ICES WKPRECISE REPORT 2009

ICES Advisory Committee

ICES CM 2009/ACOM:40

Ref. PGCCDBS

Report of the Workshop on methods to evaluate and estimate the precision of fisheries data used for assessment (WKPRECISE)

8-11 September 2009

Copenhagen, Denmark

ICES
International Council for the Exploration of the Sea
CIEM
Conseil International pour l'Exploration de la Mer

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

H. C. Andersens Boulevard 44–46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

Recommended format for purposes of citation:

ICES. 2010. Report of the Workshop on methods to evaluate and estimate the precision of fisheries data used for assessment (WKPRECISE), 8-11 September 2009, Copenhagen, Denmark. ICES CM 2009/ACOM:40. 43 pp.

Contents

Executive summary

Methods to evaluate the accuracy of fisheries statistics used for assessments are documented in the workshops WKPRECISE (held in Copenhagen from September 8-11 2009) and WKACCU (held in Bergen, Norway, 27–30 October 2008). Accuracy is determined by the amount of bias and the precision of estimates of key parameters. The WKACCU workshop focused on identifying the sources of bias in parameter estimates and data collection procedures used to assess national level fisheries data. The WKPRECISE workshop focused on sources of variability and on the procedures to estimate the precision of national level fishery statistics (quantities landed, discards, fishing effort, CPUE) and biological data collected from the fisheries. While precision of fisheries statistics can be improved by increasing the sample sizes in data collection programs, this will generally not reduce bias. It was recognized by WKPRECISE that measures of precision estimates based on fisheries data used for assessments only are meaningful for catch sampling programs that obtain representative data. A minimum requirement should be that the sampling programs pass basic checks for bias using the scorecard developed by WKACCU. Several national sampling programs were presented and reviewed during WKPRECISE. Discussions focused on survey design requirements and best practises in data collection programs that facilitate the quantification of precision of estimates based on national level fishery statistics (quantities landed, discards, fishing effort, CPUE). Procedures to assess the precision on a national level of biological data collected from the fisheries were examined. The WKPRECISE documented the complexity of typical fisheries sampling programs, including stratification and further grouping into métiers. Estimators of precision for key parameters must take into account clustering effects that are caused by multi-stage sampling. Cluster sampling is generally unavoidable in biological data collection programs for fisheries, and is also motivated by practical and administrative considerations. Catch sampling is generally done in multiple stages, with stratification at several levels. In the first stage, a stratified sample of vessels/trips, for example, may be chosen for catch sampling. For selected trips, fish from the entire catch, or from sub-samples of catches, may be sorted by species, weighed, and measured for length. Age-samples are often collected in the last stage (e.g., within a sub-sample of catch from a trip), either randomly from a sub-sample of fish measured for length, or randomly within length classes (length-stratified sampling). This multi-stage sampling requires that estimators of key parameters for total catches in a fishery appropriately accounts for the hierarchical structure of the data, and in particular clustering effects that may drastically reduce the effective sample sizes. The reason is that fish caught during one trip often have more similar characteristics then the general population of fish from all trips. Assumptions of simple random samples of fish from the total population of fish in the catches cannot be reasonably met when sampling is done in multiple stages.

Sources of variability for key parameters were examined by WKPRECISE, but only general guidelines for the estimation of precision could be documented as the detailed specification of estimators are directly linked to the particulars of survey designs employed in each program. The WKPRECISE workshop reviewed a list of key parameters & statistics used for supporting stock assessments and to identify approaches and recommendations for estimating precision. The focus was on the same tables of parameters that were documented in WKACCU (2008) for consistency, and on documenting cases where the analysis tools developed in the EU COST project were appropriate for estimating parameters and precision. The Cost project devel-

oped open source software in R that can be used to estimate fisheries statistics and associated precision for a wide range of catch sampling programs. The estimation of precision for métiers received special attention during WKPRECISE. A métier or other study domains chosen in advance may coincide with the strata adopted for stratified sampling, or may cut across them. Standard stratified estimators cannot be applied when métiers cut across strata. The estimation of precision of estimates for métiers must take into account that the statistical weights for the samples will depend on the stratum in which they belong, and often involves post adjustments of sample weights. Post-stratification of the samples (primary sampling units) may be used when the total catches in each métier is known. This is the method implemented in the COST program. The pooling of métiers also requires post-adjustments of sample weights, or imputations when data are missing in some métiers. It was noted by WKPRECISE that the practice of quota sampling to achieve a desired sample size (e.g., number of trips) in métiers may result in biased estimates of key parameters.

WKPRECISE recommends that catch sampling programs be based on statistically robust survey designs with clear definitions (and documentation) of; the sampling frame, the primary sampling units (PSUs), the stratification schemes employed, and the methods used for selecting samples in each stratum. The statistical estimation of precision requires that representative catch sampling be conducted using probability-based methods (to the extent possible within logistical constraints). Ad-hoc sampling rules out the estimation of precision and should be avoided. WKPRECISE also recommends that the precision of estimates of key parameters is given in terms of standard errors or relative standard errors (often referred to as the coefficient of variation for a parameter estimate). In addition, the number of primary sampling units observed along with estimates of the effective sample size for the associated estimate should be given. This is because the variances of key estimates are typically driven by the number of PSUs sampled, and so the effective sample size is usually much smaller than the total number of individuals sampled. If age-length keys (ALKs) are used to estimate age-distributions, then it must be noted that the precision of such estimates cannot be evaluated unless the age-length data are coupled to the primary sampling units from which the age and length data were collected. The accuracy of estimates of age-distributions based on static ALKs that do not take into account the survey design of the catch sampling programs (or ALKs derived from ad-hoc sampling) cannot be assessed.

1 Introduction

1.1 Terms of reference

A **Workshop on Methods to evaluate and estimate the precision of fisheries data used for assessment** [WKPRECISE] (Co-chairs: Jon Helge Vølstad and Sondre Aanes IMR, Norway) was held at ICES Headquarters, 8–11 September, 2009 to:

a) Review the sources of variances and establish procedures to assess the precision on national level of fishery statistics (quantities landed, discards, fishing effort, CPUE) using available data, and advice on best practices;

b) Review the sources of variability and establish procedures to assess the precision on national level of biological data collected from the fisheries.

c) Suggest quality assurance indicators for the quantities described in a) and b) to be implemented by the quality assurance framework for assessment input data.

WKPRECISE will report by 30 October 2009 for the attention of PGCCDBS and ACOM.

Supporting Information

PRIORITY:	Very high priority because variance and/or precision is crucial for guiding action. In fact, imprecise data may result in wrong decision, leading to disasterous assessments and to waste of resources. It was decided that this workshop will follow after the WKACCU workshop which solely focus on the bias component of accuracy.
SCIENTIFIC USTIFICATION AND RELATION TO ACTION PLAN:	In the current DCR and other national sampling programs and -strategies, data quality is almost solely addressed by means of target precision levels for a number of fishery-related and stock-related parameters (fishing effort, quantities landed and discarded, age composition of the landings and discards, growth curves, maturity and fecundity ogives, etc.). However, it is not because an estimate is precise that it is also accurate. The workshop will aim at establishing standardized/joint methods and indicators for evaluating and estimating the precision of submitted fisheries data. Definitions of standards (i.e., minimum requirements) should be made. Some laboratories have already developed suitable tools for such precision estimation, e.g., the EU COST-project (EU FISH/2006/15:lot 2) and the Norwegian ECA-model may contribute to this issue.
RELATION TO STRATEGIC PLAN:	
RESOURCE REQUIREMENTS:	DCR (EU Data Collection Regulation) data collection system, and other national systems and data sources.
PARTICIPANTS:	In view of its relevance to the DCR, the Workshop is expected to attract wide interest from both ICES Member States and Mediterranean EU Member States. The workshop will benefit from the attendance of managers, fishers and people from the fishing industry.
SECRETARIAT FACILITIES:	ICES HQ general facilities
FINANCIAL:	To ensure wide attendance of relevant experts, additional funding will be required, preferably through the EU, e.g. by making attendance to the Workshop eligible under the DCR.
LINKAGES TO ADVISORY COMMITTEES:	ACOM and its assessment Working Groups.

LINKAGES TO OTHER COMMITTEES OR GROUPS:	This workshop was proposed by PGCCDBS. Outcomes from this Workshop will be of interest to the Living Resources Committee and the Resource Management Committee.
LINKAGES TO OTHER ORGANISATIONS:	There is a direct link with the EU DCR and outcomes from this Workshop will be of interest to several RFOs, including NEAFC, JNRFC, GFCM and NAFO.
COST SHARE:	

1.2 Background for the workshop

For the current Data Collection Regulations (DCR) in the EU, and for other national sampling programs and sampling strategies, the quality of the resulting estimates is primarily addressed by means of setting target precision levels for a number of fishery-related and stock-related parameters (fishing effort, quantities landed and discarded, age composition of the landings and discards, growth curves, maturity and fecundity ogives, etc.). However, even if an estimate is precise it is not necessarily accurate. For example, estimates of landings that are based on sales slips will usually be very precise, but they may be very inaccurate if there are significant unreported landings. Similarly, estimates of the length distribution of the landings may be very inaccurate if they only cover a small part of the spatial distribution of the total landings. Therefore, there is a need for objective indicators of data accuracy that could be taken into account when designing and evaluating sampling programs.

Methods to evaluate the accuracy of fisheries statistics used for assessments were addressed in the two workshops WKPRECISE (held in Copenhagen from September 8-11) and WKACCU (held in Bergen, Norway, 27–30 October 2008). The accuracy of estimates of key parameters is determined by the amount of bias and precision (Lessler and Kalsbeek 1992). The WKACCU workshop focused on the identification of sources of bias in parameter estimates and data collection procedures used to assess the quality of national level fisheries' data. Bias is a systematic departure of estimates from the true value caused by non-representative data collection, poor estimators, and other persistent factors. Bias can generally not be quantified because the true value is seldom known. The WKACCU workshop therefore developed recommendations on best practices to minimize or eliminate sources of bias in field data collection procedures and in analytical methods. A practical framework for detecting and flagging potential sources of bias in fisheries data collection programs was developed. A simple score-card of indicators of bias for a suite of parameters that are important for stock assessments was constructed. The scorecard can be used to evaluate the quality of data sources used for stock assessments and to reduce bias in future data collections by indentifying steps in the data collection process that must be improved.

WKPRECISE focused on the sources of variability and the procedures to estimate the precision of national level fishery statistics (quantities landed, discards, fishing effort, CPUE) and the quality of biological data collected from the fisheries. It should be noted that measures of precision of fisheries data used for assessment only are meaningful for sampling programs that pass basic checks for bias using the scorecard developed in WKACCU.

2 Adoption of the agenda

The list of participants and the adopted agenda is provided in Annex 1 and 2, respectively. A list of working documents and a compilation of the oral presentations during the workshop is in Annex 3. All the working documents and oral presentations are available from the author(s) or the co-chairs.

3 Sampling design and estimation of precision

3.1 Best practices in fishery sampling programs

To establish some basic terminology, let us consider that we may want to conduct a survey to estimate the composition (e.g., length distribution of a species) of the total catches taken by a fleet in a particular geographic region during a quarter. We could define our sampling frame to be all the fishing trips by all vessels in that region and quarter. Then the elements of this frame (primary sampling units) would be the individual fishing trips by each vessel. The unknown population parameter that we wish to estimate is the composition of the total catch taken by all vessels that fished in the region and quarter. This can be achieved by selecting a representative sample of trips, followed by representative sampling from the catches of each selected trip.

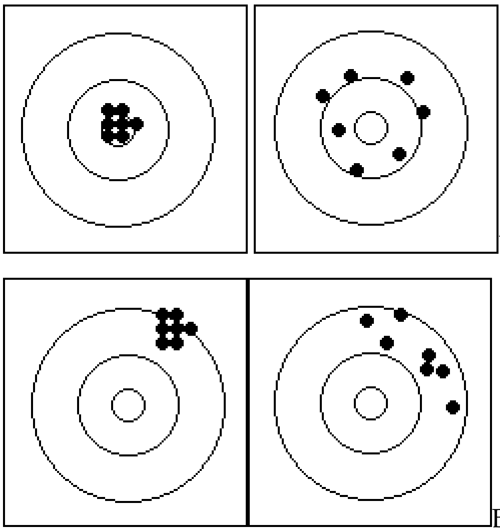
Virtually all sample surveys for the monitoring and assessment of fisheries are complex because of their multistage, stratified and clustered features (see Skinner et al. 1989; Lehtonen and Pahkinen 1996.). National- and international –level fisheries surveys often cover multiple fleets and gear types and additionally, government regulations often dictate that a wide range of fisheries statistics and biological parameters be quantified, and this adds to survey complexity.

The sampling design employed in fisheries monitoring programs should be carefully specified and documented, including a detailed description of the sampling frame, the sampling units in each stage, sample sizes at each level, detailed description of stratification of sampling units at each stage, etc. (see Cochran, 1977; Jessen 1978.) The procedures for selecting sampling units should be described and if it is not by probability-based selection the rationale for adopting an alternative procedure should be provided.

The key requirements of optimal schemes to collect biological data and statistics from fisheries are (a) to minimise bias and (b) to maximize precision for a given cost (or to minimize cost to achieve a specified precision.).

The following figure illustrates bias and precision for a parameter of interest, where the target, true value is the bull's-eye:

Precise Imprecise



Unbiased

Biased

Precise and unbiased estimates of the target values are accurate (upper left corner). It should be noted (and emphasized) that accurate estimates cannot be obtained from significantly biased sampling schemes. Indicators of bias developed by WKACCU 2008 are intended to identify the existence of bias in data collection schemes. Bias in fisheries sampling programs can arise because of problems such as fishing vessels refusing access to catches for sampling, even if the original sampling scheme was unbiased, or because the initial sampling scheme did not have a proper probability-based statistical design. A form of probability-based sampling is required to obtain reliable estimates of precision. The chance of each element (e.g., a fishing trip for a fleet-segment) being selected must be known, but need not be equal. Probability sampling in surveys ensures that the sampling errors can be estimated based on the data collected, which is not mathematically justifiable when ad-hoc sampling methods are used. Probability sampling must be used at each stage of the sample selection process in order to generate unbiased estimates with quantifiable precision. For example, the first selection stage may involve choosing vessels or trips, while the last stage involves selecting the samples of fish to be measured. There is no statistical theory to guide the use of ad-hoc samples, for which the reliability can only be judged subjectively. Failure to use probability-based sampling techniques will often lead to bias in survey estimates of unknown magnitude and direction (under-estimates or over-estimates). The precision of sampling estimates, for example as measured by their standard errors or relative standard errors, can only be estimated when probability sampling is employed.

Non-probability sampling is frequently used in fisheries monitoring programs in spite of theoretical deficiencies because of logistical constraints, cost considerations, and convenience. Judgmental sampling is an ad-hoc method that relies upon "experts" to choose the sampling units (e.g., trips or vessels), for example based on their opinions of which units are representative of the target fishery. This type of sampling is subjective and highly dependent on the choice of experts themselves, and should therefore be avoided if possible. With probability-based sampling, by contrast, the vessels/trips can first be stratified using, if necessary, whatever subjective criteria the design team wishes to impose; and then a random sample of vessels or trips can be chosen from each stratum. Effective stratification can increase the precision of survey estimates and greatly reduces the likelihood of selecting an unrepresentative sample.

Quota sampling (e.g., Jessen 1978, pp 195-197) is a procedure used by many commercial and governmental survey organizations to ensure a specified number of samples from a sub-population of interest (domain). A métier (e.g., Marchal 2008) defined by fishing area and/or gear sub-sets, and time period (i.e. month) is an example of a domain in fisheries sampling programs. The domain of interest could be the characteristics of catches from all trips within a métier. Vessels and trips may not be classified by métiers in advance of a survey, and hence are not statistical strata where sample sizes can be controlled. Quota sampling has been suggested for sampling catches in métiers so that each métier will have a desired sample size. The problem is that the sampler must use his or her judgment to select the sampling units (e.g., trips) from each métier to obtain the prescribed sample size. This subjective choice of sampling units means that quota sampling is a non-probability sampling scheme. It is difficult to determine the precision of final estimates in terms of probability, and the method has similarity with simple judgement sampling, and may only be justified for small samples where the possible reduction in sampling errors may offset the effects of unknown biases (Jessen 1978). Quota samples may produce biased estimates because not every primary sampling unit gets a chance of being selected (see Cochran 1977). For example, if vessels from two métiers often return to port at roughly the same time, then the vessel from the métier with the larger sample size quota would tend to be selected and the other vessel would have little or no chance of being sampled. A more statistically sound procedure would be to use a probability based sampling scheme whose expected number of samples in each métier is of the desired size. This can be achieved by using a balanced sampling design (see Tillé 2006) with equal or unequal inclusion probabilities for the primary sampling units. By using all information available about the fishery to guide the survey design, the sample sizes obtained for each métier will be close to the desired number. Though the sample sizes will vary from the prescribed values, the possibility of biased estimates caused by quota sampling will be eliminated.

3.2 Defining Sampling Frames

For fisheries surveys, the population of interest may be the total catches from all fishing trips. However, it is not possible to identify a list of catches from all fishing trips, and to select a random sample of catches in a forthcoming fishery. A sampling frame cannot be based directly on the catches since these are not known in advance. Instead, a sampling frame may be derived from a list or map that is used to select representative vessels from the total "population" of vessels. Sampling frames are often classified as list frames or area frames. Let us assume that the population elements are individual vessels and trips that accounts for the total catches in a fishery. A list frame is a list or registry of vessels that contains individual frame units that correspond to either individual population elements (trips) or clusters of individual population elements (trips within vessels). An area frame is comprised of a set of geographic subareas or spatial locations that may contain clusters of individual target population elements. An indirect frame such as a list of ports provides direct access to individual subsets, or clusters, of vessels and trips on selected days. The individual ports and days within such a frame are the clusters, and each cluster may contain one or more individual population elements (vessels and trips that land their catches at the selected port and day).

Available sampling frames for use in surveys of a specific fishery usually have imperfections. The ideal sampling frame provides direct access to all vessels and trips for catch sampling. If a frame does not include all vessels, for example, then it suffers from under-coverage of the target population. Under-coverage can result in serious biases in estimates of key parameters extrapolated to the entire fleet if the fishing operations of the non-covered component of the fleet differ significantly from the vessels covered by the frame. The identification of a representative frame with sufficient coverage of the fleet operations demands the judgment of experts in the particular fishery being studied. Statistical theory warns us about the uncertainties in extrapolating from a sample of vessels and trips to the frame. Incomplete or inadequate sampling frames result from failing to identify and include all vessels within a fishery and from including vessels that are not actively fishing. A sampling frame that can be used to select representative fishing trips for catch sampling may be based on a registry of vessels (preferably the entire population of vessels that participate in the fishery of interest) with appropriate contact information. Not all frames explicitly list population elements (e.g., vessels and trips). For example, a list or map of ports can be used as a frame for sampling catches from fishing trips. Although this frame doesn't show individual vessels or trips, a random sample of ports may be visited on random days, and then all vessels/trips or a subsample may be selected for catch sampling from each port and day. One advantage of such a frame is that it would include vessels that have recently entered the fishery that were not yet included in the registry of vessels available during the planning of the survey.

The goal of selecting vessels for catch sampling should be to obtain data from trips that are representative of actual fishing operations over the entire fishing season and the full geographic range of the fishery, as well as of vessel type, gear type, and targeting strategy (NMFS 2004). In observer programs, available resources typically allow for observing only a fraction of the vessels in a given fleet. Precise estimates of catch composition and key biological parameters, nevertheless, can be achieved by sampling only a small fraction of vessels in the fleet if the sampled vessels are representative and the sample size is sufficient. Ad-hoc vessel selection has the greatest potential for generating bias because this method does not guarantee that repeated selections result in samples that, on average, represent the fleet. Conducting a probability-based survey with 100% compliance (i.e., all selected vessels agree to take an observer) would also eliminate sample bias. All the methods that involve randomization (i.e., selection of vessels with known inclusion probabilities) fall in the category of 'probability-based' sampling. Probability-based selection of vessels does not guarantee that observer data can be collected representatively because various constraints can limit an agency's ability to place observers on all selected vessels. For fisheries surveys where observers are placed onboard selected vessels and trips, concerns regarding safety of selected vessels or lack of accommodations may limit the pool of sampled vessels and reduce the ability to achieve a representative sample (NMFS 2004). Bias related to deployment can sometimes nullify the benefit of a well-planned survey. In effect, an inability to place observers on selected vessels is equivalent to implementing a program with an incomplete sampling frame because a portion of the fishery fleet is eliminated from observation.

3.3 Stratification

For stratified sampling of a population, the elements in the sampling frame (e.g., list of vessels), are divided into non-overlapping subpopulations that together comprise the entire population (e.g., the whole fleet). Stratification often increases survey precision for a given sampling effort and also ensures that precise estimates can be obtained for selected subpopulations. The stratification should be specified in advance of data collections and is a way to control the number of samples in, e.g., particular subareas or vessel categories. Stratification of the fishing fleet may be based on the physical characteristics of the vessels (e.g., length, horsepower, and tonnage) and various variables used to characterize a fishing activity. The strata weights for the stratified estimates of key parameters (e.g., abundance at age) are the number of population units (e.g., number of vessels/trips) in each stratum.

If the assumption that the total catch for each trip is known without error is valid, then strata weights for the estimation of catch composition and biological characteristics may be based on the total catch in each stratum. Given that the assumption that the total catch is accurately known is usually not valid, it is probably best to use the number of primary units in each stratum to weigh the stratum estimates. Even if the total catch is known accurately in each stratum, it is not clear which set of weights would generate the most precise estimates.

A stratified sample is obtained by taking samples of predetermined size from each stratum or sub-group of the population. Frequently samples are allocated to strata in proportion to some attribute of the strata. One approach is to allocate samples to strata in proportion to strata sizes. More trips may be selected from strata that account for a large portion of the total catches, for example. Stratified sampling may produce a gain in precision for estimates of characteristics of the aggregate catch of a fishery. It may be possible to divide a heterogeneous population of vessels and trips into subpopulations, each of which is homogenous. As mentioned above, all vessels in a fleet might be grouped by characteristics such as vessel size, target species, gear type, geographic fishing area, etc. Temporal stratification may be employed in addition to this grouping of vessels to ensure that catch sampling is spread out during the fishing season, and to allow for larger sample sizes (e.g., sampling of more trips) during periods that normally account for a large portion of the total catch. If the catches in each stratum are more homogeneous than the fleetwide catches with respects to species composition, length and age composition by species, and other characteristics, then precise estimates of stratum means can be obtained from relatively small samples in each stratum. These estimates can then be combined into a precise estimate for the whole population.

Documentation of the stratification system should include stratification variables, strata boundaries, and reasons for the adopted stratification strategy. If multi-stage/phase sampling is being used, stratification for each stage/phase should be documented.

3.4 Defining study domains (sub-populations of interest)

A study domain is defined as a major segment of the population for which separate statistics are needed (Statistics Office of the United nations, 1984; Cochran 1977). A study domain would typically be a sub-population (or sub-set of the sampling frame) identified in the overall statistical plan as one for which a certain level of precision is required. An example of study domains in fisheries catch sampling programs is the métier system for grouping catches used in the current Data Collection Regulations (DCR) in the EU, where target precision levels are set for métiers. Métiers often reflect the fishing intention of vessels, e.g. the species targeted, the area visited, and the gear used, at the start of a fishing trip. However, there are many situations where fishing intention cannot be observed directly, and can only be estimated retrospectively by examining the catch profiles resulting from fishing trips, or from interview data. A métier or other study domains chosen in advance may coincide with the strata adopted for stratified sampling, or may cut across strata. Standard stratified estimators must be modified when domains cut across strata since the inclusion probabilities for the sample units (e.g., trips) will depend on the particular stratum a unit is in (Cochran, 1976, Section 5A.14).

Ideally, catch sampling programs employ statistically valid survey designs, using a clearly defined sampling frame which is known from the onset (at the time of choosing the units) and using clearly defined probability-based rules to select PSUs. In this way the probabilities of choosing the PSUs are known, and design-based estimates of the parameters of interest can be obtained. This does not mean that the notion of métier cannot be used when designing the survey. On the contrary, using all information that is available (e.g. results from last year), one should attempt to allocate sampling effort in such a way that samples are balanced across the domains of interest (e.g. métiers). In this sense, the sampling is targeted at the trip-level for a fleet, and métiers are domains of interest that are addressed at the estimation stage.

While expected sample sizes per domain of interest (métier) can be computed at the stage of the design of the survey, realized sample sizes will only be known once the survey is completed. Thus, large métiers will be covered but smaller ones may have insufficient sample sizes to allow independent estimates on the level of that métier. Because the target sample sizes for métiers cannot always be achieved, some métiers may not be sampled at all. Substituting for data in the case of missing data for métiers is referred to as imputation. A sampling frame that is available at the onset of a survey, with effective stratification, may be used to balance samples across métiers.

3.5 Probability Sampling Methods

In probability-based sampling, each population unit (e.g., vessel/trip) in the sampling frame has a known (non-zero) probability of being selected. Quota sampling is an example of a non-probability-based scheme because there is no way to achieve a non-zero probability of selection for vessels and trips that are not picked (Weisberg, Krosnick, and Bowen 1996). Levy and Lemeshow (1999) demonstrate the differences between probability-based and non-probability-based sampling schemes.

There are a wide range of probability sampling designs that can be used to draw a sample of elements from a sampling frame, and each design dictates a different specific set of estimation formulas for point estimators and point estimator variances (Særndal et al, 1992). The different possible designs offer advantages and disadvantages that should be considered in the selection of the design to be used for a particular survey.

Methods for selecting vessels and trips from a sampling frame based on a list of vessels include: (1) census (i.e., all trips from all vessels in the sampling frame are observed), (2) random sampling with replacement of vessels, (3) stratified random sampling with replacement, (4) stratified random sampling without replacement, (5) systematic random sampling, and (6) ad-hoc selection of vessels and trips. Of these methods, ad-hoc selection of vessels is the most likely to produce bias (Vølstad and Fogarty 2006).

A census of all trips by all vessels in the sampling frame would eliminate vessel selection as a potential source of bias but could be prohibitively expensive and would not eliminate bias due to errors in the frame. Random or stratified random selection of vessels is a cost-effective means of minimizing bias in general, but logistical constraints may limit representative catch sampling on randomly selected vessels. In this case, an ad-hoc selection of vessels from the frame, with full compliance, may cause no more systematic error than a random selection with poor compliance. Bias related to errors in the sampling frame (list) from which vessels are selected for observation can occur when the list fails to include all active vessels in the fishery for which inferences about catch and bycatch are to be made (Vølstad and Fogarty 2006; NMFS 2004). If the list omits an appreciable portion of vessels in the fleet for which estimates are required, then even a census (i.e., placing observers on all vessels and trips on the list) could yield poor (biased) estimates of catch and bycatch. Errors in the sampling frame can result when using lists of vessels that are not up-to-date, or if vessels are included that are not actively fishing. If the fraction of vessels not observed accounts for an appreciable portion of the total catch for a fishery, then the resulting bias in overall estimates of catch and bycatch based on observer data could be significant.

With a sampling design in place, it is important to plan in detail how data will be processed, audited, and edited for quality control and assurance. It will also be important to specify any methods to be used for imputation of missing data. The specific estimation formulas to be used for point estimators and measures of the precision of those point estimators (variance estimators) should be based on the selected sampling design.

3.6 Sample size determination

The number of primary sampling units (e.g., trips, or vessels) is of particular importance for the cost of a monitoring survey and to a large extent determines the level of precision that can be achieved for key parameters estimates. In practice the survey itself can have only one sample size for the primary sampling units (e.g., vessels, or trips). When there are many key parameters the sample size needed to achieve a target precision level for each may be calculated, and then use the largest sample size.

Sample size determination depends critically on the degree of precision wanted for the key indicators. The more precise or reliable the survey estimates must be the greater the sample size must be. Survey managers must obviously be cognizant of the impact that overly stringent precision requirements have on the sample size and hence the cost of the survey, while simultaneously being careful not to use such a small sample size that the main indicators would be too unreliable for informative analysis or meaningful planning.

Another significant factor that has a large impact on the sample size is the number of domains, defined as the analytical sub-groups for which equally reliable data are wanted (e.g., métiers). The sample size is increased approximately by a factor equal to the number of defined domains, because sample size for a given precision level does not depend on the size of the population itself (e.g., number of vessels or trips), except when a significant percentage, say, 10-percent or greater, of the population is included in the sample. Thus, the sample size needed for a domain (e.g., métier) may be nearly the same as that needed for an entire fishery. When the target precision is equal for all domains, disproportionate sampling rates must be used. It is important to note that deliberate over-sampling of sub-groups, whether for domains or strata, not only necessitates the use of compensating survey weights to form the national-level estimates, but the latter are somewhat less reliable than would be the case if the sample were distributed proportionately among the sub-groups of the population. Clearly, the number of domains has to be carefully considered when resources are limited. One strategy is to decide which domains, despite their importance, would not require equal reliability.

When setting target sample sizes for age data it is important to account for the complex multi-stage sampling schemes, and the resulting clustering of age-data. Precision in age-distributions by species is typically driven by sample sizes for the primary sampling stage (e.g., number of vessels/trips), and to a lesser extent by the number of age-readings from each trip (e.g. Vølstad et al. 1997.) It is recommended practice to document the number of primary sampling units sampled along with mean number of otoliths collected per PSU in catch sampling programs.

3.7 Methods for estimating fisheries statistics and biological parameters

For probability-based fisheries surveys, primary sample units are selected based on a random process, such that different units may have equal or different selection probabilities. These probabilities are taken into account in the analysis to construct unbiased estimates. There are basically two main approaches for obtaining statistically valid estimates of key parameters from survey data, namely design-based methods and model-based (or model assisted) methods. Design-based and model-assisted estimators refer to a class of estimators that expand or weight the observations in the sample to obtain population estimates which include the survey probability weights. Survey weights are derived from the inclusion probabilities of the samples and available auxiliary information about the target population to obtain design-unbiased estimators for the unknown population parameters of interest. The associated inferences are based on the probability distribution induced by the sampling design with the population values held fixed. In the design-based approach, the statistical inference is based on the stochastic structure induced by the sampling design. Parameter and variance estimators are derived under the concept of repeated random samples from a finite population according to the actual sampling design. This is the traditional approach of survey sampling theory (e.g., Hansen, et al. 1953; Kish 1965; Cochran 1977, Lehtonen and Pahkinen 2004). Särndal et al., (1992) expands this theory to include model-assisted inferences that use auxiliary information. In model-based estimation, the inference is based on the probability structure of an assumed statistical model, and the inclusion probabilities of the samples play a lesser role. This is the approach taken by authors like Gosh and Meeden (1997), Valliant et al. (2000), and Rao (2003).

Catch sampling programs often results in the clustering of data, for example due to repeated measurements of catches taken by selected vessels over time, or due to subsampling of the primary sampling units (trips). Since independence among sample observations is a key assumption underlying standard statistical procedures, the presence of clustering in the data should be accounted for in the analysis. That is, the primary units are generally a random sample (of the population of primary units), while the individuals selected from the primary units are not a random sample of individual fish from the entire catch population.

In fisheries catch sampling programs, data on catch composition (e.g., length and age-distribution of a species) and other parameters are often collected from clusters of fish (e.g., from catches by trip), and the variance of parameter estimates is largely influenced by the number of clusters, not the total number of fish in the sample. Therefore, precision will only be significantly improved if the number of clusters sampled is increased and not by increasing the number of fish sampled from each cluster.

Clustering as well as the effects of stratification can be measured by the so-called design effect (*deff*), which compares the sampling variance (square of the standard error) for stratified, cluster sampling, with the variance of a simple random sample of fish of the same size as the total number of fish in the cluster sample. Stratification tends to improve the sampling variance to a small degree, so that *deff* indicates, primarily, how much clustering there is in the survey sample (Kish 2003). To keep the design effect as low as possible, it is generally best that as many primary sampling units (vessels, or trips) as feasible be sampled.

The estimation of precision of parameters & statistics used for supporting stock assessments (Annex 4) was discussed in general during WKPRECISE. Where appropriate it was indicated where the EU COST software may be used to estimate the precision of these parameters.

3.7.1 Design-based estimation approaches

Fisheries scientists often employ sample survey methodology to obtain information about a large aggregate of catches from a fleet by selecting and measuring a sample of catches from a sample of fishing trips. Due to the variability of catch characteristics among items in the population, a good practice in data collections is to apply scientific sample designs in the sample selection process to reduce the risk of biased estimates of, e.g., the catch composition across all trips in the fleet. Inferences about the population (e.g., length composition of the total catch of a species in all trips, or numbers at age from length-stratified sampling within trips) are based on the information from the sample survey data. In order to make statistically valid inferences for the population (e.g., all trips), the sample design must be incorporated in the data analysis. Design-based methods for analyzing survey data can be implemented using readily available computer programs in several user-friendly software packages.

Complex sample survey designs include aspects such as unequal probability sampling, multistage sampling, and stratification. Weighted analyses are necessary for unbiased (or nearly unbiased, but consistent) estimators of population parameters. Variance estimation for estimators depends on the survey design and often requires approximate methods such as the Taylor series linearization or replication techniques.

The estimation of the precision of estimates for métiers must take into account that the statistical weights of samples will depend on the strata to which they belong, and often involves post adjustments of sample weights. Post-stratification of the samples (primary sampling units) may be used when the total catches in each métier are known. This is the method implemented in the EU COST computer program (http://wwz.ifremer.fr/cost.) The pooling of métiers also requires post-adjustments of sample weights, or imputations when data are missing in some métiers.

We very rarely, if ever have a random sample of individual animals but in practice fish are sampled from clusters of fish. For example, fish that are caught together at a station form a cluster. Other examples of sampling clusters are: the fish caught during a fishing trip, the fish in a particular market and the fish in a processing plant. From each cluster, fish for aging, measuring, *etc.* are selected, that is such data are often generated by two-stage cluster sampling. If the sample consists of a total of m fish from n clusters, then the individual animals are not a random sample from the entire population. This is because animals caught together tend to be more similar than animals in the entire population (*i.e.* there is positive intra-cluster correlation). The practical implication of positive intra-cluster correlation is that a sample of animals caught in clusters will generally contain much less information on the population structure than an equal number of fish sampled at random, that is the effective sample size is much smaller than the number of animals sampled (Pennington and

Vølstad, 1994; Pennington *et al.*, 2002; Aanes and Pennington, 2003; Helle and Pennington, 2004.) Therefore, if an estimate of the variance is based on the assumption that the sample is simple random, then the estimate will generally be highly biased. Given a random sample of n clusters and a random subsample of m_i fish from a total of M_i individuals in cluster i, then the design-based estimator

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{n} M_i \tilde{x}_i}{\sum_{i=1}^{n} M_i}$$

is an approximately unbiased and a consistent estimator of; (1) the mean age or length of the population if $\tilde{x}_i$ is the average age or length of the sample of m_i fish from cluster i or; (2) the proportion at age or length in the population if $\tilde{x}_i$ is the estimated proportion of fish of a specific age or length class in cluster i (Skinner *et al.*, 1989; Lehtonen and Pahkinen, 2004). This is a weighted average of the $\tilde{x}$'s, where the cluster sizes are the weights. Since both the numerator and denominator are random variables this is a ratio type estimator (Cochran, 1977), and an exact variance formula does not exist. The variance may be approximated using a Taylor expansion of (3) or by resampling techniques, such as nonparametric bootstrapping (*e.g.* Efron, 1983).An alternative to the design-based estimator, which in some situations may have a smaller variance than the weighted estimator, is the unweighted average of the $\tilde{x}$'s

$$\hat{\mu}_2 = \frac{\sum_{i=1}^{n} \tilde{x}_i}{n}.$$

In general, the unweighted estimator, $\hat{\mu}_2$, may be biased and this bias may not decrease with increasing sample size, but if $\tilde{x}_i$ and M_i are uncorrelated, then $\hat{\mu}_2$ may be an acceptable estimator (Cochran, 1977). If M_i and $\tilde{x}_i$ are correlated, then the expected bias of the unweighted estimator is

$$Bias(\hat{\mu}_2) = -\frac{Cov(M_i, \tilde{x}_i)}{\overline{M}},$$

where $\overline{M}$ is the mean cluster size. One reason that the unweighted estimator $\hat{\mu}_2$ is sometimes used is that the sizes of the clusters, M_i, are unknown or not recorded, and, hence, the resulting estimate may contain an unknowable bias. Therefore to avoid this source of bias, it is important to define the clusters from which the subsamples are taken, record each sampled cluster's size and use the appropriate estimator.

Standard statistical software packages generally cannot be used to analyze sample survey data since they typically assume simple random sampling of elements (equivalent to iid assumption in model-based approaches). Point estimates of population parameters in an unweighted analysis will introduce bias, and/or underestimation of standard errors for point estimates. Using the sampling weight variable with standard packages yields appropriate point estimates of population parameters. However, estimated standard errors usually are still incorrect because the variance estimation procedure typically does not take into account the clustering and/or stratification of the sampling plan. For design based methods, several software packages include procedures for estimation from complex survey data[1]:

- SUDAAN (http://www.rti.org/sudaan/index.cfm)
- WESTAT (http://www.westat.com/)
- STATA (http://www.stata.com/capabilities/svy.html)
- Survey analysis package in R (package "survey", http://faculty.washington.edu/tlumley/survey/).

These tools include several of the standard design-based methods for analyzing survey data, but we are not aware of any approaches within catch sampling utilizing these tools.

3.7.2 Model-based estimation approaches

Multilevel modelling may be used to analyze hierarchical data from multi-stage surveys. Some software packages (e.g., STATA) may be used to fit maximum likelihood models to complex survey data, using a model-based sandwich estimator that considers only the variation among model clusters (Pierre and Saidi 2008). When the model (1) coincides with the hierarchy of the survey design, (2) the sample fraction of primary sampling units is small, but still (3) a large number of PSUs (more than 50) are sampled, then a model-based sandwich estimator performs well.

One model-based approach used in routine estimation of catch at age (e.g. Northeast Arctic haddock, Northeast Arctic saithe, Northeast Arctic cod, ICES 2009) is the ECA-model implemented at Institute of Marine Research (IMR), Norway. This Bayesian hierarchical model for combining multiple sources of data to estimate catch at age was originally developed by Norwegian Computing Centre, Oslo, Norway and Institute of Marine Research, Bergen Norway (Hirst et al. 2004; 2005), and has been incorporated in the COST program (http://wwz.ifremer.fr/cost.). Later developments of the ECA-model include estimators for discards, and of stock structures based on characteristics in the otolith (in prep.).

The ECA-model can include data of the form; random length samples, random lengths and ages, and length stratified age samples, and estimates landings- and discards-at-age, and length- and weight-given-age. The model consists of submodels for proportions at age, length given age, weight given length and a discard model. The submodels include covariates for gear, season, area and year to provide estimates for a stratum or groups of strata or domains.

---

[1] U.S. Department of Education. National Center for Education Statistics. *Strengths and Limitations of Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets*, Working Paper No. 2000-03, by Pam Broene and Keith Rust. Project Officer, Susan Ahmed. Washington, DC: 2000.

Recognizing the sampling frame and that the samples consists of clusters as the primary sampling unit (Aanes and Pennington 2003), the hierarchical sampling frame with clusters such as haul/trip as the primary sampling unit and subsampling within clusters for age, length or weight is maintained by including variance components describing between and within cluster variance in the various submodels. It is assumed that all samples are drawn from underlying distributions, super populations, and probability weighing and finite population corrections are not considered. Details of the model are found in Hirst et al 2005, and on the website for the COST project (http://wwz.ifremer.fr/cost.)

3.8 Estimating age distributions

Methods for estimating age-distribution of fish from catch sample data received special attention by WKPRECISE. Age distributions of species in the catches can be estimated directly from age-samples collected from fisheries-independent surveys, or indirectly by applying Age-Length-Keys (ALK; see e.g. COST 2009, Quinn and Deriso 1999) to estimates of length-distributions from length-samples. The WKPRECISE recognized that the most common approach is the use of ALK's, but also noted that methods employed for routine estimation of age-composition of catches often are poorly documented. The use of ALK is the focus of this section.

When based on age-sample data directly, the proportions at age or length, and other biological parameters estimated from the samples are raised to the total catch using weighting factors determined by the survey design (design-based estimates), or using models under various assumptions about how the samples relates to the population of fish in the total catch. The workshop recommends a literature review of method employed in the estimation of age-distribution of fish in catches.

3.8.1 Using Age-Length Keys to estimate age composition of fish

We are aware of various practices for estimating ALK's, but many methods are poorly documented. The WKPRECISE could not evaluate an exhaustive list of practices in detail, but merely review some of the common practices. A common practice is to pool the age samples available within a stratum, or even across strata, to form an ALK. The trips are then combined by a mean of the PSU ALKs. If a trip (PSU) effect in age given length is present (i.e., a cluster effect), then the un-weighted mean will be biased. Using this approach, the multi-stage sampling design and the resulting hierarchical structure of the data is typically not accounted for. This is often done in practice if data from several PSU's (e.g., trips) are combined without the proper weighting factors. We are also aware of situations where age samples are not available, and the ALK is considered known and constant based on best guesses. If an unbiased and consistent estimator of the population distribution of age given length is required, a weighted mean estimator should be employed, where the weights are based on the total catches in each trip.

ALK's are commonly used to estimate age-distributions when lengths are sampled at random from the primary sampling units, and age samples are collected from only a subsample of the fish measured for length. An Age-Length-Key (ALK) classifies frequencies of length l and age a into n_{la} (Quinn and Deriso 1999). A common estimator of the conditional distribution of age given length from a catch is

$$p(a\,|\,l) = \frac{n_{al}}{\sum_{l} n_{al}} = \frac{n_{al}}{n_l} \quad (1)$$

(e.g. Quinn and Deriso 1999). This estimator applies to two-stage sampling, under the assumption that a random sample of lengths are taken from the catch in the 1st stage, and that a subsample of ages (usually stratified by length) are obtained from the fish measured for lengths in the 2nd stage, (see e.g. Quinn and Deriso 1999 p 303-304 for further details). The numbers at age for the population of fish in the total catch is then estimated as

$$N_a = N \sum_l p(l) p(a \mid l) ,$$

where $p(l)$ is an estimate of the length distribution of the catch. This estimator is an unbiased estimator for the population of fish, provided that L is a simple random sample from the population and that the n subsampled individuals for aging is a random sample within length groups (strata), and variance estimators are available (see for example Quinn and Deriso 1999, Morton and Bravington 2008). Notice that this also covers the situation with only one length group covering all lengths which in that case correspond to random subsampling of ages across all lengths.

Using the two stage approach described above with estimator (1), with length samples and age samples from different primary sampling units (e.g., from different trips), causes biased estimates of the age distribution (Quinn and Deriso 1999, p 311) and also introduce bias in the estimates of precision of proportions at age (see Quinn and Deriso 1999 and references therein pp 313-317 for a discussion.) It is rarely possible to collect simple random samples of fish from the total catch in a fishery. In practice, a form of multi-stage stage catch sampling is employed, where

1. A sample of PSUs (e.g., vessels/trips or hauls) are taken in the first stage
2. A random sample of L lengths are taken from the population of fish in the first stage (e.g., from the catch of the selected vessel/trip, or haul), and
3. A subsample of n ages (usually stratified by length) are obtained from the fish measured for lengths in the 2nd stage, (see e.g. Quinn and Deriso 1999 p 303-304 for further details).

Stratification is typically employed at several stages of the sampling. We refer to the standard survey sampling literature (e.g. Cochran 1977, Lethonen and Pahkinen 2004) for details on multi-stage sampling.

4 DCF métier based sampling requirements in relation to sampling design and precision estimation

Commission Decision 2008/949/EC specifies detailed rules for adopting a multi-annual Community programme pursuant to Council Regulation (EC) No 199/2008 establishing a Community framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the common fisheries policy. Central to this Data Collection Framework (DCF) is the concept of métier based sampling for estimating discard quantities and length and age compositions of landings and discards, and stock-based sampling for biological parameters such as age, sex and maturity, with target precision levels for each parameter by stock.

The métier is defined in the DCF as a group of fishing operations targeting a similar (assemblage of) species, using similar gear, during the same period of the year and/or within the same area and which are characterised by a similar exploitation pattern. The métiers are defined at Level 6 of the matrix given in Decision 2008/949/EC i.e. gear type, target assemblage and mesh size. The DCF states that "*in order to optimise the sampling programmes, the métiers may be merged. When métiers are merged (vertical merging), statistical evidence shall be brought regarding the homogeneity of the combined métiers. Merging of neighbouring cells corresponding to fleet segments of the vessels (horizontal merging) shall be supported by statistical evidence. Such horizontal merging shall be done primarily by clustering neighbouring vessel LOA classes, independently of the dominant fishing techniques, when appropriate to distinguish different exploitation patterns. Regional agreement on mergers shall be sought at the relevant regional coordination meeting and endorsed by STECF.*"

Stock-based information on age, length, weight, sex, maturity and fecundity are to be collected with corresponding information on space and time stratum, but the DCF does not specify that the fleet métier should be recorded although that does not preclude recording the trip details and associated métier with each individual fish sampled for age.

The rules contained in Decision 2008/949/EC imply sampling designs for biological parameters that are considered by WKPRECISE to be not fully in accordance with good practice for sampling survey designs and facilitating estimation of precision.

The WKPRECISE identified two main issues of concern with the DCF métier-based sampling requirements:

1. The Level-6 fleet métiers identified in Appendix IV of 2008/949/EC are in many cases too dynamic and unpredictable to constitute sampling (list) frames in which all the primary sampling units (PSUs) are identifiable in advance for allocating sampling effort. It can lead to over-stratification and under- or over-sampling (or non-sampling) of métiers. The approach needs to be modified to allow the definition of more appropriate sampling frames for implementing randomised sampling schemes.

2. The rules for collection of stock-based biological parameters allow the possibility of collection of age data independently of métier-based length sampling, and require only the recording of area and time stratum for aged fish. Work presented at WKPRECISE illustrated the need to identify the PSU for every fish sampled for age in order that the precision of age

composition estimates can be calculated. This should be specified in the DCF Guidelines.

WKPRECISE addressed both of these issues, and provides guidelines for good practice in sampling of fishery catches on shore and at sea to deliver fleet-based and stock based data specified by the DCF to support stock assessment and fleet-based management. A discussion of sampling schemes for estimating age composition of fishery or survey catches is given in Section 3.

4.1 Defining the biological data to be delivered by the DCF fleet based fishery sampling schemes

The DCF requires member states to develop statistically robust sampling schemes to deliver the following types of biological data from commercial fisheries with minimum bias and achieving predefined precision levels at a national or international level:

- Quarterly length and age compositions of the landings and discards of defined species taken by national fleet métiers (after merging) in defined fishing grounds. In most cases the quarterly landings of species by métier are from exhaustive data from EC logbooks and sales notes. These data are the primary input to catch based stock assessment models along with relative abundance indices from surveys and/or fishery CPUE.
- Estimates of the length composition of landed and discarded fish of all species or a defined subset of species taken together in individual trips of métiers contributing the top 90% of the annual national landings, or value, or fishing effort in each fishing ground. This form of concurrent length sampling is needed to predict the effect on co-occurring species of management measures affecting activities of specific fleet métiers.
- Quarterly quantities of organisms discarded from fishing boats in each national métier and fishing ground, for all métiers where >10% of the total catch is discarded. This may include métiers not selected for concurrent length sampling by the ranking system.
- Quarterly estimates of recreational fishery catches for a small number of species according to fishing ground.
- Estimates of mean length and weight at age of fish in the fishery landings and discards in each métier.

Target precision levels given by the DCF are referenced to the national level for sampling programmes defined nationally, or to the regional métier level for internationally collaborative programmes. The sampling unit is defined as the trip, and the DCF specifies that sampling effort should be proportional to the relative effort and variability in the catches of that métier, but never less than one fishing trip per month in the fishing season for trips less than two weeks and one trip per quarter otherwise.

In addition to the above data requirements for fisheries fleets, the DCF requires the estimation of stock-based biological variables that are best derived from properly designed surveys covering the full range of each stock. These include estimates of sex ratio, proportion mature at length/age, and fecundity for stocks according to multi-annual sampling schemes. There may be a need to collect such data from fishery catches if appropriate surveys are not carried out.

The fishery data collection activities of EU Member States are coordinated and harmonised through annual meetings of Regional Coordination Meetings, and recommendations from these meetings are discussed and elaborated by the annual Liaison Committee meeting comprising ICES, RCM and European Commission representatives. Recommendations related to biological sampling are then transmitted to the annual ICES Planning Groups PGCCDBS and PGMED which establish work programmes and workshops to improve the quality of fishery data and develop systems for quality assurance of data. PGCCDBS established workshops such as WKACCU and WKPRECISE to implement the ICES Quality Assurance Framework for data and support Member States to implement procedures for evaluating bias and precision building on other initiatives such as the EU COST project. As end-users of such data, ICES stock assessment Working Groups also pass recommendations to PGCCDBS/PGMED to address data issues that have arisen. Other ICES Planning Groups exist to coordinate different types of fish stock surveys funded through the DCF.

It is advised that biological data from catch sampling should not be used to estimate variables at a stock level. Estimates of average weight at length, maturity at age etc. from catch sampling are subject to bias because of selectivity in fishing gears, and because fishing operations does not cover the stock representatively. It is better to get this information from scientific surveys. Catch sampling and scientific surveys need to be compared to assess whether data from catches can be used to estimate population parameters. This must be examined for each population.

4.2 Available sampling frames:

All commercial fishing vessels are registered, and complete list frames are therefore available in each country. The fleet segmentation used for DCF economic variables (Appendix III of 2008/949/EC) identifies groups of vessels according to their overall length and predominant fishing method (e.g. beam trawlers, demersal trawls/seines, purse seiners, dredgers, drift/fixed netters). "Polyvalent" segments are also defined, comprising vessels that could use different gear types even within a trip. WKPRECISE recommends that sampling frames for biological sampling should be based on vessel lists using the predominant fishing method to define the frames.

Whilst some of the fleet métiers in the DCF matrix identified in MS National Programmes and endorsed by the RCMs are represented by more-or-less distinct populations of specialised fishing vessels (e.g. purse seiners, scallop dredges, beam trawlers, specialised Nephrops trawlers), and represent distinct strata for sampling. Other vessels may exhibit very dynamic choice of gear, mesh size, selectivity devices and target assemblage according to opportunities. In this case the métier is not an appropriate stratum and is effectively a domain relevant to desired outcomes of a sampling programme rather than the basis for random selection of trips.

5 Concluding remarks and recommendations

Representative catch sampling programs are necessary to achieve accurate estimates of fisheries statistics used for assessment. Sampling programs to estimate the composition and biological characteristics of aggregate fleet-wide catches are invariably complex. WKPRECISE recommends that catch sampling be based on statistically robust survey designs with clear definitions (and documentation) of the sampling frame, the primary sampling units (PSUs), the stratification schemes employed, and the methods used for selecting samples in each stratum. The sampling frame may be based on a list of all active vessels, with PSUs being defined as vessels/trips. In more complex surveys, the sampling frame may be a list of ports, with primary units being the combination of ports and days. Vessels/trips that land their catches in selected ports and days form secondary units. Catch samples may then be taken from a subsample of vessels and trips within a port/day in the third stage.

A métier chosen in advance may coincide with the strata adopted for stratified sampling, or may cut across strata. Standard stratified estimators cannot be applied when métiers cut across strata since the selection probabilities for the sample segments (e.g., trips) will depend on the particular strata a segment is in. Estimation of catch statistics and biological parameters for métiers (or groups of métiers) generally requires some form of sample-weight adjustments because the primary units within a métier may come from multiple strata with different sampling intensities. The pooling of data from multiple métiers without proper weighting likely causes bias, and unreliable estimates of precision.

Best practices to achieve representative data for the accurate estimation of key parameters involve a form of probability-based sampling that minimizes or eliminates systematic bias and allows for reliable estimates of precision. Probability sampling in surveys ensures that the sampling errors can be estimated based on the data collected, which is not mathematically justifiable when ad-hoc sampling methods are used. It is emphasized that accurate estimates cannot be obtained from biased sampling schemes. Ad-hoc sampling rules out the estimation of precision and should therefore be avoided. It was noted by WKPRECISE that the practice of quota sampling to achieve a desired sample size (e.g., number of trips) in métiers may result in biased estimates of key parameters. Indicators of bias developed by WKACCU 2008 are intended to identify the existence of bias in data collection schemes.

Catch samples of fish are typically collected in several stages, where the first stage involves the selection of representative vessels and trips. For selected trips, fish from the entire catch (census), or from sub-samples of catches, may be sorted by species, weighed, and measured for length. Age-samples are often collected in the last stage (e.g., within a sub-sample of catch from a trip.) This multi-stage sampling requires that estimators of key parameters for total catches in a fishery appropriately accounts for the hierarchical structure of the data, and in particular clustering effects that may drastically reduce the effective sample sizes. The reason is that fish caught during one trip often have more similar characteristics then the general population of fish from all trips. Assumptions of simple random samples of fish from the total population of fish in the catches cannot be reasonably met when sampling is done in multiple stages.

WKPRECISE recommends that the precision of estimates of key parameters is given in terms of standard errors (or relative standard errors). In addition, the number of primary sampling units observed along with estimates of the effective sample size for the associated estimate should be given. This is because the variances of key estimates are typically driven by the number of PSUs sampled, and so the effective sample size is usually much smaller than the total number of individuals sampled. If age-length keys (ALKs) are used to estimate age-distributions, then it must be noted that the precision of such estimates cannot be evaluated unless the age-length data are coupled to the primary sampling units from which the age and length data were collected. The accuracy of estimates of age-distributions based on static ALKs that do not take into account the survey design of the catch sampling programs (or ALKs derived from ad-hoc sampling) cannot be assessed. For probability-based catch sampling programs, the estimation of age distributions of fish, and the associated uncertainty, directly from the multi-stage age samples should be considered rather than using an ALK. Using an ALK formed from samples in one stratum to estimate age distributions in another will cause bias and should be discouraged. This is of particular concern when analyses are combined internationally. It is recommended that the original age-length data be shared rather than just the ALK. Ad-hoc transferring of data from one stratum to another should be discouraged and modelling or statistical imputation techniques (at the international level if data is available) should be promoted.

WKPRECISE recommends that the COST software for analysis of catch sampling data be extended to include additional design-based estimators of key parameters used for stock assessments, alongside its model-based estimators. When appropriate, estimators for complex multi-stage surveys developed in the R Survey Package should be considered. When possible, WKPRECISE recommends as a good practice that model-based estimates be compared to estimates based on sample theory for validation of model assumptions.

6 References

Aanes, S. and M. Pennington. 2003. On estimating the age composition of the commercial catch of Northeast Arctic cod from a sample of clusters. ICES J. Marine Science 60:297–303.

Cochran, W.G. 1977. Sampling Techniques, 3rd ed. Wiley, New York.

Efron, B. 1983. The Jackknife, the Bootstrap and Other Resampling Plans, 2nd ed. Society for In dustrial and Applied Mathematics, Philadelphia.

Gosh, M., and G. Meeden (1997). Bayesian Methods for Finite Population Sampling. London: Chapman & Hall.

Hansen, M.H., W.N. Hurwitz, and W.G. Madow (1953). Sample Survey Methods and Theory, Vol I and II. John Wiley and Sons. New York.

Helle, K., and M. Pennington. 2004. Survey design considerations for estimating the length com postion of the commercial catch of some deep-water species in the Northeast Atlantic. Fish eries Research 70:55–60.

Hirst, D. and Storvik, G. and Aldrin, M. and Aanes, S. and Huseby, R. B. 2005. Estimating catch-at-age by combining data from different sources. Canadian Journal of Fisheries & Aquatic Sciences 62(6): 1377-1385.

Hirst, D. and Aanes, S. and Storvik, G. and Huseby, R. B. and Tvete, I. F. 2004. Estimating catch-at-age from market sampling data using a Bayesian hierarchical model Applied Statistics 53 (1): 1-14.

Jessen, R.J. 1978. Statistical Survey Techniques. John Wiley and Sons, New York.

Kish, L. 1965. Survey Sampling. John Wiley and Sons, New York.

Kish, L. 2003. Selected Papers [Compiled and edited by Steven Heeringa and Graham Kalton]. Wiley series in survey methodology, John Wiley and Sons, Hoboken, New Jersey.

Lehtonen, R. and E. Pahkinen. 2004. Practical Methods for Design and Analysis of Complex Surveys, 2nd ed. Wiley, New York.

Lessler, J.T. and W.D. Kalsbeek. 1992. Nonsampling Errors in Surveys. John Wiley & Sons. Ne w York.

Levy, P. S., and S. Lemeshow. 1999. Sampling of populations: methods and applications. John Wiley & Sons. New York, NY.

Marchal, P. 2008. A comparative analysis of métiers and catch profiles for some French demersal and pelagic fleets. ICES J. Mar. Sci. 65: 764-686.

NMFS (National Marine Fisheries Services). 2004. Evaluation of bycatch: a national approach to standardized bycatch monitoring programs. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-F/SPO-66, 108 p. On-line version, http:/spo.nmfs.noaa.gov/tm

Pennington, M., L. M. Burmeister and V. Hjellvik. 2002. Assessing the precision of frequency distributions estimated from trawl-survey samples. Fish. Bull. 100:74–81.

Pennington, M. and J.H. Vølstad. 1994. Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. Biometrics 50:725–732.

Pierre, F. And A. Saidi. 2008. Implementing resampling methods for design-based variance estimation in multi-level models: Using HLM6 and SAS together. Section on Survey Research Methods, JSM 2008: 788-794.

Quinn II, T. J., and Deriso, R. B. 1999. Quantitative Fish Dynamics. Oxford University Press, New York.

Rao, J.N.K. 2003. Small Area Estimation. Wiley. New York.

Rago, Paul J., S.E. Wigley, and M.J. Fogarty. 2005. Northeast Fisheries Science Center Reference Document 05–09. NEFSC Bycatch Estimation Methodology: Allocation, Precision, and Accuracy. M. Smithson, 2000. Statistics With Confidence. Sage Publications, London. 446 pp., ISBN 0–7619–6031–7

Skinner, C.J., D. Holt and T.M.F. Smith (eds.). 1989. Analysis of Complex Surveys. Wiley, New York.

Statistical Office of the United Nations, "Handbook of Household Surveys, Revised Edition", (paras. 4.6, 4.7), Studies in Methods, Series F, No. 31, United Nations, New York, 1984.

Særndal, C.E., B. Swensson, and J. Wretman. 1992. Model Assisted Survey Sampling. Springer, New York.

Tillé, Y. 2006. *Sampling Algorithms*. Springer.

Valliant, R., A.H. Dorfman, and R.M. Royall. 2000. Finite Population Sampling and Inference, A Prediction Approach. John Wiley & Sons. New York.

Vølstad, J.H, W.R. Richkus, S. Gaurin, and R. Easton. 1997. Analytical and Statistical Review of Procedures for Collection and Analysis of Commercial Data Used for Management and Assessment of Groundfish Stocks in the U.S. Exclusive Economic Zone Off Alaska. Prepared for the U.S. Department of Commerce, National Marine Fisheries Service, Alaska Fisheries Science Center, Seattle, Washington. 172 pp.

Vølstad, J.H. and M. Fogarty. 2006. Report on the National Observer Program Vessel Selection Bias Workshop. Woods Hole, MA, USA.

Weisberg, H. F., J. A. Krosnick, and B. D. Bowen. 1996. An introduction to survey research, polling, and data analysis. 3rd ed. Sage Publications. Thousand Oaks, California.

Annex 1: List of participants - WKPRECISE

LIST OF PARTICIPANTS

NAME	ADDRESS	PHONE/FAX	EMAIL
Sondre Aanes Chair	Institute of Marine Research P.O. Box 1870 N-5817 Bergen Norway	Phone +47 55238627 Fax +47 55238687	sondre.aanes@imr.no
Mike Armstrong	Centre for Environment, Fisheries and Aquaculture Science Lowestoft Laboratory Pakefield Road NR33 0HT Lowestoft Suffolk United Kingdom	Phone +44 1502 524362 Fax +44 1502 524511	mike.armstrong@cefas.co.uk
Valerio Bartolino	Swedish Board of Fisheries Institute of Marine Research, Lysekil P.O. Box 4 SE-453 21 Lysekil Sweden		Valerio.bartolino@uniroma1.it
Andrea Belgrano	Swedish Board of Fisheries Institute of Marine Research, Lysekil P.O. Box 4 SE-453 21 Lysekil Sweden	Phone +46 523 18722 Fax + 46 523 13977	andrea.belgrano@fiskeriverket.se
Stijn Bierman	Wageningen IMARES Haringkade 1 1976 CP IJmuiden Netherlands	Phone +31 317481222 Fax +31 317487326	stijn.bierman@wur.nl
Åge Fotland	Institute of Marine Research P.O. Box 1870 N-5817 Bergen Norway	Phone +47 55 238682 Fax +47 55 238687	Aage.Fotland@imr.no
Dorleta Garcia	AZTI-Tecnalia AZTI Sukarrieta Txatxarramendi ugartea z/g E-48395 Sukarrieta (Bizkaia) Spain	Phone +34 946 02 9400 Fax +34 946 870 006	dgarcia@suk.azti.es
Hans Gerritsen	Marine Institute Rinville Oranmore Co. Galway Ireland	Phone 353 91 387297/353 85 1463240 Fax 353 91 387201	hans.gerritsen@marine.ie
Francesca Gravino	Malta Centre for Fisheries Sciences Fort San Lucjan BBG 06 Marsaxlokk Malta		francesca.gravino@gov.mt

NAME	ADDRESS	PHONE/FAX	EMAIL
David Hirst	Norsk Regnesental (Norwegian Computing Centre) Strathweltie, Tarland AB34 4YS Aboyne Norway	Phone +47 1339881448	david.hirst@nr.no
David Maxwell	Centre for Environment, Fisheries and Aquaculture Science Lowestoft Laboratory Pakefield Road NR33 0HT Lowestoft Suffolk United Kingdom	Phone +44 1502 524 328	david.maxwell@cefas.co.uk
Pentti Moilanen	Turku Game and Fisheries Research Institute Itäinen Pitkäkatu 3 20540 Turku Finland	Phone +358 20 57511	pentti.moilanen@rktl.fi
Pentti Moilanen	Turku Game and Fisheries Research Institute Itäinen Pitkäkatu 3 20540 Turku Finland	Phone +358 20 57511	pentti.moilanen@rktl.fi
Michael Pennington	Institute of Marine Research P.O. Box 1870 N-5817 Bergen Norway	Phone +47 55238601	michael.pennington@imr.no
Alastair Pout	Fisheries Research Services P.O. Box 101 AB11 9DB Aberdeen Torry United Kingdom	Phone +44 1224 876544 Fax +44 1224 295533	A.Pout@marlab.ac.uk
Katja Ringdahl	Swedish Board of Fisheries Institute of Marine Research, Lysekil P.O. Box 4 SE-453 21 Lysekil Sweden	Phone (+46) 523-18 753	katja.ringdahl@fiskeriverket.se
Paz Sampedro	Instituto Español de Oceanografía Centro Oceanográfico de A Coruña P.O. Box 130 E-15001 A Coruña Spain	Phone +34 981 205 362	paz.sampedro@co.ieo.es
Mario Sbrana	Centro Interuniversitario di Biologia Marina Viale Nazario Sauro 4 57128 Livorno Italy	Phone +39 0586 260723	msbrana@cibm.it

NAME	ADDRESS	PHONE/FAX	EMAIL
Nikolay Timoshenko	AtlantNIRO 5 Dmitry Donskogo Street RU-236000 Kaliningrad Russian Federation	Phone +7 4012 925554 Fax +7 4012 219997	timoshenko@atlant.baltnet.ru
Joël Vigneau	IFREMER Port-en-Bessin Station P.O. Box 32 F-14520 Port-en-Bessin France	Phone +33 231 515 600 Fax +33 231 515 601	joel.vigneau@ifremer.fr
Francesca Vitale	Swedish Board of Fisheries Institute of Marine Research, Lysekil P.O. Box 4 SE-453 21 Lysekil Sweden		francesca.vitale@fiskeriverket.se
Jon Helge Vølstad Chair	Institute of Marine Research P.O. Box 1870 N-5817 Bergen Norway	Phone +47 55238411 Fax +47 55235393	jon.helge.voelstad@imr.no
Wouter Willems	Institute for Agricultural and Fisheries Research Ankerstraat 1 8400 Oostende Belgium		wouter.willems@ilvo.vlaanderen.be
Lucia Zarauz	AZTI-Tecnalia AZTI Sukarrieta Txatxarramendi ugartea z/g E-48395 Sukarrieta (Bizkaia) Spain	Phone +34-946029400	lzarauz@suk.azti.es

Annex 2: Agenda

Workshop on Methods to evaluate and estimate the precision of fisheries data used for assessment. ICES Headquarters, 8–11 September.

8 September	10.00-11.30	Welcome. Presentation of participants. Finalizing agenda. Coffee break
	11.30-12.30	Presentation of national sampling and precision estimation examples
	12.30-13.30	Lunch
	13.30-15.00	Presentation of national sampling and precision estimation examples (cont.)
	15.00-15.15	Coffee break
	15.15-16.15	Presentation of national sampling and precision estimation examples (cont.)
	16.15-17.15	Summarizing and ToR a) discussion
	17.15	Adjourn
9 September	09.00-10.30	ToR a) presentations
	10.30-10.45	Coffee break
	10.45-12.00	ToR a) presentations (cont.)
	12.00-13.00	Lunch
	13.00-15.00	ToR a) presentations (cont.), summarizing and reporting
	15.00-15.15	Coffee break
	15.15-18.00	ToR a) summarizing and reporting
10 September	09.00-10.30	ToR b) presentations
	10.30-10.45	Coffee break
	10.45-12.00	ToR b) presentations and summarizing
	12.00-13.00	Lunch
	13.00-15.00	ToR b) summarizing and reporting
	15.00-15.15	Coffee break
	15.15-18.00	ToR c) Discussion and reporting
11 September	09.00-10.00	ToR c) reporting subgroups; 1) design recommendations, 2) ALK recommendations
	10.00-11.30	ToR c) reporting, summary guidelines
	11.30-13.00	Final summarizing, reporting, recommendations and closing of the workshop
	13.00	Adjourn

Additional considerations to take into account during the meeting:

It was agreed that quality assurance indices of catch sampling programs should be discussed. Documentation of catch sampling programs to discuss should include:

- Definition of sampling frame defined
- Definition of primary sampling units (PSUs) and of sampling units at lower stages
- Non-response documentation (proportion and number of trips not sampled because of refusals to take observers, or unsafe conditions for the observer etc.)
- Documented departures from sampling protocols
- Documented estimation methods for key parameters
- Indicators for precision: Effective samples sizes, design effects (is CV for each e.g. age class an appropriate measure?, maybe some weighted sum of CV's for frequency distributions). Should be reflecting the final estimate e.g. SSB, or F (But where do the variability come from?). OBS this only make sense if the CV is estimated correctly!
- Sample sizes: strata variances provide guidelines for number of stations necessary to meet requirements concerning precision (or cost of sampling). Bootstrapping tend to underestimate the variance.
- Small sample sizes should raise a flag.

In addition to precision estimates provided as relative standard errors (coefficient of variation) of parameters, the workshop should discuss guidelines/recommendations for additional measures or indicators of precision such as:

- Listing of the number of primary sampling units together with total number of individual measurements.
- Estimates of effective sample sizes if possible.
- Recommendations relative to COST

Annex 3: List of presentations

Stijn Bierman. Market sampling program at IMARES. Precision of catch numbers-at-age estimates.

Hans Gerritsen. Sampling program in Ireland.

David Hirst. A simulation study comparing various different approaches to estimating catch-at-age.

David Hirst et al. Estimating catch-at-age of commercial fish species by combining data from different sources

David Maxwell. Overview of UK (E&W) sampling and precision calculation

Pentti Moilanen. Marine Fishery Statistics and Biological Data Collection in Finland

Michael Pennington. An avoidable source of estimation bias.

Alastair Pout. Scottish sampling program.

Nikolay Timoshenko. Peculiarities of the ocean pelagic fish sampling in AtlantNIRO

Joël Vigneau. Overview of the COST Project

Katja Ringdahl. Swedish sampling of fisheries data

Lucia Zarauz and Paz Sampreda. Sampling program in Spain

Lucia Zarauz. The COSTsim package

Sondre Aanes and Jon Helge Vølstad. Precision in assessment - an example with Northeast Arctic Haddock

Annex 4: List of key parameters & statistics used for supporting stock assessments

To keep consistency with the WKACCU workshop we kept the same tables of parameters as was documented in WKACCU (2008). Parameters subject to bias only are included for consistency.

A - SPECIES IDENTIFICATION	SOURCES OF ERROR	PSU	BEST PRACTICE	QUANTIFICATION
1 - Species subject to confusion & trained staff	bias			
2 - Species misreporting	bias			
3 - Taxonomic change	bias			
4 - Grouping statistics	If groupings are estimated	To be documented	Avoid groupings	To be documented
5 - Identification key	bias			

B - LANDINGS WEIGHT	SOURCES OF ERROR	PSU	BEST PRACTICE	QUANTIFICATION
1 - Missing part	Bias			
2 - Area misreporting	Bias			
3 - Quantity misreporting:	Bias			
4 - Population of vessels	Bias			
5 - Source of information:			Current practice should be reported	
5.a Total landings or component of landing (stratum or domain such as métier)	No census of the catch or components of catch is available and must be estimated.	Typically haul, trip, port market and day. Harbor sampling	Documentation	See Italian method (info from David M.), and example from Alaska, and from Norway (IUU fishing)
5.b Uncertainty in allocation	The strata or group, e.g. métier are uncertain. Mostly related to métier.	Documentation is needed		Variability to be quantified.

B - LANDINGS WEIGHT	SOURCES OF ERROR	PSU	BEST PRACTICE	QUANTIFICATION
6 - Conversion factor:	When product weights (e.g. filets, gutted or fish on ice) are measured and conversion to round fresh fish is estimated using conversion factors. Relevant in presence of temporal and/or spatial product variability. The larger the conversion factor is (e.g. filet to round weight), the larger is the variance.	To be documented. Ship days/trips?	Variability to be quantified. Unkown at the moment. Information from logbooks? Spatial/temporal stratification. Spread out as much as possible	Documentation is needed
7 - Percentage of mixed in the landings;	Relevant where landings consists of a group of species, e.g. industrial catches. Species of Monkfish, cod and herring. Industrial catches. Estimation of proportions. (Landed as a group but consists of several species)	Should be based on the survey design.	To be quantified.	To be implemented in COST?
8 - Damaged fish landed:	Bias			
Final indicator				

C - DISCARDS WEIGHT	SOURCES OF ERROR	PSU	BEST PRACTICE	QUANTIFICATION
Recall of bias indicator on species identification				
1 - Sampling allocation scheme	Check with WKACCU Onboard observer. See 5.6 above	Trip/vessel		COST
1.5 Size of subsample(s)	If the sampling fractions of PSU are small, the majority of the variance is due to the variance between primary sampling units (Cochran 1977, p279 and p 305). However, there is a tradeoff between number of primary units and size of which is relevant for optimizing the sampling scheme. Haul weight is estimated from the subsample.	Given by the survey design.	Focus on the quality of the sample instead of the number of sampled individuals or number of hauls on a trip. (Because: This has a small contribution to the total variance.)	The trade off between subsample size and number of PSU will vary from case to case but is not quantified. See Alaska observer

C - DISCARDS WEIGHT	SOURCES OF ERROR	PSU	BEST PRACTICE	QUANTIFICATION
2 - Raising variable s	See ICES workshop on discard estimation procedure. Tech report for algorithm for raising procedures	Trip/vessel	According to sampling design	COST See ICES workshop on discard estimation
3 - Size of the catch effect	Bias			
4 - Damaged fish discarde d:	Bias			
5 - Non response rate:	Bias			
6 - Temporal coverage	If structures within a stratum changes through time, it is imperative that the design covers the changes appropriately.		Good coverage will reduce and improve the estimates of variance	
7 - Spatial coverage	If structures within a stratum shows spatial variability, it is imperative that the design covers the changes appropriately.		Good coverage will reduce and improve the estimates of variance	
8 - High grading	Bias			
9 - Slipping behaviour	It could account for a large proportion of mortality, but are rare events and require intensive sampling.			
10 - Management measure s leading to discarding beh aviour	Bias			
11 - Working conditions:	Will affect measurement error		Good protocol for sampling will reduce the problem.	Difficult to quanitfy. If to be quantified require experimenta l designs
12 - Species replacement:	See WKACCU for description. Can be estimated within a trip, not haul by haul. Will add to the variance.	trip		
Final indicator				

D - EFFORT	SOURCES OF ERROR	PSU	BEST PRACTICE
Re- call of bias indicator on species identification(if needed for métier allocation)			
1 - Unit definition			
2 - Area misreporting	Bias		
3 - Effort misreporting	Bias		
4 - Source of information			
Final indicator			

E -LENGTH STRUCTURE	SOURCES OF ERROR	PSU	BEST PRACTICE	QUANTIFICATION
Recall of bias indicator on discards/landings weight				
1 – Sampling protocol:	Sampling errors, measurement errors. Typing errors.	Market, port, day, box, trip,	Spread out as much as possible. Keep track of the sampling units	Require a well designed sampling scheme (i.e. sampling units are identified such that inclusion probabilities are known)
2 - Temporal coverage	Part of sampling protocol			
3 - Spatial coverage	Part of sampling protocol			
3.5 Over stratification. See 6 below.	More strata than samples. Having strata with no samples. Complicate estimating precision. Probably lower precision by too many strata. Cochran: no gain by more than "6" strata.		Avoid it! Have sufficient coverage	Modeling
4 - Random sampling of boxes/trips:				To be quantified through experiments

E -LENGTH STRUCTURE	SOURCES OF ERROR	PSU	BEST PRACTICE	QUANTIFICATION
5 - Availability of all the landings/discards				
6 - Non sampled strata:				Require modeling. See e.g. Hirst et al. 2005
7 - Raising to the trip:			Trip totals should be recorded	
8 - Change in selectivity	?			
9 - Sampled weight:	May be unknown or must be estimated by e.g. length weiht relationship		Have a weight available, or make an effort establishing good length weight relationship	
Final indicator				

Measurement error in aging results in a biased estimate of the age distribution and increase the variance.

If aging error is large and the effective sample size is low, then pooling of age groups may be advisable, for example for estimating SSB.

F – AGE STRUCTURE	SOURCES OF ERROR	PSU	BEST PRACTICE	QUANTIFICATION
Recall of bias indicator on length structure				
1 - Quality insurance protocol				
2 - Conventional/actual age validity				
3 - Calibration workshop				
4 - International exchange:	Within and between reader variability	Individual fish	An experimental setup for estimating aging error.	Need documentation
5 - International reference set:				Need documentation
6 - Species/stock reading easiness: AND trained staff				
7 - Age reading method				
8 - Statistical processing				
9- Temporal coverage				
10 - Spatial coverage				
11 - Plus group			See comment above	
12 - Incomplete ALK	Complicates estimation of age structures	Follow sampling design	Avoid using ALK's if possible	Require modeling
Final indicator				

G – MEAN WEIGHT	SOURCES OF ERROR	PSU	BEST PRACTICE	QUANTIFICATION
Recall of bias indicator on length/age structure				
1 – Sampling protocol:				
2 - Temporal coverage				
3 - Spatial coverage				
4 – Statistical processing				
5 - Calibration of equipment				
6 - Calibration workshop				
7 - Conversion factor	Weight is estimated from lengths. Conversion to live weights			Not quantified. Experiments are needed.
Final indicator				

H – SEX RATIO	SOURCES OF ERROR	PSU	BEST PRACTICE	QUANTIFICATION
Recall of bias indicator on length/age structure				
1 – Sampling protocol:				
2 - Temporal coverage				
3 - Spatial coverage				
4 – Staff trained				
5 - Size/maturity effect:				
6 - Catchability effect:	Bias			
Final indicator				

See workshop on maturity ogives.

OBS: Each of the extra parameters requires more intensive sampling to obtain precise estimates.

I – MATURITY STAGE	SOURCES OF ERROR	PSU	BEST PRACTICE	QUANTIFICATION
Recall of bias indicator on length/age structure				
1 – Sampling protocol:				
2 - Temporal coverage				
3 - Spatial coverage				
4 – Staff trained				
5 – International reference set				
6 - Size/maturity effect:				
7 - Histological reference:				
8 - Skipped spawning:				
Final indicator				

Annex 5: Recommendations

RECOMMENDATION	FOR FOLLOW UP BY:
1. WKPRECISE recommends that catch sampling should be based on statistically robust survey designs with clear definitions (and documentation) of the sampling frame, the primary sampling units (PSUs), the stratification schemes employed, and the methods used for selecting samples in each stratum.	MS, WKMERGE
2. WKPRECISE recommends that the precision of estimates of key parameters should be given in terms of standard errors (or relative standard errors). Estimates of precision should be based on estimators that take into account the actual survey design employed in the catch-sampling. Model-based estimates should accoubnt for clustering efefcts caused by hierachical data collections. In addition, the number of primary sampling units observed along with estimates of the effective sample size for the associated estimate should be given.	MS , relevant ICES stock assessment working groups
3. WKPRECISE recommends that the COST software for analysis of catch sampling data be extended to include additional design-based estimators of key parameters used for stock assessments, alongside its model-based estimators.	
4. For probability-based catch sampling programs, the estimation of age distributions of fish, and the associated uncertainty, directly from the multi-stage age samples should be considered rather than using an ALK.	MS , relevant ICES stock assessment working groups
5. When ALKs are used, it is imperative that age-length data are stored in a manner that links the data to the primary sampling units (e.g., tows or trips). This will allow bootstrap estimates of precision for probability-based sampling programs.	MS , relevant ICES stock assessment working groups
6. The pooling of data from multiple métiers should be done with caution. Estimation of catch statistics and biological parameters for métiers (or for pooled groups of métiers) generally requires some form of sample-weight adjustments because the primary units within a métier may come from multiple strata with different sampling intensities. Pooling of métiers without proper weighting likely causes bias, and may preclude estimates of precision	WKMERGE, RCM