ICES WKACCU REPORT 2008

ICES ADVISORY COMMITTEE

ICES CM 2008\ACOM:32

REF. LRC, RMC. PGCCDBS

Report of the Workshop on Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment (WKACCU)

27–30 October 2008

Bergen, Norway

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. 2008. Report of the Workshop on Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment (WKACCU), 27–30 October 2008, Bergen, Norway. ICES CM 2008\ACOM:32. 41 pp.

Contents

Executive summary **1**

1 Introduction **3**

1.1 Terms of reference 3

1.2 Background for the workshop 3

2 Adoption of the agenda **4**

3 Detecting and avoiding bias **5**

3.1 Some general comments on bias and data collection 5

3.1.1 Bias in scientific surveys estimates 7

3.1.2 Avoiding bias in estimates based on cluster samples 8

3.1.3 Detecting and controlling bias in at-sea observer programmes 9

4 Developing a scorecard for bias detection **14**

5 Conclusions **29**

6 References **30**

Annex 1: List of participants **31**

Annex 2: Agenda **33**

Annex 3. List of working documents and oral presentations **34**

Annex 4. Application of the score-card to detect sources of bias in fisheries statistics for Northeast Arctic Saithe **35**

Executive summary

This report summarizes key findings and recommendations of the Workshop on Methods to evaluate and estimate the accuracy of fisheries data used for assessment [ICES WKACCU] held in Bergen, Norway, 27–30 October 2008. The report identifies procedures and other factors that could cause bias in fisheries data used in stock assessments, and provides recommendations for improved procedures that could reduce such bias.

The accuracy of fisheries data is determined by two components: (1) Systematic errors (bias), and (2) random errors as measured by precision. The focus of this workshop was on the bias component of accuracy, whereas a second workshop (ICES WKPRECISE) scheduled for October 2009 will focus on precision. The WKACCU workshop primarily dealt with bias in fisheries-dependent data collection programs, but included a presentation and brief discussion of bias in scientific survey estimates of abundance indices and populations characteristics.

A total of 14 scientists from nine countries participated in the workshop. Two working documents on methods to evaluate and estimate accuracy of fisheries data were presented during the WKACCU workshop, and a series of oral presentations dealt with sources of bias in fisheries data collection programs. It is difficult to quantify bias in fisheries data used for stock assessment. Whereas precision in fisheries statistics can be improved by increasing the sample sizes in data collection programs, this is not the case with bias. Bias is a systematic departure from the true values, and can generally not be quantified because the true values seldom are known. To the extent possible, it is therefore important to minimize or eliminate sources of bias by developing and following sound field data collection procedures and analytical methods. Workshop participants developed a practical framework for detecting potential sources of bias in fisheries data collection programs.

The focus of the evaluation conducted during the workshop was a list of key parameters of importance in stock assessments: A) Species Identification; B) Landings Weight; C) Discard Weight; D) Effort; E) Length Structure; F) Age Structure; G) Mean Weigh; H) Sex-ratio; and I) Maturity Stages. The workshop identified several indicators to detect bias in each of these parameters. A simple score-card was then developed where each indicator was rated as green (minimal or no risk of bias), yellow (some risk of bias), and red (established sources of bias). The workshop recognized that some of the parameters identified are interconnected, and that the final bias evaluation must consider the sources of bias encountered during all the data collection and processing. The final indicator of bias should take into account the propagation of systematic errors across interconnected parameters. The scorecard is a practical tool to evaluate the quality of data sources used for stock assessments, and can help reduce bias in future data collections by identifying steps in the data collection process that must be improved The proposed scorecard was applied to the data collection program for the Norwegian Northeast Arctic saithe fishery in 2007. This case study suggested that the system is practical and useful, but it is recommended that more fisheries be evaluated to develop the scorecard further.

The workshop discussed several analytical methods and tools that could be used to assess the occurrence and magnitude of bias. These methods depend on the availability of appropriate data from multiple sources. Potential information for such assessments include data from vessel trip reports, logbooks, port sampling, and dealer landing reports (trip tickets). Fishery parameters that could be compared to assess

potential bias include proportion of sampled trips vs. trips made by the general fleet by vessel class, area, and time; average trip length for observed vessels vs. the general fleet, by vessel class, area, and time; average harvest (catch retained) per trip for observed vessels vs. the general fleet, by vessel class, area, and time (e.g. quarter); average depth of observed tows/sets vs. reported tows/sets by vessel class, area, and time (e.g. quarter); and extent of spatial overlap of observed tows/sets with reported fishing locations by the general fleet by vessel class, area, and time (e.g. quarter).

1 Introduction

1.1 Terms of reference

The Workshop on Methods to evaluate and estimate the accuracy of fisheries data used for assessment [WKACCU] (Co-Chairs: Michael Pennington and Jon Helge Vølstad, IMR, Norway will be established and will meet in Bergen, Norway, 27–30 October 2008 to:

A) Review the sources of bias and establish general parameters (indicators)/procedures to assess the bias 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 bias and establish general parameters (indicators)/procedures to assess the bias on national level of biological data collected from the fisheries by investigating (both visual and quantitative) the data coverage by stock, area, season and fleet.

1.2 Background for the workshop

For the current DCR and other national sampling programmes and sampling strategies, the quality of the resulting data is almost solely 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 ogive, 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 much 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 of objective indicators of data accuracy that could be taken into account when setting up sampling schemes.

The technical definition of an estimator's accuracy is the sum of the estimator variance plus its bias squared. For example, a thermometer that always says the temperature is 112°C is very precise (its variance is zero), but will generally have a large bias (112 minus the true value). At this workshop we do not address accuracy directly, but study the sources and causes of bias for particular estimators and data collection programs.

2 Adoption of the agenda

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.

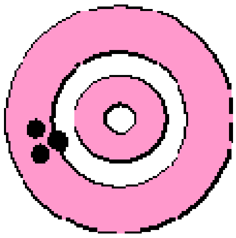
3 Detecting and avoiding bias

3.1 Some general comments on bias and data collection

Bias is the difference between the true (unknown) value and the average value of repeatedly (conceptually) applying an estimator or conducting a census to estimate some parameter: or

$$bias = \bar{\theta} - \theta_0$$

A standard pictorial example of bias is the "target example." A particular marksman shoots at a target and the bullets consistently hit the target a distance away from the bull's eye (the true value). The average distance from the bull's eye is the bias. This bias may have been caused by a faulty rifle or by the marksman having a vision problem. The difficulty with bias is that no matter how intense a biased estimation procedure is the results will be misleading.



There are basically two ways that fishery data are collected: by conducting a complete census or by using (it is hoped) a sound statistical sampling program. It is clear how a census can produce biased results. For example, fishers not reporting honestly the amount of fish caught during a trip. By contrast, bias inherent in non-census data collection program will be difficult or impossible to detect if the sampling programmes is not based on a sound survey design, for example, if it employs an ad hoc sampling scheme.

Important considerations for any sampling program are: 1) define the target population, e.g. the total commercial catch of a species; 2) carefully determine the primary sampling units, e.g. a fishing trip, fish delivered on a particular day at a port, etc.; 3) decide on the sampling intensity, e.g. how many primary sampling units should be sampled and how large a sample should be within a primary unit. Finally, it is very important that the sampling protocol is accurately and completely documented.

Some specific problems causing bias for fishery data and some solutions (not in any particular order) for sampling programmes:

- Bias may arise during sampling the length distribution from a catch if, e.g. there is size segregation in the flow of fish along the belt prior to processing and thus where the sample is taken is not a random sample of the catch. A systematic sample from each portion of the catch is one way around this.
- The availability of accurate species identification keys can be an issue, both for certain species and if there are not area specific distribution keys.
- The training of staff and adherence to established protocols are important ways of avoiding sampler's bias and for assessing the level of competence, training and staff experience.

- Bias may be caused by misreporting of landed species, e.g. when quota limits are being approached. One way for assessing this may be sudden changes in the reported species composition of landings. For processed landings, e.g. fillets, this change in species composition may be difficult to detect.
- Taxonomic changes in species nomenclature, e.g. the splitting of sandeel species.
- Grouping statistics, that is the practice of estimating catch composition from ratios rather than the identification of individual fish, e.g. for the catch composition of a discarded proportion of a catch may cause bias. The use of grouping statistics increases with increasing mixed species catches. Bias as a consequence of area misreporting, e.g. vessels misreporting fishing locations may be detected by matching logbooks with VMS data.
- Bias in catch statistics may be revealed by comparing the difference between declared landings and that recorded by on-board observers.
- Conversion factors; fillets are inherently less reliable for basing estimates of landed weight because of the uncertainty in the conversion factor used to obtain whole weight.
- High grading bias; it is unclear how to design a sampling program that detects the portion of the catch that is retained only for a short period.
- Species in mixed landings that are specified as the % of the total catch may be a source of bias.
- Working conditions may cause bias. For example, such factors include; sea conditions may affect various instruments, the layout of the vessel and uncontrollable conditions that prevent an observer from always recording the discarded catch.
- Bias that is caused by an unknown amount of slippage. It is important to acknowledging that this is a problem.
- The bias caused by the substitution of one species for another species when reporting landings. The prevalence of this switching could be determined by observers or by other means, such as interviews.
- Unknown discarding is one of the main problems that cause bias. It is not clear if there has been any attempt to quantify this problem.

Figure A gives an overview of the steps taken from sampling to stock assessment and the sources of error at each step. Gerritsen, 2007 provides a more detailed analysis of some of the main sources of error.

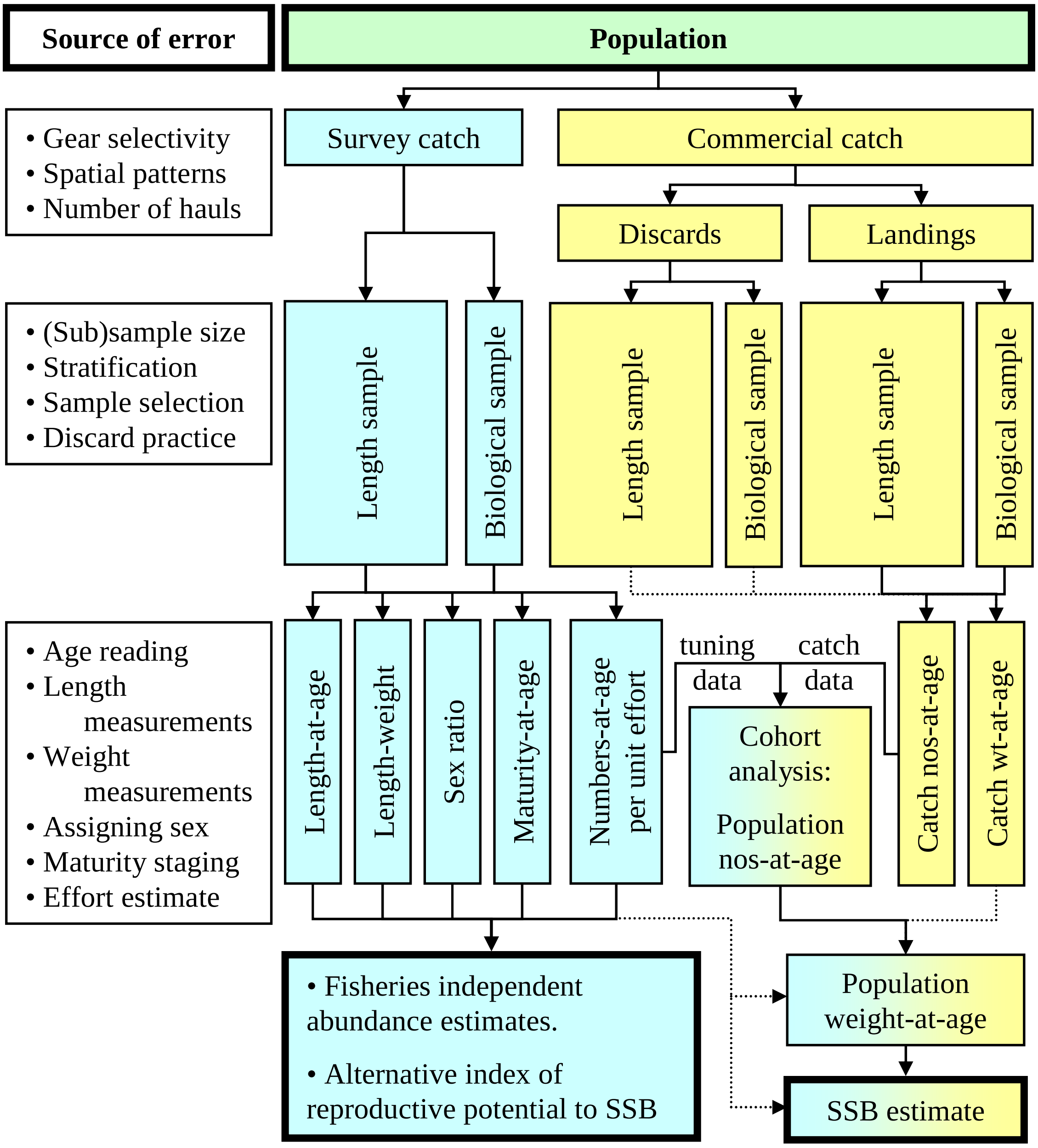


Figure A.1. Diagram of the various steps involved in data collection and analysis of fisheries data and the sources of error at each step of the process. Blue cells signify survey data, yellow cells refer to commercial data. Detailed analysis of some of the major sources of error is presented in Gerritsen, 2007.

In the following three sections we examine in more detail the possible sources of bias for three important sampling situations; scientific surveys, cluster sampling, and observer programmes.

3.1.1 Bias in scientific surveys estimates

There are many factors that may bias estimates of abundance indices and population characteristics based on data from scientific surveys. The acoustic detect ability of a stock may change from year to year because of weather conditions. Likewise, the catchability of a species by the survey trawl may vary over time as a consequence of, for example, changes in the spatial distribution, environmental conditions or in the length composition of the population. If such changes occur fairly randomly, then the yearly survey indices will be more variable, but they will still track abundance trends. If there are persistent factors that affect the estimates, then these will cause the estimates to be biased. For example, a bottom-trawl survey is sometimes used to estimate the absolute abundance of a demersal species. If it is assumed that all the fish in the area swept by the trawl are caught, but only a proportion is, on average, captured, then the estimates will be biased downwards by a factor equal to the proportion of fish that are not caught by the trawl. When this bias is constant over time, the survey series will still follow trends in abundance but will consistently underestimate the actual abundance. If the size of the bias is known, then the estimates can be adjusted accordingly to generate estimates of absolute abundance.

Large survey catches are often not completely sorted, but estimates, e.g. of length–frequency distributions, are based on a subsample of the catch. As is well known, it is difficult to take a truly random sample of fish from a large catch at sea, and therefore, estimates based on a subsample may be biased in an unknown manner. One way to protect against such "subsample bias" is to reduce tow duration, e.g. from 30 minutes to 15 minutes. Shorter tows have been demonstrated to be usually as efficient as long tows, and in general it is better to take small samples from as many locations as possible rather than take large samples from a few locations (Pennington and Vølstad; 19 91, 1994). By reducing tow duration, there will be fewer catches that need to be subsampled which will lessen this source of bias.

3.1.2 Avoiding bias in estimates based on cluster samples

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 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. Because 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 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.

3.1.3 Detecting and controlling bias in at-sea observer programmes

This section is largely drawn from a workshop held in Woods Hole, Massachusetts in 2006 to evaluate potential causes and means of controlling for vessel selection bias in NOAA observer programmes (Vølstad and Fogarty, 2006[1]). The workshop examined in depth the bias issues associated with estimates of total catch (including discard) and catch characteristics based on data collected by on-board observers for 24 fisheries representing all regions of the US marine fisheries. Although the focus of the workshop was on on-board observer programmes, the statistical points raised apply to any sampling program.

An observer program is a costly way to detect and monitor sources and causes of bias in fishery data from trip-tickets, logbooks, and port sampling. In respect to bias, a vessel and selection procedure for an observer program is considered biased if it results in catch and bycatch data that do not represent the fleet (and its fishing operations) on average (i.e. the procedure will tend to result in observer data that systematically deviates from data that would be representative of the true fleet and its fishery). Random selection is a safeguard against systematic bias in the selection procedure (i.e. on average, the samples will represent the total population of vessels in the list). A random selection of vessels, however, does not in itself eliminate sys-

---

[1] Vølstad, J.H., and M. Fogarty. 2006. Report on the Observer Program Vessel Selection Bias Workshop, Woods Hole, MA, May 17–19, 2006.

tematic bias. If observers cannot be deployed on the vessels selected by a representative method such as random sampling, or if some of the vessels selected change fishing behavior, then the resulting sample is biased. Bias resulting from logistical problems and lack of compliance is particularly difficult to quantify and control and is not likely to be reduced by increasing sample sizes.

Major causes of bias are: (1) incomplete sampling frame (i.e. incomplete list of active vessels in a fishery), (2) biased sampling procedures for selecting vessels from the sampling frame or by factors preventing the deployment of observers on all selected vessels, (3) and observer bias (i.e. measurement errors caused by changes in fishing behavior in the presence of observers).

3.1.3.1 Incomplete sampling frame

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. 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.

3.1.3.2 Sample bias related to selection of vessels from the frame and deployment of observers

The goal of selecting vessels and deploying observers should be to obtain data from trips that are representative of actual fishing effort over the entire fishing season and the full geographic range of the fishery, as well as of vessel type, gear type, and targeting strategy. Six methods for selecting vessels were documented for the 24 observer programmes evaluated during the workshop:

- census – every trip is observed for all vessels in the sampling frame
- random sampling with replacement (RS) – any vessel in the frame has a known probability (> 0) of being selected in each random sample, even if it has been previously selected (i.e. after a vessel has been chosen from the list, it is put back on the list before the next draw); this selection method includes "proportional to size" selection (i.e. selecting vessels with a probability that is proportional to their expected number of trips)
- stratified random sampling with replacement (STRS) – any vessel within a stratum has the same (known) chance of being selected, even if it has been previously selected
- stratified random sampling without replacement (STRWOR) – all vessels are covered within a selection cycle; each vessel is observed only once in each cycle (i.e. once a vessel in a stratum has been selected using RS, it is not available for subsequent draws)
- systematic random sampling – every k^{th} vessel from the list is selected, starting at a random location on the list
- ad hoc sampling – vessels are selected without known inclusion probability from all vessels in the frame

Performing a census would eliminate the potential for bias (assuming that the sample frame is complete and there is 100% compliance), but this approach usually is prohibitively expensive. Typically, available resources allow for observing only a fraction of the vessels in a given fleet. Precise estimates of catch and bycatch, 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 NMFS' ability to place observers on all selected vessels. 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. 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.

Deployment bias is equivalent to non-response error and is most often caused by logistical constraints, for example when the operators of vessels in the sample refuse to take observers, when some of the vessels selected for observer deployment are unsafe, or when selected vessels do not have space for observers. In principle, an ad-hoc selection with full compliance may cause no more systematic error than a random selection procedure with poor compliance (equivalent to a low response rate). An acceptable proportion of observable vessels (response rate) for a given observer program cannot be stated in absolute terms (e.g. 75% or higher), but will depend on the mode of data collection, characteristics of the fleet and its fishery, and the similarity between catch and bycatch rates of the unobservable vessels and those of the fleet as a whole. For a general discussion of acceptable response rates we refer the reader to Lessler and Kalsbeek, 1992. When the response rate is low, it is particularly important to evaluate what portion of the total catch is accounted for by vessels that cannot be observed and if these vessels have characteristics and fishing behavior that substantially deviate from the covered fleet. For example, if smaller vessels that cannot accommodate observers tend to operate closer to shore than the general fleet, then the catch and bycatch rates of observed vessels probably would not represent the rates of the unobserved vessels.

3.1.3.3 Observer bias

The implication of observer bias is that data recorded on selected vessels is not representative of the fishery as a whole. Observer bias can occur when vessel operators systematically change their fishing behavior, effort, and location when observers are aboard. In this case, the catch and bycatch rates for observed trips would deviate from the true typical rates. This could occur if the fisher has an incentive to lower bycatch estimates (e.g. if the fisher believes that actual bycatch estimates could result in early closure of a fishery due to in season management or changes in regulations that could restrict his future fishing opportunities). This form of sampling bias is the most difficult to evaluate and correct. Systematic errors in data collection and recording also fall into the category of observer bias, but these components were outside the scope of this workshop.

3.1.3.4 Methods for evaluating the occurrence of bias

Systematic bias in estimates of catch and bycatch are likely to be small if the observed vessels and trips have similar characteristics and fishing behavior to those of the general fleet, but would clearly be greater if the catch and bycatch characteristics of the unobserved vessels deviate substantially from the norm.

Self-reporting programmes include fishing logbooks completed by fishers; landings reports completed by fishers, dealers (i.e. buyers or processors), or both; and interviews of fishers. Determining the accuracy of observer data can be difficult unless there are methods for validating these data. Self-reporting programmes may provide reliable data on effort, length of trips, and landed catch that can be compared with estimates from observer programmes to identify potential sources of bias. These programmes are less likely to be accurate for data on bycatch and total catch, including discard. State resource agencies generally require dealers to report the amount of fish bought and sold by vessel and species; however, dealer's reports and information reported by fishers generally do not include data on at-sea discards and may be unreliable because of low rates of compliance with reporting requirements. Data on catch may be obtained by port-sampling, but there are significant concerns about the completeness and accuracy of these reports, particularly for discards, which are not observed by the port sampler.

3.1.3.5 Adequacy of sampling frame

To minimize the potential for bias, the frame used for vessel selection must cover all vessels participating in the fishery and should be based on the most current list of active vessels. When a significant number of active vessels is excluded from the frame, the vessels in the frame should have characteristics similar to those of the overall fleet (i.e. be representative). Workshop participants identified the following "diagnostics" for evaluating the representativeness of the sampling frame:

- a comparison of the characteristics of vessels included in the sampling frame with those of vessels known to be part of a fishery, but that are not included in the sampling frame (e.g. length distributions of vessel, gear type)
- an analysis of the proportion of the total catch for the fleet that was landed by vessels in the sampling frame (by area and time)

3.1.3.6 Adequacy of vessel selection and observer deployment

To diagnose selection or deployment bias, it is important, when feasible, to compare the observed vessels and trips with the general fleet using (1) self-reported data obtained from logbooks, trip reports, and dealer's reports, or (2) at-sea observations, including observers' reports and remote VMS. Comparisons can be made between vessel characteristics, areas fished, spatial distribution of effort, gears used, trip lengths, average landed harvest, and depths fished using both statistical and graphical methods. Such comparisons are particularly important in programmes using ad-hoc selection of vessels because this method is the most likely to produce biased estimates. When appropriate self-reported data or at-sea observations are available, diagnostics of bias may include comparisons of the areas and times of trips and landed catch of target species to determine significant differences in fishing operations between the observed vessels and the fleet as a whole (e.g. Liggens *et al.*, 1997; Sampson 2002; Walsh *et al.*, 2002; Rago *et al.*, 2005). An evaluation of the extent to which observed trips are representative of the general fishery may also be based on comparisons of

- average trip length for observed vessels vs. general fleet, by vessel class, area and time (e.g. paired t-test);
- average harvest (catch retained) for observed vessels vs. general fleet, by vessel class, area, and time (e.g. quarter; paired t-test);
- average depth of observed tows/sets vs. reported tows/sets by vessel class, area, and time (e.g. quarter);
- the spatial and temporal overlap of observed tows/sets with fishing locations reported by the general fleet by vessel class, area, and time (e.g. quarter).

When VMS information is available, it is also useful to compare the spatial distribution of fishing effort for vessels with VMS with the distribution of tows on observed trips by area and time.

Although observer bias is not strictly a vessel selection issue, we also recommend evaluating potential observer effects on estimates of catch and bycatch, when feasible. Comparing landed catch per trip for observed vessels with those values for unobserved vessels or trips can identify changes in fishing behavior. If fishers avoid areas where bycatch typically is high or change trip duration, length of tow, or other aspects of fishing operations to reduce bycatch when observers are aboard, then estimates of bycatch are likely to be biased. Regulations such as those associated with individual fishing quotas, in-season bycatch quotas, and marine protected areas may encourage different behavior for unobserved vessels. VMS reports display a concentration of 10% to 20% of effort within 5 km of marine closures in New England waters, indicating that fleets reallocate effort away from closed areas; however, effort appears to increase in the vicinity of protected areas because operators expect higher catch rates.

A change in fishing behavior aboard observed vessels is the most difficult source of bias to evaluate and correct. This observer bias can be eliminated only through a census (i.e. by observing all hauls or sets accurately throughout the fishery). Increasing the coverage of trips will not necessarily reduce such bias. Observer bias is usually diagnosed and quantified by comparing the behavior of vessels during observed trips or hauls/sets with the behavior of the general fleet, or by comparing the fishing operations of individual vessels during observed and during unobserved trips. Comparisons of trip or haul duration, fishing location, and catch-per-unit-effort and other metrics that characterize fishing behavior can help diagnose if the observed vessels and trips are representative of the fishery as a whole. Such comparisons generally can be made against only self-reported information from the fishing fleet; consequently, they must be interpreted with care.

In Section 4, we present a “scorecard” for rating the data available to assess a particular stock. It is assumed that the lower the “score”, the more likely that estimates based on these data will be biased.

4 Developing a scorecard for bias detection

Bias in fisheries data used for stock assessments is difficult to quantify. A major focus of the WKACCU workshop was to review and develop practical methods for evaluating potential sources of bias in fisheries data collection programs, and means of minimizing or eliminating such bias. The approach was to develop simple indicators of bias in key parameters that could be summarized in a table with a scorecard of green (minimal or no risk of bias), yellow (some risk of bias), and red (established sources of bias). 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 identifying steps in the data collection process that must be improved. The WKACCU participants chose the following list of key parameters that should be scored to evaluate potential bias in data used for stock assessment:

A) Species Identification

B) Landings Weight

C) Discard Weight

D) Effort

E) Length Structure

F) Age Structure

G) Mean Weight

H) Sex-ratio

I) Maturity Stages

Within these parameters we identified several indicators to detect potential bias. A parameter can only get a green score indicating minimal risk of bias if all the indicators are green. When one indicator is red, the total score for this parameter is red. For the orange, it is more difficult to decide how many 'oranges' are allowed, and also which combination of 'oranges' because not all indicators within a certain parameter are equally important. It was therefore decided that orange or red flags should be accompanied by the list of parameters that were given a potential or confirmed bias status.

Moreover, some parameters are interconnected and the final bias evaluation must consider the sources of bias encountered during all the processing. In the final table for the proposed bias indicator, the first reference is then the recall of the bias encountered in the linked parameters. By doing this, the final indicator is respectful of the propagation of the bias and all information is given to the user for estimating the bias on each single table. As an example, the bias parameter on the length structure of the landings (indicator E), makes first reference to the recall of the bias on the landings weight (indicator B) used for raising, which makes first reference to the indicator of species identification (indicator A). At the end, the bias indicator for the length structure of the landings cumulates the bias on three parameters.

For each indicator we specify the two extremes (green and orange OR green and red), which are easier to assess. All indicators that falls outside these extremes are given the orange status (risk of bias).

The review and practical methods to evaluate the bias for each parameter are the following:

A - Species identification

1) Species subject to confusion: The risk of bias is inherent to the species itself, depending on the difficulty of its identification. A way of evaluating the bias could be through a reference table of species to be agreed by an international forum. The setting of such a table, specific to fishing areas/regions, should be addressed by the ICES PGCCDBS.

2) Staff trained for species identification: information such as the time since the last training or information on the experience at sea are the elements to determine the risk of bias on species identification at the end of a sampling. This source of bias must be combined to the previous one as on one hand a species easy to identify do not present major risk of bias even for a novice, and on the other hand a species difficult to identify is not a source of bias if sampled by a taxonomist.

3) Species misreporting: A sudden increase of an unexpected species may occur in the statistics, thus pointing out a potential risk of species misreporting. This case is generally linked to quota consumption. Another way of detecting such a bias is dissimilarities between on-board observers reporting for the same fishing activity, or dissimilarities between on-board observers and landing statistics.

4) Taxonomic change: changes in species nomenclature over time, e.g. the splitting of sandeel species in the face of new knowledge, may impact the consistency of a time-series.

5) Grouping statistics: some commercial naming include groups of several species, e.g. lophius, megrims. It may also be the case that a commercial naming includes incidentally other species, as often encountered with the elasmobranchs (e.g. mixture of ray species in a box of *Raja clavata*). Scientific sampling surveys are generally used to quantify the percentage of species within the relevant commercial names, and if it is the case, there is no major risk of bias.

6) Existence of an identification key: photographs or sketches of species of relevance in a given fishing area are very useful tools to ensure correct species identification. The absence of such identification keys, however, is not to be considered a source of bias when the staff that conduct the species identification is trained and experienced in taxonomy.

B - Landings weight

1) Missing part: ratio between the retained fractions estimated on-board by observers and the landings of a species. A statistical test can be performed to evaluate if the slope is significantly different from one.

2) Area misreporting: like for the species misreporting, there may be a sudden increase of a species reported in an uncommon neighbouring area. This type of bias may be assessed by checking the consistency between different sources e.g. logbooks, VMS, sales notes, cpue trends of commercial vs. surveys, …

3) Quantity misreporting: known as the most current bias in fisheries statistics, this bias may be assessed together with area misreporting and with the addition of sources like economic surveys and EU control database.

4) Population of vessels: are all vessels included in the population that forms the sampling frame?

5) Source of information: it is unlikely that one source of information encompasses the statistics of all fisheries, including the temporal, spatial and fishing activity stratification. In all cases, the advantages and limitations of the sources used should provide a clear view on the related bias.

6) Conversion factor: information such as the age and the methodology used for the conversion factor, are indications on the potential induced bias. The magnitude of the conversion factor used is also an indication, e.g. estimating landing weight from fillet or from gutted fish will lead to different amplification of a bias.

7) Percentage of mixed in the landings; linked to the bias no A-5 of species identification, this potential bias is also linked to the methodology used for estimating the statistics of a species.

8) Damaged fish landed: some cases were reported of fishers proposing for sale incomplete, i.e., fish partially cut for any reason, such as bite by a shark.

C - Discards weight

1) Sampling allocation scheme: estimation of the randomness of the sampling. Is sampling pure random with a sampling protocol well followed, or is sampling allocation made on *ad hoc* or opportunistic observations? A non random sampling is clearly a source of bias which needs to be reported.

2) Raising variable: For raising to the population, different raising procedures must be compared and also raising the retained fraction to be compared with the landing statistics is a solution to assess the relevance of the variable used for raising (WKDRP, 2007).

3) Size of the catch effect: When catches are big and only a guest mated fraction has been sampled, the bias is more likely than when a significant fraction of the catch (say more than 10%) is taken for sampling. In general this information is absent even from the raw samples.

4) Damaged fish discarded: identical problem as bias no B-7 of landings weight.

5) Non response rate: the percentage of refusal is one of the most important sources of bias for on-board observers. This case discussed in general in Cochran, 1977 has also been addressed by the recent workshop on discards (Anon, 2003) in the frame of the DCR.

6) Temporal coverage: it has been discussed during the workshop that any discrepancy between the sampling and fishing effort coverage do not lead to a bias when the sampling is done randomly following a well designed protocol. In other cases, the temporal coverage in terms of mean discrepancy between proportion by units of time plus existence of non sampled strata must be evaluated.

7) Spatial coverage: identical as temporal coverage above.

8) High grading; selecting a given size range for landing a species depending on the market demand or to reduce the quota consumption automatically change the discarding ogive. High grading behaviour may be evaluated by interviews and/or on-board observers.

9) Slipping behaviour: In general, this behaviour is linked to specific fisheries such as pelagic trawling. The more or less rare occurrence of rejecting all the catch before it comes on the vessel deck needs to be evaluated.

10) Management measures leading to discarding behaviour: the specification of the measure and the date of entry into force are indications of potential bias, if not monitored through a well-designed sampling programme.

11) Working conditions: evaluating the sampled weight with a scale needs proper conditions, which are not always possible. Sampling for discards needs also good conditions for taking the sample and enough time and space for carrying the scientific work. Any constraint on working conditions may lead to a bias in the final estimates.

12) Species replacement: species thrown away (discarded) because replaced by another. This behaviour, linked to the carriage capacity, must be evaluated if it occurs, either by a well-designed sampling programme (no bias) or by external source (risk of bias).

D - Effort

1) Unit definition: Existence and follow-up of an international agreed definition and specifications. Effort statistics obtained through a census or a sampling programme.

2) Area misreporting: This bias may be assessed by checking the consistency between on-board observers, questionnaire surveys, VMS and logbooks. If there is a bias on area misreporting for the landings weight (bias no B-2), it is likely that a similar bias exists for effort.

3) Effort misreporting: similar to quantity misreporting for landings (bias no B-3). This major risk of bias is to evaluate the total effort on an incomplete population. The way of evaluating it is by checking different sources like the area misreporting above.

4) Source of information; identical to the same bias for landings weight (bias no B-4).

E - Length structure

1) Sampling protocol: Existence and adherence to a sampling protocol that yields representative selection of fish for length measurements.

2) Temporal coverage: it was discussed during the workshop that any discrepancy between the sampling and fishing effort coverage do not lead to a bias when the sampling is done randomly following a well designed protocol. In other cases, the temporal coverage in terms of mean discrepancy between proportion by units of time plus existence of non sampled strata must be evaluated.

3) Spatial coverage: identical as temporal coverage above.

4) Random sampling of boxes/trips: This bias, linked to the follow-up of a sampling protocol (bias no E-1), focuses more on the randomness of both the choice of boxes to sample (always the top box, vs. real random,) and the choice of trips (opportunistic, real random).

5) Availability of all the landings/discards: this bias is linked to the missing part (bias no B-1 of landings weight), but more focused on the special conditions linked to the auction sales conditions. The responsible for sampling are the experts having the knowledge of this information.

6) Non sampled strata: Usually, imputation rules exist for non sampled strata, thus this bias will be an evaluation of the appropriateness of the imputation rules used.

7) Raising to the trip: This bias, linked to the follow-up of a sampling protocol (bias no E-1), focuses on the raising variable used (exact knowledge of the landings weight, guest mates).

8) Change in selectivity: bias linked to the characteristics of the gear and evaluation whether the length structure sampled is representative of the exact characteristics of the gears used at the population level.

9) Sampled weight: Is the sampled box weight measured by the staff responsible for sampling, by the crew or by the port staff?

F - Age structure

1) Quality insurance protocol: Existence and follow-up of a sampling protocol.

2) Conventional/actual age validity: Existence of a validity control for the appropriateness of the reading to evaluate the true age (check with tagging or *in vivo* growing programmes).

3) Calibration workshop: Existence of a recent age reading workshop.

4) International exchange: Existence of a recent international exchange in order to compare the results of age reading by several readers from different countries on the same material. Usually, the exchange is carried out in preparation of an age reading workshop or at regular interval to assess the need of convening such a workshop.

5) International reference set: Existence and routinely use of an agreed international reference set.

6) Species/stock reading easiness: The risk of bias is inherent to the species/stock itself, depending on the difficulty of reading the age. The international calibration workshops use software able to evaluate such a bias.

7) Staff trained for age reading: information such as the time since the last training or information on the experience of the staff are the elements to determine the risk of bias on age reading. Some international calibration workshop evaluate the competence of age readers for estimating age structure for stock assessment purpose, Age readers formally approved by such a forum, would lead to an absence of bias for this parameter.

8) Age reading method: Some reading methods are known to be biased for estimating some or all ages. This information is usually found in the reports of international calibration workshops.

9) Statistical processing: when direct age reading is impossible, statistical methods may be used to estimate the age structure. This bias is to be evaluated by those responsible of carrying out the analysis.

10) Temporal coverage: identical as temporal coverage of the length structure, focused on the collection of materials for age reading.

11) Spatial coverage: identical as temporal coverage above.

12) Plus group: bias linked to the setting of the plus group, and the existence or not of international agreement.

13) Incomplete ALK: Appropriateness of the imputation rules for filling length classes without age information.

G - Mean weight

1) Sampling protocol: Existence and follow-up of a sampling protocol.

2) Temporal coverage; identical as temporal coverage of the length structure (E-2), focused on the data used for mean weight estimates.

3) Spatial coverage: identical as temporal coverage above.

4) Statistical processing: Appropriateness of the statistical method used, if any. It is often the case that a length-weight relationship is used or a Van Bertalanffy model. The time between the references used for modelling and the actual time strata is an indication on the potential induced bias.

5) Calibration of equipment: Existence of a routine calibration validation of the equipment used.

6) Working conditions: evaluating the mean weight with a scale needs proper conditions, which are not always possible. Any constraint on working conditions may lead to a bias in the final estimates.

7) Conversion factor: if a conversion factor is needed, information such as the age and the methodology used for the conversion factor, are indications on the potential induced bias. The appropriateness of the conversion factor needs also to be asserted when estimating mean weight of fish during or outside the spawning period.

H - Sex-ratio

1) Sampling protocol: Existence and follow-up of a sampling protocol.

2) Temporal coverage; identical as temporal coverage of the length structure (E-2), focused on the data used for sex-ratio estimates.

3) Spatial coverage; identical as temporal coverage above.

4) Staff trained: information such as the time since the last training or information on the experience of the staff are the elements to determine the risk of bias on estimating the sex of a species.

5) Size/maturity effect: How are immature issues being addressed? Is the method used well described and approved?

6) Catchability effect: for some species the catchability by sex vary over time. If such behaviour related change in catchability occurs, do the estimates take this into account following an agreed protocol?

I - Maturity stage

1) Sampling protocol: Existence and follow-up of an international sampling protocol (ICES WKMAT 2007, survey protocols).

2) Appropriate time period: ICES WKMAT recommended collecting maturity information during the spawning period. Is the period to collect maturity information known and agreed internationally?

3) Spatial coverage: identical as spatial coverage of the length structure (E-3), focused on the data used for maturity estimates.

4) Staff trained: information such as the time since the last training or information on the experience of the staff is the elements to determine the risk of bias on estimating the maturity stages.

5) International reference set: Existence and routinely use of an agreed international reference set.

6) Size/maturity effect: existence of a protocol for dealing with immature.

7) Calibration workshop: Existence and follow-up of the recommendations of a recent international maturity calibration workshop.

8) Histological reference: Existence and follow-up of internationally agreed references.

9) Skipped spawning: following ICES WKMAT recommendation, is skipped spawning known to happen and taken into account?

The proposal for the bias indicator is the following for each parameter:

A - SPECIES IDENTIFICATION	NO BIAS	RISK OF BIAS	CONFIRMED BIAS
1 - Species subject to confusion & trained staff	Staff trained and experienced OR Easily defined species	Any other situation	Species difficult to identify AND Novice staff
2 - Species misreporting	Checked and no problem OR checked and corrected	Any other situation	Checked + pb + not corrected
3 - Taxonomic change	No OR Yes and taken into account	Yes AND not taken into account	
4 - Grouping statistics	No groupings OR groupings and estimated	Any other situation	Groupings AND not estimated
5 - Identification key	Yes OR No and species confusion = GREEN	Any other situation	
Final indicator	All green	List of potential bias	List of confirmed bias

B - LANDINGS WEIGHT	NO BIAS	RISK OF BIAS	CONFIRMED BIAS
Recall of bias indicator on species identification	All green	List of potential bias	List of confirmed bias
1 - Missing part	Checked and Ratio = 1 OR checked and corrected	Any other situation	Confirmed missing but not corrected
2 - Area misreporting	Checked and no problem OR checked and corrected	Any other situation	Checked and problem not corrected
3 - Quantity misreporting:	Checked and no problem OR checked and corrected	Any other situation	Checked and problem not corrected
4 - Population of vessels	All covered	-	Partially covered
5 - Source of information:	Several sources considered	Only one source used	
6 - Conversion factor:	Whole fish OR appropriate conversion factor	Any other situation	CF Wrong OR Not whole and CF not used
7 - Percentage of mixed in the landings;	None OR Checked and corrected	Any other situation	Checked and problem not corrected
8 - Damaged fish landed:	No partial fish	Any other situation	problem not corrected
Final indicator	All green	List of potential bias	List of confirmed bias

C - DISCARDS WEIGHT	NO BIAS	RISK OF BIAS	CONFIRMED BIAS
Recall of bias indicator on species identification	All green	List of potential bias	List of confirmed bias
1 - Sampling allocation scheme	Well designed random sampling	Ad hoc OR opportunistic sampling	No sampling
2 - Raising variable	No raising factor needed OR follow accepted raising procedure	Any other situation	No raising factor when needed
3 - Size of the catch effect	Well designed random sampling	Any other situation	Checked and problem not corrected
4 - Damaged fish discarded:	No partial fish	Any other situation	Problem not corrected
5 - Non response rate:	High response rate/low refusal rate (figure needed)	Any other situation	Low response rate/high refusal rate
6 - Temporal coverage	Documented and OK	Any other situation	Documented and not OK
7 - Spatial coverage	Documented and OK	Any other situation	Documented and not OK
8 - High grading	no High grading OR High grading estimated	Any other situation	High grading existing but not estimated
9 - Slipping behaviour	no slipping OR slipping estimated	Any other situation	Slipping existing but not estimated
10 - Management measures leading to discarding behaviour	management not leading to impact discards behaviour OR impact corrected	Any other situation	Strong management leading to discarding and limited at sea sampling
11 - Working conditions:	good conditions OR conditions not ideal but compensated for	Any other situation	Difficult conditions and not compensated for
12 - Species replacement:	no occurrence OR corrected	Any other situation	Occurrence and not corrected
Final indicator	All green	List of potential bias	List of confirmed bias

D - EFFORT	NO BIAS	RISK OF BIAS	CONFIRMED BIAS
Recall of bias indicator on species identification(if needed for métier allocation)	All green	List of potential bias	List of confirmed bias
1 - Unit definition	Definition available	Any other situation	Problem not corrected
2 - Area misreporting	Checked and no problem OR checked and corrected	Any other situation	Checked and problem not corrected
3 - Effort misreporting	Checked and no problem OR checked and corrected	Any other situation	Checked and problem not corrected
4 - Source of information	Several sources considered	Only one source used	
Final indicator	All green	List of potential bias	List of confirmed bias

E -LENGTH STRUCTURE	NO BIAS	RISK OF BIAS	CONFIRMED BIAS
Recall of bias indicator on discards/landings weight	All green	List of potential bias	List of confirmed bias
1 – Sampling protocol:	Existence and follow-up of a well documented protocol	Any other situation	Non existing protocol OR Existing but not followed
2 - Temporal coverage	Documented and OK	Any other situation	Documented and not OK
3 - Spatial coverage	Documented and OK	Any other situation	Documented and not OK
4 - Random sampling of boxes/trips:	Representative sampling	Any other situation	Known unrepresentative sampling
5 - Availability of all the landings/discards	Known complete availability	Any other situation	Known to be unavailable and uncorrected
6 - Non sampled strata:	All strata sampled OR not all sampled but corrected by proper imputation technique	Any other situation	Not all strata sampled and problem uncorrected
7 - Raising to the trip:	Follow-up an agreed procedure	Any other situation	No raising factor when needed
8 - Change in selectivity	Checked and no problem OR problem corrected	Any other situation	Checked and problem not corrected
9 - Sampled weight:	Described and controllable	Any other situation	Known inaccurate uncontrollable procedures
Final indicator	All green	List of potential bias	List of confirmed bias

F – AGE STRUCTURE	NO BIAS	RISK OF BIAS	CONFIRMED BIAS
Recall of bias indicator on length structure	All green	List of potential bias	List of confirmed bias
1 - Quality insurance protocol	Existence and follow-up of a well documented protocol	Any other situation	Non existing protocol OR Existing but not followed
2 - Conventional/actual age validity	Checked and actual reading validated	Any other situation	Checked and invalidated
3 - Calibration workshop	Not needed OR Recently conducted	Any other situation	problem identified during a workshop not corrected
4 - International exchange:	Recently assessed and made use of	Any other situation	Recently assessed and problem not corrected
5 - International reference set:	Yes	No	
6 - Species/stock reading easiness: AND trained staff	Trained and experienced OR Easily defined Species	Any other situation	Difficult to read age, and novice staff
7 - Age reading method	Method described and appropriate	Any other situation	
8 - Statistical processing	Method described and appropriate	Any other situation	Wrong methods used
9- Temporal coverage	Documented and OK	Any other situation	Documented and not OK
10 - Spatial coverage	Documented and OK	Any other situation	Documented and not OK
11 - Plus group	No Group+ needed OR Group+ defined according to established procedures	Any other situation	Inappropriate Group+ used
12 - Incomplete ALK	Complete and validated OR Incomplete corrected following an agreed procedure.	Any other situation	Incomplete and uncorrected
Final indicator	All green	List of potential bias	List of confirmed bias

G – MEAN WEIGHT	NO BIAS	RISK OF BIAS	CONFIRMED BIAS
Recall of bias indicator on length/age structure	All green	List of potential bias	List of confirmed bias
1 – Sampling protocol:	Existence and follow-up of a well documented protocol	Any other situation	Non existing protocol OR Existing but not followed
2 - Temporal coverage	Documented and OK	Any other situation	Documented and not OK
3 - Spatial coverage	Documented and OK	Any other situation	Documented and not OK
4 – Statistical processing	Not needed OR Method described and approved	Any other situation	Problem but not taken into account
5 - Calibration of equipment	Equipment Properly calibrated	Any other situation	Known use of non-calibrated equipment
6 - Calibration workshop	Good conditions OR conditions not ideal and compensated for	Any other situation	Difficult conditions not compensated for
7 - Conversion factor	Appropriate use of Conversion factor and detection of outliers	Any other situation	Conversion factor wrong OR Not whole fish and conversion factor not used
Final indicator	All green	List of potential bias	List of confirmed bias

H – SEX RATIO	NO BIAS	RISK OF BIAS	CONFIRMED BIAS
Recall of bias indicator on length/age structure	All green	List of potential bias	List of confirmed bias
1 – Sampling protocol:	Existence and follow-up of a well documented protocol	Any other situation	Non existing protocol OR Existing but not followed
2 - Temporal coverage	Documented and OK	Any other situation	Documented and not OK
3 - Spatial coverage	Documented and OK	Any other situation	Documented and not OK
4 – Staff trained	Trained and experienced	Any other situation	Novice
5 - Size/maturity effect:	Method described and approved	Any other situation	No method OR Method available but not used
6 - Catchability effect:	No problem OR problem assessed + corrected	Any other situation	problem assessed and not corrected
Final indicator	All green	List of potential bias	List of confirmed bias

I – MATURITY STAGE	NO BIAS	RISK OF BIAS	CONFIRMED BIAS
Recall of bias indicator on length/age structure	All green	List of potential bias	List of confirmed bias
1 – Sampling protocol:	Existence and follow-up of a well documented protocol	Any other situation	Non existing protocol OR Existing but not followed
2 - Temporal coverage	Documented and OK	Any other situation	Documented and not OK
3 - Spatial coverage	Documented and OK	Any other situation	Documented and not OK
4 – Staff trained	Trained and experienced	Any other situation	Novice
5 – International reference set	Available and taken into account	Any other situation	Available and known mistakes not corrected
6 - Size/maturity effect:	Protocol exists and followed-up	Any other situation	Known problem not taken into account
7 - Histological reference:	Available and taken into account	Any other situation	Available and known mistakes not corrected
8 - Skipped spawning:	Not an issue OR Known to happen and dealt with	Any other situation	Known to happen and not dealt with
Final indicator	All green	List of potential bias	List of confirmed bias

5 Conclusions

Accuracy of fisheries statistics used for assessment is determined by amount of bias and the precision of key parameters. The terms of reference for the WKACCU workshop was limited to the identification of sources of bias in parameters and data collection procedures to assess national level fisheries statistics, and thus did not fully address the accuracy. A second workshop, WKPRECISE, to be held in September 2009 will focus on precision of key parameters and the propagation of sampling errors to the final stock assessments and will complement the WKACCU workshop.

Whereas precision in fisheries statistics can be improved by increasing the sample sizes in data collection programs, this is generally not the case with bias. Bias is a systematic departure from the true values caused by non-representative data collections and other persistent factors, and can generally not be quantified because the true values seldom are known. The workshop therefore concluded that the focus should be to minimize or eliminate sources of bias by developing and following sound field data collection procedures and analytical methods. The workshop participants developed a practical framework for detecting potential sources of bias in fisheries data collection programs. A simple score-card was applied to indicators of bias for a suite of parameters that are important for stock assessments. 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 identifying steps in the data collection process that must be improved.

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 Industrial and Applied Mathematics, Philadelphia.

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

Kish, L. 1965. Survey Sampling. Wiley, New York.

Lee, Y.W. and D.B. Sampson. 2000. Spatial and Temporal Stability of Commercial Groundfish Assemblages off Oregon and Washington as Inferred from Oregon Trawl Logbooks. Canadian Journal of Fisheries and Aquatic Sciences 57:2443–2454.

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. New York.

Liggens, G.W., M.J. Bradley and S.J. Kennelly. 1997. Detection of bias in observer-based estimates of retained and discarded catches from a multispecies trawl fishery. Fisheries Research. 32:133–147.

Murawski, S.A., S.E. Wigley, M.J. Fogarty, P.J. Rago, and D.G. Mountain. 2005. Effort distribution and catch patterns adjacent to temperate MPAs. ICES Journal of Marine Science, 62: 1150–1167.

NMFS (National Marine Fisheries Services). 2004. Evaluation of bycatch: a national approach to standardized bycatch monitoring programs. US 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. 1991. Optimum size of sampling unit for estimating the density of marine populations. Biometrics 47:717–723.

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.

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.

Sampson, D. 2002. Final report to the Oregon Trawl Commission on analysis of data from the at-sea data collection project. Oregon State University. Newport, OR.

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

Walsh, W.A., P. Kleiber, and M. McCracken. 2002. Comparison of logbook reports of incidental blue shark catch rates of Hawaii-based longline vessels to fishery observer data by application of a generalized additive model. Fisheries Research 58:79–94.

Annex 1: List of participants

NAME	ADDRESS	PHONE/FAX	EMAIL
Andrea Belgrano	Swedish Board of Fisheries, Institute of Marine Research PO Box 4 Lysekil SE-45321 Sweden	Phone +46 523 18722 Fax + 46 523 13977	andrea.belgrano@fiskeriverket.se
Mark Dimech	Malta Centre for Fisheries Sciences Fort San Lucjan BBG 06 Marsaxlokk Malta	Phone 21650933	mark.dimech@gov.mt
Inês Farias	IPIMAR Avenida de Brasilia PT-1449-006 Lisbon Portugal	Phone +351 213027108 Fax +351 213015948	ifarias@ipimar.pt
Åge Fotland	Institute of Marine Research PO Box 1870 N-5817 Bergen Norway	Phone +47 55 238682 Fax +47 55 238687	Aage.Fotland@imr.no
Hans Gerritsen By correspondence	The Marine Institute Rinville Oranmore Co. Galway Ireland	Phone 353 91 387297 Fax 353 91 387201	hans.gerritsen@marine.ie
Francesca Gravino	Malta Centre for Fisheries Sciences Fort San Lucjan BBG 06 Marsaxlokk Malta	Phone +356 Fax +356	Francesca.gravino@gmail.com
Mathieu Merzereaud	IFREMER Port-en-Bessin Station PO Box 32 F-14520 Port-en-Bessin France	Phone +33 231 515646 Fax +33 231 515601	Mathieu.Merzereaud@ifremer.fr
Kelle Moreau	Institute for Agricultural and Fisheries Research Ankerstraat 1 B-8400 Oostende Belgium	Phone +32 59 569830 Fax +32 59 330629	kelle.moreau@ilvo.vlaanderen.be
Kjell Nedreaas	Institute of Marine Research PO Box 1870 N-5817 Bergen Norway	Phone +47 55 238671 Fax +47 55 235 393	kjell.nedreaas@imr.no

NAME	ADDRESS	PHONE/FAX	EMAIL
Michael Pennington Chair	Institute of Marine Research PO Box 1870 N-5817 Bergen Norway	Phone +47 55238601	michael.pennington@imr.no
Alastair Pout	Fisheries Research Services PO Box 101 AB11 9DB Aberdeen UK	Phone +44 1224 876544 Fax +44 1224 295533	A.Pout@marlab.ac.uk
Sofie Vandemaele	Institute for Agricultural and Fisheries Research Ankerstraat 1 B-8400 Oostende Belgium	Phone +32 59 569883 Fax +32 59 330629	sofie.vandemaele@ilvo.vlaanderen.be
Joël Vigneau	IFREMER Port-en-Bessin Station PO Box 32 F-14520 Port-en-Bessin France	Phone +33 231 515 600 Fax +33 231 515 601	joel.vigneau@ifremer.fr
Jon Helge Vølstad Chair	Institute of Marine Research PO Box 1870 N-5817 Bergen Norway	Phone +47 55238411 Fax +47 55235393	jon.helge.voelstad@imr.no

Annex 2: Agenda

27 October	1000–1100	Welcome
		Presentation of participants and finalizing the agenda. Setting up computers. Appointment of reporters. List of presentations. Coffee break.
	1100–1300	Presentation of fishery data collection examples
	1300–1400	Lunch
	1400–1600	Presentation of fishery data collection examples (continued)
	1600–1630	Coffee break. Some people will attend the EU Lot8 meeting.
	1630–1800	ToR summarizing and reporting
	1800	Adjourn
28 October	0900–1100	ToR a) presentations
	1100–1215	Lunch
	1215–1500	ToR a) presentations (continued)
	1500–1515	Coffee break
	1515–1700	Discussion, summarizing and reporting
	1700	Adjourn
29 October	0900–1030	ToR b) presentations
	1030–1045	Coffee break
	1045–1300	ToR b) presentations (continued)
	1300–1400	Lunch
	1400– 1600	ToR b) presentations (continued)
	1600–1615	Coffee break
	1615–1715	Summarizing and reporting
30 October	0900–1100	Tor a) discussion
	1100–1115	Coffee break
	1115–1300	ToR b) discussion
	1300–1400	Lunch
	1400–1500	Final summarizing and reporting
	1500–1600	Recommendations, future work and closing of the workshop
	1600	Adjourn

Annex 3. List of working documents and oral presentations

All working documents and oral presentations are available from the author(s)/presenter or from the Co-chairs.

Working documents

- WD 1 - Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment. Joël Vigneau(1), Mathieu Merzereaud(1), Alastair Pout(2) *(1) Laboratoire HMMN/RHPEB IFREMER, Port-en-Bessin, France (2) FRS laboratory, Aberdeen, Scotland.*
- WD 2 - Examples of methods for investigating the accuracy of fisheries data. Hans Gerritsen and Colm Lordan. *Marine Institute, Rinville Oranmore, Co Galway, Ireland.*

Oral Presentations

Inês Farias: Black scabbard fish – The fishery in the Portuguese continental coast.

Kristin Helle: Does the reference represent the entire fleet? IMR, Bergen, Norway.

Åge Fotland: The Norwegian NeA saithe fishery in 2007.

Gier Bloom: Analysis of samples of conversion factors – saithe and cod. Directory of Fisheries, Bergen, Norway.

Joël Vigneau: The EU COST program.

Kjell Nedreaas: Bias In quantities landed – Methods to quantify unreported landing – example from the Barents Sea.

Kjell Nedreaas: Dynamics of the reference fleet.

Michael Pennington: Cluster sampling, effective sample size and bias.

Jon Helge Vølstad: Review of an appeal in the Norwegian court system where the representativeness of catch sampling from a fishery was challenged by foreign vessel owners.

Mark Dimech: An attempt to assess the accuracy of the Maltese Data Collection System.

Annex 4. Application of the score-card to detect sources of bias in fisheries statistics for Northeast Arctic Saithe

			NO BIAS	RISK OF BIAS	CONFIRMED BIAS
				O	R
Species identification		SP confusion trained staff	Trained and experienced OR Easily defined SP	all other situation	difficult to identify * Novice
		misreporting	Checked and no pb OR checked + corrected	all other situation	Checked + pb + not corrected
		taxon change	No OR Yes and taken into account	Yes and not taken into account	
		groupings	No groupings OR groupings and estimated	all other situation	Groupings and not estimated
		identification key	Yes OR No for SP confusion = GREEN	all other situation	
		Overall rating	G	O	R
			all green	names of indicators orange	1 red (name of the red)
Landings weight		species ident overall	RECALL		
		Missing part	Checked and Ratio = 1 OR checked + corrected	all other situation	Confirmed missing but not corrected
		area misreporting	Checked and no pb OR checked + corrected	any other situation	Checked + pb + not corrected
		Quantity misreporting	Checked and no pb OR checked + corrected	any other situation	Checked + pb + not corrected
		Population of vessels	All covered		Partially covered
		Sources of info ???	Several sources considered	Only one source	
		Conversion factor	Whole	any other situation	CF Wrong OR Not whole and CF not used
		% L in mixed	None OR Checked + corrected	any other situation	Checked + pb + not corrected
		Damaged fish	No partial fish	any other situation	pb + not corrected
		Overall rating	G	O	R
			all green	names of indicators orange	1 red (name of the red)
Effort		Species ident overall	(for metier allocation algorithm)	RECALL	
	Time unit	Time unit definition	Definition available	No definition	pb + not corrected
	Composite Unit	Standard format/procedure resolution demanded	fully + corrected/estimated	any other situation	incomplete
		Area misreporting	Checked + no pb/corrected	any other situation	checked + not corrected
		Source of information	Several sources considered	Only one source	
		Effort misreporting	Checked and no pb OR checked + corrected	any other situation	Checked + pb + not corrected
		Overall rating	G	O	R
Discards weight		Land/effort overall		RECALL	
		Sampling allocation scheme	Well designed random sampling	Ad hoc sampling	No sampling
		Bias on raising variable	No raising factor needed/follow accepted raising procedure	any other situation	No raising factor when needed
		Size of catch effect ??	Well designed random sampling	any other situation	Checked + pb + not corrected
		Damaged fish discarded	No partial fish	any other situation	pb + not corrected
		Non response rate	high respons rate/low refusal rate (figure needed)	any other situation	low response rate/high refusal rate (iugure needed)
		Temporal coverage	documented and OK	any other situation	documented and not OK
		Spatial coverage	documented and OK	any other situation	documented and not OK
		High-grading	no HG/HG + corrected	any other situation	HG without or out of SP
		Slipping behaviour	no SB/SB + corrected	rare slipping	significant slipping rate
		Management measures leading to discarding	management not leading to impact/impact corrected	any other situation	strong management leading to discarding and limited at sea sampling
		Catch-volume effect ??			
		Working condition	good conditions/conditions not ideal + compensated for	any other situation	hard conditions + not compensated for
		Species replacement	no occurence/corrected	any other situation	occurence + not corrected
		Overall rating	all green	1 orange	1 red

Comments:

sigbjorn: catches north of 62N are sometimes reported south of 62N and allocated to North Sea

sigbjorn: instances of underreporting of small/medium sized saithe and payment for large and better prized saithe

sigbjorn: only info from logbooks from above medium size trawlers

sigbjorn: double/trippel trawl not always reported, but small problem so far

aage: Discard not allowed and therefore no sampling scheme implemented

sigbjorn: Reports of high-grading from the international trawler fleet

sigbjorn: reports of slipping of undersized fish or large catches in the purse seine fishery

		G	O	R
Length structure	Landings overall/Discards overall		RECALL	
	Existence and follow-up of protocol	Yes and documented	any other situation	Non existing/ Existing but not followed
	Temporal coverage	documented and OK	any other situation	documented and not OK
	Spatial coverage	documented and OK	any other situation	documented and not OK
	Random sampling of trips/boxes	representative sampling	any other situation	known unrepresentative sampling
	Availability of all landings/discards	known complete availability	any other situation	known to be unavailable + uncorrected
	Non-sampled strata	all strata sampled/not all sampled but corrected by proper imputation technique	any other situation	not all strata sampled + uncorrected
	Bias on raising variable	No raising factor needed/follow accepted raising procedure	any other situation	No raising factor when needed
	Change in selectivity	Checked + no pb/corrected	any other situation	checked + pb + not corrected
	Sampled weight	described and controlable	any other situation	known inacurrate uncontrolable procedures
	Overall rating			

		G	O	R
Age structure	Length structure overall		RECALL	
	Quality insurance protocol	% double-reading and % agreement quantified to protocol standard for no bias	any other situation	checked + pb + not corrected
	Spatial coverage	documented and OK	any other situation	documented and not OK
	Time coverage	documented and OK	any other situation	documented and not OK
	Species/stock subject to reading issue	Trained and experienced OR Easily defined SP	all other situation	difficult to identify * Novice
	Trained staff			
	Agreed international reference set	Yes	No	
	Age reading methods	Method described and appropriate	any other situation	
	International exchange of otoliths	Recently assessed and made use of	any other situation	Recently assessed + pb + not corrected
	Calibration workshop	Not needed/Recently conducted	any other situation	Recently conducted + pb + not corrected
	Statistical processing for estimating age	Method described and appropriate	any other situation	Wrong ones used
	Plusgroup	No G+ needed/G+ defined according to established procedures	any other situation	Inappropriate G+ used
	Conventional actual age discrepancy	Checked and actual reading validated	any other situation	Checked and invalidated
	Completeness ALK	Complete and validated/Incomplete + corrected	any other situation	Incomplete + uncorrected
	Overall rating			

aage:
Not all Quarters are covered for all gear groups. Gear groups that fishes a small amount of the quota in some Quarters are omitted in the sampling programme.

aage:
Not all statistical areas are covered for all gear groups. Areas that represent a small amount of the quota are omitted in the sampling programme.

sigbjorn:
Method prosedures are about to be written

Mean weight	Age overall/length overall		RECALL	
	Spatial coverage	documented and OK	any other situation	documented and not OK
	Temporal coverage	documented and OK	any other situation	documented and not OK
	Use of a statistical means (source:L/W, VB Model, grouping LC for measuring weights)	Not needed/Method described and approved	any other situation	Method checked + pb + used anyway
	Calibration of equipment	Properly calibrated	any other situation	Known use of incalibrated equipment
	Working condition	good conditions/conditions not ideal + compensated for	any other situation	hard conditions + not compensated for
	Conversion factor	Appropriate use of CF AND detection of outliers	any other situation	CF Wrong OR Not whole and CF not used
	Overall rating			

igbjorn:
roblem with
verestimation of
veight at age in new
tatistical model

Sex-ratio	Age overall/length overall		RECALL	
	Spatial coverage	documented and OK	any other situation	documented and not OK
	Temporal coverage	documented and OK	any other situation	documented and not OK
	Source of information/Standard protocol	Not needed/Method described and approved	any other situation	Method checked + pb + used anyway
	Trained staff	Trained and experienced	all other situation	Novice
	Size/Maturity effect (imm)	Method described and approved	any other situation	No method/Method available but not used
	Catchability effect	No pb OR problem assessed + corrected	any other situation	problem assessed + not corrected
	Overall rating			

Maturity stages	Age overall/length overall		RECALL	
	Manual/Protocol	Available and correctly used (incl. WKMAT 2007, national/international protocols, …)	Not in compliance with protocols	Available and known mistakes not corrected
	Appropriate time period	Described and appropriate	any other situation	known inappropriate time
	Agreed intern. Ref. set	available + taken into account	any other situation	available + not taken into account
	Spatial coverage	documented and OK	any other situation	documented and not OK
	Calibration WK/Training Session	Recently conducted in accordance with	any other situation	Recently conducted + pb + not corrected
	Histological reference	Available and taken into account	any other situation	Available and known mistakes not corrected
	Skipped Spawning	Not an issue/Known to happen and dealt with (see WKMAT 2007)	any other situation	Known to happen and not dealt with
	Overall rating			

sigbjorn:
only based on IMR practice

aage:
As fishery has moved southwards in the recent years the maturity \| age can be affected by earlier maturation in the southern part of the stock.

aage:
No information of amount of skipped spawners are available for the assessment.

CPUE	bias on C * bias on E