ICES WKPICS3 REPORT 2013

ICES ADVISORY COMMITTEE

ICES CM2013/ACOM:54

REF: ACOM, PGCCDBS, RCMS, STECF-EWG ON DCF

Report of the third Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes (WKPICS3)

19-22 November 2013

ICES HQ, 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. 2014. Report of the third Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes, 19-22 November 2013, ICES HQ, Copenhagen, Denmark. ICES CM2013/ACOM:54. 109 pp.

Contents

Executive summary

This workshop, chaired by Jon Helge Vølstad (Norway) and Mike Armstrong (UK) was held at ICES headquarters, Copenhagen, from 19–22 November 2013. It was the third and final in a series of WKPICS workshops set up by the ICES Planning Group on Commercial Catches, Discards and Biological Sampling (PGCCDBS) to promote the implementation of statistically-sound designs for sampling commercial catches and help people to design and implement such schemes. The workshops focused on several classes of catch sampling schemes for estimating variables such as quantities discarded, and length or age composition of catches, taking account of the many practical problems that face people trying to obtain representative, randomized samples of catches. The Workshops have provided guidelines for good practice, and explored ways of documenting the quality of sampling designs and of the data that are collected in a way that is useful for different types of end-users.

The 2013 WKPICS3 meeting carried out the following work from its Terms of Reference:

- Some trial applications of Quality Assurance reports developed by WKPICS2 were reviewed, and plans were made to extend the trials.
- The sampling design and estimation procedures currently adopted within Europe for estimating age compositions and weight-length relationships for retained and discarded fish were reviewed based on 174 questionnaires returned from 20 countries plus the combined Baltic countries, covering 90 stocks. The potential for bias was evaluated, and advice was developed on subsampling for age and use of age length keys within design-based sampling schemes.
- Guidelines on estimation procedures for all four principal classes of catch sampling schemes, including using auxiliary data for re-weighting, were finalized
- Advice was developed for the Regional Databases concerning procedures for combining national fishery sampling data or estimates to give regional or supra-regional estimates for fisheries or stocks.
- A report for the European Commission reviewing data quality indicators used in non-EU countries and providing advice for appropriate data quality indicators to be included in the DC-MAP was reviewed and views of WKPICS were included.
- Advice and guidelines were developed in response to a recommendation from RCM-NSEA to (a) provide detailed guidance on diagnostic methods to evaluate aspects of data quality to facilitate the work of Regional Coordination Groups in coordinating regional data collection and analysis, and provide any additional Terms of Reference for the proposed WGCATCH and WGBIOP to continue this development during the transition phase of DC-MAP; and (b) to provide advice to the Steering Committee for the Regional Data Bases on development requirements for the RDB related to data quality assurance and reporting.
- The conclusions from the WKPICS series of workshops were summarized, and the next steps to providing a reference book on the design and analysis of statistical catch sampling programmes were discussed.
- The WK considered the setup of a live document (web based) to link documents and further developments in procedures etc. and concluded that this would be a task for WGCATCH.

1 Introduction

1.1 Terms of reference

WKPICS3 is the third of three workshops aimed at providing guidance on the design of fishery sampling programmes. The Terms of Reference for WKPICS3 are given below, and the background and structure of the meeting is given in Section 1.3. .

The third **Workshop on practical implementation of statistical sound catch sampling programmes (WKPICS3)**, chaired by Jon Helge Vølstad, Norway, and Mike Armstrong, UK, will meet in ICES HQ, Copenhagen, in 19 – 22 November 2013, to:

a) Evaluate the trial application of Quality Assurance reports developed by WKPICS2.

b) Review sampling design and estimation procedures currently adopted within Europe for estimating age compositions and weight-length (W-L) relationships for retained and discarded fish, evaluate potential for bias, and develop Quality Indicators related to this in QA reports.

c) Finalize guidelines on estimation procedures for all four principal classes of catch sampling schemes including using auxiliary data for re-weighting. Based on case studies, provide guidance on best practice on the estimation of discards to satisfy data calls, comparing design-based procedures and post stratification procedures.

d) Finalize recommendations for the Regional Databases concerning procedures for combining national fishery sampling data or estimates to give regional or supra-regional estimates for fisheries or stocks.

e) Summarize conclusions from the WKPICS series of workshops and consider the next steps to providing a reference book on the design and analysis of statistical catch sampling programmes. Consider the setup of a live document (web based) to link documents and further developments in procedures etc.

f) Compile a report for European Commission reviewing data quality indicators used in non-EU countries and providing advice for appropriate data quality indicators to be included in the DC-MAP.

g) Address recommendation from RCM-NSEA to (a) provide detailed guidance on diagnostic methods to evaluate aspects of data quality to facilitate the work of Regional Coordination Groups in coordinating regional data collection and analysis, and provide any additional Terms of Reference for the proposed WGCATCH and WGBIOP to continue this development during the transition phase of DC-MAP; ii) provide advice to SC-RDB on development requirements for the RDB related to data quality assurance and reporting.

The additional ToRs (f) and (g) were included following the Regional Coordination Meetings in 2013. The recommendation from RCM NS&EA to WKPICS3 is reproduced below.

Recommendation from RCM NS & EA to WKPICS3:

Specifying data quality diagnostics for fleet-based and stock-based biological data	
RCM NS & EA 2013 Recommendation 1	RCM recommends that WKPICS3 provides detailed guidance on diagnostic methods to evaluate aspects of data quality to facilitate the work of Regional Coordination Groups in coordinating regional data collection and analysis, and provide any additional Terms of Reference for the proposed WGCATCH and WGBIOP to continue this development during the transition phase of DC-MAP. In addition recommends that WKPICS3 provides advice to SC-RDB on development requirements for the RDB related to data quality assurance and reporting.
Justification	A suite of diagnostic tools will be needed by RCGs to evaluate and respond to regional data quality issues. These include but are not limited to • errors in RDB related to quality assurance and control at national level and errors during RDB data uploading • quality of fleet-based biological data in terms of coverage and numbers of samples for length and age by stock, fleet and area as needed for coordinating national data collection activities, • quality of stock-based biological data such as for estimating growth parameters, maturity ogives and sex ratios in terms of data sources, coverage of the and numbers stock of samples
Follow-up actions needed	ICES to add Term of Reference to WKPICS3
Responsible persons for follow-up actions	ICES WKPICS3
Time frame (Deadline)	November 2013 WKPICS3 meeting.

1.2 WKPICS3 participants and meeting agenda

The list of participants and the adopted agenda are in Annexes 1 and 2, respectively. All the working documents, presentations and national sampling scheme reports are located on the meeting SharePoint site.

1.3 Background to WKPICS3 and overview of report content

The data collected from fisheries have a primary function of supporting stock assessments and informing fleet-based management decisions. The data also support evaluation of ecosystem impacts of fishing. The ability to give sound fishery management advice depends critically on the accuracy (precision and bias) (Jessen 1978) of the evidence base that underpins it, and ICES has established a range of expert groups whose primary role is to coordinate and promote the collection of high-quality databased on sound scientific and statistical procedures.

During 2002, the ICES Planning Group on Commercial Catches, Discards and Biological Sampling (PGCCDBS) was established to provide support for the EU Data Collection Framework (DCF), focusing specifically on quality assurance of fishery sampling data and biological parameter estimates. Its role has been to promote best practice so that datasets and parameters supporting assessments and advice for the ICES area are based on i) statistically-sound sampling schemes; ii) correct and consistent interpretation of biological material such as otoliths and gonads; iii) technology that improves accuracy and cost-effectiveness of data collection; iv) comprehensive and easily sourced documentation, and v) efficient collaboration between PGCCDBS, expert groups and other bodies in relation to data collection. To fulfil this role, PGCCDBS has established Workshops and Study Groups to bring experts together to address specific issues. On the topic of fishery sampling schemes, these have included the Workshop on Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment (WKACCU: ICES, 2008); theReport of the third Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes; the Workshop on methods for merging métiers for fishery-based sampling (WKMERGE: ICES 2010a); the Workshop on Practical Implementation of Statistical Sound Catch Sampling Programs (ICES 2011a, 2012a, this report); the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS – ICES 2011b, 2012b, 2013); and the Workshop on Sampling Methods for Recreational Fisheries (WKSMRF: ICES 2009b) which progressed in subsequent years to the Planning Group on Recreational Fisheries Surveys (PGRFS: ICES 2010b, 2011) and then to the Working Group on Recreational Fisheries (WGRFS: ICES 2012b, 2013). There are many common threads running through all these expert groups, which together provide a comprehensive and invaluable resource on fishery sampling design, implementation and data analysis.

The overall aim for a design-based sampling strategy (e.g. Særndal 1978; Gregoire 1998) is to:

1. Collect data in a way that accuracy can be reliably assessed at national and regional level
2. Ensure that sampling intensity is allocated in a way that would minimize bias and maximize precision at the level where it matters most in the context of assessment of stocks and fisheries

The EU Data Collection Framework has required Member States to carry out fishery sampling schemes to estimate catches (mainly discards, but in some cases total catches where there are no census data on landings, such as some small-scale fisheries), and the length and age compositions of those catches. The WKPICS series has had a particular focus on helping countries design sampling schemes to meet these requirements, whilst at the same time providing recommendations on how the DCF should be structured to ensure that its requirements are based on sound statistical principles. The European Commission considers that work of WKPICS is relevant to the development of the new EU multi-annual programme of data collection (referred to at time of WKPICS3 as the DC-MAP although later called the EU MAP) in terms of providing definitions/ best-practice for implementing statistically sound catch sampling.

This year, DG Mare wanted WKPICS3 to give input to a working document to be submitted by an independent expert to the STECF Expert Working Group 13-18 (DCMAP 3) to provide:

1. Overview and compilation of discussions that have already taken place on quality indicators/measures of quality for DCF biological data (including in the context of ICES, STECF and other appropriate EU fora).
2. Overview of approaches used in other important fishing nations (e.g. USA, Canada, New Zealand, Australia, Norway...) to measure and ensure quality of biological data.
3. A reflection on whether Member States following best practice guidance for data collection is sufficient or whether measures of quality of the collected data are necessary in addition. Possible measures of quality or quality indicators that could be used and pros/cons of these.

This report is given in Annex 3.

Section 2 of WKPICS3 addresses ToRs (a), (b), (f) and (g) which all deal with aspects of evaluating and reporting on the quality of sampling schemes and outcomes:

- ToR (a) was set up to evaluate the trial application of Quality Assurance reports developed by WKPICS2. The design of the reports was considered again by PGCCDBS in 2013 and a plan was drawn up for trial applications. However, it was decided that the reports were not quite ready for such a trial, so the plan was abandoned. WKPICS3 wishes for the reports to be tested, and some proposals are given in section 2.1.
- WKPICS has addressed ToR (b) on sampling design and estimation procedures for estimating age compositions and weight-length (W-L) relationships and develop Quality Indicators related to this in QA reports by first evaluating the results of a questionnaire on ALK and WLR practices circulated to national stock coordinators within the ICES community and then considering biases associated with different practices.
- A draft report compiled for European Commission, reviewing data quality indicators used in non-EU countries and providing advice for appropriate data quality indicators to be included in the DC-MAP, was reviewed by WKPICS3 to address its ToR (f).
- WKPICS3 addressed the recommendation from RCM-NSEA to (a) provide detailed guidance on diagnostic methods to evaluate aspects of data quality to facilitate the work of Regional Coordination Groups in coordinating regional data collection and analysis, and provide any additional Terms of Reference for the proposed WGCATCH and WGBIOP to continue this development during the transition phase of DC-MAP; and ii) provide advice to SC-RDB on development requirements for the RDB related to data quality assurance and reporting.

Section 3 of WKPICS3 addresses its ToR (b) to finalize guidelines on estimation procedures for all four principal classes of catch sampling schemes identified by WKPICS2, including using auxiliary data for re-weighting and ToR (c) to finalize recommendations for the Regional Databases concerning procedures for combining national fishery sampling data or estimates to give regional or supra-regional estimates for fisheries or stocks. Some of the material for ToR (b) included in the WKPICS2 report has been revised and extended by WKPICS3, so a full report encompassing the modified guidelines and case studies from last year's report plus the new material from this year's meeting is given by WKPICS3.

As in previous years, WKPICS3 operated through a mixture of plenary sessions and breakout groups.

2 Quality indicators for fishery sampling

2.1 Evaluate the trial application of Quality Assurance reports developed by WKPICS2 (ToR a).

2.1.1 Design of QA reports for fishery sampling

The design of Quality Assurance reports for fishery sampling was considered again by PGCCDBS in 2013 (ICES 2013 a) and a plan was drawn up for trial applications. However, it was decided that the reports were not quite ready for such a trial, so the plan was abandoned.

In order to move the development of the reports along, WKPICS3 carried out a brief review of the existing versions of the reports taking into account inputs from previous PGCCDBS, SGPIDS and WGRFS meetings. A limited trial of the PGCCDBS 2013 version was carried out for Baltic Cod (see section 2.1.2), but WKPICS was not in a position to carry out a more extensive trial in time for PGCCDBS 2014.

The WKPICS3 review of QA report drafts showed one important consideration that can be improved. WKPICS2 had maintained that the QA reports are designed primarily for stock assessment, and focus on a given stock, and proposed a report structure giving simple metrics of the sampling design, implementation and sampling successes for each national sampling stratum. Nevertheless, as reflected by WGRFS 2013 (ICES, 2013b), specific ways of reporting must be considered for different end-users who have different requirements for detail.

In the case of QA reports produced for stock assessment groups there are two aspects of data quality of interest to them.

- First – information to indicate if national sampling schemes follow good practice in terms of the three key components of: statistical design, method of selecting sampling units, and estimators and method of analysis. This provides information on whether the data, in principle, are likely to be representative of the population, or if there may be a bias related to these design aspects.
- Second – no matter how good or bad a sampling design might be, sampling may not go according to plan and there may be quality issues related to implementation, for example high rates of refusal to take observers on board vessels, inadequate coverage of strata due to staffing, or other issues. These factors may lead to bias, or to poor precision due to small numbers of PSUs sampled. For stock assessment, diagnostics are needed to show where and how these problems affect subsets of data (e.g. national data used in regional assessments, or data for strata within countries) and the likely impact on overall quality of the international weighted (raised) estimates of catches and numbers at length or age for particular stocks. Tools that provide easy access to this level of detail can greatly help the task of locating the cause of unusual features in aggregated data that affect the fit of an assessment model, whilst information on biases and precision can be used directly in statistical assessment models.

At a higher level, for example the European Commission, the QA report dealing with sampling design may be the most important one for determining if good practice is being followed across a region.

WKPICS agrees with WGRFS 2013 that there is no single way to document data quality (e.g. accuracy) that is suitable for all end-users, and that a "toolkit" of reporting systems is needed that will provide different end-users with the type of information that they require (see Figure 2, Annex 3). Within Europe, the end-users and their requirements for commercial fishery sampling could include:

1. National laboratories (for documenting and monitoring national schemes)
2. Regional Coordination Groups (overviews of sampling schemes extant within the region; identification of important gaps in data; developing recommendations for optimizing sampling across countries)
3. European Commission (evaluation if Member States are meeting DCF / DC-MAP requirements for delivery of data using statistically sound methods)
4. Stock assessment expert groups (data quality in terms of precision and bias of estimates being used for assessments).

Taking this into account, the work carried out by WKPICS on its ToR on data quality indicators must be redirected to consider QA reports with two well-differentiated structures:

1. General information on sampling design and implementation.
2. Specific stock-based metrics.

The first one should address information considered useful for an assessment group to identify if any of the datasets that are influential in the assessment (perhaps a large component of the total catch, or providing data for an important fleet for which selectivity is being estimated), are derived from sampling schemes that might not yield representative data due to biased design. This is an exercise best conducted at a benchmark assessment data compilation and evaluation meeting. To have maximum impact in such a forum, the QA report on sampling design should be well focused and avoid too much detail on each national sampling design. A problem may be that a benchmark assessment meeting may sometimes have no contributors with sufficient experience of statistical sampling design to detect underlying problems. There is therefore a need for independent peer-review of national sampling schemes by experts in this field (see Annex 3). It is possible that the new WGCATCH could assume the work of reviewing and assessing on the correct sampling design and implementation procedures, or there may be a role for a separate independent review procedure. This review process should provide input to concise QA reports on sampling designs within regional seas to inform the Regional Coordination Groups, the Commission and assessment expert groups.

The second QA report gives metrics of the outcomes of the national sampling schemes for each stock being assessed. This would include the contribution of catches in each stratum to the overall catches of the stock, information related to bias (e.g. refused access to vessels or catches; evidence of non-random selection of sampling units; failures of sampling coverage), number of primary sampling units achieved in each stratum, precision of estimates etc. This information should be regarded carefully in relation to the design to determine if a failure exists in the final sampling data provided. WGRFS (ICES 2013 b) suggested the development of an archive (database or other structure) of such information allowing compilation of a history of QA information. Statistics such as numbers of PSUs sampled per stratum, precision estimates etc. have their greatest value as time-series covering the period of an assessment dataset, and should be easily extracted an tabulated from the annual QA reports along with supporting information on changes in sampling design over time.

As already stated by PGCCDBS (ICES 2013a), end-users (e.g. stock assessment or mixed fishery analysis groups) have to clearly understand that QA reports will give an overview of the general sampling situation, but information per stratum must be understood with respect to national sampling programmes, the described contributions of each country to the total catches, and the domains of interest the QA reports address.

2.1.2 Evaluation of Baltic cod QA reports

The evaluation of a trial application of quality assurance reports developed by WKPICS2 (ToR (a)) had been initiated for blue whiting but the trial was not finalized. During WKPICS3 another trial was initiated for at-sea sampling of the Western Baltic cod stock considering national data from 2012.

This exercise showed that seven countries contributed to the landings of cod2224 in 2012 (see Table 2.1.2.1 for the trial indicator table), of which four regularly provide biological data for the analytical stock assessment (SAM). These four countries (Denmark, Germany, Sweden, and Poland) filled out the required fields using national databases. The effort required to complete the table was minor since the countries already have national documentation in place from which the required data could be extracted.

The table gives a very nice overview of the contributions of national landings and fleet strata to total landing and discards of cod 2224 and, particularly, of how the sampling is balanced across the fleet strata, and, where there are potential issues, how big a contribution that fleet has to the total. This table is extremely useful for the stock coordinator to understand and handle national contributions to InterCatch (e.g. catch sampling at sea by Germany, only discard sampling at sea by DK, SWE, and POL; harbour sampling for landings by DK and SWE; the need to extrapolate discards of the Danish passive gear fleet from samples of GER and/or SWE) and will also be good for the Assessment Working group and the Regional Coordination group (RCG). Such a table may supplement stock assessment reports and could aid in RCG meetings on data quality issues because potential issues in sampling are instantly apparent by parameters given in the table, such as (i) the number of trips sampled and (ii) the number of unique vessels sampled, by country and fleet. A great advantage is that the countries are forced to report their sampling efforts in a standardized and meaningful way.

When a series of such reports across years is available, the next consideration will be summary reports that pick out the key quality indicators that can be examined at benchmark meetings in particular, to help stock assessors see if data quality is changing over time. In fact, the ongoing benchmark workshops for dab2232 and flounder2232 (WKBALFLAT) have already decided to use the quality indicator tables for the evaluation of data quality issues.

This example shows where we could start to make some adjustments to make it work even better. A few modifications of the quality indicator table were already carried out during WKPICS3. Two new rows were added for “Mean discard rate of the fleet in the year” and “Number of port samples”. The former value has to be calculated nationally anyway to produce the “contribution to national discards in fleet” and can therefore readily be provided. However, the calculation still needs to be standardized, i.e. whether a discard rate is calculated as [discard/(discard+landings)] or otherwise. The latter value provides information on the quantity of samples that cover the landings part in case of countries that do both harbour (for landings) and at-sea sampling (for discards).

The exact calculation of the “Age key quality indicator” still has to be specified to ensure that a useful/meaningful estimate is provided; at present, it is not clear what exactly this estimator intends to quantify. Alternatively, graphs could be generated showing spatio-temporal patterns of where aged fish came from (i.e. number of aged fish by fleet, quarter and ICES areas).

Theoretically, in the future the RDB could automatically provide for most of the estimates, which at present are provided by exploring national data tables. For example, information of the “Total number of vessels in the fleet” is missing in FishFrame at present. Since this information has to be calculated on the actual level of aggregation required by the QA table (in the cod2224 case, for active and passive gear), it cannot be simply added to the CE table in its current format.

Similar trial applications are needed for at-sea and harbour sampling of other stocks. The group suggested that this could be done prior to the next PGCCDBS 2014 but this was not possible and PGCCDBS should identify trial stocks to be reviewed by WGCATCH in November 2014.

Table 2.1.2.1 Trial quality indicator table of Western Baltic cod (cod2224) for the sampling year 2012.

	AT-SEA-SAMPLING											
	Stock - Species - Area - Year (Cod 2224 2012)	**Total landings 2012: 16756 t (source: FishFrame, RCMBaltic 2013)**										
		Denmark		Germany		Sweden		Poland		Finland	Latvia	Estonia
		Design		Design		Design		Design				
		Implemention		Implemention		Implemention		Implemention				
	Importance: Contribution to stock landing	53%		27%		14%		5%		2%	<1	<1
	Sampling / design effect/diagnostic for randomness... (Description according to best practice)											
	Sampling design	probability based discard sampling		probability based catch sampling		Probability based discard sampling		probability based discard sampling				
	Primary sampling unit	Vessel* trips		Vessel		Trip		Vessel				
	Sampling frame	quartely vessel list		annual vessel list		Quarterly vessel list		annual vessel list				
	Periodicity	effort is following the fishery		1-2 samples/week during fishing seasons		difficult to quantify		--				
	Contact protocol	yes		Yes		Yes		Yes				
	Sampling manual available	yes (Danish)		under preparation		No		under preparation				
	...											
	Strata from the sampling frame	Fleet 1	Fleet 2	Fleet 1	Fleet 2	Fleet 1	Fleet 2	Fleet 1	Fleet 2	Fleet 1	Fleet 2	
		active gear (Trawler)	passive gear	active gear (Trawler)	passive gear	active gear (Trawler)	passive gear	active gear	passive	active gear	passive	
	Importance: Contribution to national landing	70%	30%	67%	33%	50%	50%	47%	53%	100%	100%	
	Mean discard rate of the fleet in the year	9%	assumed low	10%	4%	14%	2%	5%	1%			
	Importance: Contribution to national discards in fleet	100%	0%	84%	16%	93%	7%	71%	29%			
	Quality indicator											
1	Total number of vessels in the fleet*	151	199	58	887	40	101	44	69	1	3	
	Number of trips sampled onboard of vessels	34	0	28	32	4	40	1	2			
	Number of unique vessels sampled	15	0	15	17	4	19	1	1			
	Total number of trips conducted by the fleet	4686	11519	3891	22156	247	4043	275	565			
	Number of trips sampled where stock occurred in the discards	34	0	28	32	4	33	0	2			
	Number of trips sampled where stock occurred in the landings			28	32	4	39	1	2			
	Number of port samples	40	40					0	68			
	Age key quality indicator (e.g. Mean number of age samples per trip sampled from this fleet)	75	75	207	63	76	14	0%	0%			
2	Non-response rate	68%		45%	53%	75%	50%	0%	0%			
	Industry decline (refusal rate)	27%		9%	3%	23%	not determined	0%	0%			
3	Goodness of fit											
	Bias 1: Spatio-temporal coverage	tested and considered all right		tested and considered all right		Few trips achieved		sampling ICES rectangle 37G4	sampling ICES rectangle 37G4			
	Bias 2: Vessel selection	6% are having a to small vessel for observers to participate		smaller passive gear vessels rejected observers		High refusal rate		no problem	no problem			
	Bias 3: ...											
4	Precision levels of e.g. parameter a, b, ...											
	e.g. CV, variance, relative sampling error											
	e.g. Input data for XSA model:											
	maturity at age											
	stock weight											
	catch weight											
	catch at age											

2.2 Sampling design and estimation procedures for estimating age compositions and weight-length (W-L) relationships, and Quality Indicators related to this in QA reports (ToR b)

2.2.1 Summary of PGCCDBS ALK WLR questionnaire

Questionnaires were circulated to ICES stock coordinators who were requested to forward them to national stock coordinators. The individual questions and multiple choice answers are given in the following text describing the returns.

A total of 174 questionnaires were returned from 20 countries plus the combined Baltic countries, covering 90 stocks. For most stocks only one country responded (49 stocks) but 22 stocks had responses from 3 or more countries. The aggregated results as well as the individual responses are available in the working documents folder of the WKPICS3 sharepoint site.

Number of responses per country

Baltic Sea Countries (BAL)	1	Germany (DEU)	18	Poland (POL)	8
Belgium (BEL)	3	Greenland (GRL)	1	Portugal (PRT)	11
Denmark (DNK)	37	Ireland (IRE)	21	Russia (RUS)	3
Estonia (EST)	2	Latvia (LVA)	1	Scotland (SCO)	21
Faroe Islands (FRO)	5	Lithuania (LTU)	4	Spain (ESP)	7
Finland (FIN)	4	Netherlnds (NLD)	5	Sweden (SWE)	9
France (FRA)	2	Norway (NOR)	7	Eng/Wales (UK EW)	4

The returns were somewhat unbalanced; some small countries like Denmark, Scotland and Ireland returned a large number of questionnaires for individual stocks, while other countries including France and the UK (EW) only returned a small number which in some cases covered a large number of stocks for which the same responses were considered appropriate. Because individual labs (or countries) are likely to use the same approach for a number of stocks, this can lead to some bias in the frequency distribution of approaches across stocks or countries. For this reason, the number of responses is given by country (as well as the total number of responses), but the results should be used primarily to indicate the range of methods adopted and not to quantify the use of methods across stocks or countries. For each question in the questionnaires, the following text summarizes the responses, provides some additional explanatory information, and then (if appropriate) gives WKPICS-3's comments.

Age–length Keys (ALKs)

1) Is this stock assessed using age based assessment models ?	Total
a) YES - age based models are used	131
b) NO - Some other criteria is used to assess this stock.	43
c) Don't know	5

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	1	3	36	1	5	4	2	11	0	14	1	0	5	7	4	5	3	10	4	9	6
b)	0	0	1	1	0	0	0	7	1	8	0	0	0	0	4	6	0	11	4	0	0
c)	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0

Most stocks in the returned questionnaires are assessed using age-based models. Most of the remaining stock assessments use length-based methods, use surveys to assess abundance directly, or are data-limited.

2) What age structures do you collect for this stock?	Total
a) Otoliths	143
b) Scales	1
c) Illicia	0
d) other	0
e) Don't know	0

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	3	36	1	5	4	2	18	0	14	1	5	5	7	4	5	3	10	5	9	6
b)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
e)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Nearly all of the age-based stocks use otoliths to age fish (scales are used in one stock). Illicia are also known to be widely used for anglerfish but these stocks are currently not assessed using age-based models.

3) How do you estimate age compositions of catches?	Total
a) Directly from random samples of fish that are aged, without reference to fish length.	44
b) Completely or partly from ALKs applied to raised length frequencies	118
c) Other	4
d) Don't know	0

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	0	18	0	0	0	2	11	0	0	0	5	0	1	0	0	1	0	0	5	1
b)	0	3	36	1	5	4	0	7	0	14	1	0	5	6	4	2	2	10	5	8	5
c)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
d)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Quite a few of the responses indicated that age compositions are estimated directly from random samples of the catch. Most of the remainder use ALKs. The Baltic countries use tag returns for the salmon stock Sal 24-31.

Properly implemented and weighted random sampling will give unbiased results. An ALK has the potential to introduce bias (see section 2.2.3 for more details)

4) Indicate the aggregation level (strata) for constructing and using ALKs.	Total
a) separately by sex	10
b) quarters or season	121
c) geographic sub-regions or fishing grounds within stock area	87
d) metiers or fleet segments	39
e) Separately for retained and discarded catch	63
f) for individual fishing trips, hauls or categories within trips	17
g) other	0
h) dont know	0

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	2	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	3
b)	0	2	36	1	5	4	2	18	0	10	1	5	5	6	4	2	1	0	5	8	6
c)	0	0	26	1	3	4	0	18	0	1	1	5	5	3	4	2	2	0	5	2	5
d)	0	2	0	0	2	0	0	7	0	7	1	5	0	6	3	0	2	0	0	2	2
e)	0	0	32	0	0	0	0	5	0	7	1	5	0	0	3	0	0	2	2	6	0
f)	0	0	0	0	0	0	0	2	0	0	0	5	0	0	0	0	0	10	0	0	0
g)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
h)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ALKs are generally constructed and applied separately by quarter or season and geographic region and often also separately by landings and discards, and in some cases also by sex, fleet segments or métiers and by fishing trips.

Arbitrary stratification or quota sampling is a practice that can lead to biased estimates. The stratification of the age sampling should follow the same (hierarchical) design of the length sampling.

5) When supplying age composition estimates for a stratum, such as a quarter or fleet metier (e.g. as entered on InterCatch), are they derived using ALKs which contain only your national length and age data from that stratum, or do you include age data from other strata or other countries?	Total
a) YES: only the length and age data from that stratum are used	76
b) MOSTLY: sometimes age data are borrowed from other strata	64
c) NO: samples from other strata, or from samples collected by a different country are commonly used	2
d) Don't know	1

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	0	22	1	0	0	2	10	0	6	1	5	5	3	4	1	1	10	2	2	1
b)	0	2	14	0	5	4	0	8	0	7	0	0	0	6	1	1	2	0	4	6	4
c)	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
d)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

About half of the responses indicate that age composition estimates for a certain stratum only contain national age data for that stratum, the other half indicated that this is mostly the case, only 2 responses indicated that data are regularly 'borrowed' from other strata or countries.

Borrowing ALK data from another stratum is an imputation method that should be used with care since it can cause bias in estimates of proportions-at-age and other parameters important to stock assessments. If the borrowing of ALK data are unavoidable, then try to validate that there are no major differences between strata e.g. for years where both strata did have sufficient data. Borrowing based on expert judgment without knowledge of the bias it may cause is considered a bad practice. Borrowing from different gears is common practice (in tools like COST this is done by default), however this can potentially also lead to bias due to different selectivity and spatial activity of gears (see section 2.2.3 for more details). Borrowing data from a different year is likely to cause strong bias in estimates of age and length compositions for target populations (stock or landings) and is an imputation method that should be avoided (unless recruitment, growth and mortality are all invariable from year-to-year, a highly unlikely occurrence).

6) In the specific case of fishery landings and discards, how are the ALKs derived and used? (indicate which combination of options applies).	Total
a) Discard ALKs are only based on samples of discards	51
b) Discard ALKs are partly or completely based on other sources	13
c) Landings ALKs are based on samples of retained fish at sea	24
d) Landings ALKs are based on on-shore samples of landings	34
e) Landings ALKs are based on sampling at sea AND on shore	33
f) Landings ALKs are partly/completely based on other sources	9
g) Don't know	3

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	0	19	0	0	0	0	13	0	0	1	0	0	0	3	0	0	7	2	6	0
b)	0	0	0	0	0	0	0	0	0	7	0	0	0	0	1	0	0	0	1	0	4
c)	0	0	0	0	0	0	0	16	0	0	0	0	0	3	0	0	3	0	0	2	0
d)	0	3	13	1	2	4	0	0	0	0	0	0	0	0	0	0	0	10	1	0	0
e)	0	0	0	0	3	0	0	2	0	14	1	5	0	3	4	1	0	0	0	0	0
f)	0	0	0	0	0	4	0	0	0	0	0	0	0	3	0	0	0	0	1	0	1
g)	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1

ALKs for discarded and retained fish are derived and used in a wide variety of ways. For the discarded catch, ALKs are generally constructed using only samples of discards but sometimes also using samples from other sources. For the retained catch, ALKs are generally constructed using only samples from the landings, either at sea on shore or a combination of both but in some cases ALK data are partly or completely derived from other sources.

Using other sources such as survey data could cause bias. (If the gear and the spatial and temporal coverage of the survey match the fishery, it may be appropriate to apply survey data to commercial catches). Even using landings data in a discard ALK could cause bias e.g. if a fish of a certain length as a different likelihood of being discarded based on a property that may correlate with age, such as condition. Arguably this is unlikely to be a major source of bias.

Sampling on shore results in some loss of spatial detail but is probably more cost-effective than sampling at sea.

7) Do age samples from fishery catches have an associated length frequency sample collected from the same catch?	Total
a) ALL age samples have associated LFDs	102
b) MOST BUT NOT ALL age samples have associated LFD data	20
c) NONE, OR ONLY SOME age samples have associated LFDs	11
d) Don't know	0

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	0	36	1	5	0	1	16	0	6	1	0	0	6	4	2	3	10	3	8	0
b)	0	3	0	0	0	4	0	2	0	0	0	5	0	0	0	0	0	0	0	0	6
c)	0	0	0	0	0	0	1	0	0	8	0	0	0	0	0	0	0	0	2	0	0
d)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Nearly all respondents stated that age samples have associated (concurrent) length samples from the same catch.

It is not good practice to have totally independent age and length sampling. This may happen when quota sampling takes place and in order to obtain age samples for e.g. small sole, the sampler obtains a (single!) box of small sole in order to fill the sampling targets for those size classes.

8) In cases where catches have an age sample as well as an LFD sample, are the age composition data from the individual age samples weighted using the numbers at length in the raised LFD for the same catches when constructing the ALK?	Total
a) YES - the age data are weighted by the raised LFD	51
b) NO - the ALKs are constructed without any weighting	80
c) Other approach used	0
d) Don't know	1

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	0	4	0	3	0	0	13	0	0	0	5	0	6	3	2	3	10	1	0	1
b)	0	3	32	1	2	4	1	5	0	14	1	0	0	0	1	0	0	0	4	8	4
c)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

In 51 cases LF data were used in weighted estimators, where the individual age data are weighted by the numbers-at-length in the length sample. However often this is not the case (80 responses). (Note that it is possible to give more than one answer to this question if different procedures are followed for discards and landings).

It is generally good practice to weight samples (see section 2.2.3 for more details).

Weight-Length Relationships (WLRs)

1) How are body weights of commercially-caught fish or shellfish derived for use in assessments?	Total
a) individual weights are recorded and used rather than a WLR	99
b) WLRs are used	88
c) combination of direct weighing and WLRs	12
d) dont know	5

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	1	0	24	2	5	4	0	19	1	6	1	5	6	1	8	3	3	0	0	8	2
b)	0	3	32	0	0	0	1	0	0	14	0	0	0	2	0	5	0	19	8	0	4
c)	0	0	0	0	0	0	0	0	0	7	0	0	0	1	0	4	0	0	0	0	0
d)	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	0	0	0

About half of the responses indicated that WLR are applied, most others indicated that individual weights are used while some use a combination of individual weights and WLR.

2) What is the source of WLR parameters applied to commercial fishery samples for this stock?	Total
a) sampling during research surveys	12
b) sampling of commercial catches	75
c) mixture of samples from surveys and fishery	10
d) parameters from another source or literature	34
e) dont know	1

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	0	0	0	0	0	0	8	0	0	0	0	0	3	0	1	0	0	0	0	0
b)	0	3	32	0	2	0	1	17	0	8	0	0	6	2	0	2	0	0	1	0	1
c)	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	0	5
d)	0	0	0	0	0	0	0	0	0	13	0	0	0	0	0	5	0	10	5	0	1
e)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

The WLR parameters are mainly sampled from commercial catches, but also regularly from the literature or from surveys or a mixture from surveys and the fishery.

Sampling from surveys to apply to commercial data might be bad practice. Fish condition may vary spatially and seasonally and there may be differences in fluid loss between freshly-caught fish on surveys and fish retained on fishing boats for a longer period. Using literature values is not good practice either, particularly if fish condition varies widely between years (e.g. between periods of high and low abundance or food supply); in most cases it should be cheap and easy to collect weight data from fish that are being aged.

3) For what types of estimates do you use a weight-length relationship (WLR) rather than direct estimates of weight.	Total
a) Factors for raising length frequency samples to known catch weights (e.g. within a zize category or total unsorted catch)	45
b) total discard catch weight for a trip, based on a raised LF	47
c) total retained catch from a sampled trip, based on a raised LF	22
d) mean weight-at-age of age composition derived from ALK	63
e) other	3
f) dont know	1

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	2	13	0	2	0	0	8	0	8	0	0	0	3	0	2	0	0	6	0	1
b)	0	0	19	0	0	0	0	13	0	7	0	0	0	0	0	0	0	0	4	0	4
c)	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	1	0	4
d)	0	2	32	0	0	0	1	8	0	8	0	0	0	3	1	1	0	0	3	0	4
e)	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
f)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

WLR are most commonly used to estimate the mean weight-at-age but also to estimate the total discard or retained catch from a trip and to raise length samples (with unknown sample weights) to known catch weights.

If WLR are used to raise samples with unknown weights to total known catch weights, be aware of potential sources of bias

4) Indicate the aggregation level (strata) for constructing and using WLRs.	Total
a) separately by sex	4
b) quarters or seasons	74
c) geographic sub-regions or fishing grounds within stock area	56
d) metiers or fleet segments	15
e) Separately for retained and discarded catch	44
f) for individual fishing trips, hauls or categories within trip	1
g) dont know	6

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
b)	0	2	32	0	2	0	1	17	0	4	0	0	0	6	1	1	0	0	2	0	6
c)	0	0	26	0	0	0	0	14	0	4	0	0	0	2	1	2	0	0	2	0	5
d)	0	2	0	0	2	0	0	5	0	0	0	0	0	5	0	0	0	0	0	0	1
e)	0	0	32	0	0	0	0	5	0	7	0	0	0	0	0	0	0	0	0	0	0
f)	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
g)	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0

The aggregation level for WLR is generally by quarter/season and also by geographic region. In quite a few cases the aggregation level was that of fishing trips, hauls or categories within trips. Some cases WLRs were constructed by métier or fleet segment or separately by sex.

WLR should be applied to the stratum they were sampled from.

5) How often are WLR parameters calculated, and over what periods are they applied?	Total
a) Annually, from samples collected that year	75
b) A block of years, based on data collected during that period	2
c) A block of years, data from a more limited sampling period	9
d) dont know	7

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	3	32	1	2	0	1	17	0	8	0	0	0	6	1	2	0	0	2	0	0
b)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
c)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	5
d)	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	1

WLR are generally applied annually in a few cases this was done over a longer period.

Using the same WLR over a long period can lead to bias if fish condition varies widely due for example to changes in abundance or food supply.

6) How are the individual length-weight data values derived for input to a WLR regression?	Total
a) Ad-hoc samples of individual fish lengths and weights	6
b) random sampling of individuals from catches	37
c) Length-stratified sampling of individuals from catches	43
d) mean lengths and mean weights from pooled samples of fish	0
e) other	0
f) dont know	7

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	4
b)	0	0	13	1	2	0	0	9	0	3	0	0	0	3	0	3	0	0	2	0	1
c)	0	3	19	0	0	0	0	8	0	5	0	0	0	3	1	1	0	0	3	0	0
d)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
e)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
f)	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	1

Samples for WLR are generally collected randomly or from a length-stratified sampling scheme. In a few cases samples were selected ad-hoc.

Quota sampling by size class could cause bias.

7) If sampling from a fishery, are retained fish weighed after gutting, freezing, filleting or other processing, requiring a conversion factor to whole (live) weight in addition to a WLR?	Total
a) No - all retained fish are sampled whole and fresh	38
b) Yes - some fish are landed whole, some are processed	15
c) Yes - generally all fish are landed processed.	31
d) don't know	1

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	0	13	1	0	0	0	17	0	0	0	0	0	0	1	3	0	0	3	0	0
b)	0	0	0	0	0	0	0	0	0	8	0	0	0	3	0	1	0	0	3	0	0
c)	0	3	19	0	2	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	5
d)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Some of the samples are from whole and fresh fish but generally all or some fish are processed (gutted etc). Generally, EU conversion factors are used (in Norway the Directorate of Fisheries supplies conversion factors).

To minimize any biases in raising numbers of fish in samples to the total reported landings given in the national landings statistics, it is important that the total sample weight from a catch is calculated using the same conversion factor that is applied to the vessel’s processed landings weight when compiling national landings statistics.

8) How are WLR data for retained fish collected from commercial catches?	Total
a) from landings on shore or observer samples brought ashore	51
b) weights are recorded at sea by observers	13
c) weights are recorded at sea by fishers during self-sampling.	4
d) don't know	5

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	3	13	1	2	0	1	13	0	8	0	0	0	3	1	2	0	0	3	0	1
b)	0	0	0	0	0	0	0	9	0	0	0	0	0	0	1	0	0	0	3	0	0
c)	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	1	0	0	0	0	0
d)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5

Weights of landings are most often measured from landings on shore or samples brought ashore by observers but sometimes weights are taken at sea.

9) How are WLR data for discarded fish collected from commercial catches?	Total
a) samples are brought ashore and then weighed	57
b) weights are recorded at sea by observers	18
c) weights are recorded at sea by fishers during self-sampling.	7
d) don't know	7

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	0	32	1	0	0	0	13	0	7	0	0	0	0	0	2	0	0	2	0	0
b)	0	0	0	0	0	0	0	8	0	0	0	0	6	0	1	0	0	0	3	0	0
c)	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	1	0	0	0	0	0
d)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	6

Weights of discards are usually calculated from samples brought ashore, but some are recorded at sea.

10) Are separate WLRs constructed for retained and discarded fish, or are the data pooled to provide a single WLR?	Total
a) Yes: separate WLRs constructed	52
b) No - samples from both sources are combined.	21
c) don't know	6

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	0	32	1	0	0	0	8	0	7	0	0	0	3	0	0	0	0	1	0	0
b)	0	0	0	0	0	0	0	8	0	0	0	0	6	0	1	2	0	0	3	0	1
c)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	5

WLRs are generally constructed separately for discards and landings but sometimes the data are combined.

11) Are any weighting factors used in calculating WLR parameters?	Total
a) Raw individual length-weight data are used, no weighting	81
b) a length frequency sample for the catch from which fish were selected for weighing is used to calculate weighting factors for individual length-weight values	1
c) another form of weighting is used.	0
d) don't know	11

	BAL	BEL	DNK	EST	FRO	FIN	FRA	DEU	GRL	IRE	LVA	LTU	NLD	NOR	POL	PRT	RUS	SCO	ESP	SWE	UK(EW)
a)	0	3	32	1	2	0	1	17	0	8	0	0	6	1	0	4	0	0	6	0	0
b)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
c)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d)	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	6

Only one reply indicated that WLR samples are weighted by numbers-at-length in the same sample; the data are mostly analyzed using unweighted estimators.

2.2.2 Case studies

The answers of the questionnaires were compiled for specific fish stocks to explore whether there is useful information to be used for assessing data quality issues. Four stocks were selected. In the case of the eastern Baltic cod (cod2532) where issues of fish weight and age readings are known, no severe concerns emerged from the overview of MS answers.

Table 2.2.2.1a, b. Overview of answers to the ALK and WLR questionnaire by country and questions for the stock cod2532; the questions and possible answers are given in section 2.2.1; the letters in the cells refer to the answers given.

Cod2532-ALK						
Question	Lithuania	Germany	Poland	Sweden	Denmark	
1	a	a	a	a	a	
2	a	a	a	a	a	
3	b	b	b	a,b	a,b	Q8 corresponds to Q3 (random otolith samples in SWE and DK)
4	a,b,c, d,e,f	a,b,c,d,e,f	b,c,d,e	b,e	b,e	
5	b	b	B	b	b	
6	e	a,c	e	a	a	
7	b	a	a	a	a	
8	a	a	a	b	b	
Comments:	Q6: GER only Catch sampling at sea	Q3: SWE directly for landings, from ALK for discards			Q3: 1 - direct sampling of age for landings, 2 - ALK for discard	

Cod2532 – WLR					
Question	Denmark	Germany	Poland	Lithuania	Sweden
1	c	c	a	a	a
2	b	b			
3	b,e	a,b,c,e			
4	b,e	a,b,c,d,e,f			
5	a	a			
6	c	c			
7	a	a			
8	a	a			
9	a	a			
10	a	b			
11	a	a			
Comments:	DK: WLR for discard only		POL: partly measure weights at sea; WLR data partly come from commercial sampling and surveys		

Denmark and Sweden are the only countries sampling directly for ages for the landings of cod2532 (i.e. port sampling where otoliths are taken from all fish in a randomly chosen sample). For discards, all countries use length-stratified sampling schemes to sample for age (questions 3, 8). Apparently, all countries have strata with missing data and therefore borrow from strata with data (question 5). An ALK should be only used for the sampling unit from which the sample was taken; however, if specific checks are in place, borrowing between sampling units may be justified (question 6).

In summary, countries should upload raw data and avoid national borrowing of ALK data. In case national borrowing is necessary, this should be documented and communicated to the stock coordinator.

In terms of the WLR of cod2532 (Table 2.2.1.b), Sweden, Poland and Lithuania reported that they record individual weights rather than a WLR. Only Denmark and Germany use a WLR. They apply sampling of commercial catches to commercial fishery samples. The difference between these two countries is mainly based on the sampling strategy; Germany samples catch at sea while Denmark samples landings in ports and discards at sea.

In summary, there seems to be no major concern related to the use of WLR in cod2532. Three countries sample weights directly and two countries that use WLR apply these in a reasonable manner (Denmark borrows between hauls while Germany also borrows between trips).

In addition to cod 2532, three other stocks were chosen to illustrate the potential diversity of catch sampling programmes for different taxa: a demersal stock with known quality issues (had-arct, arctic haddock); a pelagic stock with several landing countries (her3d22, herring in the Western Baltic); and a flatfish stock with several landing countries (Baltic flounder, fle2232).

Instead of providing major insights about the ALK or WLR issues of the stocks, it turned out that the questionnaire were not fully understood, resulting in somewhat ambiguous interpretations. This had partly emerged already when analysing the answers of cod 2532. If WLR Question 1 was misinterpreted, the country would not have to provide further information. As a consequence for the WLR of had-arct, only data from Norway and Germany were available; for her3d22 only data from Denmark; and for fle2232 only data from two (Denmark, Germany) out of five countries.

All countries provided responses to the ALK questionnaire. There are differences in the answers between the countries. However, these differences are not further discussed here because the consequences of these differences for the calculations and estimates are difficult to assess. Robust experiments are needed to compare the potential effects of different methodological approaches using real-world data. Only well-designed case studies and scenarios can illustrate the real impacts of the methodological diversity on the data quality. The current questionnaire was not designed with that purpose.

Table 2.2.2.2a,b. Overview of answers to the ALK and WRL questionnaire by country and questions for the stock had-arct; the questions and possible answers are given in section 2.2.1; the letters in the cells refer to the answers given.

had-arct-ALK			
Question	Norway	Germany	Russia
1	a	a	a
2	a	a	a
3	b	a	b
4	b,c,d	b,c	c,d
5	b	a	b
6	e	c	c
7	a	a	a
8	a	b	a
Comment	Q5, Russia: ALK data for missing strata imputed from samples collected in neighboring quarters or divisions		

(b) had-arct-WLR			
Question	Norway	Germany	Russia
1	b	a	a
2	b	b	
3	e	c	
4	b,c,d	b	
5	a	a	
6	b	b	
7	b	a	
8	a,c	b	
9		b	
10	a	a	
11		a	

Table 2.2.2.3a,b. Overview of answers to the ALK and WRL questionnaire by country and questions for the stock her3a22; the questions and possible answers are given in section 2.2.1; the letters in the cells refer to the answers given.

Her3a22-ALK				
Question	Denmark	Germany	Sweden	Norway
1	a	a	a	a
2	a	a	a	a
3	b	b	a	a
4	b,c,e	b,c,d,f		
5	b	a		
6	d	e		
7	a	b		
8	b	b		

Her3a22-WLR				
Question	Denmark	Germany	Sweden	Norway
1	b	a	a	a
2	b			
3	a,e			
4	b,c,e			
5	a			
6	b			
7	a			
8	a			
9	a			
10	a			
11	a			

Table 2.2.2.4a,b. Overview of answers to the ALK and WRL questionnaire by country and questions for the stock fle2232; the questions and possible answers are given in section 2.2.1; the letters in the cells refer to the answers given.

fle2232-ALK					
Question	Denmark	Germany	Latvia	Lithuania	Poland
1	a	b	a		a
2	a	a	a	a	a
3	a,b	a	b	a	b
4	b,c,e	b,c,d,e,f	b,c,d,e	a,b,c,d,e,f	b,c,d,e
5	a	b	a	a	a
6	a	a,c	a,e	e	a,e
7	a	a	a	b	a
8	b	a	b	a	b
Comment	Q3, DK: a - direct sampling of age for landings, b - ALK for discard; Q5: Only discard considered				

fle2232-WLR					
Question	Denmark	Germany	Latvia	Lithuania	Poland
1	a,b	a	a	a	a
2	b	a,b			
3	b,e	a,b,c,e			
4	b,c,e	b,c,d,e,f			
5	a	a			
6	c	c			
7	c	a			
8		a			
9	a	a			
10	a	b			
11	a	a			
Comment:	Q1: a - individual weight for landings, b - WLR for discard; Q2,3,4,5,6,8,9,11: Only discard considered			Q10: GER samples catch (landings + discards) at sea	

2.2.3 Use of ALKs and subsampling strategies to quantify the age composition of fish

Estimates of relative abundance by age from fisheries-independent surveys, and of catch in numbers by age from fisheries-dependent surveys, are fundamental inputs to analytical assessments of many commercially important fish stocks in the ICES area. The preferred method ("best scientific practice") for estimating numbers-at-age depends on the survey design and subsampling methods for length and age. The primary sampling units (PSUs) in trawl surveys are the trawl stations (e.g. standardized swept-area). In commercial catch sampling surveys the sampling frame of PSUs is often based on trips or a list of ports or vessels crossed with time (Design classes A – D), which results in a higher level of clustering than for trawl surveys. The subsampling of fish for length measurements and age collections is generally conducted in multiple stages, resulting in a high level of clustering.

Comparing the use of ALK with design-based estimators

The ALK method is commonly used in ICES to estimate numbers-at-age, and has been employed historically by the Institute of Marine Research (IMR) in Norway to produce input data for the stock assessment of Northeast Arctic (NEA) cod based on Extended Survivors Analysis (XSA). Aanes and Vølstad (In review; ICES CM 2013/J:08) compare estimates of age distributions from standard design-based estimators ((e.g. Særndal 1978; Gregoire 1998) for multistage cluster-sampling with estimates based on age–length keys (ALK), using data from trawl surveys and catch sampling surveys for NEA cod as an example. Analysis of trawl and catch sampling survey data, and simulation studies, were conducted to assess the efficiency of various estimators, and to evaluate subsampling strategies for age.

For the trawl surveys, the primary sampling units (PSUs) are trawl catches (standardized area-swept), while vessel-trips or fishing operations are the PSUs in the commercial catch sampling programs in Norway (see Norwegian catch sampling case studies, Annex 6). Otoliths are taken from a length-stratified subsample of fish from the scientific trawl catches, and from a random subsample of fish from the commercial catches. Aanes and Vølstad (In review) demonstrate that properly weighted ALKs and design-based estimators yield equivalent estimates of numbers-at-age when age samples are collected from each PSU. If the mean age of the cluster of fish in the PSUs is correlated with cluster size (number of fish in each PSU), then it is recommended to employ weighted ALKs or design-based estimators to minimize bias, even at the cost of a decrease in the precision as compared to unweighted estimators (see Aanes and Pennington 2003). Aanes and Vølstad (In review) use simulations studies to show that weighted estimators produce more accurate estimates of age-distributions than unweighted estimators. Estimators are tested using simulated survey data from synthetic populations of catches that match the complex covariance structures observed in the empirical data from multistage surveys.

Subsampling strategy for age

The effective sample size (Kish 2003) for estimating relative abundance-at-age as well as age-composition of the total landings of NEA cod is primarily driven by the number of PSUs with age-samples. Aanes and Vølstad (In review)demonstrate that minimal gain in precision in estimates of age-distributions for NEA cod is achieved by taking more than one age sample per 5 cm length group from each PSU in the trawl survey, or more than 20 random age samples from each PSU when sampling commercial catches.

Effect of using a fixed ALK

In practice, it is common to assume that the ALK is given without sampling errors, since it is based on a large number of fish measured for length and age (e.g. Kimura 1977, 1987, Shepherd et al. 1999; WKPICS3 questionnaires). However, since ALKs are estimated from cluster-correlated data from complex multistage sampling surveys, the resulting effective sample size for estimating age-composition of fish is typically substantially lower than the number of fish measured. Hence, ALKs may potentially be subject to large sampling errors. Aanes and Vølstad (In review, ICES CM 2013/J:08) demonstrate that if an age length key is fixed and only the variability of length compositions are allowed for, then the estimated age-distributions will appear to be more precise then they truly are since the ALK itself is subject to sampling errors. This is in agreement with Berg and Kristensen (2012) who demonstrated large spatio-temporal variability of ALK for North Sea Herring. Realistic estimates of precision of age-distributions of catches (for example through bootstrapping of PSUs) require access to data on age and length compositions at the PSU level in the database. The reason is that the precision in estimates of age-compositions primarily is driven by the number of PSUs, and not by the number of fish measured for age. Consequently, it is **recommended** that regional databases such as FishFrame be modified to allow for a higher resolution in the data.

A special case of using a fixed ALK is to "borrow" an ALK derived from sample data from another target population, for example based on data from a different area, gear, or period, and apply this ALK to the length distribution from the population of interest (e.g. landed catch). This approach will only yield an accurate estimate of age-composition in cases where the two populations surveyed have similar age–length composition and ALK (Quinn and Deriso 1999, Kimura 1977, Hoenig and Heysei 1987, Hoening et al. 2002). As a practical example (from Aanes and Vølstad, In review), consider the estimation of catch-at-age of fish caught by longline, using an ALK derived from samples of catches caught be gillnet in the same area and period. Since the size selectivity of the fishing gears are likely to be different, length-at age of landings from the two gears may also be different since faster growers may be caught at younger age than slow growers, even if the two fisheries were taking place in the same period at the same spatial locations. Aanes and Vølstad (In review) compared the estimated age distribution of the commercial catch by longline based on applying ALKs for longline and gillnet to the length distribution of longline. The resulting estimates of age composition were rather different. This suggests caution when exchanging or "borrowing"

ALK's from other populations and it is recommended that sensitivity analysis be conducted whenever possible, since the borrowing of ALK's is likely to produce bias.

In summary, when using an ALK it is advisable to perform the following checks:

- Check whether there is a correlation between cluster size (catch weight/number) and mean age. If there is, weighting the age data by cluster size (number of fish per PSU) is advisable. Unweighted estimators can result in strong bias and poor accuracy.
- Check the sensitivity of various approaches (length-stratified vs. random, reduction/increase in sample size); if this is not possible it may be useful to set up some experiments, e.g. on surveys.
- When 'borrowing' ALK data from a different stratum (gear/quarter/area), check for differences in the age structure between the strata.

2.3 Report for European Commission, reviewing data quality indicators used in non-EU countries and providing advice for appropriate data quality indicators to be included in the DC-MAP (ToR f).

A report was requested by the European Commission to review possible data quality indicators for biological data as input to STECF EWG 13-18 discussions on revision of the DCF. The requirement was to provide a review covering the following three topics related to fleet-based and stock-based biological sampling from marine fisheries:

1. Overview and compilation of discussions that have already taken place on quality indicators/measures of quality for DCF biological data (including in the context of ICES, STECF and other appropriate EU fora).
2. Overview of approaches used in other important fishing nations (e.g. USA, Canada, New Zealand, Australia, Norway...) to measure and ensure quality of biological data.
3. A reflection on whether MS following best practice guidance for data collection is sufficient or whether measures of quality of the collected data are necessary in addition. Possible measure of quality/quality indicators that could be used and pros/cons of these.

The report was drafted by the WKPICS3 co-chair (MJA) acting as an independent scientific expert for the Commission, but the Commission agreed that WKPICS3 could have valuable input. Accordingly, a draft of the report was presented at the WKPICS3 meeting, and the final submitted version reflects the input of the workshop (though it may not reflect the views of all WKPICS members or of the European Commission).

The final report is given in Annex 3.

2.4 Guidance on diagnostic methods to evaluate aspects of data quality to facilitate the work of Regional Coordination Groups and advice to SC-RDB on development requirements for the RDB related to data quality assurance and reporting (ToR g)

WKPICS3 was asked under its ToR (g) to provide guidance on diagnostic methods. Because ToR (g) has highlighted major gaps in the development of diagnostic models in the area of data management, WKPICS3 decided to focus on this area. This fits in with the request to provide advice on the development of the RDB related to quality assurance and reporting. These two issues are discussed in section 2.4.1

In section 2.4.2 some case studies are presented on diagnostic methods applied in Ireland and Norway.

Under ToR (g) WKPICS3 was also requested to consider additional ToRs for WGCATCH and WGBIOP to continue the development of diagnostic methods. Rather than add to the already extensive list of ToRs for these expert groups, WKPICS3 decided that section 2.4.1 could be used by WGCATCH as a starting point for further development of diagnostic methods. Additionally WKPICS3 provides recommendation to PGCCDBS on the focus that WGBIOP should take (section 2.4.3).

2.4.1 Overview of diagnostics methods and RDB requirements

Table 2.4.1.1. is based on a table in the RCM NSEA (2013) report; WKPICS expanded on this table and also added a column highlighting implications for the RDB. Additionally, some examples of diagnostic methods applied in national labs are given in section 2.4.2.

Table 2.4.1.1. Quality issues related to data management, example diagnostics and example mitigation procedures and implications for the RDB. This table is an expanded version of table 5.1 in the RCM NSEA (2013) report.

Stage	Quality issues	QA/QC procedures	Example diagnostics	RDB implications
National data capture	Transcription errors; data entry errors; incomplete entry; ancillary data missing (e.g. missing link between a length sample and vessel data)	Electronic data capture; range checks and other error traps in input software; cross checking of DB content and independent inventory or metadata – in relation to missing data; cross checking biological and fleet data; DB consistency checks and reports.	Outlier detection; data values beyond range checks; Differences between DB content and independent inventory or metadata; inconsistencies between biological and fleet data.	Preferably in accordance with international standards - point to the relevant documentation linking data to certain QA standard. Which procedures have been used to obtain and check this data Countries have to define, document and upload their QA tests.
National data processing	incorrect allocation of trips to métiers or strata; use of weight-length relationships; errors or undetected changes in analysis software; Problems with code lists such as vessel tables; Failure to take sampling strategy into account. Use of inappropriate auxiliary (raising) variables. Wrong species code	Quality assurance of data processing procedures and codes; checking analysis routines using standard test datasets; Following guidelines for raising data; checking for correlation with aux variable; checking species distribution/depth etc. Comparing observer data with landings on a broad scale.	Unexpected changes in processed data from previous years; Length-weight diagnostics; Comparing raised retained catch (using aux variable other than landings) to the official landings; Check number of samples in strata; Check contribution of each sample to final estimate.	

Stage	Quality issues	QA/QC procedures	Example diagnostics	RDB implications
Upload to RDB	Incomplete uploads; undetected errors in national database. Incomplete or incorrect species list or other code lists Wrong species code Species groups	Range checks and other error traps in RDB; cross checking of RDB and national DB content and ICES landings etc. checking species distribution/depth etc.	Outliers; data values beyond range checks; Differences between RDB content and national DB content.	Validation checks within the RDB are needed. Implement range checks and other checks (reference lists). Develop metadata report summarizing uploads. Document how CL and CE data has been compiled (logbook, sales slips, value). Reference to document describing this. Data fields describing what method have been used (broad elements e.g. how is value, effort, individual weights, landings, estimated)
RDB Quality checking				Data to be checked towards ICES QA standard. Check census datavs.sampling data – sampling frame/study population (needed to run quality checks) Document what kind of quality checks that have been applied (common report on results of QA checks, goes back to data supplier) Indicate if quality checks are performed and flag data if data has being checked (yes/no/checked but issues, strange but ok)

Stage	Quality issues	QA/QC procedures	Example diagnostics	RDB implications
RDB data extraction and analysis	Compatibility of national datasets (e.g. métier definitions; different forms of bias); imputation or other handling of missing data; national sampling design or cluster effects not properly reflected in data analysis; errors or undetected changes in analysis software Need to be able to identify sampling frame	Suite of diagnostic checks for RDB data; Full documentation of national sampling programmes; Cross checking data analysis procedures and national sampling design; Test datasets for analysis software.	Gaps / inconsistencies revealed in RDB diagnostic outputs or other data quality reports. Proportion of catch comprising strata with missing or imputed biological data. Differences between national survey design descriptions and analysis hierarchy. Unexpected changes in processed data from previous years.	Check data on a regional level QAs for national sampling schemes (performance, coverage, variation…). Include non-response rates in exchange format, how do you submit those? Estimation processes developed, tested, documented and implemented. Estimation processes need to be developed, documented and implemented Documentation and references, flag data

2.4.2 Examples of diagnostic methods

Observer data – Ireland

After data of an observer trip are entered, an automatic quality control report is produced that helps to identify potential problems with the data. The report identifies problems with:

- Database consistency: (duplicate data; data existing in one table but not in another, e.g. haul data without trip data; missing data)
- Raising factors (from sample to haul level): very small or large sample weights; high raising factors, discard sample weight larger than estimated total discards; large proportion of discards in the catch; low or high catch or landings rate (kg/hr); etc.
- Tow data: Excessive tow length or fishing speed (tow length is estimated from the straight-line distance between the start and end positions of the tow); Zero tow length; Impossible or unexpected shoot or haul positions; Short tow duration; Negative tow duration; Missing tow duration; Long tow duration; Tow shot before previous tow was hauled; Tow dates outside cruise dates.
- Length data: Any fish that are larger than the 99th percentile * 1.5 or smaller than 1st percentile * 0.5, are identified as outliers.

In order to identify outlying values, boundaries are defined (somewhat arbitrarily), so e.g. if a sample weight is smaller than 5kg it is flagged as an outlier (Figure 2.4.2.1). A series of SQL queries identifies outliers and passes a table to R which pasts the results into a pdf report using Sweave. Below are some examples of query results which are pasted into the report:

- Haul 23: Tow length is longer than expected (23.8nm/4.9h)
- Haul 7: Proportion of discards is high (0.83)
- Haul 1: Shoot date (2001-04-22) before departure date (2001-04-23)
- Haul 4: Haul 4 was shot before haul 3 was hauled
- Haul 3: Unexpected length for Dab (47cm)
- Haul 6: Catch rate is high (7200kg/2.5h)
- Haul 7: Raising factor is high (457.2)
- Haul 12: Sample data with missing Haul data. Sample header id: 48632

The report also generates a number of tables with haul and sample information that allow the user to drill down into any of the issues that are flagged. Additionally, two figures are produced to help put any outliers into context.

In addition to the quality control reports for individual trips, the contribution of each trip to the overall estimate of the discards or retained catch in a stratum is also investigated (Figure 2.4.2.3).

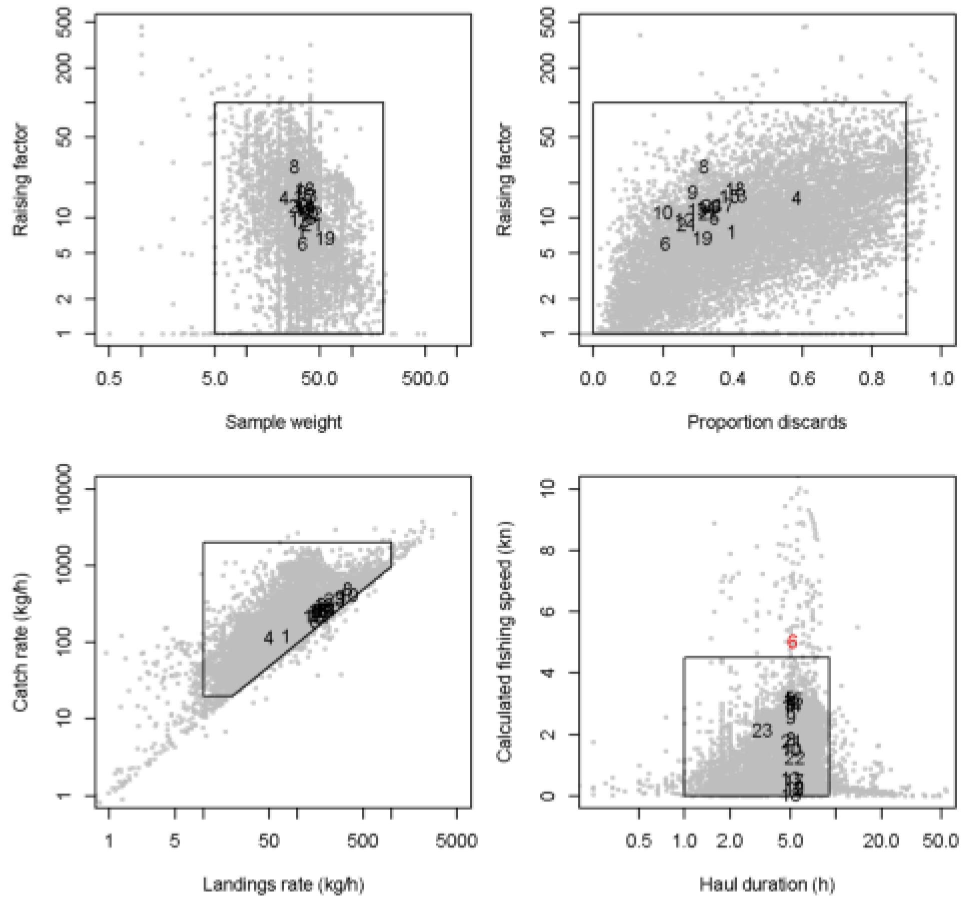


Figure 2.4.2.1. Diagnostic checks for observer data in Ireland. All observed values in the database shown in grey (each point represents a haul). The boxes represent the expected range of values, any values outside the boxes will be flagged as outliers. The black numbers refer to the haul numbers of the current survey, hauls outside any of the boxes are plotted in red, e.g. haul number 6 in the bottom-right plot which had an unexpectedly high fishing speed (which could be due to errors in the start/end time or start/end position).

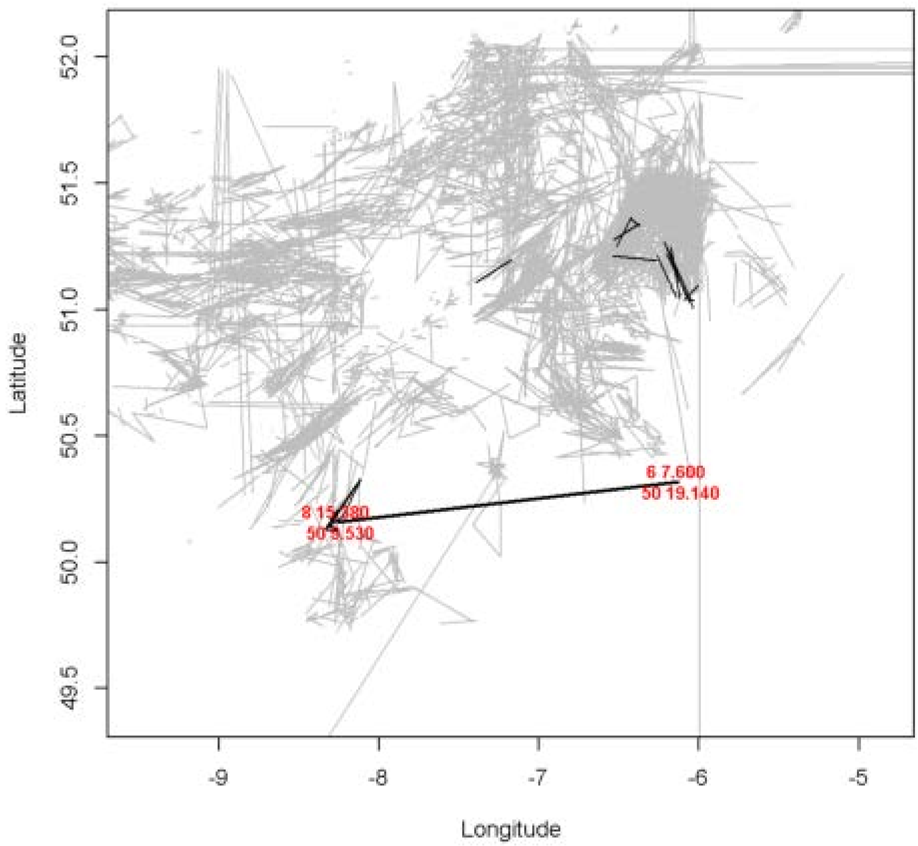


Figure 2.4.2.2. Diagnostic checks for observer data in Ireland The figure above shows the shoot and haul positions (connected by a straight line) of all the trips in the database in grey and of the current haul in black. The positional data for a suspicious haul are also given in red. In this case it appears that the longitude of the shoot or haul position was entered wrongly.

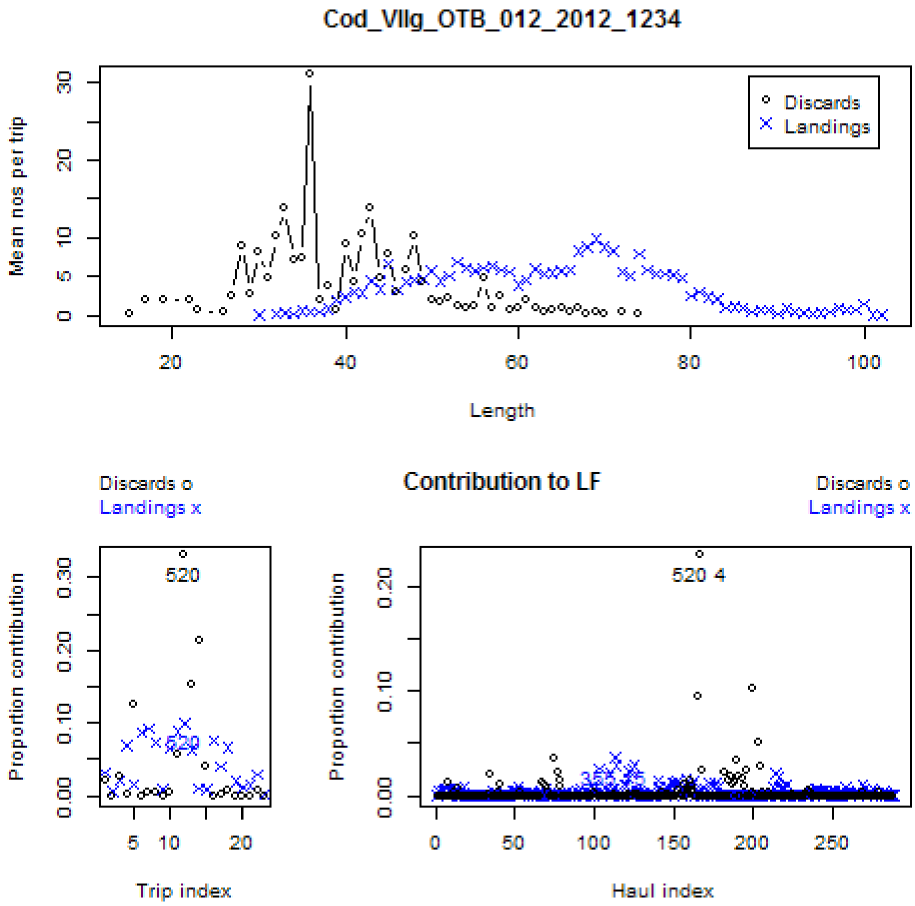


Figure 2.4.2.3. Diagnostic checks for observer data in Ireland. Figures show how much each trip (bottom left) and each haul (bottom right) contribute to the total estimate (each trip and each haul are given an arbitrary index number for plotting). So for example, if the estimated weight of discards on 23 sampled trips is 1000 kg and 250kg of that comes from a single haul, then the contribution of that haul is 25%. The figure above shows that trip number 520 contributed more than 30% of the discard estimate and that a single haul (haul 4) on that trip contributed more than 20% of the total discard estimate. Such a result would be a reason to drill down into the data for that particular trip (and haul).

Examples of diagnostics for Norwegian input data on catch-at-age for stock assesment of northeast Arctic cod, haddock and saithe are given in Figures 2.4.2.4 & 5.

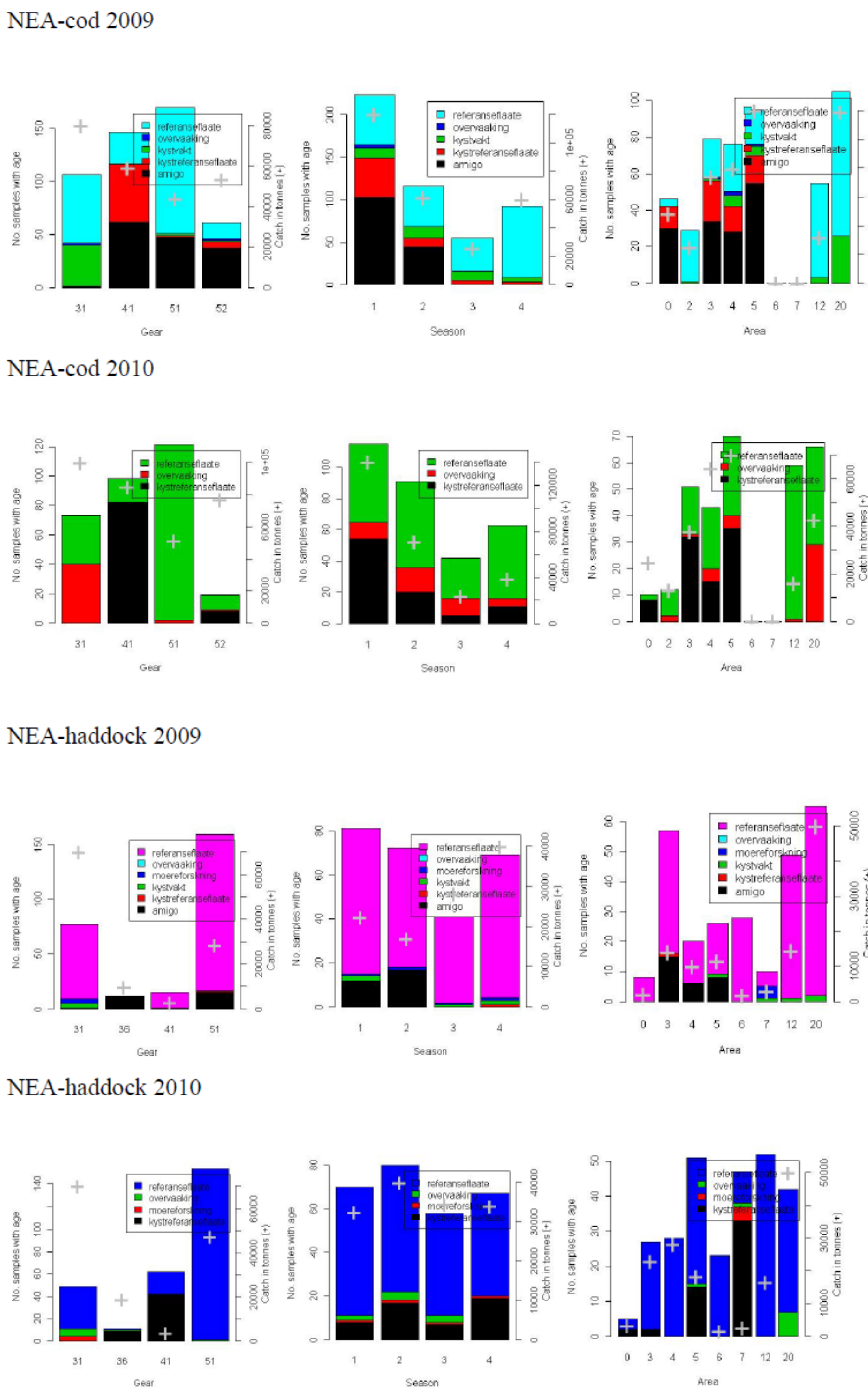


Figure 0.1. Norwegian AGE samples of commercial catches of NEA-cod and -haddock in 2009 and 2010. Note the different axes and colours. The different sampling platforms are shown by different colours, unfortunately with different colours each year.

Figure 2.4.2.4

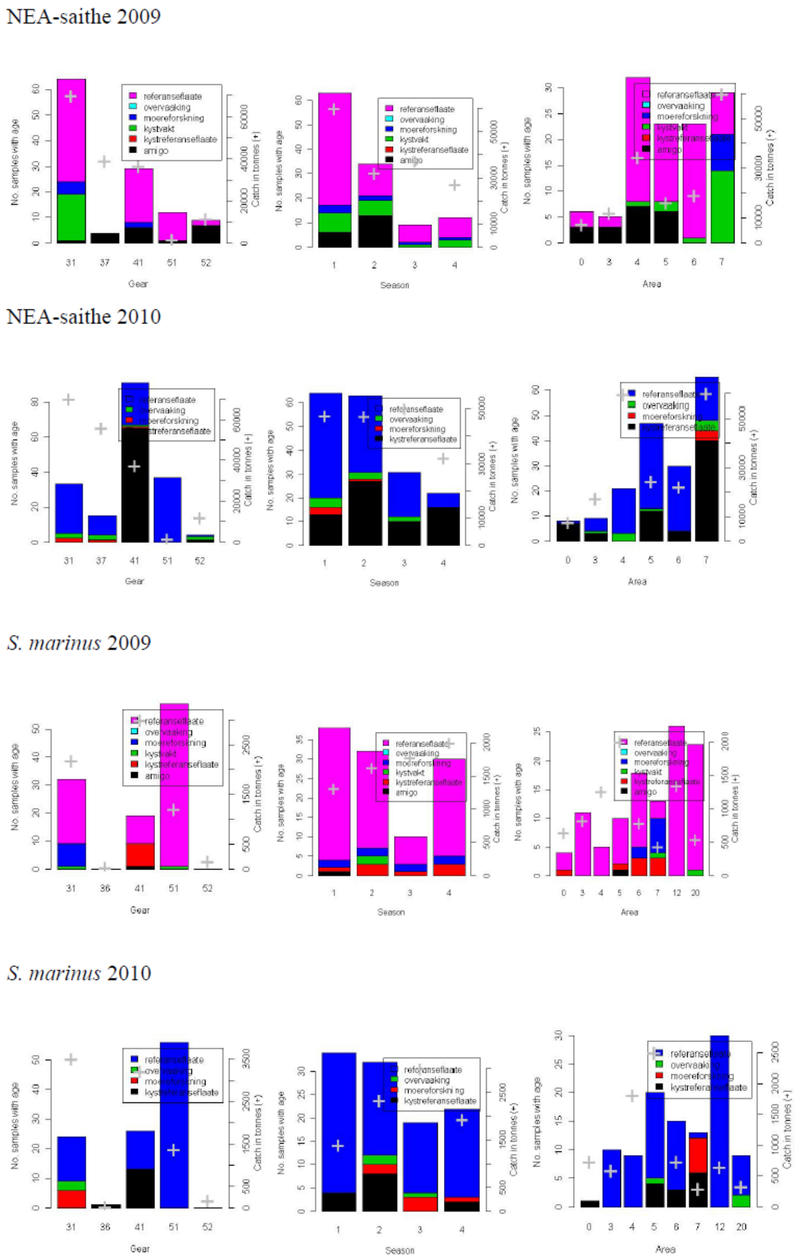


Figure 0.2. Norwegian AGE samples of commercial catches of NEA-saithe and *Sebastes marinus* in 2009 and 2010. Note the different axes and colours. The different sampling platforms are shown by different colours, unfortunately with different colours each year.

Figure 2.4.2.5.

Figures 2.4.2.4 & 5 show example of diagnostics to assess data on catch-at-age as input to stock assessments of Northeast Arctic (NEA) cod and haddock, and saithe (from ICES Arctic Working Group Reports).

Summary of age samples in 2006

Gear	Q\A	3	2	12	4	5	0	20	6	7
41	1	2284.6 4, 4, 207	0 0, 0, 0	0 0, 0, 0	10776.2 8, 8, 453	16323.9 7, 21, 1030	11991.1 8, 25, 1318	0 0, 0, 0	0 2, 8, 160	0 3, 9, 354
41	2	1559.8 4, 4, 199	0 0, 0, 0	0 0, 0, 0	3353.5 1, 1, 20	2287.1 2, 3, 101	2347.1 10, 13, 608	0 0, 0, 0	0 1, 1, 20	0 5, 6, 322
41	3	107.8 0, 0, 0	0 0, 0, 0	0 0, 0, 0	302.2 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0
41	4	141.1 0, 0, 0	0 0, 0, 0	398 2, 6, 120	599 1, 1, 24	0 0, 0, 0	0 0, 0, 0	504.1 1, 4, 80	0 0, 0, 0	0 0, 0, 0
51	1	2295.8 1, 14, 647	41.7 1, 2, 40	3951.5 4, 7, 147	1264.1 5, 6, 282	4572.2 11, 21, 814	1603.2 2, 2, 120	245.3 1, 1, 24	0 0, 0, 0	0 0, 0, 0
51	2	2655.6 9, 22, 1115	251.5 1, 2, 40	1200.4 2, 5, 108	999.8 2, 4, 86	2564.8 6, 6, 383	937 0, 0, 0	25.1 3, 6, 131	0 0, 0, 0	0 0, 0, 0
51	3	550.7 5, 6, 252	160.6 0, 0, 0	236.8 2, 3, 64	208.9 2, 2, 41	0 0, 0, 0	0 0, 0, 0	1168.1 1, 1, 20	0 0, 0, 0	0 0, 0, 0
51	4	2980.8 4, 16, 838	965.6 0, 0, 0	6455.5 4, 12, 246	558.8 5, 7, 334	0 0, 0, 0	0 0, 0, 0	2125.1 4, 22, 468	0 0, 0, 0	0 0, 0, 0
52	1	686.4 0, 0, 0	0 0, 0, 0	13.8 0, 0, 0	3916.2 2, 2, 86	14409.4 16, 15, 940	7671.9 5, 6, 410	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0
52	2	2807.2 6, 10, 563	0 0, 0, 0	0 0, 0, 0	9608.1 5, 5, 280	2573.3 0, 0, 0	1603.2 4, 4, 284	512.4 0, 0, 0	0 0, 0, 0	0 1, 1, 25
52	3	335.2 0, 0, 0	0 0, 0, 0	0 0, 0, 0	572.5 0, 0, 0	0 0, 0, 0	0 0, 0, 0	42.9 0, 0, 0	0 0, 0, 0	0 0, 0, 0
52	4	255.4 1, 1, 58	0 0, 0, 0	0 0, 0, 0	297.2 1, 1, 66	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0
31	1	1207.1 3, 3, 169	0 0, 0, 0	635.4 0, 0, 0	4660.8 6, 7, 314	6206.9 3, 6, 160	2632.2 3, 4, 79	3.5 0, 0, 0	0 1, 1, 19	0 0, 0, 0
31	2	1333.7 1, 1, 60	59.6 0, 0, 0	508.7 0, 0, 0	4292.1 6, 8, 312	1100.7 3, 6, 146	358.9 1, 1, 20	8653.3 3, 6, 120	0 0, 0, 0	0 0, 0, 0
31	3	383.9 1, 1, 95	5334.2 1, 5, 100	104.6 0, 0, 0	432.2 2, 3, 228	0 0, 0, 0	0 0, 0, 0	7759.9 2, 9, 222	0 0, 0, 0	0 0, 0, 0
31	4	696 0, 0, 0	2313.7 3, 7, 140	804.4 1, 1, 20	2587.3 3, 3, 49	0 0, 0, 0	0 0, 0, 0	13942.9 6, 24, 720	0 0, 0, 0	0 0, 0, 0
32	1	31 0, 0, 0	0 0, 0, 0	0 0, 0, 0	43.8 0, 0, 0	19 0, 0, 0	0 0, 0, 0	72.7 0, 0, 0	0 0, 0, 0	0 0, 0, 0
32	2	108.4 0, 0, 0	54 0, 0, 0	0 0, 0, 0	49.1 0, 0, 0	14.9 0, 0, 0	0 0, 0, 0	90.3 0, 0, 0	0 0, 0, 0	0 0, 0, 0
32	3	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0	13.7 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0
32	4	6.4 0, 0, 0	243.6 0, 0, 0	0 0, 0, 0	42.2 0, 0, 0	0 0, 0, 0	0 0, 0, 0	17.1 0, 0, 0	0 0, 0, 0	0 0, 0, 0

Total catch= 203117.7 #Boats sampled= 273
Sampled catch=192952.9(95)% #Serienr sampled= 424
Sampled catch required for DB estimation=184153.5(91)% #Fish sampled= 16875

Includes:

- Amigo (port-sampling)
- RF Ocean & Coast
- Coast guard

Catch in tons		
Nos boats sampled	Nos age samples	Nos fish aged

Summary of length samples in 2006

Gear	Q\A	3	2	12	4	5	0	20	6	7
41	1	2284.6 4, 4, 208	0 0, 0, 0	0 0, 0, 0	10776.2 8, 8, 453	16323.9 18, 28, 1628	11991.1 8, 27, 1473	0 0, 0, 0	0 2, 35, 1501	0 3, 9, 519
41	2	1559.8 4, 4, 211	0 0, 0, 0	0 0, 0, 0	3353.5 1, 5, 300	2287.1 2, 3, 101	2347.1 10, 13, 608	0 0, 0, 0	0 1, 1, 20	0 5, 6, 414
41	3	107.8 0, 0, 0	0 0, 0, 0	0 0, 0, 0	302.2 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0
41	4	141.1 0, 0, 0	0 0, 0, 0	398 3, 23, 1352	599 1, 1, 60	0 0, 0, 0	0 0, 0, 0	504.1 1, 29, 1611	0 0, 0, 0	0 0, 0, 0
51	1	2295.8 1, 32, 1887	41.7 1, 6, 267	3951.5 7, 37, 2159	1264.1 6, 17, 1002	4572.2 1, 86, 4282	1603.2 3, 4, 201	245.3 3, 8, 576	0 1, 1, 26	0 0, 0, 0
51	2	2655.6 9, 45, 2705	251.5 3, 15, 898	1200.4 6, 37, 3003	999.8 5, 23, 1559	2564.8 8, 10, 571	937 0, 0, 0	25.1 4, 34, 2095	0 0, 0, 0	0 0, 0, 0
51	3	550.7 6, 7, 296	160.6 1, 3, 161	236.8 4, 18, 1076	208.9 3, 5, 199	0 0, 0, 0	0 0, 0, 0	1168.1 3, 8, 477	0 0, 0, 0	0 0, 0, 0
51	4	2980.8 5, 26, 1577	965.6 0, 0, 0	6455.5 12, 80, 5160	558.8 5, 17, 970	0 0, 0, 0	0 0, 0, 0	2125.1 5, 123, 8052	0 0, 0, 0	0 0, 0, 0
52	1	686.4 0, 0, 0	0 0, 0, 0	13.8 0, 0, 0	3916.2 5, 10, 671	14409.4 19, 22, 1643	7671.9 7, 8, 518	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0
52	2	2807.2 8, 20, 1346	0 0, 0, 0	0 0, 0, 0	9608.1 15, 23, 1771	2573.3 2, 3, 221	1603.2 5, 5, 313	512.4 0, 0, 0	0 0, 0, 0	0 1, 1, 63
52	3	335.2 1, 1, 29	0 0, 0, 0	0 0, 0, 0	572.5 0, 0, 0	0 0, 0, 0	0 0, 0, 0	42.9 0, 0, 0	0 0, 0, 0	0 0, 0, 0
52	4	255.4 1, 1, 58	0 0, 0, 0	0 0, 0, 0	297.2 1, 1, 66	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0
31	1	1207.1 3, 3, 199	0 0, 0, 0	635.4 5, 5, 276	4660.8 7, 17, 1123	6206.9 7, 33, 2130	2632.2 3, 16, 934	3.5 0, 0, 0	0 1, 2, 82	0 1, 5, 225
31	2	1333.7 2, 2, 330	59.6 1, 1, 122	508.7 1, 4, 240	4292.1 10, 26, 1699	1100.7 5, 23, 1334	358.9 1, 3, 180	8653.3 9, 37, 3393	0 1, 1, 93	0 0, 0, 0
31	3	383.9 4, 5, 508	5334.2 3, 35, 2178	104.6 0, 0, 0	432.2 5, 6, 472	0 0, 0, 0	0 0, 0, 0	7759.9 14, 60, 4850	0 0, 0, 0	0 0, 0, 0
31	4	696 1, 1, 81	2313.7 3, 38, 2192	804.4 3, 4, 241	2587.3 11, 15, 1131	0 0, 0, 0	0 0, 0, 0	13942.9 11, 94, 6259	0 0, 0, 0	0 0, 0, 0
32	1	31 0, 0, 0	0 0, 0, 0	0 0, 0, 0	43.8 0, 0, 0	19 0, 0, 0	0 0, 0, 0	72.7 0, 0, 0	0 0, 0, 0	0 0, 0, 0
32	2	108.4 0, 0, 0	54 0, 0, 0	0 0, 0, 0	49.1 0, 0, 0	14.9 0, 0, 0	0 0, 0, 0	90.3 0, 0, 0	0 0, 0, 0	0 0, 0, 0
32	3	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0	13.7 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0	0 0, 0, 0
32	4	6.4 0, 0, 0	243.6 0, 0, 0	0 0, 0, 0	42.2 0, 0, 0	0 0, 0, 0	0 0, 0, 0	17.1 0, 0, 0	0 0, 0, 0	0 0, 0, 0

Total catch= 203117.7 #Boats sampled= 406
Sampled catch=197921.7(97)% #Serienr sampled= 1368
Sampled catch required for DB estimation=195679.3(96)% #Fish sampled= 86635

Includes:

- Amigo (port-sampling)
- RF Ocean & Coast
- Coast guard

Catch in tons		
Nos boats sampled	Nos age samples	Nos fish aged

Figure 2.4.2.6. Examples of summary tables to evaluate Norwegian age- and length sampling levels for northeast Arctic cod from multiple sampling platforms (see Norwegian case studies, Annex 6).

2.4.3 Recommendations for WGBIOP and WGCATCH

A significant task of the new ICES Working Group on Biological Parameters (WGBIOP) will be the continuation of quality assurance of age reading, the classification maturity stages etc. which was previously the remit of PGCCDBS. Much of this work is well-developed and it is important that WGBIOP also focuses on areas where major gaps in expertise exist, like the development of statistically sound sampling designs for stock-related parameters.

One of the major concerns in relation to sampling design is the use of commercial fishery sampling data to estimate stock parameters. Stock-based data are usually best collected using fish stock surveys covering the full range of the stock. However, in practice fishery data are often used to estimate these parameters, particularly for species with very low survey catches or where suitable surveys do not exist. In some cases this may be a valid approach but one cannot simply assume that this is the case. Fish of a certain maturity state, sex or with a certain growth rate might have different susceptibility to fishing gear than fish of the same length of a different maturity state, sex or growth rate; either through differences in behaviour, migration or condition (e.g. a fat fish is more likely to be caught in a gillnet than a fish in poor condition of the same length - which also applies to survey data).

The potential for bias can be simply illustrated by considering that (for example), the proportion mature at a given age is effectively the ratio of the total number of mature fish at that age in the stock, to the total number of all fish at that age. In practice it is almost impossible to estimate this ratio without bias, unless the mature and immature fish are mixed and fully contained within the area of a survey or fishery, and have the same probability of being caught by the sampling gear (or the differences in catchability are known). Given that ICES recommends that sampling for maturity for many species takes place close to the time of spawning, these conditions are often unlikely to be met if mature fish are undertaking spawning migrations and have a different spatial or depth distribution and behaviour to immature fish. When using large-scale surveys, there is a chance to evaluate these differences in distribution and investigate potential differences in catchability using targeted studies. However, if using fisheries as population samplers, there is much less possibility to collect data in such a controlled way. In the worst case, fisheries may specifically target spawning fish, or vary in stock coverage and selectivity over years.

Before using commercial fishery sampling data for stock-based parameters, it is therefore essential to investigate if, in principle, those parameters can be estimated from fishery samples. If there is inevitable bias, how might this change over years due to changes in fishing methods, gear design or areas fished?

One of the draft ToRs for WGBIOP asks it to continue the development of methods and guidelines for best practice in the analysis of biological data. WKPICS3 suggests that these methods could include a standard tool like COST and that WGBIOP should consider the needs for storage of these data. For example, the use of design-based weighted estimators of numbers-at-age for a regional stock would require that the Regional Data Base (RDB) (or national databases) include data at the PSU level. Additionally WKPICS3 points out that many guidelines and manuals are already finalized and that these should be made available to the working group (e.g. the Danish maturity manual for Baltic Cod).

WKPICS3 also suggests that it would be useful for WGBIOP to act as a strong link between end-users and data submitters. It may be useful to produce a table of stocks and list the assessment EG's requirements for each of the biological parameters.

Recommendation for WGCATCH:

WKPICS3 recommends that the development of statistically sound sampling designs be continued in WGCATCH, and that a strong link between WGCATCH and WGBIOP be established to ensure that statistically sound methods are developed for the collection of samples and estimation of biological parameters from commercial fisheries. For this purpose it is essential that the WGCATCH meeting is attended by experts in this field.

Recommendation for WGBIOP:

WKPICS3 recommends that a strong link between WGCATCH and WGBIOP be established to ensure that statistically sound methods are developed for the collection of samples and estimation of biological parameters from commercial fisheries. For this purpose it is essential that the WGBIOP meeting is attended by experts in this field

2.5 Finalize guidelines on estimation procedures for all four principal classes of catch sampling schemes identified by WKPICS2, including using auxiliary data for re-weighting (ToR b)

2.5.1 Principal classes of survey designs for catch-sampling programs

Fisheries catch sampling schemes considered here can broadly be categorized into four principal classes based on the number of stages in the sample selection. For at-sea sampling programs, the sampling frame is ideally constructed so that vessels, trips, and fishing operations can be selected with known probability over time. The effective sample size can be maximized by random sampling across all vessels, trips and fishing operations in each stratum. The two principal design classes for at-sea sampling are given in A and B below:

a) **At-sea sampling with trips as primary sampling units**. When trips can be selected randomly from a fleet of vessels, at least approximately, it is often reasonable to treat vessel-trips as the primary sampling units. In such cases, the list of all trips (obtained at the end of the year) makes up the sampling frame. This is a virtual frame that cannot be used in stage 1 to select the trips. The actual selection is typically based on a frame with a vessel list crossed with time. For a fleet with day-trips this can easily be achieved by randomizing the selection of days and vessels. For fleets with varying trip-length it is more difficult to selected vessels and trips with approximately equal inclusion probabilities. It can be helpful to create strata where vessels with a similar trip length are grouped.

b) **At-sea sampling with vessels as primary sampling units**. When it is not possible to approximately achieve a random sample of trips for a fleet, then another design option for at-sea sampling is to select vessels randomly in stage 1, and then select a subsample of trips throughout the year for each vessel. In this case, the vessel is the PSU, with trips as second stage sampling units and fishing operations as third stage sampling units. This design introduce an extra level of clustering, since trips and fishing operations to be sampled now are nested with a fixed number of vessels selected in stage 1. Clearly, these trips may not be considered a simple random sample from the entire fleet. A special case arises if all vessels, or a subset of vessels that are responsible for almost all of the catches, are sampled, in which case each vessel is effectively a stratum.

For onshore sampling a common approach is to conduct the sampling of catches from vessels and trips that can be accessed in ports where they land their catches. In these cases, the sampling frame is based on a list of access-sites crossed with time (for example port-days). It is common that the catches for many completed trips are sorted by market category before they are landed and can be sampled. The market categories will then form strata. Subsampling of vessel-catches may then be conducted by selecting a random sample of boxes from each market category (stratum), and then measure all fish from each box, or a subsample of fish from each box. The two principal design options for onshore sampling are given in C & D below:

c) **On shore sampling with site-days as primary sampling units**. Where the primary sampling units can be defined as site-days which can be randomly selected, there is one extra level of clustering, where site-days are selected in stage 1, trips in stage 2, boxes in stage 3 (for sorted catches), and fish in stage 4.

d) **On shore sampling with sites as primary sampling units**. Another design option is to select a sample of sites (PSUs) in the first stage, and then conduct catch sampling for a subsample of days (stage 2) within each site selected in stage 1. In stage 3, catch sampling is conducted for a sample of trips on a selected day and site. This option may be cost-effective if ports are scattered over large areas, and field samplers near a selected number of ports can be recruited. As for design class B, a special case arises if all sites, or a subset of sites where almost all catches are handled, are sampled, in which case each site is effectively a stratum.

We summarize the four primary classes of catch sampling schemes in Table 2.5.1.1.

Table 2.5.1.1. Design classes for at-sea and onshore commercial catch sampling. The level of clustering of catches typically increases from scheme A to D. Schemes A and B are at-sea sampling programs, while C and D are onshore sampling. The primary sampling units may be subsampled in multiple stages, using simple random, stratified random, or systematics sampling.

Design class	Sampling frame of PSUs	Comment, example	Examples of stratification of PSUs	Case study
A	Vessels * time	Sample a number of trips across all vessels. In the analysis, trips are treated as PSUs. Sample a number of fishing operations across all vessels/trips; fishing operations are treated as PSUs (e.g. in at-sea self-sampling programs, or at-sea intercept surveys)	Vessel-characteristics (vessel length), time (quarter)	Netherlands case study Skagerrak regional case study Norway case study
B	Vessels	Select a group of vessels and sample trips over time from each vessel. Special case: If all vessels are sampled, each vessel is effectively a stratum.	Fleets (offshore/coastal), gear, target fishery	Norway case study
C	Sites * time	Random sample of site-days (e.g. buyer-days)	Quarter, market categories	Sweden case study
D	Sites	Sample a group of ports and sample vessel/trips over time from each port. Special case: If all ports are sampled, each port is effectively a stratum.	Geographic, quarter, effort, or landings at the sites Month	Scotland case study Spain case studies

In Figure 2.5.1.1 we show an example of sampling frames based on list of vessels (a) and ports (b). It is apparent in example (a) that it is not possible to select the trips directly since they are only know completely at the end of the year. However, it may be possible to select vessels and trips randomly throughout the year so that the probability of sampling trips is approximately equal within strata. If so, it can be assumed that the PSU is a trip, and the data can be analysed according to sampling scheme A. In example (b) the scheduling of port visits is conducted by stratified sampling of site-days. This is an example of design class C.

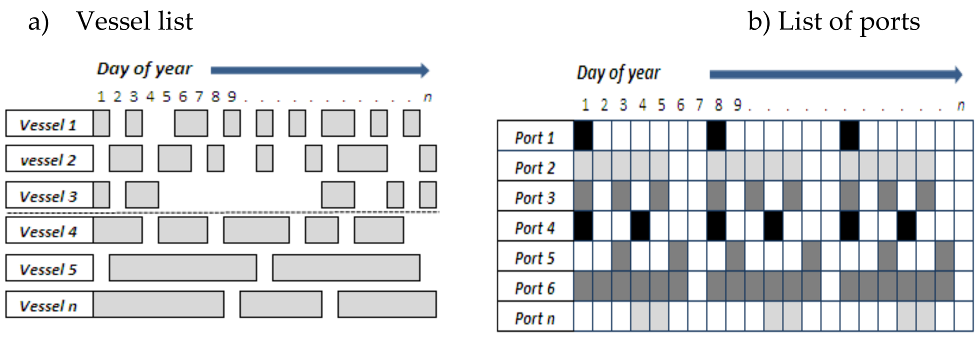


Figure 2.5.1.1. Structuring of the target population (catches) in space and time for (a) a list of vessels with trips of varying duration, and (b) a list of ports with varying frequency of days when catches are landed and can be accessed (degree of shading in boxes indicates "size" of port – e.g. number of vessel trips or size of catch landed per day.

Case studies that illustrate each design class, and examples of design-based weighted estimators are provided in Annex 6. Aanes and Pennington (2003) and Helle and Pennington (2004) provide examples using weighted-design estimators. See also Cochran (1977), Lehtonen and Pahkinen (1994), Særndahl et al. (1992), Thompson (1992), Lohr (2010), and Lumley (2010) for the theory of design-based estimators. Alternative model-based estimation methods for estimating catch-at-age are provided in Hirst et al. (2004, 2005, 2012).

2.6 Finalize recommendations for the Regional Databases concerning procedures for combining national fishery sampling data or estimates to give regional or supra-regional estimates for fisheries or stocks (ToR c)

2.6.1 RDB regional sampling, estimation procedures and data structures.

Regional Sampling designs

The sampling designs considered in WKPICS follow an envisaged regional design where strata would be defined to optimize regional goals in data collection. National data for a stock could be considered super-strata in the estimation of age and length compositions for the regional catch, for example. These regional strata would be implemented at the national level with the sampling commitment of individual countries being determined by a country's involvement in the fisheries that were deemed to be of priority at the regional level. Within each nation there is the operation of onshore and at-sea sampling schemes, determined by a regional plan. The regional strata would, or could, be augmented with nationally defined strata to meet national data needs without compromising the regional design. Such regional designs were considered at PGCCDBS 2012, with elaboration during SGPIDS 2 and WKPICS in 2013. A more detailed scoping exercise was undertaken during the RCM NS& EA and the RCM NA in 2013, using RDB data..

In a regional design (Figure 2.6.1.1) the onshore and at-sea sampling designs at the national level would consist of sampling frames for onshore and at-sea sampling following the sampling schemes A to D as defined in section 2.5 above. The main difference between a regional design and a purely national design being that, particularly in the case of onshore sampling, the ports would be chosen across nations to improve regional estimates by including ports with significant contributions to regional landings, rather than all national ports.

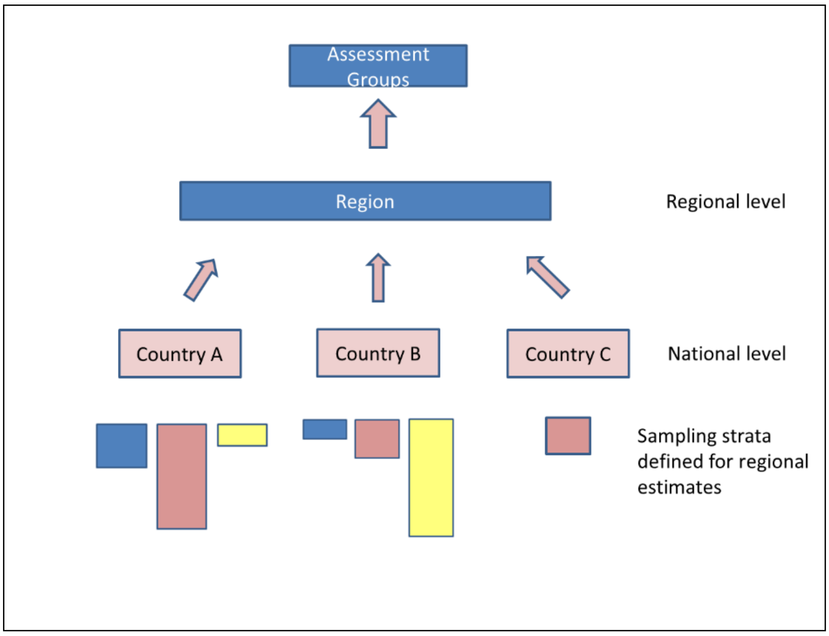


Figure 2.6.1.1. Stratification of a regional fishery sampling scheme

Under a regional design, sampling strata would be defined to achieve regional goals in data collection, so that the sampling commitments at the national level would depend on that country's importance to the regional estimate for the fishery. National estimates are aggregated at the regional level to provide estimates for the assessment groups.

Within the national schemes at-sea sampling and onshore sampling would be conducted according to the four types of design set out A- D. A characteristic of fisheries sampling is the hierarchical cluster sampling within the primary sampling units (PSU) that is needed to obtain the fish (or shellfish) that are actually measured or from which age structures are collected.

For at-sea sampling (Figure 2.6.1.2), typically the hierarchical cluster sampling within a fishing trip PSU would involve:

- Sampling of hauls (or sets) *h* from all hauls (or sets) H
    - Sampling a number of baskets *b* from all baskets of discards B
        - Species-specific sampling of individual fish for length within baskets
            - Subsampling of individual fish for age.

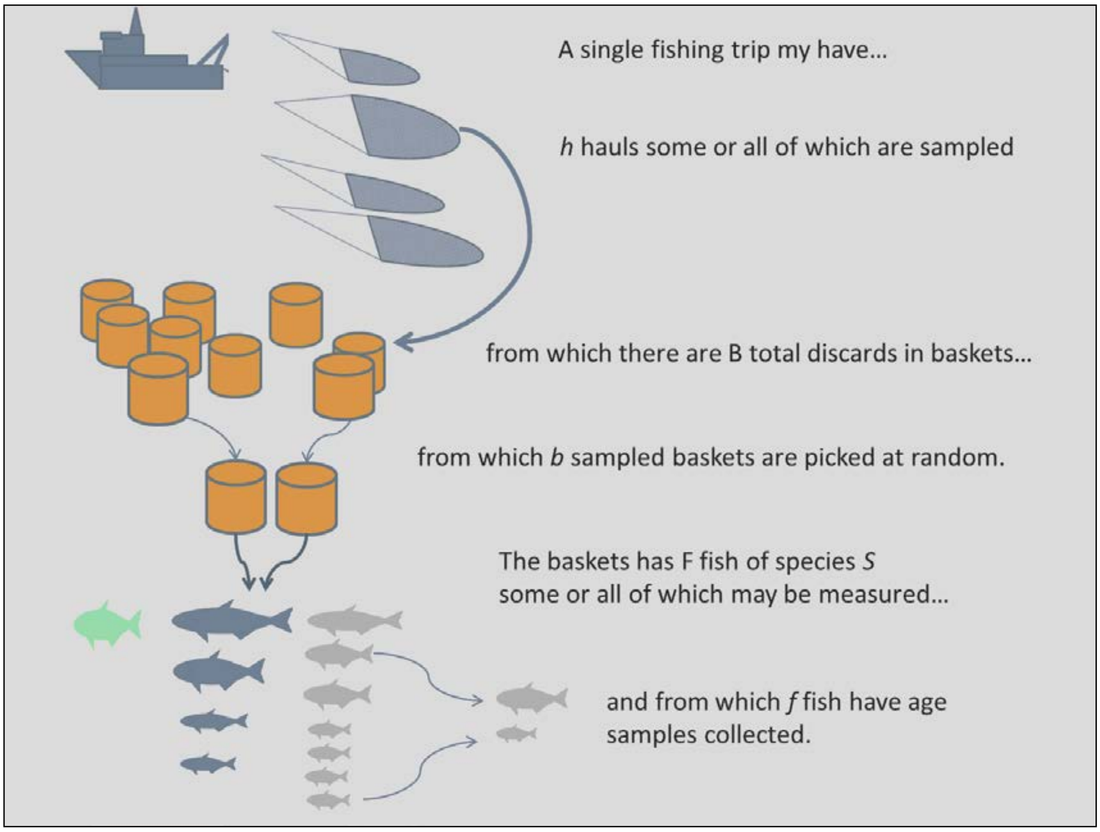


Figure 2.6.1.2 The hierarchical cluster sampling that may be involved in collecting an age samples from the discards of a single trip. Each stage of the selection hierarchy gives rise to a sampling probability which determines the overall sample weight for the measured fish.

For onshore sampling where the PSU is the port or the port and day (Figure 2.6.1.3) a typical hierarchical cluster sampling would involve:

- Sampling trips within the port on the day
    - o Sampling species-specific commercial size categories from the trip
        - ▪ Sampling of fish for length
            - Subsampling of individual fish for age.

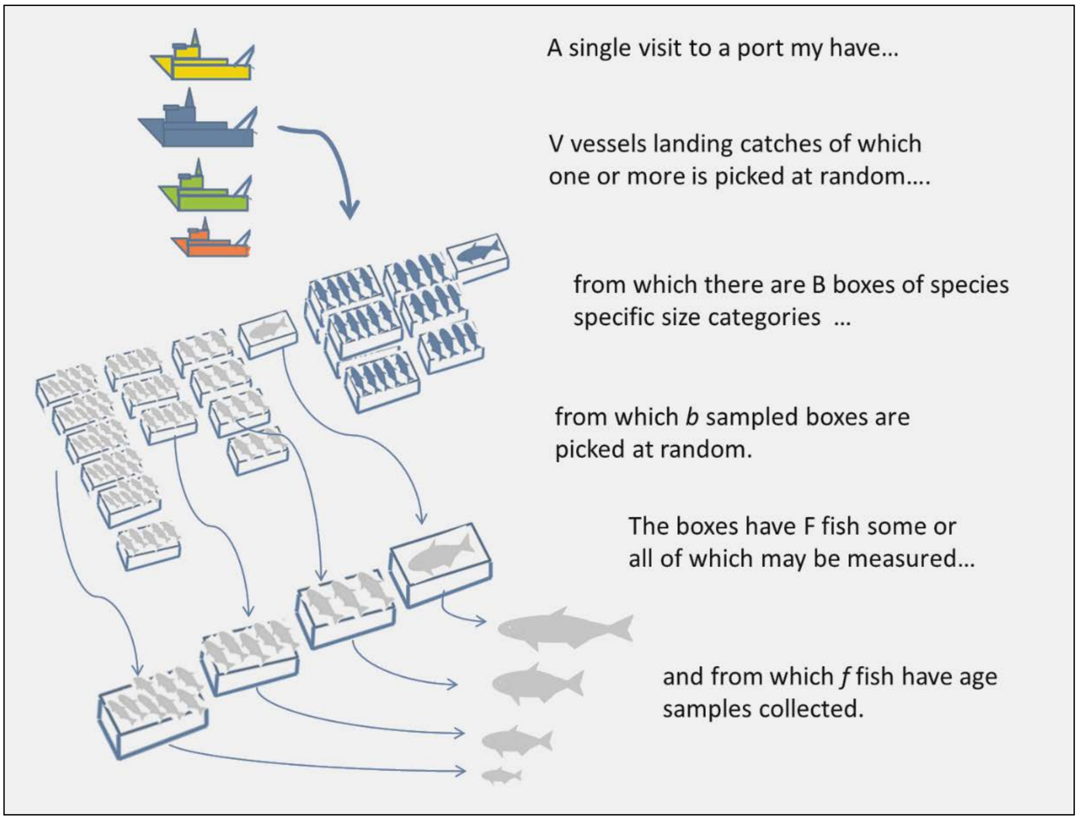


Fig 2.6.1.3.. The hierachical cluster sampling for the collection of an age samples from a visit to a port. As with at-sea sampling the sampling probabilities at each stage determine the overall sample wieight.

Estimation procedures

Unbiased estimates of numbers-at-age, or length, requires that each sample has its particular sample weight. In multistage surveys, which results in hierarchical clustered data, a base weight is applied to each sampled unit to correct for their unequal probabilities of selection. In general, the base weight is the inverse of the selection probability for the sampling unit. Sampling weights are also used to correct for non-coverage of the target population, non –response, and other factors that can cause bias.

The critical point here is that the context in which the sample was collected is required if unbiased estimates are to be derived from the sample data. This fact seems to have been largely overlooked under the quota sampling promoted under the DCF where there was an expectation that task sharing could occur which involved simply transferring samples between different national schemes, irrespective of the context in which the sampling took place. Task sharing of this type has no statistical validity and will result in biased estimates unless the fleets from the different countries involved are truly homogeneous in terms of vessels, gears, operations and area of fishing.

Estimates that are required at the regional, national, or stratum level are obtained by application of sample weights (based on the selection probabilities) to the sample data at each hierarchical sampling stage within strata then combining over strata. If estimates are required for métiers or other domains of interest, then a model-based estimation technique such as post-stratification (Holt and Smith 1979) is commonly used. Post-stratification will generally be the preferred method for obtaining estimates for domains such as a stock, grouped métiers, temporal periods, or statistical areas. Data for such domains (estimation-cells) are usually obtained by identifying the sampled units within each stratum that correspond to the domains, and determining the appropriate base weights for these Post-stratification is also used to improve efficiency of estimators. Improved precision in estimates for the target population may be achieved by adjusting the base weights so that the "sizes" of strata or estimation cells (domains) are equal to known population totals. It needs to be emphasized that, while the domain is quite distinct from the sampling stratum (being a component of the randomly sampled units and the remaining non-sampled units in a stratum), the estimation procedure for the domain must respect the sampling hierarchy in the overall sampling design.

It is also worth emphasizing that one of the particular advantages of design-based sampling is the greater utility of data obtained in this way. Provided the number of samples is sufficient, estimates for different domains can often obtained after the data are collected, using post-stratification. Hence the individual sample can be used to contribute to numerous estimates for various domains. Moreover new domains can be defined and estimates generated, the only provisos being that; (a) it is possible to ascribe a sample to a particular domain; and (b) that there are sufficient samples for that domain to generate a statistically sound estimate with sufficient precision to support the use of the estimate. The latter requirement is critical, and a common problem in relation to end-user perceptions of the resolution of estimates that can be realistically delivered by a sampling programme.

When estimates of means, rations, totals, and other population parameters are generated through design-based sampling (unlike those obtained by quota sampling), then the precision of estimates including confidence intervals can be estimated based on measures of variability between samples (e.g. standard errors, or relative standard errors). These measures of variability will ideally be calculated by bootstrap re-sampling of PSUs (assuming sufficient number of sampled PSUs per stratum), where sampling weights for domains are recalculated for each set of bootstrap samples. Also there is considerable scope with design-based sampling schemes to conduct sensitivity analyses, in which PSUs are resampled or subsampled to consider the cost–benefit tradeoff of changing the sample sizes within and across clusters (e.g. sampling rates within a haul, trip or stratum). Stock assessment expert groups operate at the regional level at which stocks exist. Hence, estimates of precision, and analysis to support the optimization of sample surveys, are needed at this scale, which in the great majority of cases involves optimization of sampling across two or more countries.

2.6.2 The Data Exchange Format used by the Regional Database (RDB)

The ability of the RDB to store national sampling data, generate regional estimates, precision levels and conduct sensitivity analysis is predicated on having a data exchange format with appropriate fields to accommodate all the necessary data recorded and enable aggregation of data at the required resolution.

The data format for the commercial catch data collected by national fisheries institutes is the csData format. Landings and effort data derived from logbook and sales slip information and collected under the control regulation is recorded in the clData and ceData formats. These data exchange formats in the RDB inherits from FishFrame and they are also the format used for the R software developed under the COST project, (ICES 2010 c). Full specifications for the Exchage format can be found in Jansen et al (2008; 2009).

The structure of the csData relates to a single fishing trip and consists of five linked hierarchical tables (data frames in R) within which are stored data on the vessel and trip (tr), the hauls (hh), the species (sl), the length frequencies (hl) and the biological variables (some or all of age, sex, maturity and length) in the ca table. The ca table is linked directly to the tr table or the hh table though it is possible to hold hh, sl and hl records for which there is no corresponding age data within the ca table.

The clData structure is a single table (data frame in R) with fields for various spatial (e.g. ICES rectangles) and temporal data (down to month) and for variables relating to the fishing operation such as vessel details, landing harbour, species-specific landed weights and values, and métier.

The ceData structure is likewise a single table with fields for similar spatial and temporal variables, and various measures of effort, such as days at sea and KW days which are related to métier. Crucially there is no link to the landed species, or any means of aggregating such data over individual trips.

The clData and ceData records are designed to hold aggregated data; thus the highest resolution in the clData format would be aggregated at the species level by gear type (level 6 métier), statistical rectangle, month, and vessel length class. Similarly ce effort data records, at the highest resolution, would be aggregated by gear type (level 6 métier), statistical rectangle, month, and vessel length class.

Limitations to the RDB operation imposed by the current data exchange format

Basically there are two principle obstacles to design based estimation imposed by the present data exchange format:

- Within the csData there are fields missing that are needed to record sampling data as it is gathered. e.g. units sampled, total units, total numbers of fish, market portions, are some of those identified at the RDB 3 workshop (ICES 2012d), SGPIDS 2013 (ICES 2013b), RCM NS&EA 2013 and RCM NA 2013.
- This arises because the range of sampling situations, and consequently the necessary protocols needed to obtain commercial catch samples, is far wider at the regional level than within the national situation within which FishFrame was originally developed. This is not to overlook the considerable strides that the adoption of a regional database, with a common data structure, has promoted in the harmonization of sampling practices.

The cl and ce data formats were developed to reflect a time, space and technical quota sampling stratification based on métier as promoted under the original DCR. They are not well suited to probability-based stratified sampling designs based on sampling frames of ports or vessel lists. The resolution at which the data can be recorded is, in critical ways, inappropriate and the ability to aggregate data over vessel groups and port groups is problematic. The formats for cl and ce have not received the same level of attention as the csData because broad adoption of revised estimation methodology has not yet occurred, however it is already apparent that some new fields are needed (for example sale location and country of first sale).

To conclude, refinement of the data exchange format is a necessary prerequisite for the RDB to develop its full potential as a means of fostering regional cooperation, implementing design based sampling, being able to generate unbiased estimates with appropriate precision, and in facilitating improvements in the data quality and the production of data quality indicators.

3 Glossary[1]

ACCURACY: In the WKPICS reports we follow Jessen (1978) and define accuracy as a measure of closeness to the targeted (true) population value. i.e. high accuracy signifies high precision and low bias.

BIAS: An estimator is regarded as biased if its expected value $\overline{y}$ estimated from repeated random samples differs from the true value (Jessen 1978).

CLUSTER: In applied survey research, the primary sampling units (PSUs) can sometimes be clustered. Given this, in a statistical population a cluster defines a "natural" sampling unit as a homogeneous grouping of PSUs (elements, observations). Such clusters may be nested within strata. In fishery science, typical examples of clusters are vessels as natural groupings of trips, or trips as a natural grouping of fishing operations. Clusters play a specific role in cluster sampling or in more complex multistage sampling techniques which may sample in subsequent steps from a cascading structure of clusters.

CLUSTER SAMPLING: In accordance with the definition of clusters, cluster sampling is a technique of sampling which involves dividing a statistical population into "natural" homogeneous groupings (or clusters). In a first step, this technique selects a simple random sample of clusters in a statistical population. In a second step, the required information is then collected from a simple random sample of the PSUs contained in each of the selected clusters. The difference between cluster sampling and stratified sampling is that in cluster sampling the cluster is treated as the sampling unit where the analysis is done on a population of clusters. Furthermore, in cluster sampling only the selected clusters are studied whilst in stratified sampling all strata are sampled. A common motivation for cluster sampling is to reduce the sampling effort in order to save money and/or time which results in an increased efficiency and contrasts with stratified sampling where the main objective is to increase precision. For more complex forms of cluster sampling information see MULTISTAGE SAMPLING.

DOMAIN: In many surveys, estimates are desired for a number of classes of the target population. Such subpopulations of special interest have been given the name domain of study by the U.N. Subcommision of sampling in 1950 (Cochran 1977, p. 34). Data to identify a domain of interest such as a métier are typically obtained after samples have been collected. This is often done by cross-classifying the sample data from PSUs by one or more predictors such as gear, target species, statistical area, and quarter. For domains that are not strata, sample sizes cannot be specified in advance. Special analysis techniques of domain estimation will be required (e.g. Lumley 2010, p. 32).

FLEET: A physical group of vessels sharing similar characteristics in terms of technical features and/or major activity.

FISHERY: A group of vessel voyages targeting the same (assemblage of) species and/or stocks, using similar gear, during the same period of the year and within the same area.

FLEET SEGMENT: a group of vessels with the same length class (LOA) and predominant fishing gear during the year, e.g. according to the Appendix III of the EU-

[1] Many of the statistical terms are based on Elsevier's Dictionary of Biometry

DCF. Vessels may have different fishing activities during the reference period, but are classified in only one fleet segment.

MÉTIER: 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 characterized by a similar exploitation pattern. The catches for such a sub-population of fishing operations in a fishery (domain) cannot generally be sampled with known probability since a list of PSUs is not available in advance. Estimates of catch characteristic for a métier (domain) are therefore often based on stratification after selection of PSUs (post-stratification.) EU Commission Decision 2008/949/EC (DCF) provides detailed requirements for Member States to collect economic data by fleet segment, and biological data by fleet métier.

MULTISTAGE SAMPLING: A sampling method in which the target population (e.g. total fleet-wide national landing) in principle first is divided into a number of groups or primary units from which samples can be drawn (sampling stage 1). Examples of first stage units (primary sampling units) are trips, vessels, and port-days. The fish in each primary sampling unit in the sample can then be subsampled in the secondary stages, etc..

POST STRATIFICATION: Post-stratification is a model-assisted estimation method used to improve efficiency of estimators, and to provide estimates for domains. Survey weights are adjusted so that the number of sampling units or size of estimation cells (post-strata or domains) is equal to known population totals. The resulting weights are then used in forming estimates of means, totals, or ratios of variables collected in the survey for the target population, or for domains. For example, in a sample survey of commercial landings the estimation cells may be a stock, gear, statistical area, or quarter, and the number of trips, or the total landings in each cell used to adjust sample weights may come from the most recent population census (trip-ticket data or logbooks). See Holt and Smith (1979) for a nice description of the method and the rationale for its use.

PRECISION: In the WKPICS reports we follow Jessen (1978) and refer to precision as a measure of reliability, or closeness of each observation to its own average over repeated trials. Precision is often measured by the standard error or the relative standard error (RSE).

PRIMARY SAMPLING UNIT: A sampling unit in the first stage in multistage sampling is called a primary sampling unit. Primary sampling units in the most common catch sampling schemes can be trips, vessels, or site-days.

RAISING: In the ICES community, the term 'raising' refers to the use of weighted estimators, for example where estimates from a weighted ratio estimator (e.g. mean discard or ratio of discard to total catch) based on a sample is expanded to the total target population (e.g. fleetwide catch).

RELATIVE STANDARD ERROR (RSE): We follow Jessen (1978) and define the relative standard error of some estimator, say $\overline{y}$, as a measure of precision where $RSE = SE(\overline{y}) / \overline{y}$.

SAMPLE DESIGN: The totality of instructions, protocols, and rules that govern a sampling method.

SAMPLING FRAME: In statistics, a sampling frame is the list of sampling units or device from which a sample is drawn. The sampling frame comprises all the primary sampling units and any stratification of these, and may be based on a vessel registry or list of ports.

SAMPLING UNIT: In order to take a sample from a population, the target population must consist of, or be divided into non-overlapping parts (units). Sampling can then be conducted by selecting units according to a defined sampling scheme. These units are called sampling units in the survey sampling literature. The units that can be selected in catch sampling schemes are typically groups (clusters) of fish, such as the cluster of fish in a landing from a fishing trip, or cluster of fish caught in a fishing operation. |

STRATIFICATION: The advance decomposition of a **finite population** of sampling units of size N into k non-overlapping subpopulations (strata) of size N_i.

STRATIFICATION AFTER SELECTION: If a simple random sample is taken from a **finite population** of sampling units of size N the sample may be treated as a stratified sample during the analysis if the post-strata sizes N_i are known. Stratification after selection (post-stratification) is usually applied if the strata to which the selected sampling units belong are only known after the sample is taken. This is often the case for métiers. Standard stratified estimators cannot generally be applied when métiers cuts across strata.

WEIGHTS: Sampling weights in design-based estimation are the inverse of the likelihood of being sampled (inclusion probabilities). A weighting adjustment technique is sometimes used in domain estimation (e.g. applying post-stratification) when proper auxiliary variables are available. Such variables must have been measured in the survey, and the population values must be available. Typical auxiliary variables used in catch-sampling programs are total catch or total number of trips by stratum or métier. The population distribution of such variables can usually be obtained from national statistical institutes. Global Adult Tobacco Survey (GATS) (2010) provide nice examples that illustrate the principle of weighted estimators and sample weights.

4 References

Aanes, S. and Vølstad, J.H. Efficient statistical estimators and sampling strategies for estimating the age-composition of fish (In review).

Aanes, S. and J.H. Vølstad. 2013. Evaluating sampling strategies and comparing statistical estimators with the use of age-length keys to estimate age-distributions of Northeast Arctic Cod. ICES 2013 ASC/Session J, J:08.

Aanes, S., and Pennington, M. 2003. On estimating the age-composition of the commercial catch of Northeast Arctic cod from a sample of clusters. ICES J. Mar. Sci. 60(2): 297–303. doi:10.1016/S1054-3139(03)00008-0.

Anon. 2013a Report of the Regional Coordination Meeting for the North Sea & Eastern Arctic. Vigo, September 2013

Anon. 2013 b Report of Regional Coordination Meeting for the North Atlantic. Sukarrieta, September 2013

Berg, C.W. and K. Kristensen. 2012. Spatial age-length key modelling using continuation ratio logits. Fish. Res. 129–130: Pages 119–126

Cochran 1977. Sampling Techniques. John Wiley & Sons.

EU 2009. Common tool for raising and estimating properties of statistical estimates derived from the Data Collection Regulation. Project no SI2.467814.

EuroStat. 2000. Assessment of the quality in statistics. Eurostat/A4/Quality/00/General Standard Report, April 4-5, Luxembourg.Global Adult Tobacco Survey (GATS). 2010. Sample Weights Manual, Version 2.0. Atlanta, GA: Centers for Disease Control and Prevention.

Global Adult Tobacco Survey (GATS) (2010). http://www.ncbi.nlm.nih.gov/pubmed/23311796 and other links

Gregoire, T.G. 1998. Design-based and model-based inference in survey sampling: appreciating the difference. Can. J. For. Res. 28: 149-1447.

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.

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

Hirst, D.J., Storvik, G., Aldrin,M., Aanes, S. and Huseby, R.B 2005 Estimating catch-at-age by combining data from different sources. Can. J. Fish. Aquat. Sci. 62: 1377–1385.

Hirst, D.J. , G. Storvik, H. Rognebakke, M. Aldrin, S. Aanes and J.H. Vølstad. 2012. A Bayesian modelling framework for the estimation of catch-at-age of commercially harvested fish species. Can. J. Fish. Aquat. 69: 1–13. doi:10.1139/CJFAS-2012-0075

Hoenig, J.M. and D.M. Heisey. 1987. Use of a Log-Linear Model with the EM Algorithm to Correct Estimates of Stock Composition and to Convert Length to Age. Trans. Amer. Fish. Soc. 116:232-243.

Hoenig, J.M. R. Choudary Hanamura, and D.M. Heisey. 2002. Double and Triple Sampling for Repeated Surveys and Partial Verification. Biometrical Journal 44 (5), 603-618

ICES 2007a. Report of the Workshop on SexualMaturity Sampling (WKMAT). ICES CM 2007/ACFM:03

ICES. 2007b. Report of the Workshop on Discard Raising Procedures (WKDRP). ICES CM 2007/ACFM:06, 55pp. 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.

ICES 2008a. Report of the Workshop on Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment (WKACCU). ICES CM 2008/ACOM:32, 41 pp.

ICES 2008b. Report of the Workshop on Maturity Ogive Estimationfor Stock Assessment (WKMOG). ICES CM 2008/ACOM:33

ICES, 2009a. Report of the third Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes, 19-22 November 2013, ICES HQ, Copenhagen, Denmark. ICES CM2013/ACOM:54. 114 pp.

ICES 2009b. Report of the workshop on sampling methods for recreational fisheries surveys (WKSMRF). ICES CM 2009/ACOM:41

ICES 2010a. Report of the ICES Workshop on methods for merging métiers for fishery-based sampling (WKMERGE). ICES CM 2010 / ACOM:40

ICES 2010b. Report of the Planning Group on Recreational Fisheries Surveys (PGRFS), 2–6 May 2011, Esporles, Spain. ICES CM 2010/ACOM:34.

ICES 2010c. Report of the Workshop on implementation of the Common Open Source Tool (COST), ICES CM 2010/ACOM:42, 20 pp.

ICES 2011a. Report of the Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes, 8-10 November 2011, Bilbao, Spain. ICES CM 2011/ACOM:52

ICES. 2011b. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS). ICES 2011 /ACOM:50.

ICES 2011c. Report of the Planning Group on Recreational Fisheries Surveys (PGRFS), 2–6 May 2011, Esporles, Spain. ICES CM 2011/ACOM:23, 111 pp.

ICES 2012a. Report of the Second Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes, 6 – 9 November, Copenhagen. ICES CM 2012/ACOM:54

ICES 2012b. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS). ICES 2012 /ACOM:50

ICES 2012c. Report of the Working Group on Recreational Fisheries (WGRFS). ICES CM 2012/ACOM:23.

ICES 2012 (d) Report of RDB workshop 3. Copenhagen November 2012

ICES 2013a. Report of the Planning Group on Commercial Catches, Discards and Biological Sampling; PGCCDBS 2013. ICES CM 2013/ACOM: 49.

ICES 2013 b. Report of the Working Group on Recreational Fisheries 2013 (WGRFS). ICES CM 2013/ACOM 23.

ICES 2013c. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS). 24-28 June 2013, Lysekil, Sweden. ICES 2013 /ACOM:56

Jansen, T,. Degal, H., Hakansson, K., Egekvist, Dalskov, J. and Koster, F. W. (2008). FishFrame 5.0: a web based data warehouse application for management access, and integration of fisheries and stock assessment data. ICES CM 2008/R:26

Jansen, T. (Ed). 2009. Definition of Standard Data Exchange Format for Sampling,Landings, and Effort Data from Commercial Fisheries ICES Cooperative Research Report No. 296 . 43pp.

Jessen, R.J. 1978. Statistical Survey Techniques. John Wiley & Sons. New York. 520 pp.

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

Kimura, D.K. 1977.Statistical assessment of the age- length key. J. Fish. Res. Board Can. 34: 317-324.

Kimura, D.K. and Chikuni, S., 1987: Mixtures of Empirical Distributions: an Iterative Application of theAge-Length Key. Biometrics 43, 23-35.

Kish, L. (2003). Selected papers [Compiled and edited by Steven Heeringa and Graham Kalton]. John Wiley & Sons, Hoboken. 356 pp. .

Korn E. and Graubard B. 1999. Analysis of health surveys. John Wiley & Sons.

Lehtonen, R. And E.J. Pahkinen. 1996. Practical Methods for Design and Analysis of Complex Surveys. 2nd Edition. John Wiley & Sons. 344 pp.

Lohr, S.L. 2010, Sampling: Design and Analysis. 2nd Edition. Brooks/Cole CENGAGE Learning. ISBN-13: 978-0-495-11084-2; ISBN-10:0-495-11084-1

Lumley, T. 2010. Compex Surveys. A Guide to Analysis Using R. John Wiley & Sons. Hoboken. 276 pp.

Nedreaas, K., Stransky, C., Jardim, E. and Vigneau, J. 2009. Quality assurance framework—the concept of quality assurance applied to fisheries data and its operationalization under the ICES scope. ICES CM 2009/N:06.Westrheim, S. J., and W. E. Ricker. 1978. Bias in using an age-length key to estimate age-frequency distributions. J. Fish. Res. Board Can. 35: 184-189.

Quinn, T.J., and Deriso, R.B. 1999. Quantitative fish dynamics. Oxford University Press, New York.

Shepherd, J. G. 1999. Extended survivors analysis: An improved method for the analysis of catch-at-age data and abundance indices. – ICES Journal of Marine Science, 56: 584–591.

Særndal. 19878: Design-based and model-based inference in survey sampling [with discussion and reply]. Scan. J. Stat. 5: 27-52.

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

Thompson 1992. Sampling. John Wiley & SonsAanes, S., and Pennington, M. 2003. On estimating the age composition of the commercial catch of Northeast Arctic cod from a sample of clusters. ICES J. Mar. Sci. 60(2): 297–303. doi:10.1016/S1054-3139(03)00008-0.

Annex 1: List of participants

Full name	E-mail
Sondre Aanes	sondre.aanes@imr.no
Maciej Adamowicz	madamowicz@mir.gdynia.pl
Mike Armstrong	mike.armstrong@cefas.co.uk
Casper Willestofte Berg	cbe@aqua.dtu.dk
Patrik Börjesson	patrik.borjesson@slu.se
Sofia Carlshamre	sofia.carlshamre@slu.se
Chun Chen	chun.chen@wur.nl
Liz Clarke	e.d.Clarke@marlab.ac.uk
Jon Elson	jon.elson@cefas.co.uk
Ana Claudia Fernandes	acfernandes@ipma.pt
Hans Gerritsen	hans.gerritsen@marine.ie
Joachim Gröger	joachim.groeger@ti.bund.de
Ryszard Grzebielec	rysiek@mir.gdynia.pl
Beatriz Guijarro	beatriz.guijarro@ba.ieo.es
Kirsten Birch Håkansson	kih@aqua.dtu.dk
Edwin van Helmond	Edwin.vanHelmond@wur.nl
Thomas Hervé	homas.herve@ifremer.fr
Uwe Krumme	uwe.krumme@ti.bund.de
Estanis Mugerza	emugerza@azti.es
Nuno Prista	nmprista@ipma.pt
Ana Maria Aranha Ribeiro Santos	ana.ribeirosantos.cefas.co.uk
Katja Ringdahl	katja.ringdahl@slu.se
Jose Rodriguez	jose.rodriguez@st.ieo.es
Marijus Spegys	marijus.spegys@gmail.com
Romas Statkus	statrom@gmail.com
Marie Storr-Paulsen	msp@aqua.dtu.dk
Agurtzane Urtizberea	aurtizberea@azti.es

Didzis Ustups	Didzis.Ustups@bior.gov.lv
Jon Helge Vølstad	jon.helge.voelstad@imr.no
Susan Wigley	susan.wigley@noaa.gov

Annex 2: Agenda

Tuesday November 19th

9:00 – 09:30 Introduction and Workshop Objectives

Rest of morning: Plenary Presentations

Lunch 12:30-13:30

Afternoon: Subgroups to review any work done in advance of meeting, develop a work schedule for the meeting, allocate tasks and start work:

Wednesday, November 20

09:00 – 10:30: Plenary: Subgroups to summarize approaches being adopted; resolve any difficulties; and any initial outcomes.

10:30 onwards: Continue with subgroup work

Lunch 12:30-13:30

13:30 – 18:00:

- Subgroups– continue work and start report drafting

Thursday, November 21

9:00 – 12:30. Subgroups– presentation and discussion of work completed

Lunch 12:30-13:30

13:30 – 16:00: Subgroups–report drafting

16:00 – 18:00:` Plenary: brief update by Subgroups on progress; review of text already drafted

Friday, November 22

9:00 – 12:30.

- Review text drafted by subgroups (continued).
- Agree deadlines and responsibilities for any post – WK work

General meeting closes at 12:30

Annex 3: Report to European Commission: “Data quality indicators for biological data as input to discussions on revision of the DCF”.

This report was prepared for the European Commission by Dr M. Armstrong, as an independent expert, with subsequent input from the third ICES Workshop on Practical Implementation of Statistical Sound Catch Sampling Programs (WKPICS3), 19-22 November 2013, Copenhagen. The author accepts all responsibility for the views contained within the report, which may not reflect the views of all WKPICS3 members or of the European Commission. Report prepared 24 November 2013

Summary

This report was requested by the European Commission to review possible data quality indicators for biological data as input to STECF EWG 13-18 discussions on revision of the DCF. The requirement was to provide a review covering the following three topics related to fleet-based and stock-based biological sampling from marine fisheries:

1. Overview and compilation of discussions that have already taken place on quality indicators/measures of quality for DCF biological data (including in the context of ICES, STECF and other appropriate EU fora).
2. Overview of approaches used in other important fishing nations (e.g. USA, Canada, New Zealand, Australia, Norway...) to measure and ensure quality of biological data.
3. A reflection on whether MS following best practice guidance for data collection is sufficient or whether measures of quality of the collected data are necessary in addition. Possible measure of quality/quality indicators that could be used and pros/cons of these.

This review proposes that the present system of reporting data quality in DCF programmes is inappropriate. Experience in other countries is that quality evaluation should be through a well-structured peer-review process supported by clear documentation of all components of the sampling programmes and the sampling outcomes. This type of review is a complex process that may be carried out in stages within Institutes and through external peer review, and requires appropriate experts in statistical survey design and practical implementation.

Quality of a sampling survey programme should be evaluated in relation to two aspects of sampling: 1) the ability of the programme to (in principle) deliver data that are fit for purpose, by reviewing the design of the programme against guidelines and standards for best practice; and 2) evaluation of the quality of the data following implementation of the sampling survey, covering each of the two components of accuracy: bias and precision.

Some specific Quality Indicators for each of these aspects are discussed. These relate to i) design of the sampling programme (e.g. coverage of the sampling frame), ii) bias arising during implementation (e.g. non-response rates; proportion of total landings in strata with missing samples), and iii) indicators related to precision (e.g. relative standard error - referred to in DCF texts as CV; effective sample sizes (ESS); numbers of primary sampling units sampled). Quality indicators should be examined in the context of a broader review of a sampling programme, as on their own they may be uninformative or even misleading, and should be clearly distinct from indicators of compliance to DCF legal requirements. Quality standards for fishery sampling are as yet still in development and are incomplete.

For a well-designed, probability-based sampling survey, the detailed outcomes will reflect the sampling intensity and coverage as well as factors beyond the control of the samplers, including changes in abundance of fish stocks and in fishing activities, gears and non-response rates. The outcomes should feed back into improvements in design.

Sampling programmes should be designed in consultation with end-users, particularly at a regional scale, so that the level of disaggregation of estimates that can be supported is clearly understood, and the cost of acquiring more detailed estimates can be considered.

Overview and compilation of discussions that have already taken place on quality indicators/measures of quality for DCF biological data (including in the context of ICES, STECF and other appropriate EU fora).

Sampling surveys are widely used to collect information in all walks of life, and there is a large body of literature dedicated to the design and interpretation of such surveys, and evaluation of their quality (Cochran, 1977; Lessler and Kalsbeek, 1992; Levy and Lemeshow, 1999; Lohr, 2010.). Lohr (2010) concludes that the definition of survey quality as "fitness for use" recognizes the multiple purposes of survey data, and when referring to seven dimensions of quality given by EuroStat (2000) also concludes that data accuracy is the most important aspect of data quality. The two key components of accuracy – bias and precision -are examined in Appendix 1 with reference to the outcomes of two ICES workshops (WKACCU – ICES 2008a and WKPRECISE – ICES 2009a). See also Lessler and Kalsbeek (1992) for thorough coverage of systematic (nonsampling) errors in sample surveys.

Accuracy of data are not always clearly and objectively considered when compiling data for stock assessments carried out by ICES, partly because the accuracy has not been formally evaluated, and no, or limited, indicators of quality are supplied. Problems with the fit of an assessment model cannot be traced back to individual datasets when good-quality and poor-quality data such as catches-at-age are combined across countries. Decisions on whether to include, exclude or down-weight particular datasets cannot be made in an informed way.

Discussions within the ICES community on data quality

A substantial investigation into the quality of fisheries sampling programmes, data and associated analysis has been conducted by the ICES Planning Group on Commercial Catches, Discards and Biological Sampling (PGCCDBS), in their role to promote the ICES Quality Assurance Framework (Nedreaas *et al* 2009), and by workshops and study groups established by PGCCDBS. In addition to establishing protocols and standards for fish ageing and maturity determination, the PGCCDBS and its workshops and study groups have covered topics such as sampling and estimation for maturity ogives (WKMAT: ICES 2007a; WKMOG: ICES 2008b), accuracy of sampling data (WKPRECISE: ICES 2009a; WKACCU: ICES 2008a), discard raising procedures (WKDRP: ICES 2007b); design of commercial fishery sampling schemes (WKMERGE: ICES 2010b; WKPICS: ICES 2011a, 2012c; SGPIDS: ICES 2011b, 2012a, 2013b) and recreational fishery surveys (WKSMRF: ICES 2009b; WGRFS: ICES 2012b, 2013a).

These ICES initiatives have had a progressive impact since the late 2000s in increasing the awareness within the ICES community of the need for statistically sound sampling design rather than ad-hoc methods, and have developed an important and well-documented body of knowledge of fishery sampling design, implementation and analysis. An important component of this has been the development of guidelines for best practice as well as proposals for ways in which the quality of sampling programmes and the data gathered from them can be documented for a range of end-users such as stock assessment scientists, regional coordination groups and the European Commission.

An overview and compilation of discussions that have already taken place on quality indicators or other measures of quality for DCF biological data within the ICES community since 2007 is given in Appendix 2. Some important conclusions are given below:

- Data quality evaluation is a complex process as it encompasses the statistical soundness of the sampling design, the outcomes of implementing the scheme, how the data are managed, and how the data are analysed. Aspects of quality related to bias and precision need to be considered separately. Several ICES groups dealing with commercial and recreational fisheries sampling have devoted considerable efforts to designing reporting systems that can identify quality issues at all the stages from design to analysis, as would be required for a full audit of survey design and data quality. These can show, through suitable diagnostics, how quality problems propagate from national sampling strata through to final combined international data, so that sampling can be improved in a well targeted way. National sampling schemes need not have identical design, if they follow best practice standards and have correctly calculated, representative estimates with associated variance.
- Data end-users must not expect estimates at a higher level of disaggregation than the survey was designed for. A recurrent example is the unrealistic expectations to post-stratify fishery sampling data into highly resolved fleet métiers, when the inevitable outcome is many métiers having no or very few samples. It is essential that end-users work with survey experts to ensure that surveys and end-user needs are properly aligned at a national and regional scale
- Precision for a given survey design is an outcome related to sampling effort. Sampling schemes should be designed to deliver the desired precision at the scale of aggregation needed by end-users – e.g. for catches-at-age for a stock, it is the precision of the combined international estimates. The process of coordination of sampling between countries should identify the sampling needed at a national scale to deliver the desired precision for combined international data. Sampling programmes should then evolve in response to achieved precision relative to the desired precision.
- Estimation of precision is only meaningful if sampling has been designed around the basic principles of random sampling. It follows that the primary requirement is to adopt good practice in designing a sampling scheme so that biases are minimized, and to have procedures for evaluating any biases that may arise during the implementation phase.
- Assuming that a statistically sound sampling scheme is in place, the calculation of precision should take into account the sampling design and any cluster sampling effects which are common in fisheries sampling.
- A key to effective quality evaluation is full and accurate documentation of national sampling programmes. It is vital that such documentation is stored in repositories providing easy access to all users who need them.

Discussions within STECF on DCF and DCMAP data quality indicators

The Scientific, Technical and Economic Committee on Fisheries (STECF) has established a series of expert working groups (EWG) over the lifespan of the Data Collection Regulation and the Data Collection Framework to advise the Commission on the content of these regulations and to review the achievements of Member States against their national obligations as laid out in the relevant Commission Decisions. During 2012 and 2013, the attention of STECF has been focused on the structure, content and operation of the new Multiannual Programme for Data Collection (DCMAP), including how the quality of data can be enhanced through a revised Regional Coordination process. Of particular interest is how the quality and performance of national sampling programmes can be evaluated, and what types of quality indicator (QI) should be defined as part of this purpose.

An overview of the recent STECF discussions on DCMAP and quality indicators is given in Appendix 2. Some important conclusions are given below:

- STECF fully acknowledges and agrees with ICES proposals demonstrating the need for statistically sound fishery sampling programmes, and for collaboration within regions to ensure that these principles are pursued within a regional sampling programme driven by end-user needs.
- A key aspect of quality evaluation is adherence to best practice guidelines, which implies the need for guidelines and standards and appropriate documentation of national sampling schemes to allow evaluation against these standards. In general, STECF EWGs have highlighted a need for two components of quality evaluation: design vs best practice, and quality indicators (e.g. as listed by EuroStat (2000)) to demonstrate the quality of supplied data.
- The DCMAP should not contain prescriptive precision targets such as target CV values, as have previously been included in the DCR and DCF, but it is important that the precision of estimates needed by end-users can be evaluated.

Conclusions from review of ICES and STECF discussions on data quality evaluation

The conclusions of ICES and STECF discussions on reporting of quality of fleet-based and stock-based biological data in the DCF can be distilled down to two core elements:

i) An evaluation of whether national sampling programmes are designed and implemented, and the data managed and processed, in a way that follows agreed sets of standards. A national programme meeting these standards is in principle capable of providing the desired standard for data quality.

ii) An evaluation of the quality of the data that have been collected, using diagnostics and quality indicators that identify potential (or known) bias, and those that provide estimates or indices of achieved precision. ICES groups such as WKPICS, SGPIDS and WGRFS have proposed that this should be an evaluation of national contributions as part of a regional sampling programme, because quality indicators for national programmes are of limited value in isolation as you cannot easily see how they impact the estimates at a regional or stock scale, or how they can be optimized to improve data quality for stocks or regional fleets. The Regional Data Bases are seen as a work in progress towards facilitating regional data quality evaluation.

Overview of approaches used in other important fishing nations (e.g. USA, Canada, New Zealand, Australia, Norway...) to measure and ensure quality of biological data.

In response to a query circulated to contacts in fisheries laboratories in the USA, Canada, New Zealand and Australia, detailed responses were obtained from the USA, New Zealand and Norway. Detailed responses are given in Appendix 3, and a summary of key points related to data quality evaluation is given below. When viewing these responses, consideration must be given to factors such as the extent to which there are shared stocks with other countries, or to the existence of legal requirements for peer review of data collection and assessments (as in the USA) which are not applicable to the EU.

Data quality assurance in New Zealand

New Zealand has developed its fisheries data quality evaluation further than Europe by having published a "Research and Science Information Standard for New Zealand Fisheries". Key elements are that:

- Data must be collected according to documented procedures, and in a manner that reflects standard best practices generally accepted by the relevant science and technical communities. Data and information sources must be identified or made available upon request.
- Data collection methods, systems, instruments and statistical sampling designs must be designed to meet the requirements and objectives of the research projects concerned, and should be validated before use. Instruments must be calibrated using applicable standards or fundamental engineering and scientific methods.
- Data must undergo internal or external quality assurance prior to being used in analyses that are intended or likely to inform fisheries management decisions.

There is emphasis on the need for independent peer review to ensure the relevance, integrity, objectivity and reliability of information, and the science quality assurance and peer review processes are required to use a quality ranking system with four categories: 1 = High Quality (which should essentially be anything that is good enough to be used in an assessment or to inform management decisions in other ways); 2 = Medium or Mixed Quality (data that might be used but would have many associated caveats); 3 = Low Quality (data that should not be used at all because it is not reliable and may produce misleading results); U = Unranked (has not been reviewed – and therefore should be used with caution if at all).One of the key purposes of the science information quality ranking system is to inform fisheries managers and stakeholders of those datasets, analyses or models that are of such poor quality that they should not be used to make fisheries management decisions (i.e. those ranked as "3" or "U"). The NZ Science Working group processes involve "staged technical guidance" on data quality, for example evaluating a survey design, evaluating the preliminary analyses, suggesting sensitivity analyses, and ensuring that the conclusions are justified by the data and analyses.

Data quality assurance in the USA

In the USA, there is national coordination of the NOAA Fisheries activities regarding implementation of the Data Quality Act. Activities to strengthen the integrity of scientific information include science program reviews of the NOAA Fisheries science centers and the scientific peer review process. External peer reviews are also conducted through the Center for Independent Experts (CIE). As part of the Northeast Regional Stock Assessment Workshop (SAW) process, SAW working group members routinely review and evaluate data inputs used in stock assessments. Major independent peer reviews of sampling surveys include the recent National Academies review of the marine recreational fisheries survey program, which led to a major revision of the program (now known as the Marine Recreational Information Program MRIP) with stronger emphasis on aspects of statistical design.

Data quality assurance in Norway

Norway has a national strategy to develop the Norsk Marint Datasenter (NMD: the Norwegian Marine Data Center), which will manage all research data from research surveys and fisheries sampling programs such as the Norwegian Reference Fleet conducted by the Institute of Marine Research, Norway, in accordance with national requirements, standards, and international agreements. As part of this development, IMR is currently refining the data handling, management and dissemination of data and data products through a large infrastructure project called Sea2Data. IMR is developing the infrastructure to facilitate easier access to data, improve the quality control and quality assurance (QA/QC) of data from their collection, through data entry, data storage, analysis, and dissemination. The relational database used for data storage, as well as modules for data analysis to provide stock assessments and other end-products is integrated in the Sea2Data framework, using open-sources programming tools and analysis packages (such as R). The QA/QC includes documentation of sampling protocols for research surveys and fisheries monitoring programs, instructions for data punching, and a range of checks to minimize data entry errors and other sources of errors.

A reflection on whether MS following best practice guidance for data collection is sufficient or whether measures of quality of the collected data are necessary in addition. Possible measure of quality/quality indicators that could be used and pros/cons of these.

Evaluation of quality against best practice guidelines

The process of evaluating the quality of fisheries data includes quality assurance, such as comparison with documented standards, monitoring of processes, and error prevention to ensure data are "fit for purpose", and quality control using systems to detect and correct errors in the data. The desired quality is determined by the end-users. There are parts in the process of data collection and processing within an Institute where formal quality accreditation through, for example, ISO 9000, may have been awarded (e.g. quality systems in a fish ageing laboratory). These provide a part of the evidence of adherence to quality standards, but for many other key aspects of fisheries sampling such as design and implementation of sampling surveys, there is no consolidated set of standards for best practice. Elements of this are contained in a diverse range of ICES reports, including the guidelines for best practice for fishery sampling surveys given by ICES WKPICS2 (ICES, 2012c) and WGRFS (ICES, 2013a), IBTS manuals, guidelines for sampling for maturity given in ICES WGMAT/WGMOG (ICES

2007a & 2008b), guidelines for discard raising given in WKDRP (ICES, 2007b) etc. Some of these need updating and expanding.

There is a clear need to develop a consolidated, updated and more complete set of guidelines for best practice, and quality standards, for fishery sampling programmes. These will help countries to develop statistically sound sampling programmes, allow the quality of those programmes to be properly evaluated, and ensure that data collected by different countries for a stock or fishery are compatible and can be combined.

A possible process of conducting such an evaluation is shown in Table 1. This is purely illustrative. Such an evaluation is a technically and statistically complex process, and can only be done through peer review by people with appropriate competences. There is strong emphasis on peer review in the USA and New Zealand, and this occurs at several stages including internal reviews within Institutes, and reviews involving external experts. Within Europe, the establishment of the ICES benchmark system involving data compilation and evaluation meetings prior to assessment meetings, involves external experts and peer review, but does not consistently or fully adopt the procedures shown in Table 1. The current process of evaluation of Member States annual DCF reports by STECF in no way constitutes a peer review as described.

Table 1. Possible elements of quality evaluation of a fishery sampling programme (illustrative)

Programme stage	Existing guidelines and standards ("best practice")	Quality evaluation procedure	Performance measures	Possible Quality Indicators
Design of sampling scheme	e.g. WKPICS & WGRFS best practice guidelines; IBTS protocols etc.	Review of documentation on sampling design relative to quality standards	Indicators of bias potential due to design.	Score against quality standards, e.g. frame coverage, sample selection procedures etc.
Implementation of sampling scheme	e.g. WKPICS & WGRFS best practice guidelines; IBTS protocols etc.	Review of sampling <u>outcomes</u> – e.g. diagnostics of coverage, refusal rates, sample numbers and precision etc.	Indicators of extent of bias (e.g. low, medium, high, unknown); Indicators of precision.	Number of primary sampling units sampled in each sampling stratum; CV; frame coverage; refusal rates.
Data archiving and extraction	To be done.	Review of documentation of QA/QC procedures relative to quality standards. e.g. use of electronic data capture; error traps; ………	Indicators of extent and effectiveness of QA/QC procedures.	Score against quality standards
Data analysis	e.g. WKPICS & WGRFS best practice guidelines; IBTS protocols; etc.	Review of documentation of estimation procedures relative to quality standards.	Indicators of extent of bias (e.g. low, medium, high, unknown)	Score against quality standards, e.g. analysis follows design

Quality indicators (QI)

The core of a quality evaluation is a peer review of all aspects of a sampling programme against documented standards, and the critical requirement is to have accurate and complete documentation of all components of the programme, including key assumptions in the processing and analysis of the data. The existence of this documentation is an important aspect of quality evaluation. A range of QIs are possible for used in the overall quality evaluation procedure, to deal with 1) aspects of bias related to design; 2) aspects of bias related to implementation, and 3) precision. Design-related indicators are a direct indicator of quality of the sampling programme, whilst implementation bias and precision are aspects of data accuracy (uncertainty). This distinction must be clear in the quality evaluation process. Quality Indicators should also be clearly distinguished from any metrics to indicate compliance with DCF legal requirements.

QI's for quality of design should relate to guidelines and standards for best practice such as those developed by WKPICS2 (Appendix 6). A simple but important QI is:

i) Coverage of the sampling frame (e.g. how much of the landed catch of each species into a country is into the ports included in an onshore sampling scheme)

This should be known at the design stage, and the potential impact evaluated then.

QI's for bias related to implementation error could include:

i) Non-response rates (e.g. refusal to allow access to vessels or catches for sampling). This also needs to be backed up with documentation of reasons, and any analysis to indicate if these vessels or sites have different characteristics and activities to those sampled.

ii) Proportion of total landings in strata with missing samples (a problem of over-stratification)

Non-response is of concern if it is suspected that the actual bias may be large. If the non-respondents have the same catch rates or catch compositions as those who provide access, there is no bias. A QI based on a figure for non response may not be sufficiently informative on its own, and should be an indicator derived as part of a specific evaluation of non response and its effects. Non-response may be impossible to control.

QIs related to precision could include

i) Relative standard error RSE (referred to in DCF texts as CV, referring to coefficient of variation of the mean)

ii) Effective sample sizes (ESS)

iii) Numbers of primary sampling units sampled, ideally by stratum

The advantage of RSE/CV values is that they are a direct measure of precision, and can easily be incorporated into statistical assessment models.

Effective sample sizes provide a meaningful index of precision, having accounted for cluster sampling effects. The alternative common practice of reporting actual numbers of fish measured or aged is highly misleading except perhaps for rare species where only one or two are present in the catch from any PSU. The downside of ESS as an indicator is that it is not widely used and would require development of skills and software in each lab to carry out the estimation.

Numbers of PSUs sampled can be considered as a proxy for ESS. It is likely to be smaller than the ESS, but much closer to ESS than to numbers of fish sampled. Simulations in Norway (Aanes and Vølstad, In review) demonstrated that ESS is closely associated with number of PSUs.

The current DCF requires MS to report data quality as achieved sample numbers vs expected sample numbers, and achieved precision estimates (CV) vs target CV values (Appendix 5). For some variables such as length-at-age or maturity-at-length or age, additional rules are specified for calculating the average CV over a range of length or age groups (see footnote 2 to Appendix 5). It is implicitly expected that MS will have calculated the CVs correctly, in accordance with a sampling design that yields meaningful CVs. Some MS will have used COST tools for this purpose (for example for estimating fleet raised discards and associated CV). Where MS are not collaborating within a region, they are expected to achieve the target CV specified in the DCF Commission Decision. If collaborating with other MS, it is the combined estimates that must meet this requirement (i.e. a lower precision is needed for each MS).

There are several major shortcomings of the current DCF reporting of precision. For example, the required precision levels are arbitrary and do not reflect any agreement or analysis of the desired precision for combined national estimates for supporting assessments or management advice, and calculated CVs only reflect the true precision if the sampling scheme has adopted a probability-based design and there are no major biases in design and implementation such as inadequate frame coverage or extensive non-response. Reporting of CVs or sample numbers for estimating growth parameters, maturity ogives or sex ratio on their own provide no information on the quality of the data. Such estimates are critically dependent on the design of the sampling to achieve unbiased data for a stock over its full range. The quality of such data and estimates can only be evaluated through expert peer review of the entirety of the sampling scheme within and between countries, and of the adherence to protocols for ageing and maturity staging. In some cases, the data are from collaborative surveys such as IBTS and the sampling achieved by individual participating countries is not very informative on its own.

Combined indicators

A possible approach to "scoring" the quality of a sampling programme is to develop some form of combined indicator. The idea behind the ICES WKACCU traffic-lights score card was to have an overall bias score based on all the component sources of bias. This has proved a difficult concept to put into operation as the biases need to be weighted somehow. Lohr (2010) presents the ideas of **total survey error** and **total survey design** proposed by other authors**:**

- **Total survey error** = coverage error + nonresponse error + measurement error + processing error + sampling error
- **Total survey design**: designing a survey to reduce errors in general, not just sampling errors. This needs an understanding of where the major error components are, so that steps can be taken to reduce them. As Lohr (2010) states, this calls for an interdisciplinary approach. For fisheries sampling this would need experts in statistical sampling design, and experts in implementation.

The idea of "data quality reports" being developed through PGCCDBS, WKPICS, SGPIDS and WGRFS are a move towards Total Survey Design and provision of diagnostics that can highlight elements of total survey error. The WGRFS 2013 proposal for a "Quality Assurance Toolkit" valuably extends this to addressing the needs of different end-users – for example the diagnostics needed to evaluate the quality of survey programmes in terms of design quality and uncertainty related to implementation, and simpler quality indicators needed by stock assessment scientists. These reports are an important concept and need to be tested and developed further.

Quality standards

Currently, documentation of quality standards is patchy and incomplete for all the stages in Table 1 relevant to DCF / DC-MAP data collection on fleet-based and stock-based biological variables. Available documentation includes:

- The "best practice guidelines" for fishery sampling schemes produced by WKPICS2 (2012c) and WGRFS (2013a), which are an important step forward but represent guidelines rather than agreed quality standards.

- Standards and protocols for age reading and maturity staging developed and documented by ICES PGCCDBS and its workshops including the workshop for national age reading chairs (WKNARC). A further workshop (WKSABCAL) is planned to improve the methods of estimating and reporting the quality of age readings.
- Standards and protocols for aspects of design and implementation of research trawl surveys given by ICES groups such as IBTSWG
- Documentation of methods of data analysis by classes of catch sampling schemes, given by WKPICS2 & 3.

Some important omissions are:

- Quality standards for data archiving and management - i.e. validation of data through quality assurance and quality control procedures to avoid or trap errors, and to identify and correct errors in databases. Some Institutes may have existing protocols or standards for this. The RCM NS&EA asked WKPICS3 to initiate a process of developing such standards, and progress is reported in WKPICS3.
- Quality standards / best practice for collection of data to estimate biological parameters such as growth parameters, maturity ogives, weight-at-length, sex ratio. Some guidelines on maturity ogive estimation were provided by earlier ICES workshops (WKMAT, WGMOG) but there is a clear role for the new ICES Working Group on Biological Parameters to develop the necessary quality standards although this group will not meet until 2012.

There are many other sources of error, such as incorrect species identification or non-compliance to sampling instructions. The use of training schemes, and other schemes such as temporary exchanges of sampling staff between laboratories to ensure consistency of methods, should be encouraged.

Annex 3 – Appendix 1. Components of data quality

ICES groups such as WKPICS, SGPIDS and WGRFS have focused on three main components of the quality of data and the estimates derived from them: i) The design of a sampling scheme, ii) The implementation of the sampling, and iii) The analysis of the results. Within each of these elements, two different aspects of data quality and uncertainty in estimates have been explored:

- Systematic errors (bias)
- Random errors as measured by precision.

Systematic errors: The Workshop on Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment (WKACCU: ICES 2008a) focused on aspects of bias, how to document it in an informative way, and considered approaches to reduce such bias. The workshop noted that 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**. WKACCU examined sources of bias inherent in fishery data collection that relate directly to elements of the EU Data Collection Framework: a) species identification; b) landings weight; c) discard weight; d) fishing effort; e) length structure; f) age structure; g) mean weight ; h) sex-ratio; and i) maturity stages. The workshop identified several indicators to detect bias in each of these parameters. A 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). ICES has promoted the use of the scorecard in the data compilation and evaluation part of benchmark stock assessments, but this approach turned out to be too complex and difficult to implement and combine for a fish stock across several countries which may also have different sampling schemes.

Precision of estimates: The Workshop on methods to evaluate and estimate the precision of fisheries data used for assessment (WKPRECISE: ICES 2009a) 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 based on fisheries data used for assessments are only meaningful for catch sampling programmes that obtain representative data. The workshop advised that **a minimum requirement should be that the sampling programmes pass basic checks for bias using the scorecard developed by WKACCU**.

An important concept is the trade-off between precision and bias, which is a core issue for the design of sampling surveys, and for estimating biological parameters such as fish age. This comes down to issues such as the cost of reducing bias and increasing precision, and the relative impact that bias and precision have on stock assessments and quality of advice based on them.

Annex 3 – Appendix 2: Developing a body of knowledge within ICES on statistical sampling design for fisheries sampling

Since the late 2000s, there has been a rapid and important increase in awareness within the ICES community of the need for statistically sound sampling and advice on how to do it. A series of ICES expert groups has developed advice and guidelines to help national scientists adapt their fishery sampling schemes from what has in many cases been an *ad-hoc* approach to data collection, to one that is more firmly grounded in statistical sampling theory. These ICES groups have devoted considerable effort to develop formats for reporting data quality, and guidelines for good practice. In the area of trawl surveys, international coordination has been the norm, and the need for clear guidelines and standards has been addressed for many years by ICES groups such as the International Bottom-trawl Survey Working Group (IBTSWG). Considerable work has also been devoted by the ICES Planning Group on Commercial Catches, Discards and Biological Sampling (PGCCDBS) since 2002 to establish guidelines and standards for good practice in fish ageing and maturity estimation, based on the results of many workshops, exchanges, and studies, and the PGCCDBS has established a repository of reports on this. Two ICES workshops established by PGCCDBS on estimation of maturity ogives, which produced guidelines for best practice, were also carried out (WKMAT - ICES 2007a and WGMOG - ICES 2008b)

For commercial fisheries sampling design, implementation and estimation, the first significant developments that went beyond basic descriptions of national practices were the workshop on discard raising procedures (WKDRP – ICES 2007b) and WKMERGE (ICES, 2010b), which was set up initially to provide guidelines for merging of fleet métiers, a concept introduced into the DCF from 2008 onwards. The workshop focused instead on the statistical problems introduced by the métier approach, including:

- Inappropriateness of defining sampling strata according to dynamic métier characteristics such as gear type, mesh size, target species, due to problems in controlling sampling probabilities;
- Incentives for "quota sampling" whereby samplers abandon any random, probability-based approach (if one existed) to deliberately fill sample quotas for specified métiers for a specified quarter or other period.

The WKMERGE report triggered ICES PGCCDBS to instigate a series of Workshops on Practical Implementation of Statistical Sound Catch Sampling Programmes (WKPICS1 -3: ICES 2011a, 2012c & 2013 in prep) and the Study Group on Practical Implementation of Discard Sampling Programmes (SGPIDS 1-3: ICES 2011b – 2013b). These groups documented principles for statistical sampling design and its implementation in practical conditions, for different classes of sampling schemes. Methods for reporting data quality were explored. WKPICS2 (ICES 2012c) developed Guidelines for Best Practice for sampling of commercial fisheries for biological variables at the request of the European Commission, and a version of these for recreational fishery surveys was developed by ICES WGRFS in 2013 (ICES, 2013a).

The series of ICES meetings on commercial and recreational fishery sampling design since 2008 have hugely raised awareness of sampling survey concepts within the participating countries, and within linked processes such as the Regional Coordination Meetings. Input from experts from non-EU countries including Norway, USA, Australia and New Zealand have been highly influential in this, particularly in the field of recreational fishery survey design. The ICES groups have also been influential in some changes to the way in which DCF sampling achievements for fleet based biological sampling are now reported, including the definition of sampling frames.

In practice, the ability of EU Member States to report the achieved precision of métier related and stock related biological variables has been problematic where data have (a) not been collected according to a probability-based design; and (b) the analysis has not necessarily followed the typically hierarchical cluster sampling structure of the data, or appropriate software has not been available. Development of analysis routines in the COST project (ICES 2010c) has helped for estimation of precision and for providing visual diagnostics of sampling coverage, but may not always cope with individual sampling designs and some countries have developed separate software for this. The matching of the analysis method to the design of the sampling remains an ongoing challenge for the Regional Data Bases set up to facilitate regional coordination of sampling.

Data quality indicators and data quality reports

Since the introduction of the EU Data Collection Regulation in 2002, and its successor the Data Collection Framework in 2008, EU Member States have been required to include metrics of achieved data quality in their annual reports of sampling completed in the previous year. The regulations included the concept of "precision levels" corresponding to 95% confidence intervals of +/- 5%, 10% or 25% for estimates of fishery discards, recreational fishery catches, length and age compositions and biological variables such as length-at-age or maturity-at-length or age.

The idea of data quality reports was developed by PGCCDBS in 2011 (ICES, 2011c) following a request from ICES WGCHAIRS 2011 to develop some templates for reporting on quality of input data for stock assessments, mainly for ICES assessment Review Groups. There was a need for easily comprehended overviews of how data quality has varied over time, and a range of such templates would be needed according to the nature of the data (e.g. landings; discards quantities; length or age compositions). PGCCDBS included the concept of WKACCU scorecards for bias in its proposals. Inspired by the formal review system for stock assessments conducted in the US through the Centre for Independent Experts, PGCCDBS also proposed a system of "data compilation and evaluation" workshops to be carried out in advance of benchmark stock assessments, where data for the assessment would be compiled and evaluated for bias and precision. Simple diagnostics such as tabulation of numbers of trips sampled for length or age, by country and stratum, were proposed.

The concept of the WKACCU bias scorecard and its utility has since then been discussed in several ICES meetings of PGCCDBS, WKPICS, SGPIDS and WGRFS in an attempt to develop data quality reports that more explicitly highlight bias issues around sampling design and incomplete sampling coverage. The proposed reports also considered precision issues such as small numbers of samples overall or within individual national sampling schemes for particular stocks. A developing concept was towards reports that document types of bias at different levels in the hierarchy of design, implementation and analysis in each national sampling scheme and in the final international data supplied to ICES stock assessment Working Groups. Methods of indicating precision achieved were considered, either direct estimates (CVs) or proxies such as effective sample sizes or just numbers of primary sampling units sampled.

During 2012, the idea of data quality reports was developed further by three ICES groups: SGPIDS, WKPICS and WGRFS. Their findings are summarized below.

The ICES SGPIDS meeting (ICES, 2012a) examined potential quality indicators for at-sea observer sampling, based around:

- The number of unique vessels and fishing trips in the total population, the study population and the planned and realized samples;
- The non–response rate (proportion of all attempted contacts that ultimately failed to provide a sample, for whatever reason)
- The refusal rate (the proportion of vessel skippers who, having been successfully contacted, ultimately failed to allow the observer to go on board to obtain the sample).

WGRFS 2012 addressed a Term of Reference on recreational fishery surveys to "*Develop and implement a score card system (see for example: WKACCU – Workshop on Methods to Evaluate and Estimate the Accuracy and Bias) in order to evaluate country survey programs*." Their approach was to develop a logical, hierarchical framework for documenting the accuracy of recreational fishery catch estimates combined over countries for individual fish stocks, and for tracing the source and type of errors at each stage from the design and implementation of national surveys through to the compilation of international estimates (Fig 1). The two components of accuracy (precision and bias) were considered: The proposed QA scorecard framework for recreational fishery data included numerical metrics such as catches, precision, numbers of primary sampling units sampled etc. and WKACCU-type traffic lights highlighting bias at each of the three stages of sampling design, implementation and bias. Different detail would be provided for the different levels of aggregation from national survey components to the combined international data:

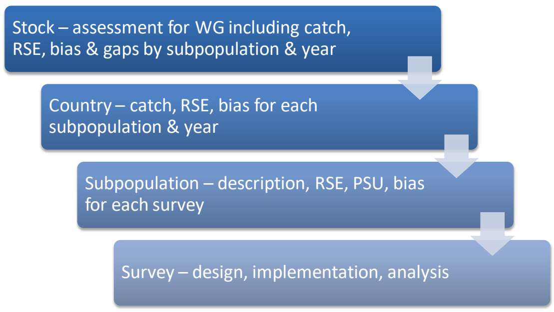


Figure 1: WGRFS 2012 proposal for a nested schema for the assessment of recreational fishing data for stakeholder use such as in stock assessments.

WKPICS2 (ICES 2012c) reviewed the proposals of SGPIDS 2012 and WGRFS 2012, and proposed a simple one-page form that can be used to evaluate quality of biological data used for stock assessments. They suggested the following four Quality Indicators (QI): type 1 – Target vs. sampled population (frame coverage); type 2 – Response rates (e.g. refusals to take observers); type 3 – "Goodness-of-fit" (diagnostics on how representative the data are of the population on a temporal and spatial scale); type 4 – Precision estimates. It was suggested that these indicators, together with other information on the sampling, should be included in a quality assurance (QA) report. It was envisaged that the QA report could eventually be automatically provided via the Regional Data Base.

WKPICS2 suggested that QA reports should describe the contribution each country makes to the total catches (discards and landings) of that stock, and the proportion caught or landed within each stratum of the national sampling frame. Given the particularities of each region or the stocks within a region, the Regional Coordination Groups (in the new DC-MAP) and/or assessment groups should develop the quality indicators further according to their specific needs and concerns. WKPICS2 also produced a set of "best practice guidelines" for fishery sampling, at the request of the European Commission.

PGCCDBS 2013 proposed sending the WKPICS2 QA reports for a trial on a few stocks, but it was later felt that more development was needed.

The WGRFS meeting in 2013 (ICES, 2013a) further explored and tested the scorecard system developed by the group in 2012. "Best practice" guidelines for recreational fishery sampling were also developed based on WKPICS2, covering the design, implementation and analysis of sampling schemes whilst also providing information on the existence and possible magnitude of biases. The conclusion from this exploratory work was that there is no single way to document data quality that is suitable for all end-users, and a "toolkit" of reporting systems was needed to provide different end-users with the information they require (Figure 2). A fundamental requirement was to have **detailed documentation of national sampling and estimation schemes, structured in line with the elements of the "best practice" guidelines, highlighting specific bias and precision issues with design, implementation and estimation.**

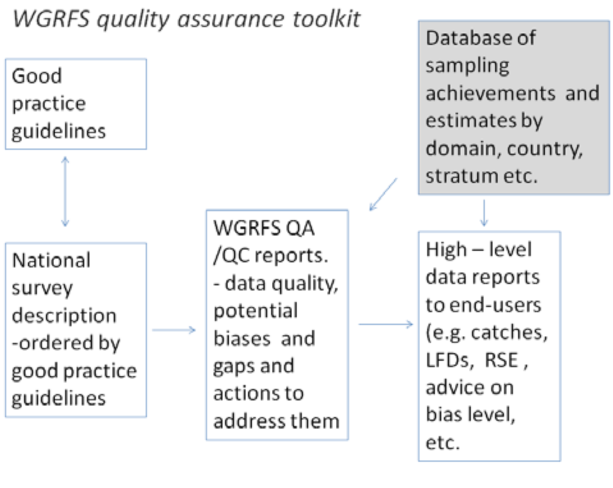


Figure 2. Scope of a "Quality assurance toolkit" proposed by ICES WGRFS 2013.

Finally, the SGPIDS 2013 meeting used some case studies to generate Quality Indicators (QI) based on the numbers of vessels in the national fleets, and the number of trips they conduct, in relation to the planned and realized number of trips sampled. Spatial mapping of fleet activities and sampled vessel locations was carried out. The quality indicator table developed by WKPICS2 was modified (Figure 3). It was aimed at investigating potential bias caused by non-successful contact attempts, improving the national sampling efforts, and documenting and providing a meaningful and transparent overview of the quality of the sampling. The group agreed on the usefulness of the quality indicator table for different potential end-users. Possible end-users would include: stock assessment working groups, auditors of annual reports (DCF/STECF/RCGs), EU commission. At the national level the quality indicators would be of use to ministries, national administrations, and fisheries as well as for in-house evaluation at national fishery institutes. For stock assessment purposes, it was recognized that part of the information has to be completed at the stock coordinator level, and that the national fishery institutes would provide data on the sampling scheme and its operation.

AT-SEA-SAMPLING						
Stock - Species - Area - Year (e.g. Cod - North Sea - 2011)						
	Country A			Country B	Country C	...
	Design			Design		
	Implemention			Implemention		
Importance: Contribution to stock landing	40			60		
Sampling / design effect/diagnostic for randomness... (Description according to best practice)						
Sampling design	probability based discard sampling			quota sampling of catches		
Primary sampling unit	Vessel			Trip		
Sampling frame	quarterly vessel list			annual vessel list		
Periodicity	ca. 1 sample per week during fishing season			1 sample per month		
Contact protocol	yes			no		
Sampling manual available	under preparation					
...						
Strata from the sampling frame	Fleet 1	Fleet 2	Fleet 3	...	...	
	e.g. active gear (Trawler)	e.g. passive gear	e.g. seine netters			
Importance: Contribution to national landing	75%	20%	5%			
Importance: Contribution to national discards	95%	1%	4%			
Quality indicator						
1 Total number of vessels in the fleet	60	300	5			
Number of trips sampled onboard of vessels	30	20	0			
Number of unique vessels sampled	5	17	0			
Total number of trips conducted by the fleet	1000	8000	6			
Age key quality indicator (e.g. Mean number of age samples per trip sampled)	100	50	0			
2 Total number of vessels contacted in the year	81	77	not determined			
Not available	32	48	not determined			
No contact details	1	0				
No answer	0	0				
Observer decline	9	7				
Industry decline	9	2				
Successful sample	30	20				
3 Goodness of fit						
Bias 1: Spatio-temporal coverage	tested and considered all right					
Bias 2: Vessel selection	smaller vessels rejected observers					
Bias 3: ...	comment					
4 Precision levels of e.g. parameter a, b, ...						
e.g. CV, variance, relative sampling error						
e.g. Input data for XSA model:						
maturity at age						
stock weight						
catch weight						
catch at age						

Figure 3 SGPIDS (ICES 2013b) proposal for a quality assurance report for regional assessment data from at-sea sampling, modified from a version designed by WKPICS2 (ICES, 2012c). Sections in green are likely to be completed by the national fisheries institutes, those in yellow by stock coordinators.

WKPICS3 (ICES 2013 in prep) is currently reviewing the state-of-play with development of data quality reports and is planning a trial on some stocks as had been planned by PGCCDBS in 2013.

Annex 3 – Appendix 3. STECF views on data quality indicators in DCF and DCMAP

A feature of the original Data Collection Regulation and its successor the Data Collection Framework has been the requirement for Member States to annually report their achieved sampling in terms of numbers of samples, numbers of fish collected or precision (CV). When planning the DCF in 2007, the STECF Study Group on Research Needs (SGRN 06-03) saw no need to change the precision levels in the new DCR, but was very much in favour of a strictly pragmatic approach with regards to the their use. In SGRN's opinion, precision levels should primarily be used as a guide when setting up sampling programmes (how many samples should be taken, when and where), and not as a compulsory threshold for financing purposes. Reaching the required precision levels was a national responsibility although for a number of parameters, (such as ALKs, sexual maturity, fecundity, etc.), there was room for regional, co-operative data collection systems and a regional approach to the calculation of precision levels. The SGRN 06-03 was very supportive of such moves and recommended that the new DCR had provisions for promoting the regional, co-operative approach to achieving precision.

STECF in its recent meetings dealing with the development of the DC-MAP have considered the implications of the revision of the roles and work programmes of the current Regional Coordination Meetings (to be re-designated as Regional Coordination Groups; RCGs) as proposed by STECF 12-07 (2012). The STECF 12-07 report proposed that the RCGs would develop regional work plans in which end-user priorities are ranked to ensure work plans operate within (limited) capital and human resources. For example, it would be for the RCGs in close liaison with the end-users to determine whether for a given resource base it was preferable to take fewer samples from more species or vice versa. Assuming that Member States develop statistically sound schemes for sampling commercial fisheries (as emphasized through the "Oostende Declaration" produced by the North Sea & Eastern Arctic RCM in 2012), regional coordination would revolve around the stock/species-orientated sampling priorities based on regional assessment and advisory needs. A national catch-sampling scheme could be seen as comprising sampling frames and strata within the overall regional sampling activity, but with priorities and sampling levels coordinated at the regional level. STECF 12-07 also considered the possibility of defining appropriate sampling frames and strata that could cross national borders, and also of accommodating nationally important issues that may have a lesser priority in regional terms.

The STECF "Review of Proposed DCF 2014-2020, Part 1" (STECF-12-07 – EWG 12-01 April 2012), report emphasized that it is essential that the quality of data are known when it is used for analysis by end-users, because management actions based on poor data should be avoided. However in its report, **EWG 12-01 no longer advocated predefined quality targets (e.g. precision levels) as at present there was no basis for setting such targets**. In many cases, it would also be impossible to evaluate how many sampling resources would be needed to meet predefined targets. Instead EWG 12-01 proposed to set a minimum sampling target, remaining at least at the present level. However, **it would be required to evaluate the quality of the data every year at the regional level (RCM) and end-user aggregation level**. **If it appears that this would lead to unacceptable quality, there should be provisions to adjust the minimum sampling level in consultation with the end-user**. These proposals by STECF also identified a need for:

- clear documentation and prioritizing by end-users of the estimates needed to support regional assessment and advisory needs;
- implementation of best practice in designing and running statistically sound sampling schemes;
- a need for some degree of optimization of sampling across countries to achieve the most cost-effective data collection supporting assessments and advice.

The STECF "Review of Proposed DCF 2014-2020 PART 2" (STECF-13-01 Jan 2013: EWG 12-15) discussed the need to include quantitative targets for sampling effort to ensure maintenance of sufficient sampling by the MS. **Such quantitative targets could be motivated by quality requirements**. They discussed how quality could be evaluated and assured. In the present DCF precision (Coefficient of Variation of the mean, CV) is the "stand-alone" indicator of data quality. Even if data are precise they could be corrupted by bias. **Quality indicators could relate to the design, performance and documentation of the sampling programme as well as to the output data.** Quality indicators need to be developed by relevant expert groups. The DCMAP needs to assure that MS are obliged to report on the quality of the data in accordance with the indicators. **The indicators themselves do not need to be included in DCMAP but have to be listed somewhere. Annual work plans should be evaluated against a best practice**. Guidelines on the application of best practice in statistically sound sampling programmes in a national as well as in regional sampling designs need to be developed. In relation to sampling intensity, MS should be obliged to sample the stocks that appear on the priority list. **The number of samples should be based on an aspirational precision level agreed with the end-users at the RCG for each stock and variable**. The planned number of samples by stock should be included in the annual work plan. Reference list should be made available at a repository. MS should report on achieved quality for the performance of the sampling programmes as well as the sampled data. The quality assessment should be done using different quality indicators. The quality indicators should be made available at the repository.

The Expert Working Group on "Review of DC MAP- Part 1" (STECF 13-06 April 2013 EWG-13-02) noted that in the past DCR and present DCF, quality targets for biological variables had been defined in the form of coefficient of variation (CV) of the estimates. In practice, problems have been experienced by this approach. **The target CV values listed in the DCF are questioned because they seem to be arbitrary choices and are not based on any pre-analyses or advice.** EWG 13-02, after reviewing the present requirements of the DCF and the related problems, proposed the following **framework for data quality requirements**. This proposal has to be considered for all type of data (biological, economic and transversal):

1) The DC-MAP should not include any predefined quality targets
2) MS should design sampling schemes in accordance with best practice guidelines
3) MS should provide quality indicators (QI) in the annual report according to international standards (i.e. EuroStat) and as specified in the guidelines for annual reports

All national sampling schemes should clearly **document** the sampling frame, sample selection procedures, response rates (e.g. refusals to take observers), imputation methods for missing data and weighting procedures employed to derive national estimates.

EWG 13-02 suggested that the DC-MAP should include the obligation for MS to apply best practices guidelines and Quality Indicators (QI) as provided by STECF or RCGs.

On the topic of Quality indicators (QI); EWG 13-02 referred to EUROSTAT standards for quality reports (Anon 2009a) that provide a list of potential Quality and performance indicators. In particular, EUROSTAT standards for quality reports advocate the CV, a range of CV or confidence intervals as the most appropriate indicators to quantify sampling errors. This is consistent with WKPRECISE (ICES, 2009a) which recommended that the precision of estimates of key parameters should be given in terms of standard errors (or relative standard errors)

In the follow-up meeting "Review of DC-MAP – Part 2" (STECF-13-12 July 2013 & EWG-13-05), it was again emphasized that biological data collection must be aligned to the specific assessment or management requirements of end-users. The EWG recommended that for commercial and recreational fisheries, **Member States should be responsible to ensure best practice in design and implementation of statistically sound catch sampling schemes**. Best practice can be defined as sampling designs, implementation and data analysis that lead to minimum bias and an accurate estimate of precision, and which make the most efficient use of sampling resources. The EWG also proposed the following requirements:

- All national surveys should document the sample frame, sample selection procedures, response rates, imputation methods for missing data and weighting procedures employed to derive national estimates. Deviation from the best practice guidelines (*as given by WKPICS2*) should be described to allow the identification of possible bias in the final estimates.
- For commercial fisheries, a minimum sampling threshold (not target) should be set rather than precision targets, remaining at least at the present level of activity and consistent with best practice in terms of statistical robustness. Regional coordination should ensure that national sampling programmes are organized to satisfy the end-user requirements within the operational constraints of the sampling programmes.
- Countries with a very low share of the recreational catches of target stocks in a region should have correspondingly lower survey effort and precision requirements for the delivery of data. Regional coordination should ensure that national sampling programmes are organized such that they satisfy the end-user requirements within the operational constraints of the sampling programmes.

Annex 3 – Appendix 4. Overview of approaches used in other important fishing nations (e.g. USA, Canada, New Zealand, Australia, Norway…) to measure and ensure quality of biological data.

Data quality assurance in New Zealand

Information was provided by Pamela Mace (New Zealand Ministry of Fisheries). In New Zealand there are a number of different standards for different types of data. For example, it is usually expected that research trawl surveys (which tend to be random-stratified) should provide biomass estimates with a CV of the mean no greater than 20% (or, in some cases 30% where there are other data that informs an assessment to the extent that a lower precision is OK). Should a survey have poorer precision than this, the indices will be down-weighted accordingly in the assessment, so it is not a case of whether the data should be used or not, it may be more of a case of whether it is cost-effective to collect such data if this is the best you can do.

The NZ Science Working group processes involve "staged technical guidance" on data quality, for example evaluating a survey design, evaluating the preliminary analyses, suggesting sensitivity analyses, and ensuring that the conclusions are justified by the data and analyses. This process was formalized a few years ago in the New Zealand Government document "Research and Science Information Standard for New Zealand Fisheries" published in May 2011. An extract is given later in this section. The document outlines a system now used for ranking the quality of science information:

1 = High Quality (which should essentially be anything that is good enough to be used in an assessment or to inform management decisions in other ways).

2 = Medium or Mixed Quality (data that might be used but would have many associated caveats)

3 = Low Quality (data that should not be used at all because it is not reliable and may produce misleading results).

U = Unranked (has not been reviewed – and therefore should be used with caution if at all).

One of the key purposes of the science information quality ranking system is to inform fisheries managers and stakeholders of those datasets, analyses or models that are of such poor quality that they should not be used to make fisheries management decisions (i.e. those ranked as "3" or "U"). Most other datasets, analyses or models that have been subjected to peer review or staged technical guidance in the Ministry's Science Working Group processes and have been accepted by these processes should be given the highest score (ranked as "1"). *Uncertainty*, which is inherent in all fisheries science outputs, should not by itself be used as a reason to score down a research output, unless it has not been properly considered or analysed, or if the uncertainty is so large as to render the results and conclusions meaningless (in which case, the Working Group should consider rejecting the output altogether). A ranking of 2 (medium or mixed quality) should only be used where there has been limited or inadequate peer review or the Working Group has mixed views on the validity of the outputs, but believes they are nevertheless of some use to fisheries management.

One expected issue was nobody ever wanting to give anything a "1" because all fisheries data are uncertain. However, after three years of using this classification scheme, people stopped equating Quality and Uncertainty (except at the extreme of course). The following link also has a document on fish ageing protocols and catch sampling protocols in New Zealand.

http://www.fish.govt.nz/en-nz/Publications/Research+and+Science+Information+Standard.htm

Extracts from the New Zealand Government document "Research and Science Information Standard for New Zealand Fisheries. May 2011"[2] are given below:

"Fisheries 2030 is the Government's goal and plan of action for New Zealand fisheries…. Internationally and locally there is an increasing move towards ensuring that high-quality evidence is used for policy formulation and decision-making, with increasing emphasis on the need for independent peer review to ensure the relevance, integrity, objectivity and reliability of information. These key principles for science information quality have been integrated into the Research and Science Information Standard."

In relation to Key Principles for Science Information Quality:

The quality of research and science information relates primarily to relevance, integrity, objectivity and reliability. The primary, internationally accepted mechanizm for evaluating the quality of research and science information is peer review and, as such, peer review is both a principle and a mechanizm. These five key principles should underpin all quality assurance processes for research and science information. Ideally, the key principles should be satisfied PRIOR to research and science information being used to inform fisheries management decisions.1

Peer Review – Is the principal process used to ensure that the quality of scientific methods, results and conclusions meet the accepted standards and best practices of the science community. Peer review is an organized process that uses peer scientists with appropriate expertise and experience to evaluate the quality of research and science information.

Relevance – Scientific research must be relevant to the fisheries management question(s) 1 being addressed, contributing directly to answering those management questions and addressing management objectives for that fishery.

Integrity – Refers to the security of information, and to the protection of information from inappropriate alteration, selective interpretation or selective presentation. It must be ensured that the information is not compromised or biased, particularly with regards to presenting the uncertainty of that information, to ensure that information remains complete throughout the science-to-decision process.

Objectivity – Refers to whether the information presented is accurate, impartial and unbiased. Objective interpretations or conclusions do not depend upon the personal assumptions, prejudices, viewpoints or values of the person presenting or reviewing the information. Scientific methods must be used in the collection and analysis of data, and science processes must be free of undue non-scientific influences and considerations. Data must be obtained from credible and reliable sources. To the extent possible, data and analyses must be accurate and unbiased.

[2] Ministry of FisheriesTe Tautiaki i nga tini a Tangaroa. www.fish.govt.nz. ISBN 978-0-478-11927-5 (print) ISBN 978-0-478-11928-3 (online)

Reliability – Relates to the accuracy and reproducibility of information. Research and science information must be accurate, reflecting the true value of the results being reported within an acceptable level of imprecision or uncertainty appropriate to the data and analytical methods used. Information should not be biased and should not suffer from such a high level of imprecision that the results and conclusions are rendered meaningless. Methods and models used to produce science information must be verified and validated to the extent necessary to demonstrate that results may be reliably reproduced by an independent scientific expert using the same data and analytical methods.

In relation to data collection:

- Data must be collected according to documented procedures, and in a manner that reflects standard best practices generally accepted by the relevant science and technical communities. Data and information sources must be identified or made available upon request.
- Data collection methods, systems, instruments and statistical sampling designs must be designed to meet the requirements and objectives of the research projects concerned, and should be validated before use. Instruments must be calibrated using applicable standards or fundamental engineering and scientific methods.
- Data must undergo internal or external quality assurance prior to being used in analyses that are intended or likely to inform fisheries management decisions.
- Science quality assurance and peer review processes implemented in accordance with this Standard are required to assess the quality of information by applying the following quality ranking system:
- **1 – High Quality** is accorded to information that has been subjected to rigorous science quality assurance and peer review processes as required by this Standard, and substantially meets the key principles for science information quality. Such information can confidently be accorded a high weight in fisheries management decisions.1
- **2 – Medium or Mixed Quality** is accorded to information that has been subjected to some level of peer review against the requirements of the Standard and has been found to have some shortcomings with regard to the key principles for science information quality, but is still useful for informing management decisions. Such information is of moderate or mixed quality, and will be accompanied by a report describing its shortcomings.
- **3 – Low Quality** is accorded to information that has been subjected to peer review against the requirements of the Standard but has substantially failed to meet the key principles for science information quality. Such information is of low quality and should not be used to inform management decisions. Where it is nevertheless decided to present such low quality information in fisheries management decisions, the quality shortcomings of the information should be reported and appropriate caution should be applied.
- **Unranked – U** is accorded to information that has not been subjected to any formal quality assurance or peer review against the requirements of this Standard. Where unranked information is used to inform fisheries manage-

> ment decisions,1 it should be noted that the information has not been reviewed against the Standard, and that the quality of the information has not been ranked and cannot be assured.

Fisheries managers particularly need to be informed when information is unranked (U), or is ranked as being of low quality, so that the uncertainties or shortcomings regarding information quality can be noted, and appropriate weight given to such information when used to inform fisheries management decisions

Data Quality assurance in the USA

Fishery dependent sampling in the Northeast (NE) region of the USA is a shared responsibility of two major institutions within the National Marine Fisheries Service (NMFS), numerous state agencies, and for recreational catch data, a national office in Washington DC. Collectively these groups provide the raw data that are used in stock assessments and management advice.

Commercial fishery discards and/or landings

The NE region uses a dual system of estimating total landings. Dealers who sell federally regulated species are required to report landings on a weekly basis. Individual fishers with federal permits are required to report landings by stock area. These logbooks are known a Vessel Trip Reports (VTR). Matching of VTR and Dealer records is required by end-users and this requires significant reliance on imputation methods. Potential errors of imputation have been estimated but such data are not routinely reported for landings. Fishery discards are based on a comprehensive at-sea sampling program for all fleets. The sampling design and allocation of observers to vessels is updated annually under the provisions of a fishery amendment known as the Standardized Bycatch Reporting Methodology (SBRM). The SBRM evaluate the precision of discard estimates by species or groups of species. Sampling requirements for the next year are based on the sample size necessary to achieve a standard level of precision, defined as a coefficient of variation of 30% of the estimate. Since each fleet captures and discards multiple species, the sampling requirements for the fleet are based on the sampling requirements for a species or group of species. A formalized algorithm is used to reduce sampling requirements by taking into account the magnitude of the estimate in relation to the total catch and total discard of the species. This ensures that sampling effort is not inappropriately targeting elusive estimates of precision for small quantities. Sampling precision and discard estimates are provided to stakeholders on an annual basis.

Recreational fishery catches

Recreation catches are based on a two-stage sampling design that independently measure fishing effort and catch per unit of effort. Estimation of fishing effort was, until recently, based on a random digit dialling phone survey of households in coastal counties. Catch per unit of effort is measured via intercept sampling where individual fishers are interviewed as they complete their fishing trips on shore or when landing their boats at boat ramps and other locations. Charter boats / head boats are also sampled. The design of the recreational fishery survey programme in the USA was recently subject to a major peer review by the National Academies at the request of NMFS, and the remit of the review can be found at: ftp://ftp.gulfcouncil.org/Ecosystem%20Folder/NRC%20Summary%20of%20Review%20of%20Rec%20Fisheries%20Survey.pdf

Current survey methods and recommended alternatives were compared with relation to costs, sources of bias, precision, and timeliness. Criticisms of the programme included the freedom that the survey staff had to target particular sites or times of day, and inadequate coverage of the day. The revised Marine Recreational Information Program (MRIP) is based on a more sophisticated approach that fully corresponds to the actual sampling design, and places greater emphasis on adherence to strict protocols for statistical design, particularly randomization of sites and days

Length compositions of fishery catches (landed; discarded)

Length samples are routinely taken via a port sampling program that relies on stratification by geographic region, species, stock area, market category and season. Sampling requirements for each species are determined annually by individual analysts. For species that have multiple stocks, extra care is required to ensure that samples are properly attributed to stock area. Length compositions of discarded fish are based on samples taken by at-sea observers.

Age compositions of fishery catches (landed and discarded)

Age samples are routinely taken via a port sampling program that relies on stratification by geographic region, species, stock area, market category and season. For most species, age samples for landings are processed by NMFS; but for some species, Canada and a number of other states provide additional processing capacity. Ageing standards are validated by cross validation among various laboratories, and occasionally by direct validation methods. Results of age estimation samples and comparisons among readers are available on a web page that provides measures of precision. The rationale, methods, data presentation and statistical measures of quality assurance and quality control estimates for the production ageing of Northwest Atlantic species are given at http://www.nefsc.noaa.gov/fbp/QA-QC/index.html. While age samples of landed and discarded fish are routinely taken by at-sea observers, most of these samples are not processed. Instead, estimates of age compositions of landed fish are derived from port samples and age composition of discards are based on age-length keys derived from fishery-independent surveys.

Growth parameters

Growth parameters using von Bertalanffy growth models are usually derived by analysts and are typically the products of academic theses rather than routine sampling efforts. Most peer-reviewed articles provide some measure of precision of derived parameters and their covariance.

Maturity ogives (proportion mature at age or length)

Maturity ogives are routinely derived for fishery-independent data. Maturity ogives for landings and discards are difficult because many species are not landed whole. Subsampling for maturity status increases the costs of sampling (due to destructive cutting of fish) and is therefore not provided. Instead, we rely on measures of maturity derived from fishery-independent surveys or special studies for species whose maturity status cannot be reliably determined during our spring or fall bottom-trawl surveys.

Sex ratios

Sex ratios are important for a number of fisheries particularly spiny dogfish, and increasingly, in various flatfish and monkfish assessments. Empirical evidence suggests greater mortality rates for males than females although direct experimental confirmation is lacking. Obtaining sufficient samples to derive length specific sex ratios is difficult especially when external sex determination cannot be done. For these species, special sampling programs have been devised (e.g. summer flounder).

Science Quality Assurance

There is national coordination of the NOAA Fisheries activities regarding the implementation of the Data Quality Act. Activities to strengthen integrity of scientific information include science program reviews of the NOAA Fisheries science centres and the scientific peer review process. External peer reviews are also conducted through the Center for Independent Experts (CIE).

As part of the Northeast Regional Stock Assessment Workshop (SAW) process, SAW working group members routinely review and evaluate data inputs used in stock assessments.

Further information is available at http://www.st.nmfs.noaa.gov/science-quality-assurance/index and http://www.nefsc.noaa.gov/saw/ .

Information that might be of interest to working group members:

- Northeast Fisheries Science Center's Northeast Fisheries Observer Program (NEFOP) http://www.nefsc.noaa.gov/fsb/program.html
- Standardized Bycatch Reporting Methodology (SBRM) http://www.nefsc.noaa.gov/fsb/SBRM/
- The NMFS Science Program Reviews on Stock Assessment Data Collection Programs and Management occurred in 2013. Information pertaining to each region can be found at http://www.st.nmfs.noaa.gov/science-program-review/
- Optimization model used as a tool to guide sea day allocation

http://www.nefsc.noaa.gov/publications/crd/crd0509/ (Rago et al. 2005). One important aspect of using the optimization model to allocate sea days is that it explicitly incorporates a regular feedback mechanizm for continuously improving the performance of the bycatch monitoring and thus, can be viewed as a set of quality assurance/ quality control measures that provide a formal way of updating and improving the sampling design as new information is obtained (Figure 12). {Note: this optimization tool is no longer applied in the Northeast region due to changes in fishery management regulations}

- Pre-Trip Notification System (PTNS) – a recently developed system that used a self-adjusting probability-based, tiered selection process to randomly assign observer coverage across the groundfish fleet on a proportional basis for the purpose of monitoring discards.

http://nefsc.noaa.gov/publications/crd/crd1321/crd1321.pdf (Palmer et al. 2013)

Data quality assurance in Norway

As a part of the national strategy for developing the Norsk Marint Datasenter (NMD) (Norwegian Marine Data Center), a national marine data centre, IMR is currently refining the data handling, management and dissemination of data and data products through a large infrastructure project called Sea2Data. The main objective of this project is to prepare the institute to be able host a wide suite of marine data, and to make them readily available to researchers and other users. As a first step along this route, a general infrastructure is developed and applied for our field operations. However, the technical solutions, the strategy and work flows are general and will be used as a template for other types of data.

The project is organized in well defined tasks to: Improve and operationalize the operational infrastructure; incorporate and quality testing historical data; and improving tools to extract data/products from the data model. The project consists of seven work packages.

NMDwill manage all research data from research surveys and fisheries sampling programmes (such as the Norwegian Reference Fleet) conducted by the Institute of Marine Research, Norway, in accordance with national requirements, standards, and international agreements. Through Sea2Data, IMR is developing the infrastructure to faciliate easier access to data, improve the quality control and quality assurance (QA/QC) of data from their collection, through data entry, data storage, analysis, and dissemination. The relational database used for data storage, as well as modules for data analysis to provide stock assessments and other end-products is integrated in the Sea2Data framework, using open-sources programming tools and analysis packages (such as R). The QA/QC includes documentation of sampling protocols for research surveys and fisheries monitoring programs, instructions for data punching, and a range of checks to minimize data entry errors and other sources of errors.

Annex 3 – Appendix 5. Current DCF data quality indicators

Standard report table number	Quality indicators [1]	Precision estimation
Table III.C.3 - Expected sampled trips by metier	Achieved vs expected nos. trips sampled by region, fishing ground and metier, at sea and on shore (DCF specifies minimum 12 samples per metier per year). Only metiers in top 90% ranking as agreed by RCMs expected to be sampled.	
Table III.C.4 - Metier sampling strategy	Achieved vs expected nos. trips sampled by region and sampling frame, at sea and on shore	
Table III.C.5 – Sampling intensity for length compositions (all metiers combined)	Precision (CV) of length compositions by stock, region and fishing ground from sampling on shore or at sea, in relation to DCF required annual precision. Total numbers of fish measured.	CV (relative standard error) based on sampling design
Table III.C.6 - Achieved length sampling of catches, landings and discards by metier and species	Numbers of fish measured from landings and discards, by species, region, fishing ground and metier	
III.D.1 recreational fisheries	Planned and achieved numbers of samples.	
Table III.E.3 - Sampling intensity for stock-based variables	Number of fish sampled and CV for length@age, weight@age, maturity@age, sex ratio in relation to planned numbers and required precision target, by species, region and fishing ground. Results can be given for the individual country, or as a collaborative sampling between countries in a region (CV target the same as for individual country).	CVs calculated for individual length or age groups and averaged over groups[2]

Notes:

[1] Shortfalls of less than 10 % from the plan are considered to be an acceptable operational margin for length and age sampling, and need not be justified.

[2] Precision estimates should be calculated as the weighted average of CVs over all length/age classes. The weight to be used is the total estimated number of individuals per length/age classes. Precision estimates should be calculated following the provisions of the DCF (Commission Decision 2010/93/EUC section B.B2.4).

For stocks of species that can be aged, average weights and lengths for each age shall be estimated at a precision level 3, up to such an age that accumulated landings for the corresponding ages account for at least 90 % of the national landings for the relevant stock.

(2) For stocks for which age reading is not possible, but for which a growth curve can be estimated, average weights and lengths for each pseudo age (e.g. derived from the growth curves) shall be estimated with a precision of level 2, up to such an age that accumulated landings for the corresponding ages account for at least 90 % of the national landings for the relevant stock.

(3) For maturity, fecundity and sex ratios, a choice may be made between reference to age or length, provided that Members States which have to conduct the corresponding biological sampling, have agreed the following:

(a) For maturity and fecundity, calculated as proportion of mature fish, precision of level 3 must be achieved within the age and/or length range, the limits of which correspond to a 20 % and 90 % of mature fish;

(b) For sex ratio, calculated as proportion of females, precision of level 3 must be achieved, up to such an age or length that cumulated landings for the corresponding ages or lengths account for at least 90 % of the national landings for this stock.

Annex 3 – Appendix 6: WKPICS2 Guidelines for best practice in catch sampling schemes

Documentation of sampling design, performance of sampling and production of estimates

Process that need to be described	Best practice	Comment	Bad practice
Target population	The target population needs to be identified and described. Access to the target population for sampling purposes need to be analysed and documented.		
Primary sampling units (PSUs)	Choice of PSUs should be identified, justified and documented. PSUs could be trips, vessels*time or sites*time (harbours, markets, access points). Size of PSUs should be documented	If PSU is something else than trip, vessel or site the choice need to be thoroughly explained.	
Sampling frame	The sampling frame (list of PSUs) should be a complete list of non-overlapping PSUs. The sampling frame should ideally cover the entire target population.	If it is not possible to cover the entire target population with the sampling frame it is good practice to clearly describe how large the excluded part of the population is and the reason for excluding it.	To exclude large parts of the target population in an ad-hoc way.
Stratification of the sampling frame	Strata should be well defined, known in advance and fairly stable. Clear definitions and justifications of strata should be available. One PSU can only be in one stratum. The minimum number of samples within a stratum depends on objective, PSU and variance and needs to be calculated. The number of samples within a stratum needs to be justified, in particular if it is below 10.	If the desired minimum number of samples per stratum is not analytically assessed, the choice needs to be justified and described. Care needs to be taken to avoid over-stratification.	To over-stratify (few or no samples in each strata) the sampling schemes. Over-stratification results in increased risk for bias, particularly for ratio estimates, and a need to impute data.

Distribution of sampling effort	The way sampling effort is distributed between strata needs to be described. In accordance with best practice, this can be based on analysis of variance or just distributed proportionally. The different sampling inclusion probabilities/weighting need to be documented.	If other methods, such as expert judgment are used, this should be explained and justified.	
Sample selection procedure	In accordance with good practice, the selection of PSUs to sample should be done in a controlled way allowing for estimation of sampling inclusion probabilities for the different samples. In principal this mean that samples shall be chosen randomly (probability based sampling). Random sampling can be either simple random sampling or systematic random sampling. The selection procedure needs to be justified and described	If it is impossible to use probability-based sampling, the samples need to be thoroughly validated for how representative they are. This process need to be described. If a non-probability based sampling design is applied, this needs to be accounted for in the estimation process (e.g model based estimations). This needs to be thoroughly explained. For small-scale fisheries where there is no census information on the target population, the only way to sample in accordance with good practice is randomly.	Ad-hoc based sampling, without proper documentation to allow estimation of bias, where the sampling inclusion probabilities cannot be estimated.
Hierarchical structure in the sampling	All the levels in the hierarchical structure of the sampling scheme need to be documented. Sampling should be random at all levels. Sampling probabilities should be worked out at each level, and information for this needs to be collected (e.g number of boxes)		Failure to account for the different levels of sampling units in the design and estimation processes. (Risk for bias as well as hiding true variation)
Protocol for selection of samples at lower sampling levels (SSU, etc.)	Such protocols should exist in a national repository		

System to monitor performance of sampling schemes - Quality Indicators	Non-response rates should be recorded. Precision of estimates (relative standard error) should be calculated, where relevant. Effective sample size (or appropriate proxy such as number of vessels or trips sampled) should be calculated and recorded.
Documentation of raising/weighting procedure for national estimates	Data analysis methods should be fully documented, covering: (1) how the multistage sample selection is accounted for in the raising/weighting procedures; (2) ancillary information (for example from fleet census data), that is used to adjust sample weights to correct for any imbalance in samples compared to the population; (3) methods of adjustment for missing data and non-responses.

Annex 4: WKPICS3: Terms of Reference for the next meeting

The WKPICS series is now completed. A new Working Group on Commercial Catches will meet from November 2014. Further details can be found in the 2013 report of the ICES Planning Group on Commercial Catches, Discards and Biological Sampling.

Annex 5. List of presentations and working documents

All presentations, Working Documents and national sampling scheme summaries are archived on the WKPICS SharePoint site.

Annex 6. Case studies that illustrate the four design classes

Annex 6.1. A Dutch case study for design-class A: Dutch at-sea discard sampling by observers on-board shrimp trawlers

Dutch at-sea discard shrimp on-board sampling, Design class A

Design	Sampling frame	Stage 1 (PSU)	STR 1	Stage 2 (SSU)	STR 2
A	Vessel * time	trip	vessel-characteristics, quarter	Hauls/boxes	

This is the Dutch national discard observer sampling programme (Figure 1), aiming at estimating the total number/weight of discards in shrimp fishery at the fleet level, with possible stratification of quarter, area or vessel characteristics.

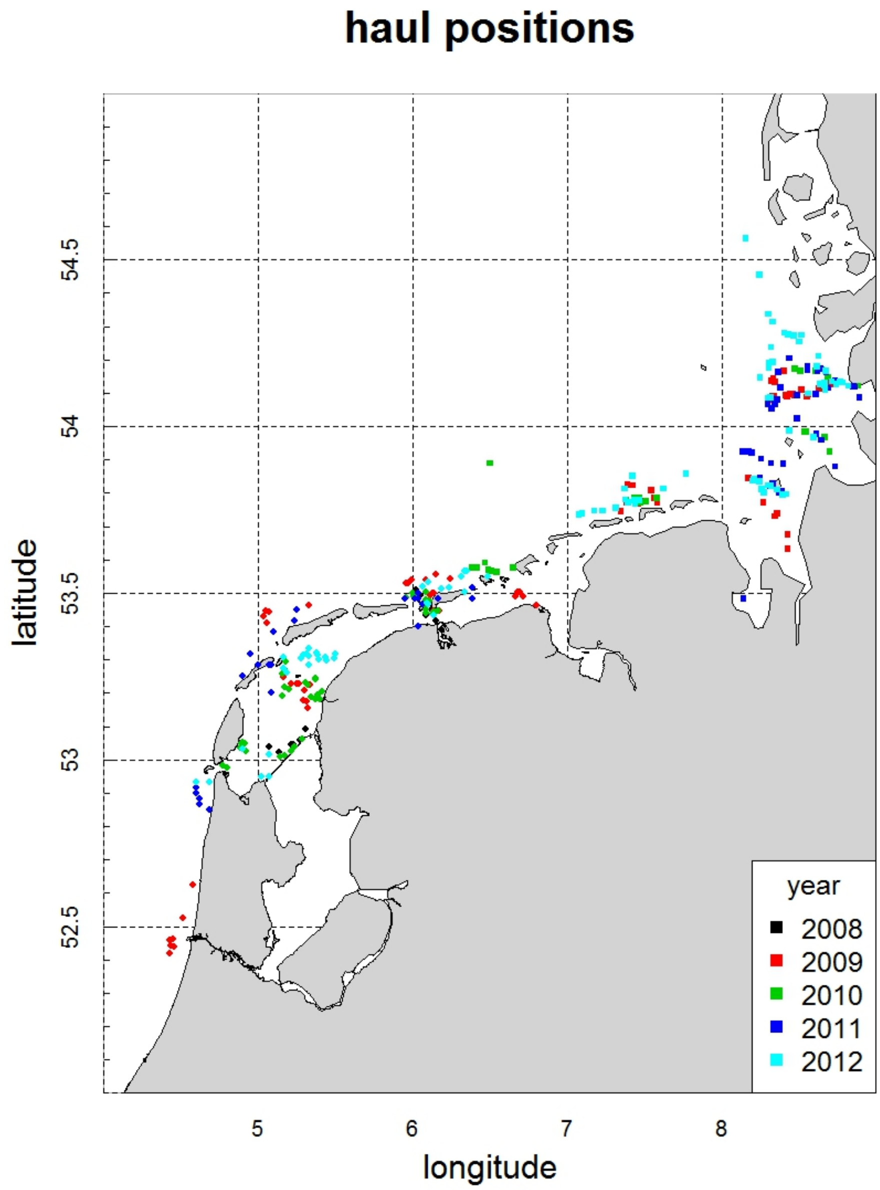


Figure 1. Sampling haul locations of the observer on-board shrimp discard sampling, from 2008-2012.

In this sampling scheme, all Dutch vessels, targeting solely for shrimp, have gear type TBB (beam trawl) and mesh size between 16-32mm. A shrimp fishing trip usually last for 3 days. During the year, the number/length of trips varies among vessels, quarter and area. A summary of trip distribution for all the 186 vessels are plotted in Figure 2 and 3.

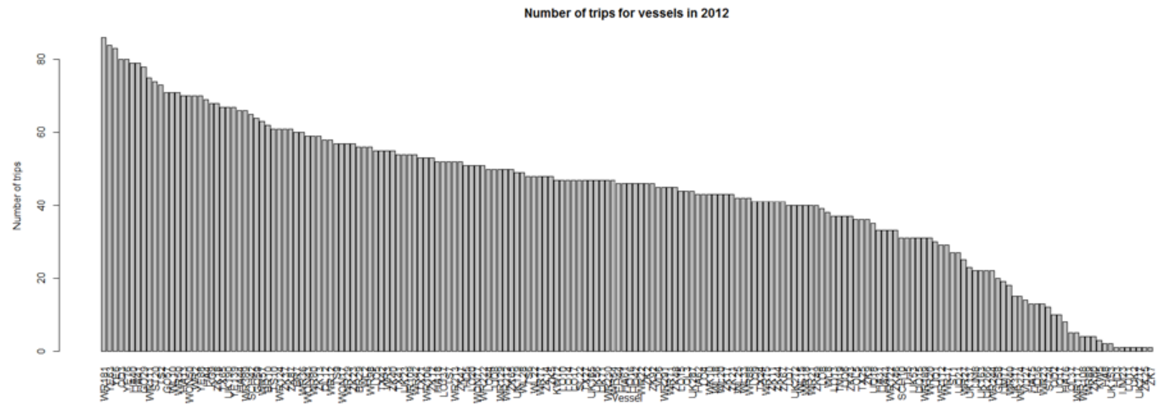


Figure 2. Number of trips for each vessel in 2012. The maximum number of trips per vessel is 86, while the minimum number of trips is 1.

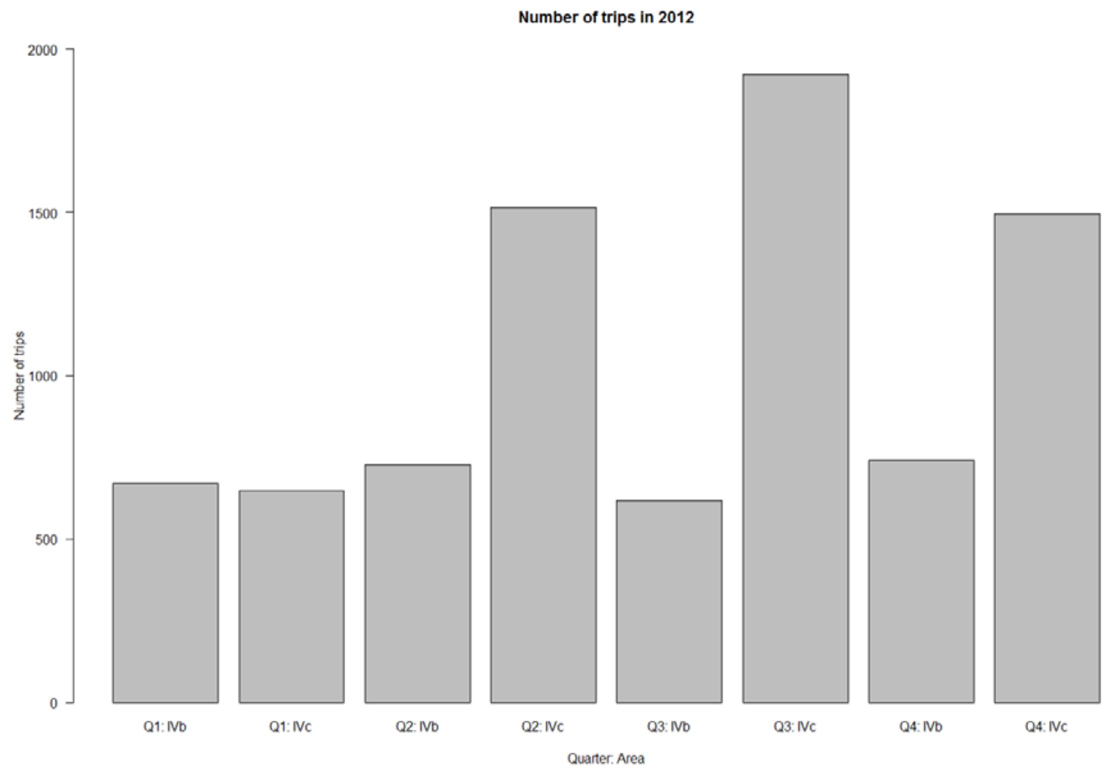


Figure 3. Number of trips by quarter×area in 2012.

Trips are the primary sampling unit. It is reasonable in the analysis phase to treat the list of all trips (obtained at the end of the year) as the sampling frame. Unfortunately, information about the list of all trips is always unknown at the design phase. Therefore, the actual selection is based on a frame with a vessel list crossed with time.

How the sampling is done theoretically: We first obtain a vessel-time availability map obtained from the previous year (Figure 4). This map indicates whether a vessel is at sea at a specified time. Afterwards, we conduct simple random sampling (SRS) on the vessel-time available population. The trip corresponding to the selected vessel-time slot is then the selected trip. If all trips have equal duration, we end up with equal probability of trips being sampled.

Sampling in practice: In practice, problems arise regarding the logistic to select the trips for the entire year. An entire year vessel-time table is usually not known in advance. Even if we use the information from previous years, it seems impractical to get the response of the skipper sooner enough about an incoming randomly selected observer trip in December. Thus, resampling for the substitute sample would be biased and inefficient. Practical solutions would be to stratify the sampling scheme by quarter. Thus, number of samples required for each quarter would be proportional allocated according to prior knowledge. Within each quarter, the vessel-time table selection can be more logistically efficiently implemented (see ICES SGPIDS Report 2013).

After the trip is selected, the corresponding skipper of the trip is informed. The response (accept/refusal) as well as some characteristics variables (responding time, form of response, reason of refusal, vessel/skipper information, time, etc.) are then recorded. In case the skipper refuses the on-board observer, a new trip is selected from the vessel-time table following the same procedure and the skipper is informed. We repeat such process until a skipper accepts the on-board observer. And we repeat such process until we have selected all the budgets of *n* samples. Once the observer is on-board, the secondary sampling unit are fishing operations (e.g. hauls, boxes). The observer systematially samples (10 hauls or all hauls) per day so that the bias of haul-time selectivity is minimized. At the end of the year, the census data of number of trips per vessel/quarter/area are obtained.

Analysis: Before raising, a non-response analysis is conducted to check the response rate and whether bias exits in the accepted observer trips (see ICES SGPIDS report 2013). If no bias appears in the data, raising is conducted. Using census data of the trip population, several estimators can be applied to estimate the average/total variable of interest:

1) estimator from SRS of vessel-time table ≈ SRS of trips (unbiased, higher precision):

$$\hat{\tau} = \frac{N}{n}\sum_{i=1}^{n} y_i$$

2) if needed and sample size allows, post-stratified estimator (unbiased, possible to improve precision)*: ; $\hat{\tau}_{st} = \sum_{k=1}^{L} N_k \overline{y}_h$

3) Horvits-Thompson estimator (unbiased, precision??): $\hat{\tau}_{\pi} = \sum_{i=1}^{v} \frac{y_i}{\pi_i}$, where v refers to the number of distinct units in the sample.

Suppose the probability of selecting trip i in each draw is p_{trip_i}, the probability of having sample trip_i in the v sample is: $\pi_i = 1-(1-p_{trip\_i})^v$

4) If somehow the census of total number of trips is unknown, or somehow the variable of interest is better correlated with some other variables, for instance, (we think) the number of catch is highly correlated to the horsepower (power*duration_trip), we can use ratio estimator with horsepower as the auxiliary variable. Ratio estimator is biased but possible to improve precision. Note that the ratio estimator improves the precision only when the variable of interst is linearly correlated with the auxilliary variable through origin.

$$\hat{\tau}_r = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} \tau_x$$

Estimator of the variance: Our sampling fraction is almost zero, so we don't need to apply finite population correction factor (approximate as sampling with replacement) to estimate the variance. Because we have a large number of hauls/boxes per trip, the variance at the secondary sampling level is negligible. Thus, the variance of the estimator is estimated at the PSU level.

1) analytical unbiased/approximated estimator

Post-stratified estimator (chapter 11.6, Thompson 2012)

Horvits-Thompson estimator (Ch 6.2, Thompson 2012)

Ratio estimator (Ch 7.1, Thompson 2012)

2) empirical method: bootstrapping

Reference:

Steven K. Thompson, 2012. Sampling, 3rd Edition. ISBN: 978-0-470-40231-3. 472 pages March 2012, Wiley.

Annex 6.2 Norwegian case studies for design-classes A and B.

Case Study 1, design class A (Table 2.5.1.1 in main body of WKPICS3 report). Trips and fishing operations are sampled at sea by contacting randomly selected vessels on a given day. The unprocessed catch is then brought to port, where the catch is sampled by IMR staff.

Design	Sampling frame	Stage 1 (PSU)	STR 1	Stage 2 (SSU)	Stage 3 (TSU)
A	Vessel*time	Vessel and trip	Port sampling. Gear, area, quarter, demersal coastal fisheries and pelagic fisheries	Trip /fishing operation	Fish (age, length, weight)

The Institute of Marine Research runs a port-sampling scheme using a boat to visit ports along the coast north of 62 °N. IMR staff contacts a stratified random sample of vessels operating in the statistical area around a port. IMR staff samples fish (length and age by species) from the randomly selected fishing trips when the fish is landed in the port. For practical and funding reasons, trips cannot be selected from all combinations of regions, seasons and gears. In particular, samples are, mostly, taken from vessels operation in statistical areas near the coast. When a sample from a fishing trip is landed, the weight and length of each fish are recorded as well as the size of the catch taken during the trip. In this work we assume that we have a random sample of catches (fishing trips) from the population of catches (trips) in each of a set number of gear categories, in four quarters of the year, and in a set number of regions. We also assume that the fish were chosen randomly from each catch and ages were determined without error.

Case study 2, design class A. Biological sampling conducted by the Coast Guard, with random sampling of fishing operations at sea.

Design	Sampling frame	Stage 1 (PSU)	STR 1	Stage 2 (SSU)	Stage 3 (TSU)
A	Vessel*time	Vessel and fishing operation	Gear, area, quarter, high seas fisheries	Catch operation	Fish (age, length, weight)

The Norwegian Coast guard's fleet consists of 13 different vessels with various capacities. The Coast guard's main tasks are related to fisheries surveillance, protection of the marine environment, search and rescue and customs. The coast guard collects data on catch composition and length distribution of commercial species from each fishing operation (haul or set) during inspection of fishing vessels at sea. Ideally this could be looked upon as a random sampling of fishing vessels and fishing operations. This is, however, only partly the case as the Coast guard's choice of inspection and sampling is partly decided from a risk of violation of fisheries regulations. However, if the fishing gear used is representative for the whole fleet, and there are catches reported from the actual stratum, then the Coast guard data are used for assessment purpose.

Case study 3, Design class B. The offshore and coastal reference fleets (RF).

Design	Sampling frame	Stage 1 (PSU)	STR 1	Stage 2 (SSU)	Stage 3 (TSU)
B	Vessels	Vessel	2 Fleets (Offshore vs Coastal Fleet), Gear, Target fishery	Catch operation	Fish (age, length, weight)

The Reference Fleet is a small group of Norwegian fishing vessels that provide the Institute of Marine Research (IMR) with detailed information about their fishing activity and catches through trained self-sampling on a regular basis. The Norwegian reference fleet represents an example of at-sea sampling where vessels are primary sampling units (principal design class B). Here vessels are selected as the primary units. Within vessel, catch operations are sampled systematically through time at a constant rate. Within catch operation fish are sampled at random for ages, lengths and weights. It should be noticed that the hierarchy within boat also include the trip level, but since the sampling design dictates sampling catch systematically through time (regardless of the trip unit) we omit this level in this example, and thus leaves this as an example of a three stage sampling design. In this sampling program vessels are stratified according to gear and area.

To illustrate the principles of implementing a design based estimator we show weighted estimators for means, totals and ratios within a stratum. The stratified estimators to obtain estimates across strata may be outlined following the same principles. Let Y and Z be variables for which data has been collected and y_{ijk} and z_{ijk} values observed for the kth fish, in the jth haul at the ith boat corresponding to the sampling levels

1) Samples b of total B boats
2) Samples n_i hauls of total N_i within boat i, $i = 1, \dots, b$ at an approximately constant rate τ.
3) Samples m_{ij} fish of total M_{ij} within n_i, $j = 1, \dots, n_i$

In this example three estimators from classical sampling theory are considered; the total, mean and ratio which can be written generally as:

$$\hat{\theta}_{tot} = \sum_{ijk} \omega_{ijk} y_{ijk},\ \hat{\theta}_m = \frac{\sum_{ijk} \omega_{ijk} y_{ijk}}{\sum_{ijk} \omega_{ijk}},\ \text{and}\ \hat{\theta}_r = \frac{\sum_{ijk} \omega_{ijk} y_{ijk}}{\sum_{ijk} \omega_{ijk} z_{ijk}},$$

where ω_{ijk} is the sampling weight for fish ijk. If the sampling weights are the inverse of the sample inclusion probability $\omega_{ijk} = {}^{1}/_{\pi_{ijk}}$, the first of these estimators estimates the population total of the variable Y. This is a Horwitz-Thompson estimator (cf Lumley 2010) which is a generalized estimator accounting for inclusion probabilities. Estimates of the mean and ratio are derived from the population total, and the estimator for the mean is also a ratio estimator if the population total size is unknown.

The joint inclusion probability can be written as a product of the conditional probabilities at each level in the sampling hierarchy: Define the observation variable I_i as a binary variable indexing the data that are sampled at stage i. Then the joint inclusion probability is

$$\pi_{ijk} = P(I_i = 1, I_j = 1, I_k = 1) = P(I_i = 1)P(I_j = 1|I_i = 1)P(I_k = 1\,|I_i = 1, I_j = 1)$$
$$= \pi_i \pi_{j|i} \pi_{k|ij}$$

In this example it is assumed that the units are sampled by simple random sampling at each stage. Consequently the level inclusion probabilities are

$\pi_i = \frac{b}{B}$, $\pi_{j|i} = \frac{n_i}{N_i} = \tau$, and $\pi_{k|ij} = \frac{m_{ij}}{M_{ij}}$, respectively, such that $\pi_{ijk} = \frac{b}{B}\frac{n_i}{N_i}\frac{m_{ij}}{M_{ij}}$, or $\omega_{ijk} = \frac{B}{b}\frac{N_i}{n_i}\frac{M_{ij}}{m_{ij}}$.

First notice that for this design an estimator of the total number of fish M is given by $\widehat{M}_\omega = \sum_{ijk} \omega_{ijk}$. An estimator for the total numbers-at-age is then given by $\widehat{M}_\omega(a) = \sum_{ijk} \omega_{ijk} y_{ijk}^{(a)} = \sum_{ijk} \frac{B}{b}\frac{N_i}{n_i}\frac{M_{ij}}{m_{ij}} y_{ijk}^{(a)}$, where $y_{ijk}^{(a)}$ is a binary variable taking values 1 if the age is a and 0 otherwise.

Furthermore, the mean estimate of $Y^{(a)}$ is the proportion $p(a)$ of the population with $Y^{(a)} = 1$ such that $\hat{p}(a) = \frac{\sum_{ijk} \omega_{ijk} y_{ijk}^{(a)}}{\sum_{ijk} \omega_{ijk}} = \frac{\widehat{M}_\omega(a)}{\widehat{M}_\omega}$.

To increase the precision in the estimate of catch-at-age we used the auxiliary information of reported catch weights made available by the official landing statistics in retrospect of the sampling and scale the estimates accordingly. First realize that the numbers-at-age is the proportion at age times the numbers $M(a) = p(a)M$. In the preceding example and estimator of M was given utilizing the sampling weights. However an obvious estimator of M frequently used is the ratio of total reported catch weight W to mean fish weight, $\widehat{M}_w = \frac{W}{\bar{w}}$.

This suggests:

$\widehat{M}_w(a) = \hat{p}(a)\widehat{M}_w = \hat{p}(a)\frac{W}{\bar{\bar{w}}}$,

i.e. the proportion at age by mean weight scaled to the total catch weight. The proportion at age by mean weight is a ratio estimator, and substituting for the sampling weights we obtain:

$\widehat{M}_w(a) = \frac{\sum_{ijk} \frac{BN_iM_{ij}}{bn_im_{ij}} y_{ijk}^{(a)}}{\sum_{ijk} \frac{BN_iM_{ij}}{bn_im_{ij}} w_{ijk}} W = \frac{\sum_{ijk} \frac{M_{ij}}{m_{ij}} y_{ijk}^{(a)}}{\sum_{ijk} \frac{M_{ij}}{m_{ij}} w_{ijk}} W$, i.e. knowledge of number of vessels and number of catch operations are not necessary. This is the estimator used for the comparison of design based and model based estimation of catch-at-age in the Norwegian case study presented at this workshop (Aanes and Hirst, 2012).

Comparisons of $\widehat{M}_\omega(a)$ and $\widehat{M}_w(a)$ are of interest, but has not been done for the Norwegian data because data on total number of vessels B and total number of catch operations within each vessel i, N_i, has not yet been compiled.

The above estimators are implemented using R survey package (CRAN, Lumley 2010). This includes standard approaches to estimate means, ratios and totals for a probability based design. It also offers standard methods for estimating precision and correlation structures by various methods including analytical estimators for variance where they exist, approximation by linearization (e.g. for ratio estimators), and re-sampling methods such as bootstrapping. Of particular interest in this setting is that a ratio estimator is approximately unbiased if the sampling size is sufficiently large. The variance of the ratio estimator is based on linearization (Taylor expansion) or re-sampling methods. Both approaches depend on sufficient sample sizes and generally suffer if sample sizes (#PSU's within stratum) are small. This is illustrated in Aanes and Hirst (2012) who showed by simulations that the estimates of standard error and the coverage of 80% confidence level both decreased with increasing stratification (<40% coverage for the full stratification for the Norwegian data), keeping the number PSU's constant but varying the stratification. Increased level of stratification thus means a reduced number of samples in each stratum.

For the Norwegian data, domain estimation has not been considered but the principles are outlined as follows: Estimates of a subpopulation domain $\mathcal{D}$ of interest (e.g. a specific métier) is obtained by assigning sample weights of zero for observations outside the domain, whereas sampled individuals within $\mathcal{D}$ retain their original sample weigths. This is achieved by including the indicator function $I[(ijk) \in \mathcal{D}]$ that equals 1 if the kth sampled individual in haul j in PSU i is in $\mathcal{D}$, and 0 otherwise. Effectively this reweights the data according to its actual design to achieve appropriate weights for the domain. This will also affect the variance estimate see for example Korn and Graubard (1999) pp. 207-211 for details. The obvious restriction by domain estimation is that samples for the domain of interest must be present.

Model-based estimation of catch-at-age based on data from all sampling platforms.

A Bayesian modelling framework is used for the estimation of catch-at-age of commercially harvested fish species from the three Norwegian sampling platforms described above (Hirst et al. 2012). The following assumptions are made:

- There is a well-defined PSU, and the PSUs in the data are a simple random sample from some population of units (i.e. collection of all PSUs). The PSUs in the three sampling programs that provides input to the model are vessel; fishing trip, or individual hauls, sets, etc. The fraction of the total PSUs sampled is very low, so sampling with replacement can be assumed;
- a random subsample of fish is taken from each PSU;
- simultaneous measurements of length and age, and of length and weight are taken for the subsample; and
- all biological data are identified to a PSU.

The model can use length-stratified age data or length-only data as long as ages are measured on some of the samples. The most common type of data that cannot be used is an ALK that has no link to the PSUs.

Annex 6.3. A Swedish case study for design-class C: The onshore catch sampling program for Baltic cod.

The Swedish onshore catch sampling program for Baltic cod.

Design	Sampling frame	Stage 1 / (PSU)	STR 1	Stage 2 (SSU)	STR 2
C	Sites * time	Buyer-day	Time stratified by quarter	Box	Size category

Cod in the Baltic Sea is separated in two different stocks: the western Baltic stock (subdivision 22-24), and the eastern Baltic stock (subdivision 25-32). Cod is mainly caught in three métiers; bottom trawls, gillnets and longlines. For assessment purposes these métiers are grouped in active (bottom trawls) and passive gears (gillnets and longlines). Catches from gillnets and longlines are usually landed the same day (i.e trip length = 1 day) whereas the typical trip length for bottom trawlers is 1-5 days. Upon landing, the catch is sorted in accordance with EU standard S1-S5. Due to the scarcity of large cod the size categories 1-3 are usually combined to S1-3. The sampling takes place after the buyers have received the catch from the individual vessels. At the buyer, catches from several trips and gears are combined and arranged by size category. Various practicalities makes it difficult to perform a representative sampling of vessels, trips or gears when boxes of fish by landing size category are selected. For example the accessibility of the landings may vary due to timing of the landing of the day; the practical processing performed by the buyer may influence accessibility to part of the catch, etc. To overcome this, the use of *buyer-day* as a unit of sampling (*PSU*) has been suggested.

A complete list of registered buyers will be compiled at the beginning of the year which will be used to generate a random selection of 10 *buyer-days* per quarter. Within the proposed scheme a complete random sample of buyer-day could be replaced by stratified selection or weighted probability, e.g. by the amount of cod processed by the buyer during the same quarter of the previous year. In this way, larger buyers are able to process a larger amount of the total catch would have a higher probability to be sampled and a higher relevance on the composition of the total catch than smaller buyers. Similarly, the sampling could be spread over the quarter according to an evaluation of prior time distribution of the landings. At each sampling occasion a representative sample of boxes within each size category will be sampled. Initially, following the earlier sampling protocol, the planned number of individuals to sample per size category is; S1-3=250; S4=100; S5=130.

The estimation procedure account for the fact that the *buyer-day* combination is our *PSU*, and a two-steps raising is carried out accordingly. In practice, samples from a sampling occasion (= same *PSU*) are first raised to the whole landing processed by the sampled buyer the day of the sampling occasion, and then to the landings processed by all the buyers in the quarter (= sampling frame). The two-step calculation is done separately for each commercial category. Because of the direct sampling for age strategy (i.e. all measured fish are also aged), no ALK is employed in the catch statistics calculations. Catch statistics for different gears (= domain of interest) will be carried out by post-stratification based on logbook information available by the end of the year.

Annex 6.4. A Scottish case study for design class D

Design	Sampling frame	Stage 1 (PSU)	STR 1	Stage 2 (SSU)	STR 2	Stage 3 (TSU)	Stage 4 (QSU)	STR3
D	Scottish markets	Market-day	4 Markets	Trip	Species, size category	Box	Fish	Stratification for length by age

Here we consider how the estimators described in section 3.3.1 of WKPICS2 would apply to estimation of total Scottish landed numbers-at-age for key demersal species by sampling landed fish at markets. As this is for illustrative purposes, we gloss over some of the practicalities of sampling which complicate the issue, such as: markets which are not sampled; the fact that, for smaller markets, we cannot predict whether a market will take place or not on a particular day; and that the selection of fish sampled for age is not random over sales categories. The sampling frame is Scottish markets. Stratification levels are market and sales category within a trip. Sampling units are: primary sampling unit (SU1) – market-day within the stratum market, secondary sampling unit (SU2) – trip within market-day, SU3 – box within stratum sales category, SU4 – fish measured for length (and age). Fish are measured for age within a stratified sample of fish measured for length, however the lengths are currently stratified over the whole trip rather than each sales category and so we first estimate numbers-at-age at the trip level.

First we define some notation. Let *N* represent the total number in the population, and *n* represent the number sampled, with subscripts to denote the sampling unit or stratum of interest and suffices to denote the variable of interest. For example *N*(*a*) is the number-at-age, and *NM* is number of markets. The sampling procedure is as follows:

1. Sample n_M of a total of N_M markets. (Here we set $n_M = N_M$.)
2. Sample n_m market-days of a total N_m market-days within market *m*.
3. Sample n_{mk} trips of a total N_{mk} within market-day *k* of market *m*.
4. Sample all N_{mki} sales categories of trip *i* within market-day *k* of market *m* (to ensure a complete length distribution).
5. Sample n_{mkih} boxes of a total N_{mkih} within sales category *h* of trip *i* within market-day *k* of market *m*. (Usually, but not always, n_{mkih}=1.)
6. Measure for length n_{mkihb} fish of a total N_{mkihb} fish in box *b* of sales category *h* of trip *i* within market-day *k* of market *m*. (Here $n_{mkihb} = N_{mkihb}$.)
7. Age n_{mkil} fish of a total $n^{(l)}_{mki}$ of length *l* in trip *i* within market-day *k* of market *m*, within a total of n'_{mki} fish sampled in trip *mki*. (Note that $n'_{mki} = \sum_{h=1}^{N_{mki}} \sum_{b=1}^{n_{mkih}} n_{mkihb}$).

To perform the estimation, we first estimate the landed length distribution for that trip, $\hat{N}_{mki}^{(l)}$, where $N_{mki}^{(l)}$ is the total landed number-at-length l for trip i within market-day k of market m. Historically, the usual practice is to first raise sampled numbers-at-length in a category to total numbers-at-length in the category $\hat{N}_{mkih}^{(l)}$, using the inverse proportion of boxes sampled in that category, N_{mkih}/n_{mkih}, then sum over all categories to aggregate to the trip level.

Define $y_{mkihbj}^{(l)}$ such that $y_{mkihbj}^{(l)}=1$ if fish j in box b of sales category h of trip i within market-day k of market m is of length l and 0 otherwise:

$$y_{mkihbj}^{(l)} = \begin{cases} 1 \text{ if fish } mkihbj \text{ is of length } l \\ 0 \text{ otherwise} \end{cases}$$

Then the above procedure can be written as two equations:

$$\hat{N}_{mkih}^{(l)} = \frac{N_{mkih}}{n_{mkih}} \sum_{b=1}^{n_{mkih}} y_{mkihbj}^{(l)} \quad \text{and} \quad \hat{N}_{mki}^{(l)} = \sum_{h=1}^{n_{mki}} \hat{N}_{mkih}^{(l)}$$

which can be condensed into:

$$\hat{N}_{mki}^{(l)} = \sum_{h=1}^{n_{mki}} \hat{N}_{mkih}^{(l)} = \sum_{h=1}^{n_{mki}} \sum_{b=1}^{n_{mkih}} \frac{N_{mkih}}{n_{mkih}} y_{mkihbj}^{(l)}$$

This historical raising process is actually equivalent to the application of equation (3) from section 3.3.1 at the trip level, with the sampling weight for fish j in box b of sales category h (of trip i within market-day k of market m), ω_{mkihbj}, given by $\omega_{mkihbj} = \frac{N_{mkih}}{n_{mkih}}$.

Next we estimate a proportional age-length key, $\hat{p}_{mki}^{(a,l)}$, for trip i, apply it to the length distribution for the trip to get an age-length distribution for the trip, and sum over lengths to get an age-distribution for the trip. The age-length distribution for the total catch in a stratum is then estimated as a weighted average of age-length distributions across the trips within the stratum. This is an alternative approach to estimating catch-at-age, where age-length keys are used only within primary sampling units (see also Hirst et al. 2012). The age data within a trip (e.g. 1 or 2 otoliths per length group) is applied to the length frequency for that trip. This is an alternative approach to the common practice of applying a pooled ALK for many trips (for example by strata) to the raised telngth-frequency distributions (LFDs). The proportional age-length key is simply based on the number-at-age a for a given length l of the stratified sample of length-measured fish over the trip:

$$\hat{p}_{mki}^{(a|l)} = \frac{n_{mki}^{(a,l)}}{n_{mki}^{(l)}}$$

and we apply this to the length distribution to give an age-length distribution for the trip:

$$\hat{N}_{mki}^{(a,l)} = \frac{n_{mki}^{(a,l)}}{n_{mki}^{(l)}} \hat{N}_{mki}^{(l)} = \frac{1}{n_{mki}^{(l)}} \sum_{u=1}^{n_{mki}^{(l)}} y_{mkiu}^{(a,l)}$$

which we then sum over length to estimate an age-distribution, i.e. numbers-at-age a for each trip *mki*, $\hat{N}_{mki}^{(a)}$.

$$\hat{N}_{mki}^{(a)} = \sum_l \hat{N}_{mki}^{(a)} = \sum_l \frac{n_{mki}^{(a|l)}}{n_{mki}^{(l)}} \hat{N}_{mki}^{(l)} = \sum_l \frac{\hat{N}_{mki}^{(l)}}{n_{mki}^{(l)}} n_{mki}^{(a,l)}$$

Return now to the market, and consider the number-at-age of each trip, $\hat{N}_{mki}^{(a)}$. The total landed number-at-age a, $N^{(a)}$, sold at the markets sampled, can be estimated by the weighted estimator given by equation (3) of section 3.3.1:

$$\hat{N}^{(a)} = \sum_{mki} \omega_{mki} \hat{N}_{mki}^{(a)} = \sum_{m=1}^{n_M} \sum_{k=1}^{n_m} \sum_{i=1}^{n_{mk}} \omega_{mki} \hat{N}_{mki}^{(a)}$$

where

ω_{mki} is the sampling weight for trip *mki*.

Now $\omega_{mki} = \frac{1}{\pi_{mki}}$, where π_{mki} is the inclusion probability of trip i on market-day k at market m, and this is given by the probability of selecting trip i on market-day k, multiplied by the probability of selecting market-day k for market m, multiplied by the probability of selecting market m, *i.e.*

$$\pi_{mki} = \pi_m \pi_{k|m} \pi_{i|mk}$$

All markets are selected, then market-days are selected and then trips at the market, so $\pi_m = 1$, $\pi_{k|m} = \frac{n_m}{N_m}$ and $\pi_{i|mk} = \frac{n_{mk}}{N_{mk}}$ and hence $\omega_{mkihblj} = \frac{1}{\pi_{kihblj}} = \frac{N_m}{n_m} \frac{N_{mk}}{n_{mk}}$.

Thus $$\hat{N}^{(a)} = \sum_{m=1}^{n_M} \sum_{k=1}^{n_m} \sum_{i=1}^{n_{mk}} \omega_{mki} \hat{N}_{mki}^{(a)} = \sum_{m=1}^{n_M} \sum_{k=1}^{n_m} \sum_{i=1}^{n_{mk}} \frac{N_m}{n_m} \frac{N_{mk}}{n_{mk}} \hat{N}_{mki}^{(a)} = \sum_{m=1}^{n_M} \frac{N_m}{n_m} \sum_{k=1}^{n_m} \frac{N_{mk}}{n_{mk}} \sum_{i=1}^{n_{mk}} \hat{N}_{mki}^{(a)}$$

Now compare this to the historical practice of "raising". Here the number-at-age at trip level $\hat{N}_{mki}^{(a)}$ is raised to the number-at-age on market day *mk*, $\hat{N}_{mk}^{(a)}$, by summing over trips and then raising by the inverse proportion of trips sampled on that market day, N_{mk}/n_{mk}:

$$\hat{N}^{(a)} = \sum_{m=1}^{n_M} \frac{N_m}{n_m} \sum_{k=1}^{n_m} \hat{N}_{mk}^{(a)}, \quad \text{where} \quad \hat{N}_{mk}^{(a)} = \frac{N_{mk}}{n_{mk}} \sum_{i=1}^{n_{mk}} \hat{N}_{mki}^{(a)}.$$

Next, the numbers-at-age for the sampled market-days, $\hat{N}_{mk}^{(a)}$, are raised to numbers-at-age for the market, $\hat{N}_{m}^{(a)}$, by summing over sampled market-days, and raising by the inverse proportion of market-days sampled from each market *m*, N_m/n_m :

$$\hat{N}^{(a)} = \sum_{m=1}^{n_M} \hat{N}_{m}^{(a)}, \quad \text{where} \quad \hat{N}_{m}^{(a)} = \frac{N_m}{n_m} \sum_{i=1}^{n_m} \hat{N}_{mk}^{(a)}$$

and finally we sum over markets.

So estimation by means of sample weights is a formalization of the historical raising process, with two main differences. First, historically, samples have been pooled together into groups of samples by, for example, gear and quarter, rather than acknowledging each stage of the actual sampling process, such as sampling at markets on a particular day. Second, the inverse proportion of units sampled to total number of unit (*i.e.* the selection probability at each stage if random sampling is used) is used as the raising factor at each stage, rather than the proportion of landed weight sampled to total landed weight which has often been used historically. Proportions of landed weights can be used to estimate the proportions of units sampled if these are not available (see Norwegian case study), but for the Scottish case study we have not ascertained which approach is better when both are available.

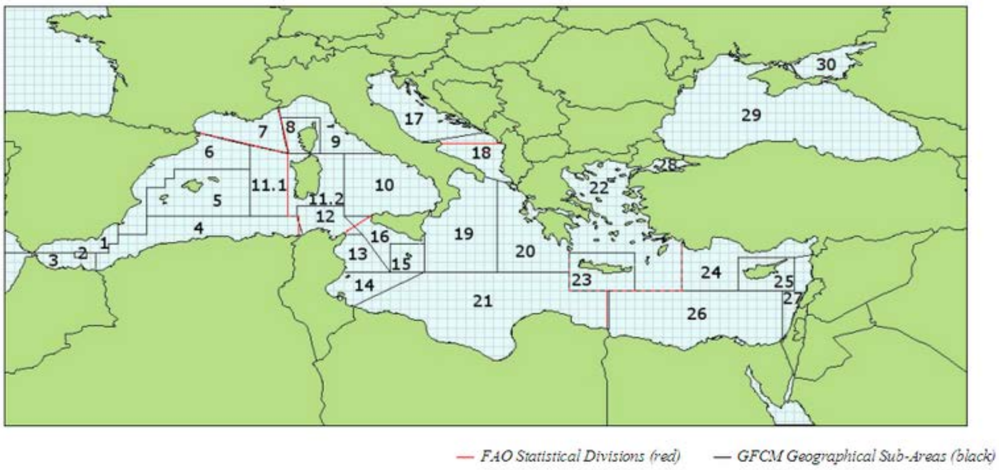
Annex 6.5. Spanish case studies for design class D

Case study 1: Onshore sampling for trammelnets for demersal species in the Balearic Islands

Design	Sampling frame	Stage 1 (PSU)	STR 1	Stage 2 (SSU)	STR 2	Stage 3 (TSU)	Stage 4 (QSU)	STR3
D	Site	Trip	Month					

Population sampled: landings of demersal species by trammelnets.

The sampling design in the Spanish Mediterranean is done at the métier level (level 5), for those métiers selected by the ranking system, as state in the DCF Regulation. This ranking system is performed for each Geographical Sub-Area (GSA, see map). The GSAs are described by the General Fisheries Commission for the Mediterranean (GFCM) as geographically defined zones, in the Mediterranean, Black Sea and connecting waters, used to compile data, monitor fisheries and assess fisheries resources in a geo-referenced manner.



01 - Northern Alboran Sea	07 - Gulf of Lions	13 - Gulf of Hammamet	19 - Western Ionian Sea	25 - Cyprus Island
02 - Alboran Island	08 - Corsica Island	14 - Gulf of Gabes	20 - Eastern Ionian Sea	26 - South Levant
03 - Southern Alboran Sea	09 - Ligurian and North Tyrrhenian Sea	15 - Malta Island	21 - Southern Ionian Sea	27 - Levant
04 - Algeria	10 - South and Central Tyrrhenian Sea	16 - South of Sicily	22 - Aegean Sea	28 - Marmara Sea
05 - Balearic Island	11.1 - Sardinia (west) 11.2 - Sardinia (east)	17 - Northern Adriatic	23 - Crete Island	29 - Black Sea
06 - Northern Spain	12 - Northern Tunisia	18 - Southern Adriatic Sea	24 - North Levant	30 - Azov Sea

Concurrent sampling of trammelnets for demersal species (GTR_DEF métier) in GSA 5 (Balearic Islands) is performed onshore in a unique fish market (Palma de Mallorca). All the landings coming from all the ports of Mallorca Island are sent daily to this single fishing market. Afterwards, they are kept refrigerated during all the night long and that is the moment when the sampling takes place. The sampling happens two days per month during the fishing season (approximately 6 months). Each day, trips (which last one day) are selected randomly. And then, since the landings usually are not very high, all the landings for each selected trip are sampled, taking into account size categories if they are available.

We estimate the annual size distribution for each assessed stock regularly in the GFCM Working Group on Stock Assessment or for the Data Calls to the STECF-EWG on Mediterranean Stock Assessment.

We have weights (w_s) per size class coming from length-weight relationships calculated from previous samplings on board oceanographic vessels during scientific surveys or from samplings carried out in the laboratory. We also have total landings per month in biomass (B_m).

1. We sample N_m months (fishing season).

2. Sample $n_{m,d}$ market-days of a total of N_d market-days in month m, with a total of n_m months.

3. Sample $n_{m,d,t}$ trips of a total $N_{m,d,t}$ trips within market-day d.

4. We define N_s as the number of size categories.

5. Measure for length $n_{m,d,t,s}$ individuals in the size class s from a total $N_{m,d,t,s}$ ($n_{m,d,t,s}$= $N_{m,d,t,s}$).

We calculate the number of fish sampled in each size class s per month:

$$\sum_{d=1}^{n_d}\sum_{t=1}^{n_{d,t}} n_{m,d,t,s}$$

We raise the number of fish sampled to the total number of fish landed per size class for year:

1. We calculate the percentage in biomass of each size class per month:

$$\frac{w_s \times \sum_{d=1}^{n_d}\sum_{t=1}^{n_{d,t}} n_{m,d,t,s}}{\sum_{s=1}^{N_s}\sum_{d=1}^{n_d}\sum_{t=1}^{n_{d,t}} n_{m,d,t,s} \times w_s}$$

2. We multiply by the total landings in biomass B_m per month.

$$B_m \times \frac{w_s \times \sum_{d=1}^{n_d}\sum_{t=1}^{n_{d,t}} n_{m,d,t,s}}{\sum_{s=1}^{N_s}\sum_{d=1}^{n_d}\sum_{t=1}^{n_{d,t}} n_{m,d,t,s} \times w_s}$$

3. We sum the estimated number of individuals landed per size class per month to obtain an estimated annual size distribution.

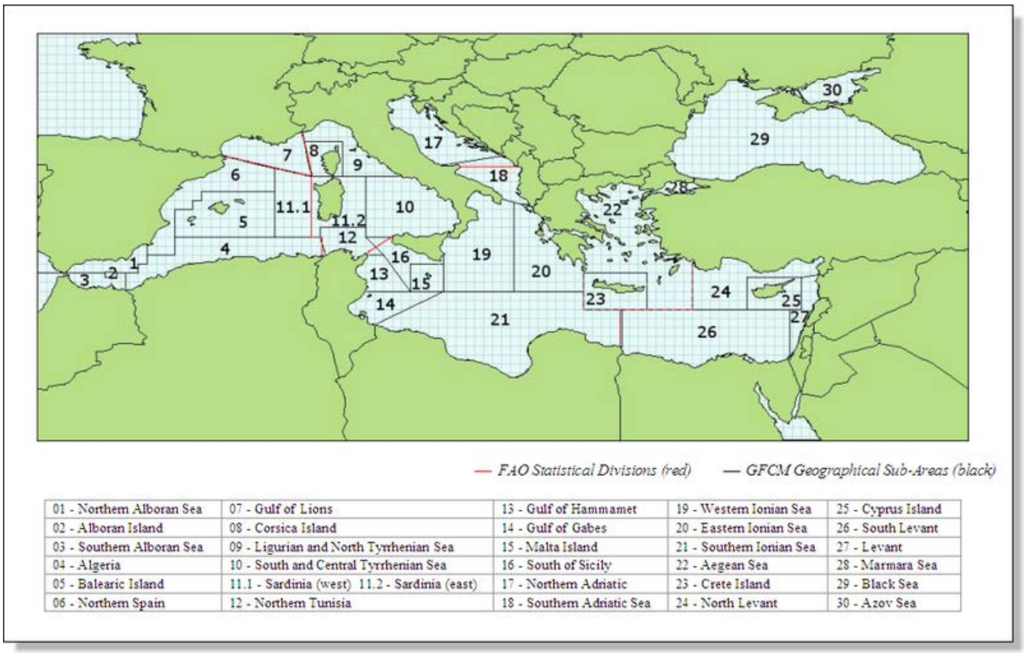
$$\sum_{m=1}^{N_m} \frac{w_s \times \sum_{d=1}^{n_d}\sum_{t=1}^{n_{d,t}} n_{m,d,t,s}}{\sum_{s=1}^{N_s}\sum_{d=1}^{n_d}\sum_{t=1}^{n_{d,t}} n_{m,d,t,s} \times w_s}$$

Case study 2, design class D. Onshore sampling for purse-seine of small pelagic species in the Mediterranean (Northern Spain area)

Design	Sampling frame	Stage 1 (PSU)	STR 1	Stage 2 (SSU)	STR 2	Stage 3 (TSU)	Stage 4 (QSU)	STR3
D	Site	Port	Month	Trip				

Population: landings of small pelagic species by purse-seine.

The sampling design in the Spanish Mediterranean is done at the métier level (level 5), for those métiers selected by the ranking system, as state in the DCF Regulation. This ranking system is performed for each Geographical Sub-Area (GSA, see map). The GSAs are described by the General Fisheries Commission for the Mediterranean (GFCM) as geographically defined zones, in the Mediterranean, Black Sea and connecting waters, used to compile data, monitor fisheries and assess fisheries resources in a geo-referenced manner.



The sampling of purse-seine of small pelagic fish (PS_SPF métier) in the GSA 6 (Northern Spain) is performed onshore. Twenty ports in the GSA6 have landings from PS_SPF and 4 of them were selected, due to their geographical distribution (trying to cover the widest area), large number of boats and landings, and because of port facilities. Based on 2012 data, these 4 ports are among the 6 most important ones in biomass landed (40% of biomass), and among the 8 most important ones in fishing effort (25% of trips, one trip lasts a single day). In each port, 2-3 sampling days are carried out monthly. For each sampling day, trips are randomly selected. From each trip sampled, a box (or more) is randomly selected, taking into account size categories if they are available.

Annex 7. Recommendations

Recommendation	Addressed to
WKPICS3 recommends that the development of statistically sound sampling designs be continued in WGCATCH, and that a strong link between WGCATCH and WGBIOP be established to ensure that statistically sound methods are developed for the collection of samples and estimation of biological parameters from commercial fisheries. For this purpose it is essential that the WGCATCH meeting is attended by experts in this field.	WGCATCH
WKPICS3 recommends that a strong link between WGCATCH and WGBIOP be established to ensure that statistically sound methods are developed for the collection of samples and estimation of biological parameters from commercial fisheries. For this purpose it is essential that the WGBIOP meeting is attended by experts in this field	WGBIOP