLE DU MOUHOUN</td>
<td>9</td>
<td>75</td>
<td>84</td>
</tr>
</tbody>
</table>

Total: 178, 508, 686

Source: Authors using fishery data of Burkina Faso, 2006-2008

Dominant fishing time by region

The fishing season typically starts in January and ends in December.

Figure 9. First fishing month by region.

Source: Authors using fishery data of Burkina Faso, 2006-2008
Figure 10. Last fishing month by region.

<table>
<thead>
<tr>
<th>Region</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOUCLE DU MOUHOUN</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>15</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>55</td>
</tr>
<tr>
<td>CASCADES</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>CENTRE</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>CENTRE EST</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>CENTRE NORD</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td>3</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>CENTRE OUEST</td>
<td>3</td>
<td>9</td>
<td>8</td>
<td>13</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>CENTRE SUD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>EST</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>12</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>46</td>
</tr>
<tr>
<td>HAUTS-BASSINS</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>42</td>
</tr>
<tr>
<td>NORD</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>PLATEAU CENTRAL</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>SAHEL</td>
<td>2</td>
<td>6</td>
<td>14</td>
<td>14</td>
<td>16</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>SUD OUEST</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>17</td>
</tr>
</tbody>
</table>

Source: Authors using fishery data of Burkina Faso, 2006-2008

**Dominant species by region**

Each fishing site has three variables for the dominant species. Dominant species 1 is the most common species, dominant species 2 is the second most common, and dominant species 3 is the third most common species. The vast majority of fishing sites list *carpes* (carps) as the most common species and *silures* (silurus) as the second most common. The third most common species is typically *capitaines* (captain fish) or other (*autres*). It should be noted that species information is only available for 684 of the 686 sites.

Figure 11. Dominant species 1.

<table>
<thead>
<tr>
<th>Region</th>
<th>carpes</th>
<th>silures</th>
<th>capitaines</th>
<th>crevettes</th>
<th>autres</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOUCLE DU MOUHOUN</td>
<td>62</td>
<td>16</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CASCADES</td>
<td>37</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CENTRE</td>
<td>29</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CENTRE EST</td>
<td>23</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CENTRE NORD</td>
<td>49</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CENTRE OUEST</td>
<td>54</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CENTRE SUD</td>
<td>27</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>EST</td>
<td>56</td>
<td>14</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HAUTS-BASSINS</td>
<td>47</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>NORD</td>
<td>27</td>
<td>12</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>PLATEAU CENTRAL</td>
<td>53</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SAHEL</td>
<td>49</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SUD OUEST</td>
<td>20</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Source: Authors using fishery data of Burkina Faso, 2006-2008
2.2.2 Fishing site survey

The second data source is from an on-site survey of fishers at fishing sites conducted in 2008. The census described in section 2.2.1 provided the frame for the fishing site survey. All fishing sites with at least seven fishermen were included in the sample, which included approximately 80 percent of the fishers. In the sample design, each sampled site was visited twice a month during 2008. The time-points were selected using a one-per-stratum sample design where the
strata were 15-day intervals. For each visit, the data collector recorded the total number of landings and interviewed the fisher for three landings selected by systematic sampling. The figures below display a subset of the variables recorded on the fishing site survey.

**Average catch weight against survey time-point by region with 95% confidence intervals (CIs)**

The regions with the highest average catch weights were Hauts-Bassins, Sahel and Nord.

*Source: Authors using fishery data of Burkina Faso, 2006-2008*
Average catch weight against survey time point by species with 95% CIs

The two most common species, carpes and silures, typically had the highest average catch.

Source: Authors using fishery data of Burkina Faso, 2006-2008
Average catch weight by waterbody type (yes = permanent, no = not permanent)

Average catch weights for non-permanent water exhibited more variation than average catch weights for permanent water.

Source: Authors using fishery data of Burkina Faso, 2006-2008

Conversion factors from heaps to kg by species and time point

For fishing sites without weighing stations, catch was originally measured in non-standard units. The most common non-standard unit was heaps. If a non-standard unit was used, the data collector also provided a conversion factor relating the non-standard unit to kg. The conversion factor applied was the number of kg in one non-standard unit. For heaps, the conversion factor was the number of kg in a heap.
Figure 14. Conversion factors from heaps to kg (number of kg per heap), by species and time point.

Source: Authors using fishery data of Burkina Faso, 2006-2008

Demographics

The majority of fishers were men. Among the sampled landings where the fisher was the household head, 0.3 percent of fishers were female. Among the sampled landings where the fisher was not the household head, 4.6 percent of fishers were female.
Figure 15. Number of household heads by sex and region (unweighted).

<table>
<thead>
<tr>
<th>Region</th>
<th>Homme</th>
<th>Femme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boucle du Mouhoun</td>
<td>1558</td>
<td>11</td>
</tr>
<tr>
<td>Cascades</td>
<td>397</td>
<td>8</td>
</tr>
<tr>
<td>Centre</td>
<td>580</td>
<td>0</td>
</tr>
<tr>
<td>Centre - Est</td>
<td>405</td>
<td>1</td>
</tr>
<tr>
<td>Centre - Nord</td>
<td>528</td>
<td>0</td>
</tr>
<tr>
<td>Centre - Ouest</td>
<td>164</td>
<td>0</td>
</tr>
<tr>
<td>Centre - Sud</td>
<td>366</td>
<td>0</td>
</tr>
<tr>
<td>Est</td>
<td>1061</td>
<td>0</td>
</tr>
<tr>
<td>Hauts - Bassins</td>
<td>1393</td>
<td>3</td>
</tr>
<tr>
<td>Nord</td>
<td>167</td>
<td>0</td>
</tr>
<tr>
<td>Plateau Central</td>
<td>595</td>
<td>1</td>
</tr>
<tr>
<td>Sahel</td>
<td>544</td>
<td>0</td>
</tr>
<tr>
<td>Sud - Ouest</td>
<td>296</td>
<td>0</td>
</tr>
</tbody>
</table>

Source: Authors using fishery data of Burkina Faso, 2006-2008

Figure 16. Number of non-household heads by sex and region (unweighted).

<table>
<thead>
<tr>
<th>Region</th>
<th>Homme</th>
<th>Femme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boucle du Mouhoun</td>
<td>842</td>
<td>171</td>
</tr>
<tr>
<td>Cascades</td>
<td>149</td>
<td>9</td>
</tr>
<tr>
<td>Centre</td>
<td>229</td>
<td>0</td>
</tr>
<tr>
<td>Centre - Est</td>
<td>298</td>
<td>0</td>
</tr>
<tr>
<td>Centre - Nord</td>
<td>393</td>
<td>0</td>
</tr>
<tr>
<td>Centre - Ouest</td>
<td>364</td>
<td>2</td>
</tr>
<tr>
<td>Centre - Sud</td>
<td>183</td>
<td>0</td>
</tr>
<tr>
<td>Est</td>
<td>569</td>
<td>0</td>
</tr>
<tr>
<td>Hauts - Bassins</td>
<td>362</td>
<td>15</td>
</tr>
<tr>
<td>Nord</td>
<td>38</td>
<td>0</td>
</tr>
<tr>
<td>Plateau Central</td>
<td>313</td>
<td>0</td>
</tr>
<tr>
<td>Sahel</td>
<td>215</td>
<td>0</td>
</tr>
<tr>
<td>Sud - Ouest</td>
<td>170</td>
<td>2</td>
</tr>
</tbody>
</table>

Source: Authors using fishery data of Burkina Faso, 2006-2008
2.2.3. Population census

The final data source is the 2006 population census. The data set contains households that reported participation in fishing on the 2006 population census. The census collected demographic characteristics as well as the types of agriculture and plants grown at the household.

Figure 17 below shows that the proportion of fishing households with a female household head was slightly higher than the proportion of female fishers in the fishing site survey. In the census, 4 percent of household heads were female.

**Figure 17. Sex of head of household by region (1 = male; 2 = female).**

<table>
<thead>
<tr>
<th>Region</th>
<th>1</th>
<th>2</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boucle du Mouhoun</td>
<td>5236</td>
<td>226</td>
<td>0</td>
</tr>
<tr>
<td>Cascades</td>
<td>1294</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Centre</td>
<td>803</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>Centre-Est</td>
<td>1383</td>
<td>77</td>
<td>0</td>
</tr>
<tr>
<td>Centre-Nord</td>
<td>2804</td>
<td>84</td>
<td>0</td>
</tr>
<tr>
<td>Centre-Ouest</td>
<td>3821</td>
<td>246</td>
<td>0</td>
</tr>
<tr>
<td>Centre-Sud</td>
<td>1057</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>Est</td>
<td>3626</td>
<td>60</td>
<td>0</td>
</tr>
<tr>
<td>Hauts-Bassins</td>
<td>4435</td>
<td>183</td>
<td>0</td>
</tr>
<tr>
<td>Nord</td>
<td>966</td>
<td>62</td>
<td>1</td>
</tr>
<tr>
<td>Plateau Central</td>
<td>1301</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>Sahel</td>
<td>1239</td>
<td>46</td>
<td>0</td>
</tr>
<tr>
<td>Sud-Ouest</td>
<td>1770</td>
<td>161</td>
<td>0</td>
</tr>
</tbody>
</table>

*Source: Authors using fishery data of Burkina Faso, 2006-2008*
3. Implementation of illustrations

The illustrations below relate to issues that emerge from the Burkina Faso data and approaches used to collect fisheries data. For each illustration below, the connection to data collection for fisheries in Burkina Faso is first explained. Then, the methodology and results are described. Each illustration is concluded with a brief discussion of the main implications.

3.1. Illustration 1

Illustration 1 relates to two issues arising the Burkina Faso survey procedures. The first is sustainability of the survey plan over time. The second is variance estimation. First, problems associated with sustainability over time and variance estimation in the Burkina Faso survey plan are explained. Then, a simple illustration using the Burkina Faso data is devised to illustrate how sampling can address these two issues. In particular, in the illustration, sampling allows for a more efficient allocation of monetary resources and variance estimation.

3.1.1. Sustainability and variance estimation in the Burkina Faso survey procedures

In the Burkina Faso survey plan, a full census of fishing sites was completed in 2007, and a survey was conducted in 2008 that used the information from the fishing site census as the frame. In the 2008 survey, all fishing sites with at least seven fishermen were visited twice a month. This effort presumably required a heavy investment for the two years 2007 and 2008 and allowed estimation of level for 2008. In the framework of a regional project of the West African Economic and Monetary Union, another census was implemented in 2011 to set up a monitoring system, with the support of French Research Institute for Development. An important limiting factor in developing monitoring programmes and ensuring the continuity of survey programmes, which will be addressed here, relates to resource allocation. Rather than invest heavy resources in a tremendous survey effort for a single time-point, resources can be distributed across multiple activities. Visiting a sample of fishing sites instead of a census of fishing sites can reduce the cost of the survey. The savings can then be spent on other activities, such as multiple surveys over time, frame updates or both.

A second issue in the 2008 survey relates to variance estimation. The sample design consisted of a one-per-stratum sample of time-points followed by a systematic sample of landings. Because second-order inclusion probabilities are zero for one-per-stratum designs and systematic samples, variance estimation is
not possible for the Burkina Faso sample design. Only sample designs for which a design-consistent variance estimator can be constructed are considered here.

### 3.1.2. Implementation of illustration 1

An illustration using the Burkina Faso data is devised to demonstrate how sample designs that allow for variance estimators can reduce costs and enable the savings to be used for other activities. Two parameters are considered: average catch (defined in the appendix to this publication) and the proportion of landings for which the fisher’s primary occupation is agricultural. The aim is to construct a sample design that will allow for sufficient resources to update the frame in the year preceding the survey.

The sample for the 2008 survey provides the basis for this illustration. In the Burkina Faso data, the sites for which demographic variables are recorded differ slightly from the sites for which catch-related variables were recorded (this difference was probably due to nonresponse). The sites that contain information for both demographic and catch-related variables are considered, resulting in 276 sites. In the 2008 survey protocol, each of these 276 sites was visited 24 times in a year, which would require 6,624 total visits to fishing sites. The 276 sites are treated as the population of sites for the purposes of this illustration. It is supposed that the cost structure is such that implementing the survey protocol for a single time-point at a selected fishing site costs twice as much as implementing the census protocol. With these assumptions, it is possible to update the frame containing the population of 276 fishing sites if the number of visits in the sample can be reduced by 138.

Using an outdated frame introduces the potential for bias. Three values for the bias are considered. For the bias value called “Random” in table 3 below, a random sample of 10 percent of fishing sites was removed from the set of possible fishing sites. For the bias value called “Large sites”, 10 percent of the fishing sites with the largest average catch was removed. For the bias value called “Small sites”, 10 percent of the fishing sites with the smallest average catch was removed. The bias for the two parameters’ average catch and the proportion with primary occupation of agriculture was calculated, the bias being the difference between the average catch or proportion with primary occupation of agriculture based on the subset and the corresponding mean based on the full set of fishing sites.

Table 3 below shows the biases for the two parameters and the three methods of calculating bias. The bias is multiplied by 10,000 to simplify the presentation. As expected for the parameter average catch, the “Random” bias is close to zero,
while the biases for the “Large sites” and “Small sites” procedures are positive and negative, respectively. A comparison of the biases for the random and large sites methods for the proportion in agriculture indicates that sites with a high average catch tend to have lower proportions of agricultural involvement. This finding seems reasonable, because individuals reaping the benefits of a high average catch may have relatively less need to participate in agriculture.

Table 3. Biases (multiplied by 10 000) for two parameters and three methods.

<table>
<thead>
<tr>
<th></th>
<th>Random</th>
<th>Large sites</th>
<th>Small sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average catch</td>
<td>587</td>
<td>-6 281</td>
<td>3 926</td>
</tr>
<tr>
<td>Proportion</td>
<td>-92</td>
<td>178</td>
<td>-33</td>
</tr>
<tr>
<td>agricultural</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors using fishery data of Burkina Faso, 2006-2008

A sample design that reduces the number of visits to sites by approximately 138 is defined. For each region, a random site is selected. That site is removed from the sample for half of the survey time-points. For the target number of 24 visits per site, this would reduce the total number of visits by 156. For each month, this sampling procedure yields a random sample of sites, such that variance estimation is possible.

The benefit of selecting a sample from an updated frame is a reduction in bias. It is supposed that a survey from an outdated frame is biased and a survey from an updated frame leads to unbiased estimators. The sampling procedure results in a smaller MSE than visiting all sites 24 times per year if the variance of the survey estimator is smaller than the squared bias. Although it is possible to calculate the variance analytically for this design, using Monte Carlo is simpler. Therefore, this sampling procedure is repeated 100 times and the average catch and proportion with primary occupation of agriculture are estimated for each of the 100 samples. The Monte Carlo bias and Monte Carlo variance of the estimators are estimated.

The table below shows the Monte Carlo bias, standard deviation, and root mean squared error (RMSE) of the estimators of the two parameters: average catch and proportion with an agricultural principal occupation. As for table 3, the bias and variance in table 4 are multiplied by 10 000. The bias, as expected, clearly makes a negligible contribution to the MSE. The sampling standard deviations in table 4 are smaller than the biases in table 3. For this population configuration and design, selecting a sample and updating the frame leads to more efficient estimators than can be calculated by visiting all sites at every time-point.
Table 4. Monte Carlo bias, standard deviation, and root mean squared error (multiplied by 10 000).

<table>
<thead>
<tr>
<th></th>
<th>Bias</th>
<th>Standard deviation</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average catch</td>
<td>75</td>
<td>189</td>
<td>203</td>
</tr>
<tr>
<td>Proportion</td>
<td>-5</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>agricultural</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors using fishery data of Burkina Faso, 2006-2008

3.1.3. Summary and discussion of illustration 1

This illustration compares two approaches to sampling. For the first approach, all sites are visited at all time-points, and at the expense of the large sample, the frame is not updated. This approach is represented through estimators that are biased but have zero variance. As an alternative, a procedure that spreads resources between the processes of conducting a survey and updating the frame is considered. A cost structure in which the two approaches have the same cost is considered. The Monte Carlo MSEs in table 4 represent the MSEs from the procedure that involves both updating the frame and selecting the sample. The absolute values of the biases in table 3 represent the RMSEs from the procedure in which no funding was reserved for updating the frame. For the three biases considered, sampling and updating the frame was more efficient than visiting all sites contained on an outdated frame.

In this illustration, the problem of updating the frame and conducting a survey for a single time-point is considered. The same principle applies to conducting surveys over time. Reducing the sample size for a single time-point can provide resources for conducting multiple surveys over time.

3.2. Illustration 2

Illustration 2 relates to two issues arising in the data for Burkina Faso. The first is the problem of estimating both demographic and catch-related statistics in a coherent fashion. The second is the problem of obtaining complete coverage of the population. Bayley and Petrere (1983) propose a sample design procedure in which villages are selected from a frame of villages and fishing sites are then associated with the sampled villages. An extension of this approach is considered, whereby a sample of villages is first selected. For each sampled village, two further subsamples are then selected: the first is a sample of households, and the second is a sample of landing sites associated with each
village. It is proposed to estimate demographic characteristics from the household sample and catch-related statistics from the landing site sample. First, an explanation is provided of how issues in the Burkina Faso data motivate this investigation. The methodology and results of the illustration are then described.

3.2.1. Estimating demographics and coverage for Burkina Faso

Both the landing site survey and the population census record the sex of the household head. The proportion of female household heads in the data from the landing site survey is 0.003. The proportion of female household heads in the data from the population census is 0.042. The estimate from the population census is an order of magnitude greater than the sample proportion from the landing site survey. One possible reason for the difference is that these two surveys estimate two different population parameters. The landing site survey estimates the proportion out of all landings, while the census estimates a proportion out of all households. This illustrates difficulties documented elsewhere (Pollock, Jones and Brown, 1994) associated with estimating demographic characteristics from landing site surveys.

The second issue is related to coverage. Ideally, the landing site frame census covers all fishing sites. Nonetheless, the possibility exists that some sites are omitted. If the census truly contains all villages, then enumerating all fishing sites where individuals fish in sampled villages has the potential to identify sites that have been excluded from the landing site census.

3.2.2. Implementation of illustration 2

A scenario that takes into consideration the different strengths of the household survey and the landing site survey involves a two-phase design. First, a sample of villages is selected. Then, all landing sites where individuals fish in each sampled village are enumerated. Next, two second-phase samples are designed. The first is a sample of households, and the second is a sample of landings.

Because the Burkina Faso data do not support an investigation of the full procedure, a simplification is considered. First, villages in all of the following three datasets are identified: the census file, the survey dataset with catch-related variables and the survey dataset with demographic variables. This set of villages is considered the population.

Two population parameters are considered: the proportion of household heads that are female and average catch. The population parameter for the proportion of females is deemed the proportion of households in the census with a female household head. This population proportion is 0.028, which differs from the
figure of 0.042 cited above because we subset the villages with demographic and catch-related variables in the fishing site survey. The population average catch is defined as the average catch from the landing site survey data, which is 4.67.

A simple sample design is considered. Although the simple design is not particularly realistic, it allows for illustration of the point. The design is an equal-weighted two-stage design. In the first stage, a stratified random sample of villages is selected, where 50 percent of villages are selected from the population of villages in each region. Then, two samples are selected: one of households and the second of fishing sites. To select the sample of households, 50 percent of the households in each sampled village are selected. To select the sample of fishing sites, 50 percent of the fishing sites associated with each sampled village are selected and all data associated with each sampled fishing site are used. The proportion of females and average catch is estimated from the landing site survey (the proportion of females is used for the sake of simplicity). The proportion of female household heads is estimated from the household survey. Because this is an equal-weighted design, estimates are simple means.

A Monte Carlo study is conducted where this sample design is implemented 100 times. The figure below presents histograms of the 100 estimates. The black vertical line corresponds to the population parameter of interest, and the red vertical line is the Monte Carlo mean of the parameter of interest. The landing site survey is approximately unbiased for the average catch, and the household survey is approximately unbiased for the proportion of female household heads. The landing site survey produces an estimator with a negative bias for the proportion of females. This negative bias is expected because the proportion of females in the population defined by the landing site data set is 0.025 after restricting to the subset of villages in all three required data sets.
3.2.3. Summary and discussion of illustration 2

Figure 18. Histograms of 100 estimates from village sample.

This illustration may be viewed as somewhat artificial; however, it demonstrates the feasibility of using two-stage design to overcome the issues identified in section 3.2.1. A single frame can be used – the village frame – to locate both households and fishing sites. Catch-related statistics can be estimated from the landing site sample, and demographic statistics from the household sample. It is possible to identify fishing sites that have been omitted from a list of sites in the first-stage village sample. It may be possible to save costs if maintaining the village list is less expensive than maintaining a list of fishing sites. Maintaining a village list may enable sharing costs between other activities, such as population and agricultural censuses.

One potential inefficiency associated with the village list is overcoverage. Coupling the survey from a village list with a survey from a list of fishing sites through a dual-frame design is one way to improve efficiency. Dual-frame designs are considered in the next illustration.

3.3. Illustration 3

In many agricultural multiple-frame surveys, one sample is selected from an area frame and the other sample is selected from a list frame. The area frame may cover the entire population but may be inefficient. The list frame may have undercoverage. The two samples could be used jointly to estimate the undercoverage of the list frame and to construct dual-frame estimators of population parameters. A dual-frame survey design is considered in this illustration.
3.3.1. Relevance of the dual-frame design to the Burkina Faso scenario

One problem discussed in illustration 2 is the possible overcoverage of the village frame. In practice, many villages may have zero fishing households, which could cause the design illustrated in section 3.2 to be inefficient. In a multiple-frame survey, independent samples from different frames are combined to produce a single estimator of a population parameter. One frame may be a relatively inefficient frame; an example is the village frame from illustration 2, that covers the entire population but also contains ineligible elements. The other frame may be a more efficient frame, such as the frame of fishing sites. The efficient frame may be relatively rich with auxiliary information and contain few ineligible population elements; however, it may also fail to cover the entire population. Rather than consider the village frame and the fishing site frame, as for illustration 2, a dual-frame design with an area frame and a list frame is considered.

3.3.2. Construction of the population, samples and estimators for the dual-frame design

The first frame is an area frame. To define an area frame, the starting point is the land cover map in figure 19. To approximately isolate the waterbodies, locations with color code less than or equal to 60 are identified. Figure 19 below shows the resulting map, in which green and yellow interior locations represent water. The area frame is defined to consist of the collection of water locations. It is assumed that the water locations cover all possible fishing sites. This results in 1 640 possible fishing sites.

The second frame is the list frame of fishing sites. It is assumed that the list frame of fishing sites contains 80 percent of all fishing sites, and that the area frame covers 100 percent of fishing sites. To construct this situation with the Burkina Faso data, fishing site IDs are randomly assigned to the yellow and green locations on the map in figure 19. This is done by taking 1 640 independent draws with replacement from the set of possible site IDs. The activity is restricted to the fishing site IDs for which catch statistics have been recorded for the 2008 survey. Then, an indicator variable to represent presence or absence on the fishing site frame is randomly assigned. The indicators are independent Bernoulli trials with a success probability of 80 percent.
Independent samples are selected from the two frames. A sample of size 100 is selected from the area frame. Independently, a sample of size 100 is selected from the set of locations contained on the list frame. This second sample represents the process of selecting an independent sample from the list frame.

Estimating total catch is considered. The population parameter of total catch for the constructed population is 681 245.8 kg. Hartley’s (1962) dual-frame estimator and variance estimator are used. Table 5 compares the estimate and 95 percent confidence interval to the population parameter for one selected sample. The population parameter is contained in the 95 percent confidence interval.

**Table 5. Estimate and 95 confidence interval for total catch from dual-frame design.**

<table>
<thead>
<tr>
<th>Population parameter</th>
<th>Estimate</th>
<th>Lower interval endpoint</th>
<th>Upper interval endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>681 245.8 kg</td>
<td>686 611.6 kg</td>
<td>584 483.9</td>
<td>788 739.3</td>
</tr>
</tbody>
</table>

Source: Authors using fishery data of Burkina Faso, 2006-2008
To verify that the procedure works as it should, a Monte Carlo study is conducted in which 100 samples are selected using this estimation procedure and 100 estimates are constructed. The ratio of ten times the Monte Carlo bias to the Monte Carlo standard deviation (a t-statistic for bias) is -0.27, supporting the theory that the estimator is approximately unbiased. The empirical coverage of the 95 percent confidence intervals is 97 percent.

This illustration focuses on demonstrating the dual-frame approach and validating the approach through simulation. Alternatives to the dual-frame approach are possible. One alternative would select a sample using only the area frame. If surveying a unit from the area frame is more expensive than surveying a unit from the list frame, then the sample size for a survey from only the area frame is necessarily smaller than the sample size from the dual-frame design to maintain a fixed cost. In addition, selecting a sample from only the area frame does not provide a mechanism for estimating the undercoverage of the list frame. This illustration is a simplification of reality, in that the area frame and the list frame are perfectly matched prior to sample selection. For this illustration, one could consider an approach that uses this information on linkages – for instance, selecting one sample from the list frame and a second sample from the complement of the list frame. Because a situation where the two frames are perfectly linked in advance of the survey is unlikely to arise in practice, this comparison is not pursued for the purposes of this illustration. Further study of the dual-frame design, with comparisons to alternatives, in the context of fishery statistics is an area for future research.

3.3.3. Discussion of illustration 3

This illustration demonstrates the process of constructing a dual-frame estimator using the Burkina Faso data. In the illustration, an approximately unbiased estimator of the total annual catch is obtained and confidence intervals with empirical coverage close to the nominal level is constructed. The dual-frame estimator allows for taking advantage of the relative strengths and weakness of different types of frames.

3.4. Illustration 4

The final illustration is related to measurement but has connections to MSF. When a weighing station is available, measuring yield directly in kg is possible. When a site does not have a weighting station, a data collector records a subjective measurement of yield. If the data collectors’ subjective determinations differ systematically from the direct measurement of weight in kg, then subjective assessments can lead to biased estimators of catch.
3.4.1. Measurement issues in the data for Burkina Faso

In the fishing site survey for Burkina Faso, catch is ultimately reported in kg. If the fishing site does not have a weighing station, then catch cannot be directly measured in kg. Instead, catch is measured first in units of either “heaps,” “baskets” or “other”. The data collector then establishes a conversion factor that relates the number of kg per unit (that is, kg per heap). The majority of measurements that are not in kg are measured in heaps; therefore, the population is defined to be the subset of survey data in which the original measurement is in units of heaps. This results in a population of size 10,843. Systematic differences between measurements originating in heaps and direct measurements of kg lead to biased estimators of catch. This illustration mimics the process of determining catch using two methods for a subset of landings. Then, a relationship between actual catch weight (in kg) and the data collector’s determination of catch weight is estimated. This estimated relationship is then applied to the full sample to obtain an approximately unbiased estimator of average catch.

3.4.2. Methodology for illustration 4

To mimic a situation in which two measurements of weight for a sample of sites are available, “true weights” for the population of 10,843 are simulated. A model for the relationship between the conversion factor (kg/heap) that the data collector establishes subjectively and the actual conversion factor is postulated. Letting \( x \) and \( y \) be the subjective and true conversion factors, respectively, it is assumed that \( y = 0.95x + e \), where \( e \) has a normal distribution with mean 0 and standard deviation 0.02. Then, the true catch is defined to be the product of \( y \) and the number of observed heaps. The means of the directly determined catch weight and the indirect catch weight based on the data collector’s assessment are, respectively, 5.68 and 5.39.

A total sample of size 200 is selected. For a random subsample of the total sample of size 100, it is supposed that two conversion factors are obtained: a true conversion factor and a conversion factor based on a data collector’s subjective assessment. The ordinary least-squares regression is fit with the true conversion factor as the response and the observed conversion factor as the explanatory variable. The estimated relationship for the subsample is used to predict the true catch weight for the remaining 100 sampled elements. The estimator of average catch is defined to be the mean of the 200 values, where 100 are observed values for true catch and the other 100 are the predicted values of true catch. This modified estimator is compared to a naïve estimator. The naïve estimator is the mean of the 200 values, where 100 are observed values for true catch and the
other 100 are the observed catch measurements from the subjective assessment. This second estimator is analogous to the procedure used for the Burkina Faso 2008 survey.

The sampling and estimation procedure are repeated for a Monte Carlo sample of size 100. A t-statistic for the bias of the modified estimator is 0.796, indicating that the modified estimator is essentially unbiased. In contrast, the t-statistic for the naïve estimator is 3.24, showing a significant bias.

### 3.4.3. Summary of illustration 4

In fisheries surveys, measurement issues are closely tied to the choice of the frame. Pollock, Jones and Brown (1994) discuss these connections at length. In a canonical case considered in this illustration, an inexpensive measurement differs systematically from an expensive measurement that targets the true value of interest. For this situation, estimates based on the inexpensive measurement are biased. This illustration mimics the process of using both measurement techniques for a subset of samples. The relationship between the two measurement methods for the subsample is estimated. Extrapolating that relationship to the entire sample allows for obtaining approximately unbiased estimators of catch weight.

### 4. Summary and recommendations

Fisheries surveys target diverse populations with unique characteristics. The strengths and weaknesses of different types of frames depend on the types of parameters and populations of interest. An MSF should make an effort to take advantage of the relative strengths of different types of lists for different situations. Different types of lists are often associated with different contact methods. Thus, the choice of the frame is tied not only to the sampling variance but also to important measurement issues.

Combining different types of lists, contact methods and measurement techniques through the MSF can be an effective way to take advantage of the relative strengths of different approaches. In illustration 2, a list of villages is used as the basis for both a household survey and a landing site survey. In illustration 3, a dual-frame design is considered. In illustration four, two different measurements of catch are obtained—one using an inexpensive method and the other using a more expensive method—for a subsample.

Another theme in the illustrations is the value of using probability samples to reduce the cost of the survey. This can enable the survey organization to devote
remaining resources to reducing non-sampling errors. Illustration 1 constructs a situation in which reducing the survey sample size to allow for more frequent frame updates can lead to more efficient estimators than estimators based on a large sample from an outdated frame. Illustration 4 shows that devoting resources towards obtaining two determinations of catch weight may reduce the bias relative to an estimator based only on subjective determinations. Using probability samples can also enable spreading resources over several years, which can improve the temporal continuity of survey estimation programmes.

The messages from the illustrations above reinforce well-known principles of survey sampling. The four illustrations demonstrate the advantages of using multiple lists, multiple measurement instruments and probability samples. Implementing the procedures using the data for Burkina Faso exemplifies the relevance of fundamental statistical principles for developing countries.
Annex references


Appendix

Definition of average catch

Let $y_{itjk}$ index the catch weight (kg) for fishing site $i$, time-point $t$, landing $j$ and species $k$. The average catch is

$$N^{-1} \sum_{i=1}^{I} \sum_{t=1}^{T} \sum_{j=1}^{J} \sum_{k=1}^{K} y_{itjk} \delta_{itjk},$$

where $I = \sum_{i=1}^{I} \sum_{t=1}^{T} \sum_{j=1}^{J} \delta_{itjk}$, and $I$ is an indicator variable that is equal to 1 if a catch weight is recorded for the combination $itjk$ and is zero otherwise. Typically, $T = 24$ and $J = 3$. The specific values of $I$ and $K$ depend on the nature of the illustration and the context.