ICES PGDATA REPORT 2015

ACOM/SCICOM Steering Group on Integrated Ecosystem Observation and Monitoring

ICES CM 2015/SSGIEOM:06

Ref. SCICOM & ACOM

Report of the Working Group on Data Needs for Assessments and Advice (PGDATA)

30 June–3 July 2015

Lysekil, Sweden

ICES
CIEM
International Council for the Exploration of the Sea
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. 2016. Report of the Working Group on Data Needs for Assessments and Advice (PGDATA), 30 June–3 July 2015, Lysekil, Sweden. Eirini Glyki. 103 pp.

https://doi.org/10.17895/ices.pub.8627

Contents

Executive summary

The ICES Planning Group on Data Needs for Assessments and Advice (PGDATA) met for the first time in Lysekil, Sweden, from 30 June–3 July 2015. The main focus for the group in its first year was the end-use of data and information on data quality in the ICES stock assessment process, particularly the benchmarking of singe--species stock assessments. The PG reviewed previous benchmark stock assessment meeting reports going back to 2009, and also the responses of ICES stock assessment expert groups to data quality questionnaires for discards estimates supplied by Member States in the 2015 ICES data call. An extremely variable approach between expert groups to evaluating and acting upon the quality of data available for the assessments was found. PGDATA drafted, using this background, detailed guidelines for the data compilation and evaluation stage of ICES benchmark stock assessments to encourage a more consistent, transparent and objective approach for data evaluation. The guidelines will be tested using a full data evaluation process for Irish Sea whiting in the Irish Sea benchmark assessment (WKIRISH) in 2016. PGDATA discussed its role in relation to InterCatch, the Regional Data Bases (RDB) and the ICES Data Group, and suggested a prioritized list of work to be conducted in InterCatch. The PG recognizes the potential huge value of the RDB as a tool for end-users to scrutinise the coverage and quality of fishery sampling data, including the evaluation and documentation of data quality for benchmark and update assessments at ICES. PGDATA recommends that funding be made available for further development of the RDB including analysis routines to provide estimates needed for stock assessments or other end use together with diagnostics of the quality of data and estimates. During the meeting the PG addressed a European Commission request on the needs for recreational fishery data, and supported the detailed response of the 2015 ICES Working Group on Recreational Fishery Surveys, but further emphasizing role of RCG / ICES in defining regional needs and sampling plans. Feedback on the role and work programme of PGDATA was sought at the meeting from the chairs of ICES Expert Groups (WGBIOP, WGCATCH) and the regional coordination meetings (RCMs), and the work programme for 2015–16 was reviewed and adapted.

1 Introduction and Terms of Reference

The first PGDATA meeting was hosted in Lysekil Sweden and had 18 participants from ten countries.

PGDATA evolved from the ICES Planning Group on Commercial Catches, Discards and Biological sampling following the restructuring of the group into three separate expert groups, two of which deal with collection, interpretation and quality assurance of data on commercial catches (WGCATCH: ICES, 2014) and on biological parameters (WGBIOP: ICES, 2015). The remit of PGDATA differs from these two EGs in focusing on end-user needs for data and information on data quality.

The background and three-year ToRs and workplan for PGDATA are given in Annex 1, and the detailed ToRs for the meeting are included in the "Summary of the Work Plan" section of the annex and extracted below. The agenda for the meeting is given in Annex 2. The participation at the first meeting (Annex 3) reflected its large focus on documentation of data quality in the ICES benchmark stock assessment process. Attendees included one of the current chairs of the ICES Benchmark Steering Group, as well as people involved in the benchmarking system including one of the responsible scientists for the 2016 Irish Sea Benchmark process (WKIRISH) that is being used as a test case for PGDATA guidelines on data evaluation developed at the meeting. Chairs from the regional coordination meetings (RCMs) also participated in PGDATA. As one of the strong tools to improve data quality is a regional database, time was also allocated to discuss both the regional database as well as the ICES InterCatch database used for most stock assessment in ICES. The two extra ToRs G and H on recreational fishing data and ICES InterCatch and Data Group were added closer to the meeting.

Detailed ToRs for PGDATA 2015:

The **Planning Group on Data Needs for Assessments and Advice** (PGDATA), chaired by Mike Armstrong*, UK, and Marie Storr-Paulsen*, Denmark, will meet in Lysekil, Sweden, 30 June–3 July 2015, to work on ToRs and generate deliverables as listed in the Table below.

a) Review all or a representative selection of previous ICES benchmark and associated data compilation and evaluation meetings to determine how these were implemented, focusing particularly on how (if at all) data quality was evaluated, how this information was utilized at the benchmark assessment meeting, how proposals for new work or data collection were arrived at and prioritized, and where there were shortfalls that need to be addressed through establishing a clearer framework for each type of benchmarking process.

b) Review the responses to the data-quality questionnaires for discards estimates included in the 2015 data call for stock assessment EGs, and how the information was used by the EGs.

c) Using the planned benchmark meeting for the Irish Sea (WKIrish) as a test case, work with the assessment team to identify the data needed, and use this as a test case to develop an initial draft framework and guidelines for compilation and evaluation of relevant data for benchmark assessments, including provision of time-series of data quality indicators (bias and precision) that can be incorporated directly in assessment models or used as supporting information.

d) Clearly define the scope and working practices of PGDATA and identify the working relationships that PGDATA should establish within ICES (e.g. ICES SCICOM/ACOM Steering Groups; survey and other data collection EGs; assessment EGs; ICES DataCentre) and with external bodies.

e) Review and adapt the work programme for the next two years of PGDATA, and develop the ToRs for the 2016 meeting.

f) Consider the need for specific workshops prior to the 2016 core-group meeting, or study proposals to address PGDATA goals.

g) Identify what data on recreational fishery that should be collected, with focus on the spatial and temporal resolution that is needed to support the fisheries management advice.

h) InterCatch and role of ICES Data Group in PGDATA

2 ToR A Review of previous ICES benchmarks

ToRs A–C of PGDATA 2015 address the generic three–year ToR(a) to "*Design and test a Quality Assurance Framework for assessment EGs to evaluate data quality and its impact on assessments, particularly within the benchmarking process, and test this in regional case studies*" (Annex 1). To support this ToR, a review of previous ICES benchmark reports, particularly in relation to documentation of data compilation and data quality was performed.

The ICES benchmarking process is a means of developing peer-reviewed datasets and analysis methods needed for ICES to provide advice to clients. For stock assessment, a full benchmark is carried out for regional stocks at intervals to develop assessments based on the best available data and the most appropriate analytical procedures. The data and methods are then documented in the stock annex as the "recipe" for subsequent update assessments conducted annually or at longer intervals. Currently, the stock assessment benchmark process typically includes two physical meetings, one to evaluate and recommend data inputs for the assessment, and one to agree on an analytical or other stock assessment method using these data. PGDATA is specifically interested in the data evaluation component, to ensure that stock assessments can be carried out with detailed understanding of data quality, and that there is transparency in ICES advice regarding data limitations. From a practical point of view, there are benefits in statistical stock assessment modelling of having objective information on precision, bias or relative accuracy of different datasets being used. This can guide internal weightings where needed based on a-priori knowledge, or in the case of suspected or known biases, to develop alternative data input scenarios for sensitivity testing in the assessment. To facilitate these goals, PGDATA has developed detailed guidelines for the data compilation and evaluation stage, and this is covered under ToR C. However, an initial step was to investigate how well the previous benchmarks have completed the initial data evaluation stage.

In total, 34 completed reports of benchmark workshops were reviewed, covering the years 2009–2015 (Table 2.1). This excludes inter–benchmark protocols. The aim of this particular review was not an in-detail analysis of the benchmark workshops, but to check if particular information regarding data use and data quality were documented in the reports and were easily identifiable. PGDATA developed a short list of questions for a review of all the reports, and a more detailed list of questions that were addressed for a more detailed review of a subset of reports. It is possible that in some cases more detailed information on data quality was available in files on SharePoint but not referred to in the final benchmark report, and was missed in this review.

The short question list was:

i) Do the ToRs refer to a separate data compilation meeting, and is there a specific ToR(s) asking for evaluation of data quality?

ii) Is there an issues list highlighting specific data problems?

iii) Are there Working Documents summarized at the end of the report that deal with the different aspects of data?

iv) The benchmark assessment report will deal sequentially with each type of input data or biology assumptions such as M – do these just analyse aggregated data given to the stock assessor (e.g. evaluating data according to correlations between or within datasets or how well they fit the assessment model), or is there an independent a-priori evaluation of data quality in terms of how the data were collected, quality indicators etc. (which may be in a WD or a reference to an earlier report)?

v) Is there any evidence that datasets are being excluded or weighted according to a-priori evaluation of data quality?

vi) Are there constructive recommendations for improving data quality where there are issues identified?

vii) Have stakeholders attended the workshop and did they provide data?

The results in a nutshell:

i) 24 mention a data compilation workshop in the ToRs, and 4 have reports on them in the actual benchmark report

ii) 22 have the benchmark issue list in the report

iii) 14 deal with specific data issues in Working Documents; 9 have no Working Documents

iv) 31 mention evaluation of data quality; 15 have filled out bias scorecards (WKACCU, 2008)

v) Difficult to evaluate how a-priori evaluation of data quality has been used to exclude or weight datasets in most cases.

vi) 31 documents discuss needs for data improvement, of which 9 are specific, the rest only general

vii) In 24 meetings at least one stakeholder was present

Thus in most cases information on data quality was not easily identifiable, or it was unclear how data quality was evaluated and how low-quality data were dealt with in the benchmarks assessments. The most important document from a benchmark is the updated Stock Annex, describing the data and methods used for a given assessment. Unfortunately it does not document necessarily the changes over time in quality of time-series data used.

Table 2.1 Number of benchmark workshop reports reviewed per year using the short list of questions.

YEAR	NUMBER OF BENCHMARK REPORTS
2009	4
2010	4
2011	3
2012	4
2013	6
2014	6
2015	6

A more detailed list of questions was produced before the meeting and all participants did an evaluation of one or two benchmark reports. The specific questions on this analysis can be found in Table 2.2. Fourteen reviews were carried out, in some cases multiple stocks were partially assessed in one review. In those cases, conflicting answers have been counted individually. The stocks/benchmarks groups in this more detailed review were: WKNSEA2015 (Sole IV), WKCOD2011 (Cod IV), WKHAD2014 (Haddock IV, IIIaW, VIa), Baltic Plaice, WKARCT, WKBALTCOD, WKFLAT2010, WKSPRAT (Sprat IV), WKPELA02 (Sardine VIIIc), WKROUND2013 (Irish Sea, haddock), WKROUND2013 (Whiting, North Sea, 2 reviews), WKBALFLAT2014, WKICE (reports can be searched on the ICES website).

Table 2.2. Summary results of the detailed review of a selection of 14 ICES Benchmark assessment reports

QUESTION	RESPONSE	RESULTS (X=NO REPLY)
Was there an issues list documented in the report ("Benchmark information per stock")	1: yes 2: no	q1: 11 q2: 3 x: 0
Was there a separate data compilation and evaluation workshop (DCWK)?	1: yes, in a separate report 2: yes, in an annex to assessment report 3: No	q1: 5 q2: 3 q3: 6 x: 0
Is there evidence that the data compilation and evaluation was done in advance of the meeting e.g. presented as working documents?	1: yes 2: no 3: partially	q1: 8 q2: 3 q3: 3 x: 0
Did stakeholders attend the meeting and contribute data?	1: yes 2: no	q1: 8 q2: 6 x: 0
Were references /links provided to documentation of data collection methods, quality assurance procedures, analysis methods, operational manuals?	1: yes 2: no 3: partially	q1: 2 q2: 3 q3: 9 x 0
Did the group collate and describe new datasets and quality indicators that have become available?	1: yes 2: no 3: partially 4. N/A	q1: 5 q2: 4 q3: 4 q4: 1 x: 0
Did the group delete old datasets from the assessment as a response to bad quality?	1: yes 2: no 3: N/A	q1: 4 q2: 7 q3: 3 x: 0
Was there an evaluation of the quality of historical landings data including underreporting and misreporting, with advice on how to deal with this in assessments?	1: yes 2: no 3: partially 4. N/A	q1: 6 q2: 6 q3: 2 q4: 0 x: 0
If historical landings data are very uncertain – were alternative plausible scenarios presented that could be used by the assessment WG?	1: yes 2: no 3: partially 4. N/A	q1: 2 q2: 5 q3: 3 q4: 5 x: 0
Was there a description of the design of port-sampling schemes to estimate length/age composition of landings, and how they have changed over time?	1: yes 2: no 3: partially 4. N/A	q1: 1 q2: 8 q3: 4 q4: 0 x: 1

QUESTION	RESPONSE	RESULTS (X=NO REPLY)
What methods were used to document the quality of historical length / age compositions of fishery landings for the full time-series?	1: Numbers of trips sampled 2. Numbers of fish sampled 3: Standard errors or equivalent 4: Bias indicators 5. None 6. N/A	q1: 4 q2: 4 q3: 0 q4: 1 q5: 8 q6: 1 qx: 0
Was there a description of the design of discard sampling schemes and how they have changed over time?	1: yes 2: no 3: partially 4. N/A	q1: 1 q2: 4 q3: 9 q4: 0 qx: 0
What methods were used to document the quality of historical discards estimates for the full time-series?	1: Numbers of trips sampled 2. Numbers of fish sampled 3: Standard errors or equivalent 4: Bias indicators 5. Not given 6. N/A	q1: 1 q2: 2 q3: 3 q4: 0 q5: 6 q6: 4 x: 0
If appropriate, were historical recreational fishery catches documented and their quality evaluated?	1: yes 2: no 3: partially 4. N/A	q1: 1 q2: 3 q3: 2 q4: 8 x: 0
Was there a description of how the design and vessels for fishery-independent surveys such as trawl surveys have changed over time, including vessel intercalibration studies where applicable?	1: yes 2: no 3: partially 4. N/A	q1: 3 q2: 2 q3: 9 q4: 0 x: 0
How was the quality of research vessel surveys documented over the time-series?	1: Numbers of tows / stations / transects 2: Standard errors or equivalent 3: Bias indicators such as coverage 4. Not given 5. N/A	q1: 3 q2: 1 q3: 1 q4: 7 q5: 1 x: 2
If used, were the methods for deriving fishery-dependent (CPUE or LPUE commercial tuning fleet) data over the full time-series clearly explained?	1: yes 2: no 3: partially 4. N/A	q1: 3 q2: 2 q3: 3 q4: 6 x: 1

QUESTION	RESPONSE	RESULTS (X=NO REPLY)
Was advice given on potential biases in LPUE/CPUE series, for example trends in vessel efficiency or effects of legislation?	1: yes 2: no 3: partially 4. N/A	q1: 3 q2: 2 q3: 3 q4: 6 x: 0
If an age based assessment has been conducted has there been an analysis of the age readings performance (between and within countries)?	1: yes 2: no 3: partially 4. N/A	q1: 7 q2: 6 q3: 1 q4: 0 x: 0
Was there a review of candidate values for natural mortality with full explanation of the evidence in support values proposed (e.g. life history; multispecies models.)?	1: yes 2: no 3: partially 4. N/A	q1: 6 q2: 3 q3: 5 q4: 0 x: 0
Was evidence provided for the choice of maturity ogive (proportion mature at age / length) with a description of data sources and how representative the sampling is of the population?	1: yes 2: no 3: partially 4. N/A	q1: 3 q2: 2 q3: 7 q4: 0 x: 2
Have any changes in the environment that could explain changes in the stock development been described?	1: yes 2: no 3: partially	q1: 5 q2: 5 q3: 4 x: 0
Are any large changes in the regulation (gears, mesh sizes, closed areas) that would affect CPUE, catch-at-age or other fishery data described?	1: yes 2: no 3: partially 4: N/A	q1: 5 q2: 5 q3: 3 q4: 0 x: 1
Did the group provide advice on how to improve future data collections if data quality is seen to be inadequate?	1: yes 2: no 3: partially	q1: 7 q2: 4 q3: 3 x: 0

3 ToR B Review the responses to the data-quality questionnaires for discards estimates

Since 2014, ICES has been committed to provide total catch advice, except when the rate of discarding is known to be significant but not quantified. To be able to provide catch advice, discards estimates are needed. This information could be either on the knowledge that discards are negligible, or on the estimated discard rate obtained from sampling. For several stocks a long time-series of good quality discard estimates is available and included in the analytical assessment. However, for many other stocks the knowledge of discard rates is limited. In 2015, ICES launched a data call requesting the necessary data to conduct the update assessments. In the data call the discard estimates for the last three years (2012–2014) were requested, to ensure that the catch advice was provided also for stocks that do not include discards in their assessment.

The ability to give total catch advice also depends on the quality of the discards estimates, and ICES therefore asked the chairs of PGDATA to develop a data quality questionnaire that would have to be completed by all stock coordinators, both for stocks where discards are currently used, and those stocks for which discards estimates are available but not used. The approach of the PGDATA chairs was to develop two tables to be completed. The first describes the designs of national sampling schemes for discards estimation and how these have changed between years (if at all), and highlighting potential for bias in the sampling design. The second table provides statistics on sampling coverage such as numbers of trips sampled and how many of those trips had catches and discards of the species in question, and includes estimates of precision where available. The two tables and the instructions for completion are included in Appendix 1.3 of the Annex 4 (benchmark data evaluation guidelines) in this report.

PGDATA 2015 carried out an analysis of how these tables were completed and used by stock assessment Expert Groups (EGs). This analysis was based on all actively assessed stocks in 2015 (some stocks assessed in 2014 were not assessed in 2015) in the remit of the ICES Working Group for the Bay of Biscay and the Iberic Waters ecoregion (WGBIE), the Working Group for the Celtic Seas Ecoregion (WGCSE), the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK) and the Baltic Fisheries Assessment Working Group (WGBFAS). A complete analysis of all the ICES stocks from all EGs was not possible due to time limitation and because some ICES stocks would not be assessed until after PGDATA in 2015. The four EGs represent the majority of the stocks that are or will be under the EU Landing Obligation, for which discards are significant (defined in this document as more than 5% of total catch).

This section compares the use of discard information to provide catch advice in 2015 (advice on fishing opportunities for 2016) in relation to advice given in 2014 (advice for 2015), from the four EGs reviewed. From a total of 75 stocks, for which ICES provided in 2014 and in 2015, only 46 stocks have the same discard information in 2014 and 2015. Note that the advice for some stocks (mainly in the Celtic Seas ecoregion) will not be addressed until October 2015. The use of discard data in the ICES advice was categorized as "used", "not used, unknown or unquantified", and "not used, negligible". The number of stock for which discard information was used has increased in 2015 in all EGs considered, except the WGBIE.

For almost all the stocks in each EG in 2015 where the discard data were used for the first time, the self-evaluation of potential bias on the discard data were given as 1 or 2 (from a scale of 1–3, being 1 the best) (Table 3.1). However, there are a few countries /

stocks for which the self-evaluation was 3 (higher possibility of bias) and the discard data were still used to provide catch advice. For most of the stocks for which discard data were used for the first time to provide catch advice in 2015 (advice for 2016), the last three years, or the last year (2014), of data was used to "top up" the landings advice.

Table 3.1. Overview of the use of discard data in 2014 and 2015. The analysis was performed based on four ICES Expert Groups. The Ecoregion is a "proxy" of each EGs but a few stocks cover more than one Ecoregion.

		Ecoregion	Baltic	North Sea	Celtic Sea	Bay of Biscay and Iberian	Total
		Expert Group	WGBFAS	WGNSSK	WGCSE	WGBIE	
		Total number of stocks in this analysis	16	23	21	15	75
		Number of stocks with the same information in 2014 and 2015	10	10	16	10	46
Use of discard data	used	2014	4	9	8	4	23
		2015	8	22	10	4	44
	not used, unknown or unquantified	2014	8	10	7	6	31
		2015	3	0	5	5	13
	not used, negligible	2014	6	4	6	5	21
		2015	5	1	5	6	17

Ideally, sufficient time-series of discards data should be included directly in an analytical stock assessment model to allow the fishing mortality and selectivity associated with discarding to be estimated and forecasted, and changes over time to be detected. However if discard rates are relatively high but the quality of the time-series data are poor (e.g. due to small sample sizes, biases in sampling schemes, large changes in sampling design over time), or if the dataseries only cover a few recent years requiring extensive hindcasting of recent apparent discard rates, assessments may be badly degraded by inclusion of discards data. On the other hand, where high rates of discarding occur over several age classes, including where there is highgrading that may have altered over time, exclusion of discards data may result in assessments that are biased and fit poorly to relative abundance indices. This emphasizes the need for a detailed evaluation of the quality of discards data, both in the benchmarking system and also for ad-hoc use of data for topping up to give total catch advice. Under its ToR C, PGDATA includes detailed guidelines for evaluation of the quality of discards estimates in the benchmarking process.

Landings Obligation requirements to land all catches of designated species, and to self-report all discards of species with exemptions due to *de minimus* or high survivability, raise a new set of concerns over the accuracy of catch data for components that are discarded or landed as damaged or below minimum size. This must be considered when using the data in assessments in future, and for evaluating any changes in selectivity that the LO is intended to induce.

4 ToR C Develop guidelines for data quality indicators in Benchmarks using the WKIRISH as test case

4.1 The ICES benchmarking process flow

Comprehensive guidelines were developed by PGDATA to provide ICES benchmark data evaluation teams with suggestions for tasks that should be completed prior to and during the benchmark data evaluation meeting. The guidelines cover all types of data and biological parameters commonly used in stock assessments. For some benchmark assessments, only some of the data types and parameters may require full evaluation depending on the issues list for the benchmark, or if previous benchmark data evaluation workshops have carried out a full evaluation which only requires an update with more recent data. The guideline headings are included in this section but for a full documentation of the guidelines see Annex 4. The guidelines include a description of tasks required for each data or parameter type, which will also relate to a specific Term of Reference for the benchmark data evaluation, and a set of supporting information in a set of appendices.

However, not only the work tasks but also the benchmark process was discussed during the PGDATA meeting. As a benchmark process is rather time consuming, the data need to be ready and available for a full process to be conducted. The two key meetings – the data evaluation meeting and the benchmark assessment meeting – are typically only 4–5 days long, and can only be fully effective if the ICES and external experts at the meetings are able to work with documentation of data quality evaluation and trial assessments completed in advance of the meetings.

In some cases the work required to document data quality may require skills not present in the laboratories carrying out the benchmarking, for example stock identification, statistical survey design, fishing technology or evaluating quality of age data. In these cases, ICES Expert groups on the relevant topics should be able to provide advice and should be consulted at an early stage. This can require an initial benchmark issues list to be developed in which the questions to be posed to the EGs are given, and a revised issues list is produced after the EGs have provided their advice (See Annex 4 for an example on biological parameters).

The co-chair of the ICES Benchmark Steering Group attended PGDATA and developed a work process flow for the entire benchmark exercise, indicating also the possibility to terminate the process if the expected progress has not been met (Figure 4.1). This highlights the extended nature of the process and that it is not just a couple of workshops where all the work is done.

A major issue for benchmark data compilation and evaluation is the supply of data from national laboratories, in time for the work to be done and in the format needed. Data may already be captured in InterCatch and DATRAS for example. However, to dig deeper into the designs, implementation and analysis of data collection schemes, and to evaluate the quality of the data, there will often be the need for metadata, documentation and detailed data not held by ICES. Data that have not previously been used may also become available but with no previous record on ICES databases. To facilitate this process, there is a clear need for software systems that can readily provide the information and diagnostics needed for the compilation and evaluation of data. Examples include the development of Regional Data Bases from which national fishery sampling metadata and detailed data can be sourced, GIS or other mapping tools, the WebGR software to facilitate evaluation of quality of age data, and other diagnostic tools such as the from the EU COST project, which may also be linked to or built into the RDB development.

PGDATA fully supports the development of such enabling technologies to improve the efficiency and scope of data evaluation tasks, and includes a recommendation in Annex 7 for continued funding support for the RDB and WebGR.

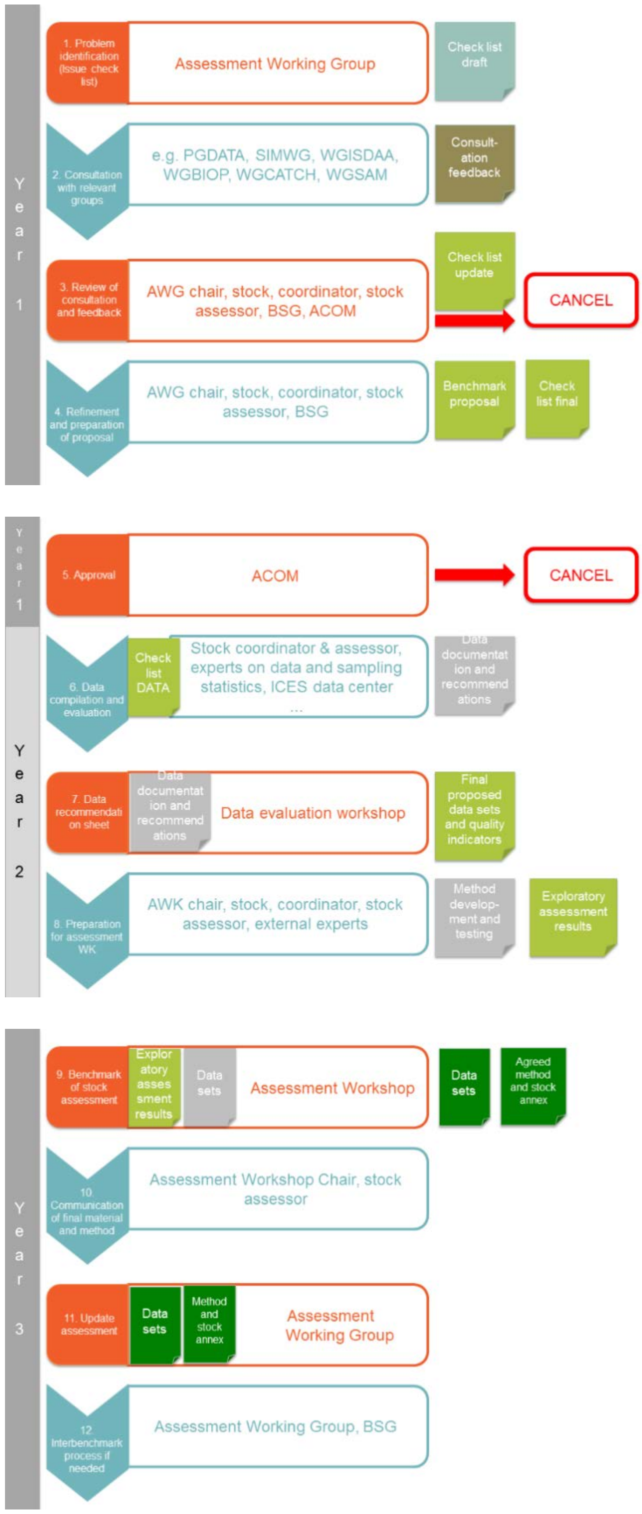


Figure 4.1. The suggested 3 year ICES benchmark process for stock assessments

4.2 Summary of data types for benchmark data evaluation guidelines

The following is a summary of the individual data types and processes that are covered in the PGDATA guidelines given in Annex 4. These guidelines will be trialled in detail for Irish Sea whiting as part of the WKIRISH2 benchmark data evaluation process which will culminate in the data evaluation workshop in autumn 2016. However the use of the guidelines is not restricted to this and can be used for any other benchmark assessments before then, so that more extensive feedback for improvements can be made.

4.2.1 Stock Identification

This is to explain the basis for existing assumptions on stock structure and mixing rates between stock areas, or proposed new assumptions which form the basis for spatial aggregation of fishery and survey data and/or adjustments to datasets to account for stock mixing. This is an area where the ICES Working Group on Stock Identification Methods (WGSIM) can provide advice. However, revision of stock boundaries can have major implications for collation of data and for the quality of data for the new stock definitions and this has to be accounted for as well.

4.2.2 Review and recommend life-history parameters

Life-history parameters (e.g. growth parameters, maturity ogives, fecundity, natural mortality), are used extensively in assessment models and for calculating biological reference points. This section of the guidelines deals documentation of the basis for the choice or range of parameters, including models to describe growth, maturation, and fecundity by age, sex, or length, and providing quality diagnostics. The ICES Working Group on Biological Parameters (WGBIOP) may be approached for advice.

4.2.3 Describe the history of fishery management regulations

Information is needed on management regulations and actions that are expected to have caused changes in the quality of fishery catch data or the selectivity patterns of fisheries that are of relevance for the scientific assessment of the stocks and provision of advice. The guidelines cover the documentation of regulations that are generic to all species taken by the fleets in the stock area, or are specific to individual species. Expert advice from the ICES Working Group on Fishing Technology and Fish Behavior (WGFTFB) may be needed.

4.2.4 Develop time-series of catch estimates with bias and precision indicators.

Fishery catch data are key inputs to most stock assessment procedures. This guideline addresses the retained or discarded fishery catch for all types of commercial and recreational fishing. Separate data evaluations are needed for catches that are recorded exhaustively (e.g. landings logbooks), and for those estimated through sampling schemes (e.g. discards and recreational catches). Clear documentation of data collection methods is needed, and how these may have changed over time. In many cases the fishery catch data have uncertainties caused by biases related to the design of the data collection schemes or to implementation error such as non-response, misreporting, species identification errors or mixed species categories, or conversion factors. The guidelines cover such issues as well as provision of precision estimates or indicators. The benchmark team may need to seek advice and assistance from ICES EGs with expertise in statistical sampling survey design and analysis, such as the Working Group on Commercial Catches (WGCATCH) and the Working Group on Recreational Fishery Surveys (WGRFS).

4.2.5 Estimate the length and age distributions of fishery landings and discards, with bias and precision indicators.

Fishery age compositions, and in some cases length compositions, are also a key input to many stock assessment procedures. As with estimation of catches by surveys, the description and evaluation of additional sampling surveys to estimate length and age composition, and evaluation of data quality, can be complex. This aspect of data collection can be combined with catch estimation when seeking assistance from ICES Expert Groups dealing with such surveys such as WGCATCH. Assistance from the ICES Working Group on Biological Parameters (WGBIOP) should be sought in relation to quality of age estimates, where needed.

4.2.6 Fishery selectivity (pattern of catchability at length or age) in the assessment model.

Most age-based or length-based stock assessment models require some assumptions about selectivity i.e. how catchability varies with size or age in fisheries. Selectivity in this context is a combination of the selectivity properties of fishing gears of different design, and factors influencing the probability of fishing operations encountering fish of different sizes and ages, for example related to distribution of fishing or behavior patterns of the fish. Knowledge of the selectivity characteristics of the gears, and the distribution of fishing relative to the population of fish of different sizes, can help identify plausible selectivity patterns where this is needed in the assessment model. The ICES Working Group on Fishing Technology and Fish Behaviour, WGFTFB, may be able to provide advice on this.

4.2.7 Recommend values for discard mortality rates, where appropriate, and indicate the range of uncertainty in values.

ICES assessment EGs have, for most assessed stocks, assumed that all discards die. The potential for dispensations from the EU landings obligation for species with high discard survival has resulted in a range of studies on the mortality rates of fish and shellfish discarded or released alive from fishing operations. Many recreationally caught fish are also released alive after capture and have variable survival rate depending on a range of factors such as deep hooking, bleeding and water temperature. There are numerous published studies on post release survival of marine species, though relatively few are from Europe. Increasingly, ICES assessment EGs will need estimates or inferences of mortality of discarded or live-released fish caused by the fishing operation. Post-release mortality in recreational fisheries is an ongoing topic covered by the ICES WGRFS which can provide advice. The ICES workshop on Methods for Estimating Discard Survival, WKMEDS, has covered some of the issues.

4.2.8 Review all available and relevant fishery-independent and dependent data sources on fish abundance

4.2.8.1 Fishery-independent data

Assessment EGs make extensive use of research surveys to provide absolute estimates of abundance, or more commonly, relative abundance indices, for tuning length or age based stock assessments. In many data-limited assessments, surveys provide the main source of information on stock trends. Survey data may be used as size/age-aggregated indices or as length or age based indices. Some assessment models require the parameters of the selectivity pattern of a survey at length or age to be fixed or estimated, and for indicators of data quality such as CVs or effective sample sizes to be input to the model separately for the total abundance indices and the length or age compositions. The guidelines indicate how surveys currently used in the stock assessment or new relevant surveys should be reviewed and recommendations given of which series are considered adequate and reliable for use in stock assessments. As with survey estimates of fishery catches and catch compositions, the evaluation of fishery-independent survey data can be complex and will require outputs and support from expert groups dealing with design and implementation (e.g. International Bottom Trawl Survey Working Group, IBTSWG) and those dealing with interpretation and end-use of survey data (e.g. Working Group on Improving use of Survey Data for Assessment and Advice, WGISDAA).

4.2.8.2 Fishery dependent data

Fishery dependent abundance indices continue to be used for some stocks, with or without fishery-independent data, and may be the only information available on stock trends for some data-limited stocks. Assessment and advisory groups need to understand the limits imposed by the quality and resolution of such data, and guidelines on aspects to be covered in a benchmark data evaluation are given. Advice from the ICES Working Group on Fishing Technology and Fish Behavior (WGFTFB) should be sought in evaluating the suitability of a fleet for providing abundance indices and for evaluating issues such as technology creep.

4.2.9 Ecosystem changes

Longer term or episodic/transient changes in environmental drivers known to influence distribution, growth, recruitment, natural mortality or other aspects of productivity and which are relevant to assessments and forecasts should be considered. There are potential circumstances where the data inputs to an assessment model, or the assumptions in the model, need to take into account environmental drivers. These may be episodic or transient phenomena such as mortality or changes in fish distribution caused by low-oxygen water or lethal temperature events, or longer term trends in environmental conditions. The data evaluation team should source and review existing information and make recommendations on how this information should be used by the assessment team.

4.2.10 Recommendations for research

Research and development is a key aspect for improving scientific advice. It is important that the benchmarking process has access to the results of completed studies and information on progress in research that has been recommended at an earlier stage to support the assessment process. The data evaluation team should provide a review of existing recommendations for research to develop and improve the input data for the assessment, and what has been achieved. If work is still ongoing, describe progress, problems encountered, how these will be resolved and expected finalization of the work. If this cannot be progressed, consider a recommendation that the work should be stopped.

During the data evaluation workshop, proposals for changes to data collection or needs for new data or studies may be identified. The workshop must identify the relative priorities of the recommendations and expected impact on the quality of the assessment, and take into account feasibility.

4.2.11 Overview of recommendations of the Data Workshop.

It is important that the data evaluation workshop provides a clear and accurate documentation of all the datasets and input parameters that are recommended and supported by the documented evaluations. The guidelines recommend developing a spreadsheet of assessment model input data and parameters that reflects the decisions and recommendations of the data evaluation workshop, covering all aspects of data and parameter estimates covered in the guidelines. This will include quality indicators such as age-error matrices and time-series of CVs or sample sizes that are needed for input to the assessment model, in addition to plausible ranges of parameters such as M, and alternative catch histories where needed. The spreadsheet should also document any data that were evaluated by the data evaluation team but not recommended for use. This is a key output of the data evaluation process. The benchmark assessment workshop will use this table to indicate which data were used, and explain why any of the data are not used or are modified.

4.2.12 Data evaluation workshop report

The guidelines recommend that the report of the data evaluation workshop should be finalized and agreed, and the spreadsheet of recommended assessment input data completed, within two weeks of the end of the workshop. This is to allow the stock assessment team time to evaluate the recommendations, seek any clarification from the data evaluation team, or conduct any of their own analyses if they disagree with the findings of the data evaluation workshop. The report should provide complete documentation of workshop actions, decisions, list of working documents, other information used by the workshop, and a list of any additional tasks to be completed following the workshop with dates and responsibilities for completion.

The data evaluation workshop report and Excel table of recommended inputs should stand as separate documents alongside the assessment workshop report with both being available from the same ICES web page.

5 ToR D define the scope and working practices of PGDATA

PGDATA had in advance of the meeting asked relevant working groups within ICES on what role they could imagine PGDATA could have in relation to their own group. A short summary from each presentation is included in the report. At the meeting the advantage of mapping the connections between different ICES Working Groups was discussed, particularly within the ICES Steering Group on Integrated Ecosystem Observation and Monitoring (SSGIEOM) which covers the data EGs dealing with surveys, fishery data, fishing technology and fish biology.

5.1 Working Group on Biological Parameters (WGBIOP)

A main objective of WGBIOP will be to support the development and quality assurance of regional and national provision of biological parameters as reliable input data to integrated ecosystem stock assessment and advice, while making the most efficient use of expert resources. All National Age Reader/Maturity Stager Coordinators (ICES and GFCM) have been invited as well as relevant external experts such as statisticians, sampling designers or specific EG members. The participant-list, however, do not at first glance represent the needed expertise to answer the ToRs for WGBIOP 2015 and thus the chairs approached PGDATA for advice on how to attract the necessary expertise. A suggestion was to map out the needed expertise for the ToRs for 2016 with the experience from the 2015 meeting and address ACOM/SCICOM fora in due time before the deadline for registration for WGBIOP 2016.

The WGBIOP chairs presented the current detrimental status of the calibration tool WebGR, which will cease to work by the end of 2015 unless appropriate action is taken. Keeping this tool viable is fundamental for the quality assurance of biological parameters and is used pan-European by all institutes participating in calibration exercises and also within National institutes for internal quality assurance. An immediate rescue of the system is estimated to be rather low budget demanding (in the area of €6000). However in order for ICES to take over the system, which is the way to ensure future use of the system, a larger amount of funding is needed. PGDATA decided to advocate the rescue of WebGR both within the ICES recommendation system (see recommendations in Section 10), but also at the RCMs WGBIOP will further develop a plan for the continuation and upgrading of WebGR in the format of a proposal for a two year project.

5.2 ICES DATA centre

5.2.1 InterCatch

InterCatch is the backbone of the data used for the full analytic stock assessment. This year there has been an increased use of InterCatch, which mean more stocks have good documentation of the data at the level being reported. More than 55 stocks had data in InterCatch in 2015. The following assessment working groups were using InterCatch: AFWG, HAWG, NWWG, WGBFAS, WGBIE, WGCSE, WGDEEP, WGNSSK, WGEF, WGHANSA, WGWIDE, WGMIXFISH.

5.2.1.1 Developments

The environment around InterCatch is changing and new demands are always present. Through the last year the following functionality have been developed and implemented

- Status of data in InterCatch
- Copy allocation for a different year
- Rewrote copying under extract (safer)
- View imported data
- Creation of the validation file
- Import delete more user friendly

5.2.1.2 New developments

The new requests for functionalities are prioritized in the list below

- New landings obligation (discard brought to the harbour for some species and areas) – how will this affect IC, and how should IC deal with the new landings obligations.
- Make it possible to see and delete all wrongly Import data
- Overviews of stock progress in InterCatch for EG chairs
- Easy and multiple deletions of set up of unsampled strata's allocations for age and length distributions
- User guided information in IC, to easily guide and help the users
- Quality checks: 1) Age and length range checks. 2) Mean weight and mean length range check per age or length per stock. Make the validation file interactive when processing the data though IC. View Sums of products (SOP) with stratum overview and in allocations
- Make it possible to use age and length in parallel (dealing with potential CATON (catch numbers-at-age) differences)

It is of high priority to make sure InterCatch fulfil the requirements of the working groups, when they are depending on the data in InterCatch for the assessment. Therefore the focus is to concentrate on the changes regarding the new landing obligation.

5.2.1.3 The Role of PGDATA in connection with InterCatch

Previously the Planning Group on Commercial Catches, Discards Biological Sampling (PGCCDBS) had InterCatch under its Terms of Reference, but after the changes of groups under ACOM, InterCatch is now under PGDATA. This means that status of InterCatch and information of priorities of new developments are given to the PGDATA. The PGDATA should support InterCatch and help in making sure InterCatch can support the working groups. PGDATA can suggest extra developments and in dialog with ICES Secretariat change the order of prioritization of the development, to make sure InterCatch support and fulfil the needs of the working groups.

5.2.2 Regional Data Base RDB

The Regional Data Base (RDB) should be used for raising of the national data before it is transferred to InterCatch. The ICES Secretariat is funded by the European Commission (EC) to host and maintain the RDB, but the EC are so far not funding developments. The Regional Coordination Meetings (RCM) for the Baltic Sea, for the North Sea & Eastern Arctic and for the North Atlantic are all using and depending on the data from the RDB. The RCMs send out a data call each year for upload of data into the RDB. The RCMs use the data in the RDB to plan and coordinate sampling in the three regions. The RDB can also have a key role in facilitating the evaluation of fishery data and provision of diagnostics and quality indicators for the ICES benchmark data evaluation process, representing a deeper level of detail and potential functionality than in InterCatch which accepts data already raised by fleet. The use of software routines integrated with the RDB can free up the ability to interrogate data and provide diagnostics of data quality and statistically raised estimates with greater flexibility. This would represent a major advance in the efficiency and effectiveness of the data compilation and evaluation process for fishery data.

5.2.2.1 ICES funding development of the RDB

ICES Council approved in September 2014 funding for an extra person working on the RDB for 1½ year. The focus has been to develop and implement a new check to validate métiers depending on areas. New reports used by the RCM have also been developed. The implementation of consistent harbour codes and changing to the Location codes LOCODE under the EC Master Data Register will soon be completed. The change have been implemented but there are still some harbours to map to the new codes. Later the species code system will change to the WoRMS species codes.

5.2.2.2 Governance of the RDB

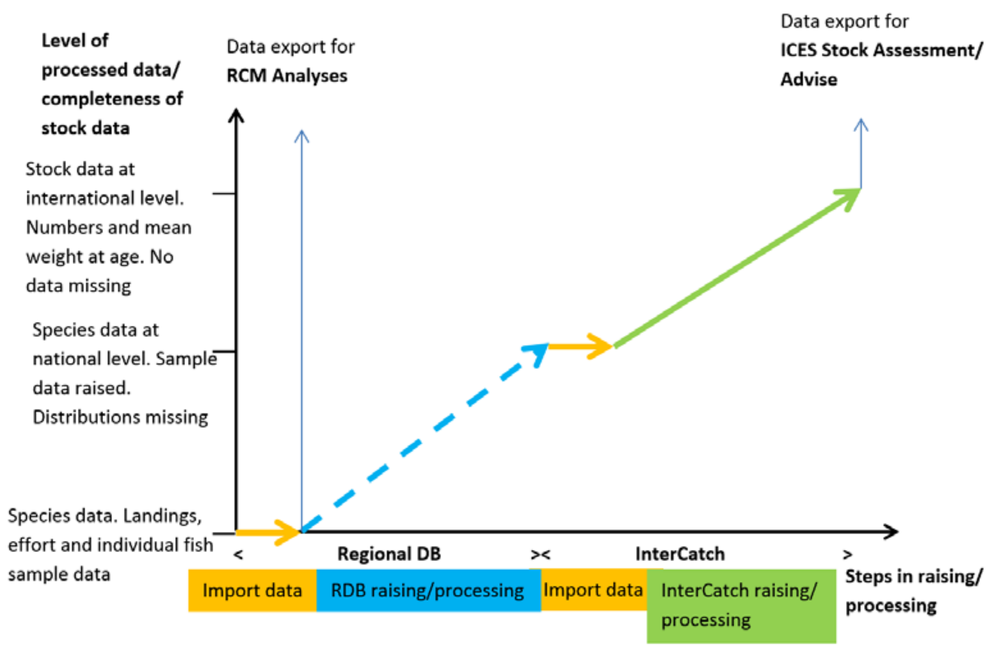
The Steering Committee of the RDB (SCRDB) has the following governance of the RDB: Technical governance, Strategic planning, Operational issues and Estimates of costs. The Liaison Meeting (LM) have the following input to the SCRDB: Prioritize between the suggestions for development from the RCM's and were needed formulate some of the ToRs on the SCRDB agenda. The RCMs have the following tasks: Content governance, Prioritize and develop road maps for data uploads, Monitor general problems with data uploads/data processing and report that to SCRDB for action, Suggest areas for development and Appoint people to SCRDB. The RCMs write recommendations to the LM.

5.2.3 EC's feasibility study on data systems

In 2013 EC started a feasibility study on "Scientific data storage and transmission under the 2014–2020 Data Collection Multi-annual Program". A consultant company looked into all systems and organizations involved in fisheries data in 2013 and 2014. From ICES Secretariat's side DATRAS, InterCatch and the RDB were analysed. The conclusion from the feasibility study was that a 'Scenario 4' was the preferred scenario, where the RDB and InterCatch was in ICES Secretariat, with a structure close to the one existing today, but with some streamlining of data flows within a data hub. The EC published a follow-up call for tenders in 2016 for a "study on availability and dissemination of DCF data" to develop a prototype data hub in which thematically specialized databases (biological data - RDB, fleet economic data and fisheries activity data) would be linked so that biological and economic data can be aggregated in the forms needed to supply a wider range of end-users. In the ICES community it was a big question, when this proposal was first mentioned by the EC, why such a study must be completed before any further, badly needed development of the existing RDB can be funded. It had been expected that after the first feasibility study the EC would support development of the RDB, but this will now be delayed further.

5.2.4 The RDB and InterCatch together

The overall goal is to combine the RDB and InterCatch, so there is full documentation of data and processes behind the ICES advice. Today we have two systems, which are not completely streamlined (see the overview of the two systems below). The reason for the blue dashed RDB rising line is because, so far the RDB is not yet widely used for data raising from samples to the fleet.



The overall idea is to streamline the data flow going from the RDB to InterCatch, which would ensure a full documentation of data and ensure only approved standardized raising methods have been used throughout the processes.

PGDATA is strongly supportive of the continued development of the RDB for the reasons given above, and recommends continued funding to improve the functionality (see recommendations Section 10).

5.3 ICES Secretariat

PGDATA was recommended by the Planning Group of Commercial Catches, Discards and Biological Sampling (PGCCDBS). The driving decision for splitting PGCCDBS into the three separate EGs, two of which (WGCATCH and WGBIOP) previously worked as subgroups, was to address workload issues at PGCCDBS. The recommendation for a new expert group with a different scope focusing on end-use of data was well received by the ICES Advisory Committee (ACOM). It was important to recognize that the first PGDATA meeting should reflect the scope and role of the new Expert Group (EG) and plan future work.

The key topic for PGDATA is data quality analysis for the end use of data; both on fisheries dependent and independent data. The impact of the data quality in the assessment and the advice should also be addressed in the incoming years. PGDATA should also play a key role on setting the priorities and advice for development on the ICES databases that store such data (e.g. InterCatch, Regional Database). As it was already this year's case, PGDATA is a key expert group to provide the basis for the advice on data collection (in relation to the special request on recreational catch).

PGDATA is closely linked to the ICES Strategic Plan, essentially in the area <u>of improving data collection and use</u>:

a) Coordinate and integrate surveys;

b) Develop guidelines for best practice in design and implementation of statistically sound catch sampling schemes;

c) Identify the data required to provide advice on fisheries and environmental issues and communicate the requirements to those responsible for the collection of data;

d) Promote efficient and effective data storage through integration of data in regional databases, including making data available for experts through InterCatch.

and on <u>developing the scope of advice</u>:

a) Integrated ecosystem assessment

b) "New" areas, such as aquaculture

c) Integrate considerations of bycatch in the advice for fisheries (including elasmobranchs, mammals, and seabirds)

d) Facilitate transition from single-stock benchmarks to regional benchmarks.

The link of PGDATA and the other EGs under the Steering Groups on Integrated Ecosystem Observation and Monitoring (SSGIEOM), the fish stocks assessment groups, and the benchmark process are crucial.

6 ToR E Review and adapt the work programme for the next two years of PGDATA

It was decided during the PGDATA meeting to focus next year's work plan on the data needs for the quality assurance framework (QAF) and how this needs to be implemented in ICES Expert groups. Furthermore the group should work on cost benefit analysis of data both during a workshop (See ToR F) and also with the focus on the impact of the data in stock assessment.

Year 2016	29 February–4 March	San Sebastian Spain	Interim report by 4 April 2016 to SSGIEOM, SCICOM & ACOM	1 more day – (5) lunch to lunch
Year 2017			Final report by "DATE" to SSGIEOM, SCICOM & ACOM	

Summary of the next two years Work Plan

Year 2	Planning of future work, including a workshop, to develop procedures for evaluating data needs in terms of impacts of data quality and cost–benefits of data through development of MSE or similar tools, focusing on the relative impacts of data improvements on quality of stock assessment estimates and associated advice; Guidelines to other SSGIEOM EGs on QAF implementation; Further development and testing of QAF procedures in benchmarks; Consultations with data end-users such as stock assessors on role and operation of PGDATA. Terms of Reference for the second annual meeting of PGDATA will be: i) Plan the June 2016 PGDATA Workshop on cost benefit analysis of data collection in support of stock assessment and fishery management (WKCOSTBEN), taking into account outcomes of the EU project DG MARE/2014/19 "Strengthening Regional Cooperation in the Area of Fisheries Data Collection" ii) Review outcomes of consultations, to be done prior to PGDATA meeting, with ICES SSGIEOM chair and EGs on implementing the SSGIEOM ToR to "Promote the development within EGs of standards and guidelines for good practice in data collection covering the design and implementation of surveys, fishery and other related data collection programmes, the archiving and interpretation of data and samples, the analysis of data, provision of data quality indicators, and the documentation of procedures". iii) Using the 2015 benchmark data evaluation meeting for the Irish Sea (WKIRISH) as a test case, work with the assessment team to identify / and review the benchmark process and modify the guidelines for benchmark data evaluation meetings if required. iv) Develop actions in response to pre-meeting consultations with end-users on PGDATA role, including the potential roles for PGDATA to provide expert support to the Regional Coodination Group process under the revised Data Collection Framework v) Respond to recommendations and requests for advice from other ICES Expert Groups, RCMs or other bodies. vi) Plan the ASC theme session on "when is enough – enough" in connection to the ongoing activities in PGDATA. vii) Map the skills required for the PGDATA future work programme. viii) Develop a strategy for collaboration between PGDATA and WGISDAA (ICES WG on integrating survey data in assessments and advice) on topics of common interest. ix) Develop the PGDATA workplan for 2017.
Year 3	Review of progress / results in implementing QAF; further implementation in benchmarks; Methodological Workshop – developing and testing criteria for evaluating data needs and requests; consultations with end-users on data needs; 3rd PG meeting; evaluate future PGDATA workplans.

7 ToR F future workshops

The following workshop was proposed by PGDATA, now accepted, on the subject one cost benefit analysis of data collection in support of stock assessment and fishery management (WKCOSTBEN). This was one of the topics for the second year of PGDATA. Subsequent to the meeting, some PGDATA members also submitted a proposal for a theme session in the 2016 ICES Annual Science Conference entitled "When is enough, enough?" which will seek papers from the wider marine science community dealing with the propagation of errors in assessments and advice, and the cost–benefit of data collection. Outcomes of the WKCOSTBEN can feed into this ASC session.

2015/2/SSGIEOM:06

The **Workshop on cost benefit analysis of data collection in support of stock assessment and fishery management (WKCOSTBEN)**, chaired by Mike Armstrong*, UK and Jon Helge Vølstad*, Norway, will meet in ICES HQ, 28 June–1 July 2016 to:

a) Propose options and analytical methods for an objective framework to evaluate the benefitsvs.costs of datasets used to support stock assessment and fishery management advice, where the benefits are in terms of accuracy (bias and precision) of assessment results and derived management variables, and risks to stocks associated with management under uncertainty. This framework should be able to evaluate existing datasets, new data requests from end-users, and options for focusing elements of funding, survey design, spatial and temporal coverage, and sampling effort towards components of data collection that have greatest influence on quality of assessments and management decisions for particular stocks or groups of stocks.

b) Identify a range of stocks for detailed case studies, including those with full analytical age-based assessments and data-limited assessments, and contrasting stock status and biology. Describe the data used in the assessments, the design of fishery-dependent and fishery-independent sampling surveys providing the data, including hierarchical cluster sampling designs and analytical methods for quantifying precision reliably. Evaluate sampling rates and allocation for given survey designs that are required to derive estimates with adequate precision. Specify how simulations of the sampling schemes could be used to relate precision to sampling intensity and costs.

c) Develop a proposal for a longer term (3-year) project to develop a general methodological framework and open-source software to carry out cost–benefit analysis and provide proof of concept using the case study stocks. Identify potential sources of funding.

d) Identify the need for follow-up workshops in 2017 onwards in the event of no funding for a dedicated project.

WKCOSTBEN will report by 7 August to the attention of the ACOM, SCICOM, and PGDATA.

Supporting Information

Priority	This workshop is considered to have a very high priority for establishing data requirements under the DCF and for ensuring the cost effectiveness of data collection.
Scientific justification	International agreement to exploit all stocks at MSY means that a range of assessment methods is needed to determine MSY reference points and stock status relative to these, including for many data-limited stocks. This will lead to requests for improved or additional data that may not be feasible within existing DCF and national budgets for data collection. It is imperative that objective methods are developed to allow the most cost-effective use of data collection funds to help achieve these management goals. This may involve identifying areas of data collection that have relatively large influence on ability to assess the stocks and those that have relatively little influence, and the costs of collecting these data. Where new data are requested, it must be possible to make an informed judgement on the benefits these will bring to the assessments and management in relation to the feasibility and costs of data collection. Without such a decision framework, the ability to achieve MSY goals may be unnecessarily impeded. This framework will help the European Commission and its Regional Coordination Groups to make informed decisions on regional data needs under the revided DCF and help coordination between countries.
Resource requirements	The principal resource requirements are people with the skills needed for the workshop. Historical data needed for the case study evaluations are already collected and must be made available.
Participants	To be arranged
Secretariat facilities	Some secretarial support will be needed.
Financial	Member States may fund this through their EMFF programme..
Linkages to advisory committees	ACOM and SCICOM
Linkages to other committees or groups	PGDATA, WGCATCH, WGRFS, WGBIOP, WGISDAA.
Linkages to other organizations	RCMs

Theme session proposal

The Theme session proposal for the 2016 ICES Annual Science Conference entitled "When is enough, enough?", chaired by Mike Armstrong, UK, Marie Storr-Paulsen, DK, Jim Ianelli , USA and Jon Helge Vølstad, Norway.

Description

In recent years, there have been increasing demands on data with documented quality to support stock assessments and advice as well as to support the ecosystem approach to management. An increase in the volume and complexity of data collection and the need to meet quality standards within logistical and economic constraints requires prioritization and optimization of the national as well as regional sampling programs through:

- a better regional coordination,
- improved survey sampling designs and analytical methods,
- development of new technology
- quality documentation and,
- cost–benefit analysis
- uncertainty in assessments

The main focus of this theme session is the wide range of data including new datasets that ICES uses to support its stock assessment and advisory process. The session will invite papers on approaches to evaluate the quality of datasets, how to use them as effectively as possible, and objective methods to identify and prioritize data needs. It is essential that the quality of the data, and how this impacts the accuracy of key parameter estimates that are the basis for advice, is understood before they are used to support fisheries stock assessments and advice as well as other end-users.

Stock assessments are based on data from fisheries-independent as well as fisheries-dependent sampling surveys with inherent uncertainty due to sampling errors and various sources of systematic errors (bias). It is important to quantify how errors in input-data propagate through assessments to help identify the most cost-effective data collections and sampling efforts that adequately support assessments and advice or other management processes.

In later years statistical assessment models (such as SAM) have been developed which can account for sampling errors and the high degree of complexity in the input data.

This session aims to bring together fisheries scientists and statisticians with expertise in survey sampling design and analysis, practical experience with data collections, stock assessment modelling, harvest control rules, simulation studies and statistical analysis to assess our current ability to quantify uncertainty in input data, and to track how uncertainty in input data propagates through stock assessment models to affect harvest rules.

Papers are welcomed in the following areas:

Data collections to support the ecosystem approach with cost-effective designs and documented quality. Simulation studies to test and develop sampling designs, particularly in a regional context.

Sampling and analysis methods for fisheries-dependent and fisheries-independent surveys that follow best scientific practice to provide data for stock assessments

Methods that reflect assessment uncertainty by allowing for uncertainty in the historical catch estimates and other key population parameters typically assumed as fixed or measured without error.

Incorporation of sampling errors in input data in the assessment model, evaluation of model fit to observation data, and how this can be integrated in a stock assessment outputs.

Demonstration of how management decision is affected by uncertainty in survey data and stock assessment. Central to this will be papers that demonstrate a feedback loop showing how improved survey design can lead to reduced variance, improved stock assessment model prediction and a subsequent reduction in uncertainty associated with implementing a specific harvest rule, such as one based on MSY.

Suggested theme session format

A good mixture of talks and discussions sessions is proposed for this session over a period of 1 day. A keynote presentation will be used to define the challenges, and followed by a set of oral presentations covering different themes to address these challenges (optimization through a better regional coordination, survey design, statistical assessment modelling, and cost–benefit analysis

Expected participation *

This session will appeal to a wide range of participants including: fisheries scientists involved in monitoring and assessment, statisticians working on survey methods and statistical assessment models, fisheries managers, policy-makers, and non-governmental organizations (e.g. producer organizations). This should attract national, European and international stakeholders.

Linkages to ICES Strategic Plan *

This theme addresses three pillars from the ICES strategic plan: building a foundation of science (Pillar 1), producing the information and advice decision-makers need (Pillar 2) and underpinning the science and advice through data and information service (Pillar 3).

This has a very strong link to Goal 2 through coordination of data collection, and Goal 3 through ensuring quality assurance, transparency, and documentation of data and goal 4 through promoting the data and data quality for science and advice needs on both regional and subregional levels.

Linkages to ICES Steering Groups and/or Advisory Committee (if relevant) *

The theme is linked to several SCICOM/ACOM Steering Groups such as The Benchmark Steering Group (BSG) through the focus of documentation on the data quality, on Integrated Ecosystem Observation and Monitoring (SSGIEOM) through the optimizing monitoring needs and ACOM/SCICOM Integrated Ecosystem Assessments (SSGIEA) through the validation of the quality in the data to be incorporated. It will also feed into ACOM through the assessment groups where the quality in data needs to be documented and validated to support of a trustworthy advice

8 ToR G Special request from EU Identify what data on recreational fishery that should be collected.

8.1 Details of request from the Commission

PGDATA was provided with an additional Term of Reference to help ICES develop advice to the European Commission on the need for, and use of, recreational fishery data. The request from the Commission was as follows:

DG MARE kindly requests that ICES provides advice on how data needs for monitoring the recreational fisheries should best be defined to meet expected end-user needs. This advice should be delivered the latest by 21st August 2015 and address the following questions in detail:

1. What are the drivers for the collection of recreational fishing data?

2. What recreational fishery data (biological, economic & fisheries activity) are needed to support the scientific advice?

3. How will these data be used in stock assessment and fishery management advice?

4. What spatial and temporal resolution of data is needed to support fisheries management?

8.1.1 Response by ICES WGRFS in June 2015

This request was considered during June 2015 by the ICES Working Group on Recreational Fisheries Surveys (WGRFS; ICES 2015) which developed a detailed response to the four questions which can be read in the report of that meeting. The summary in the WGRFS response gives the following recommendations:

- The need to include recreational fishery data in a stock assessment procedure should be evaluated on a case-by-case basis, according to the known magnitude of catches compared with commercial catches based on previous surveys or pilot studies. This should be reviewed regularly as recreational catches can fluctuate significantly between years and recreational effort can remain high even where stock are depleted.
- The types of surveys being conducted for the successful management of shared stocks need to cover the stock area and thus need to be agreed at a regional level. Precision targets should be set at the overall stock level for combined international estimates, and bias in data collection and estimates should be documented. Data collection requirements should be evaluated by regional coordination groups and WGRFS before being ratified by the European Commission. This approach mirrors regional coordination of commercial fishery sampling.
- Where recreational fishing surveys exist, multispecies data should be collected as the costs are not significantly greater than for single species data collection.
- To facilitate the inclusion of recreational fishery data in stock assessments, an annual frequency of data collection is needed over a number years to develop time-series of recreational mortality that comprises of both kept and released components of the catch.

- Biological data on catches (size or age composition) is required both for caught and released components if catch at size or age is needed for an assessment model.
- In order to make optimum allocation decisions between commercial and recreational sectors, it is also necessary to collect information on the economic value and social benefits of recreational sea fishing. However, this is unlikely to change quickly, so a frequency of every five years is appropriate and be driven by end-user needs.

The rationale underpinning these recommendations is discussed in detail in the WGRFS document.

8.1.2 PGDATA response

PGDATA reviewed the WGRFS document (ICES 2015) and agrees with the content and the overall recommendations of WGRFS. However PGDATA has the following additional comments:

- PGDATA considers that the frequency of recreational fishery surveys in a specified region (e.g. annual, biennial or longer intervals), addressed in WGRFS recommendation 4, should be evaluated on a case by case basis. For some stocks, recreational fishery removals may be at such a level that having estimates at longer intervals than a year may unacceptably degrade the quality of the assessments and the advice, and impede achievement of management goals such as MSY. At the other extreme, there may be stocks where recreational fishing mortality (F) is very low compared to the total catch, and it would be sufficient to have some form of survey at longer intervals. This would detect any trend of increasing recreational fishing mortality or proportion of total F attributable to recreational fishing, particularly on a stock that is declining in abundance. Where recreational catch estimates are small but very imprecise, adding recreational catch estimates may reduce precision of the assessment for only gaining a small reduction in bias due to exclusion of the data. This would need to be evaluated. It is likely that the frequency of surveys in a region will be determined by the data needs for stocks that are most impacted by recreational fishing, however adding more stocks to an already conducted survey will result in relatively small additional costs.
- PGDATA agrees strongly with WGRFS that recreational catch estimates are needed for a much wider range of stocks than currently listed in the DCF requirements. Recreational fishery impacts may be greatest on species occurring mainly in inshore waters, and could include species that are desirable targets but may be at relatively low abundance or are vulnerable due to aggregation in areas or around seabed features easily located and targeted by fishing. This could include protected, endangered or threatened species. Regional integrated ecosystem assessments will require knowledge of such impacts.
- ICES may need to include estimates of recreational fishery removals in data-limited stocks as well as data-rich stocks. A key difficulty at present is that, with few exceptions: 1) estimates are only available for a very few recent years; 2) are provided for only the small number of species for which the DCF requires estimates; 3) have incomplete international cover-

age (e.g. for Baltic cod and Area IV/VII sea bass); and 4) are not well coordinated between countries. This poses a great challenge for including recreational fishery catch data in assessments and forecasts. As an alternative recreational data can be used for "topping up" catch forecasts where the recreational fishery data cannot be included in an assessment - this is currently done for a number of stocks in connection to discards – but loses the ability to correctly apportion historical F where this is estimated within an assessment model from age or length composition data. There is an urgent need to develop and test robust methods for including short time-series of recreational catches in assessments and advice. This could be done by an expert group or by funding a contract. For the few cases were a time-series of recreational catches is available, ICES will be able to give separate advice on the recreational fishery.

9 ToR H InterCatch and the priorities off future work

It is the opinion of PGDATA that a database with access to raw data is preferred (a function that already exists in the RDB). Several functions requested below for InterCatch (IC) are currently available in RDB. IC could then be used for storing the final data used for assessment but all data quality checking and raising could then be produced in the RDB. Data handling for benchmarks and exploration of data would be enhanced by having access to raw data. Therefore, PGDATA agreed that the supply of a fully functional RDB by ICES should have highest priority (See recommendations Chapter 10). Further improvements on IC should not be carried out if they detract effort that can be used to advance the development of the RDB. Currently data uploaded to the RDB is not used by the ICES expert groups but only by the RCMs, where the data call deadline is after the main part of the expert groups. If RDB could be used on a regular basis for the expert groups a large part of the recommended improvement for IC would not be needed. To avoid duplicate work PGDATA suggests that a test stock (Baltic cod – WGBFAS) should be used to investigate if RDB can be used for some of the quality data checking already in the 2016 data call.

9.1 InterCatch tasks in a prioritized order

1. **New landings obligations (discard brought to the harbour)** – how will this affect IC, and how should IC deal with the new landings obligations. (Relevant to all WG). This has been taken into account by the data centre and will be implemented in IC in 2016

2. Make it possible to see and delete **all wrongly Import data.** (Relevant to all WG). A large improvement has happened on this issue in 2015

3. **Raise discards in strata with no landings**. (Relevant to all WG). This is currently possible in RDB were several raising methods are available.

4. Make it possible to **use age and length in parallel** (dealing with potential CATON differences). (Relevant to WGBFAS).

5. **Standard output with five tables/graphs diagnostics for each stock. Numbers of age readings per country, weight at age per country** (stock coordinators to show the chair). (Relevant to all WG). This is currently possible in RDB.

6. Overviews of **stock progress in InterCatch** for EG chairs (Relevant to all WG).

7. Easy and **multiple deletions** of set up of unsampled strata's **allocations** for age and length distributions. (Relevant to all WG).

8. **Quality checks**: 1) Age and length range checks. 2) Mean weight and mean length range check per age or length per stock. Make the validation file interactive when processing the data though IC. View SOP with stratum overview and in allocations. See more details under section "data quality". (Relevant to all WG). This is currently possible in RDB.

9. **User guided** information in IC, to easily guide and help the users. (Relevant to all WG).

9.2 Data requirements

- The new landing obligation has been introduced in 2015 in the Baltic and will in the coming years be introduced in all EU waters. This will require a new function in IC – below minimum reference size (BMS) on the same level as landing and discard. If this function is not present the landed fraction will not be comparable between years. (ICES is currently developing this and seeking advice on how best to handle BMS and damaged fish)
- Data of different formats (e.g. different CANUM types) for the same strata should be able to be stored so that they can be used for different assessment models, e.g. length and age based data for Eastern Baltic Cod
- In cases with zero landings but discards, a function should be applied to raise discard data to another variable such as (all landings, effort etc.), a function which is already available in RDB. This is important for some economically less important stocks and as ICES is increasing the numbers of stocks to be assessed this will be an increasing requirement
- Possibility to raise discards using more than the ratio of the LAN/DIS of the same species in the same stratum (e.g. effort and all species, as already provided in RDB Figure 9.1). This is especially important in stocks, where the discard ratio is independent of the landed amount.

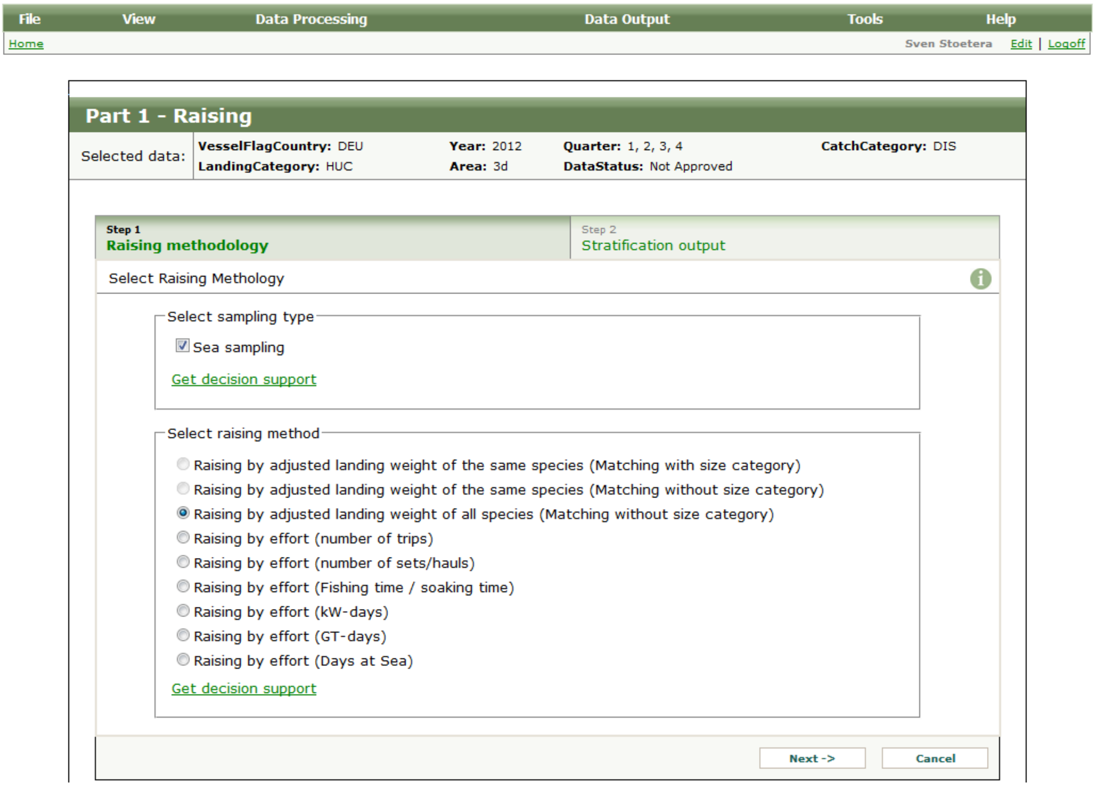


Figure 9.1. Example demonstrating a variety of borrowing options for discard raising in RDB

9.3 Data quality

- Check functions should be provided by InterCatch after having uploaded the data for the national data submitter (NDS). This could be in the form of a table with landings uploaded per area (e.g. Subdivision and Quarter. The NDS should check the correctness of landings. It is currently possible to see the amount per uploaded stratification, but it would be very nice with a more arrogated view e.g. just per area or just per quarter for checking.
- A comparison (plot function) of landings of current year vs. previous year.
- Re-uploading data should be followed by a message to Stock Coordinator on the details of the changes.
- Saving of discard allocation scheme should be possible.
- The evaluation of the data quality of harbour sampling programmes could be improved by changing the way the sampling is reported. The number of boxes sampled per EU size sorting category per stratum should be reported, not only the total number of boxes sampled. This can be cross-checked with the landings information of EU size sorting category level which will be requested in the data call (see above). Again the function is available in RDB
- Overview plots should be provided, giving an overview of missing information (e.g. discard weights or age samples) and a rough differentiation in the amount of biological samples (e.g. above/below 20 ages in a given stratum). This could be done using a traffic-light plot (Figure 9.2).In the RDB there are 4 report with this – called extrapolation options (target/source)

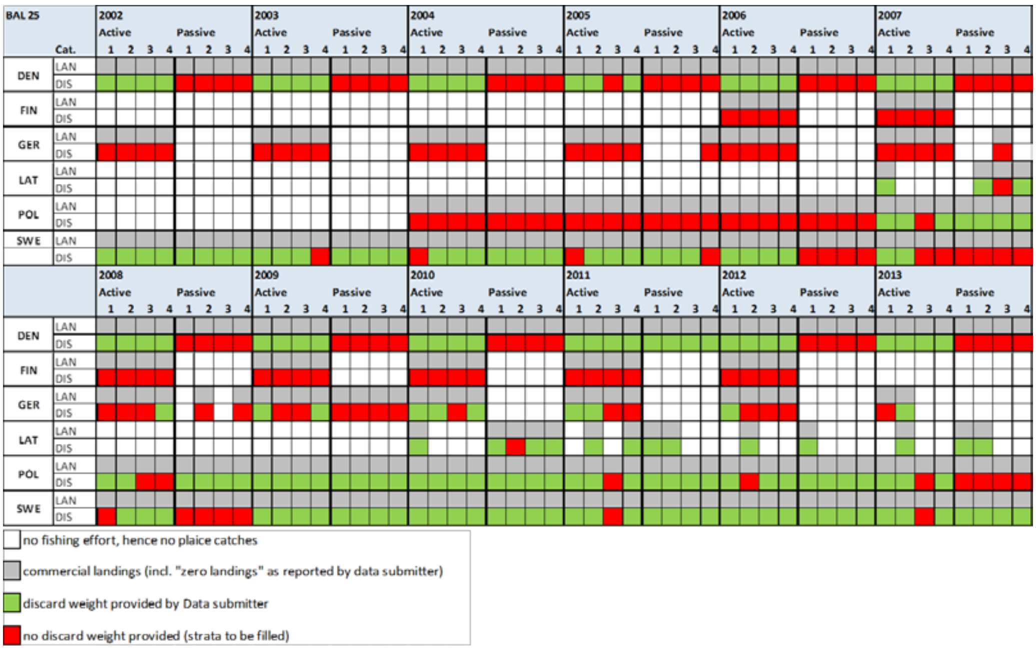


Figure 9.2. Example for a traffic light plot to give an overview about sampling coverage in discard estimations.

9.4 Functionality

- Interface should be more user-friendly, for instance all windows should be made resizable to view full data on larger screens.
- The order of working steps in the menu should follow the order of execution (e.g. the "Finalize" step should come at the end).
- Easy downloads of multi-annual data (e.g. CANUM, CATON by quarter, fleet, subdivision for several years back in time) should be possible, e.g. to facilitate the preparatory work of SC prior to benchmarks.
- Language of instructions and warning messages should be checked for correctness and easiness of understanding. See Figure 9.3 for an example of an unclear message.

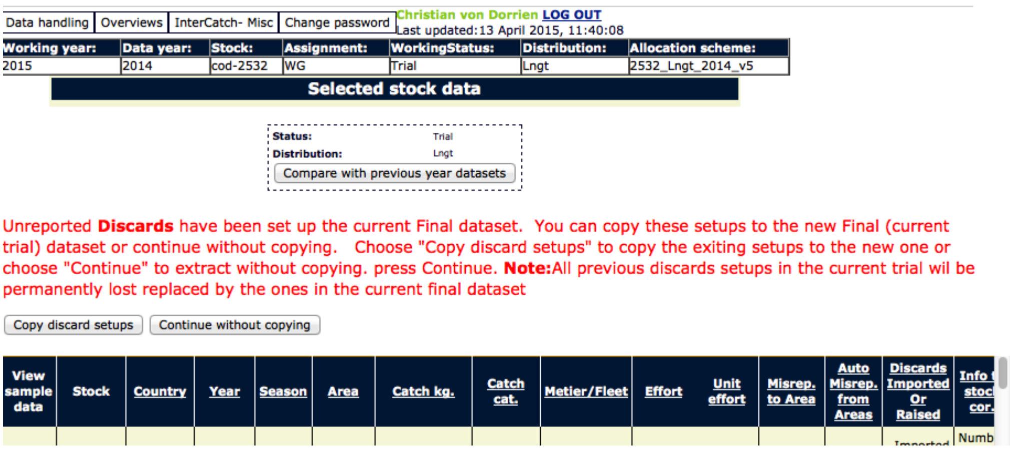


Figure 9.3. Example of an error message where wording as well as instructions to solve the issue are hard to understand.

- Sorting of viewed data should be possible directly in interface, e.g. allowing to view and delete double entries.
- Output tables and standard graphs should be developed in accordance to the need of the AWG. This might be developed similar to RDB, where a Pivoting function exists (see Figure 9.4), as well as a select-tree to generate graphs and tables in an efficient and fast way directly within the database

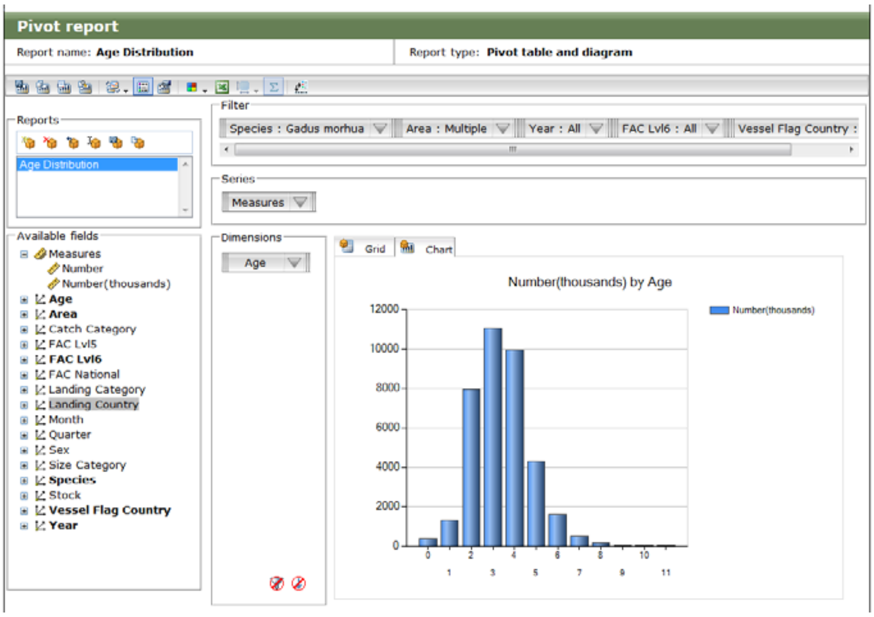


Figure 9.4: Example of the Pivot report function in RDB (here: Age-distribution of cod-2224 for all countries and years combined). Left hand selection tree allows multiple possibilities for further exploration of data (e.g. ages per country and quarter in a given SD or size and landing category)

- Test datasets. To have access to more than one final version of data, e.g. in cases of sensitivity analyses.
- A function is needed that allows holding and merging of more than one dataset (stratum) per year, together with the possibility to select the needed strata when extracting data. This possibility would be required for sensitivity analyses, e.g. alternative datasets with discards twice or three-times those reported, age information raised by data from only one country, landings raised by data from one country
- It should be possible to de-activate selected sampled strata when the data from a country are considered unreliable (e.g. due to poor sampling level etc.) in order to allow the allocation of data from sampled strata from other countries which a more reliable sampling scheme. These changes of course have to be documented by the program.

10 Recommendations

PGDATA has one recommendation.

Recommendation	Addressed to
PGDATA recommends that the European Commission considers the need to fund software tools initiatives when developing calls for proposals for regional studies in support of the CFP. PGDATA stresses the vital importance of quality assurance procedures for the collection and interpretation of data to ensure that fishery management decisions under the CFP are based on the best possible and most cost-effective evidence. Further to ensure that the underlying stock assessment and supporting analyses are carried out with clear documentation of the quality and limitations of input data. Such quality assurance and documentation can be complex and require sophisticated methods and expert knowledge. The development of software tools to facilitate this process is an essential component of the quality assurance and quality control process, and for providing data quality indicators and a framework to support regional cooperation in data collection. Examples already supported by EU funding include the COST tools, the fish ageing tool WebGR, and the Regional Databases in their current form. The Regional Databases require further investment to develop analysis routines supporting statistically sound sampling design and provision of data diagnostics and quality indicators needed for regional coordination and stock assessment. WebGR is an image analysis tool that has proved extremely effective for supporting quality assurance of fish ageing during international workshops and exchanges but requires funding to migrate to ICES as a host, to fix known bugs.	European Commission

11 References

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

ICES. 2015. Report of the Working Group on Commercial Catches (WGCATCH), 10–14 November 2014, ICES HQ, Copenhagen, Denmark. ICES CM 2014/ACOM:36. 295pp.

ICES 2015. Report of the Working Group on Recreational Fishery Surveys (WGRFS). ICES CM 2015/SSGIEOM:10

ICES. 2015. First Interim Report of the Working Group on Working Group on Biological Parameters (WGBIOP), 7–11 September 2015, Malaga, Spain. ICES CM 2015/SSGIEOM:08. 67pp.

Annex 1: Terms of Reference for PGDATA 2015

The **Planning Group on Data Needs for Assessments and Advice** (PGDATA), chaired by Mike Armstrong*, UK, and Marie Storr-Paulsen*, Denmark, will meet in Lysekil, Sweden, 30 June–3 July 2015, to work on ToRs and generate deliverables as listed in the Table below.

	Meeting dates	Venue	Reporting details	Comments (change in Chair, etc.)
Year 2015	30 June–3 July	Lysekil, Sweden	Interim report by 31 July 2015 to SSGIEOM, SCICOM & ACOM	
Year 2016			Interim report by "DATE" to SSGIEOM, SCICOM & ACOM	
Year 2017			Final report by "DATE" to SSGIEOM, SCICOM & ACOM	

ToR descriptors

ToR	Description	Background	Science Plan topics addressed	Duration	Expected Deliverables
a	Design and test a Quality Assurance Framework for assessment EGs to evaluate data quality and its impact on assessments, particularly within the benchmarking process, and test this in regional case studies.	The ACOM/SCICOM assessment and advisory process needs to be based on a better understanding of the impacts of data quality. Build on experience in PGCCDBS, WKPICS, SGPIDS and other EGs; Establish close working with case study benchmark workshops; consult with WGCATCH, WGBIOP, WGISDAA, ICES DataCentre, other relevant SSGIEOM EGs & ACOM.		Year 1–3	Review of processes and outcomes of previous ICES benchmark data compilation and evaluation meetings, particularly in relation to data quality and how this was addressed in the subsequent assessment benchmark meetings. Draft Quality Assuarance Framework for ICES benchmark assessments, with associated guidelines, examples and tools; Reports on case study evaluations.
b	Develop and test analytical methods for identifying improvements in data quality, or collections of new data, that have the greatest impacts on the quality of advice	Objective procedures are needed to identify where data quality improvements will have greatest impact on quality of advice. Build links with statistical experts within and external to ICES; establish workshops to develop and test methods. Consult with the intergrated assessment working groups		Year 2	Workshop Methods & software Case study results

ToR	Description	Background	Science Plan topics addressed	Duration	Expected Deliverables
c	engage with end-users to raise awareness of what types and resolution of management decisions (e.g. by fleet or area) can realistically be supported by present or proposed data collections	Assessment and advisory groups need to understand the limits imposed by the quality and resolution of data. Consultation needed with ICES EGs & SSGs, RCMs/RCGs; stakeholder Advisory Committees, European Commission and other RFMOs.		Year 1–3	Consultation reports Documented guidelines
d	Advise on objective methods for evaluating requests by end-users for new or amended data collections within the new DCF/DC-MAP	Essential to prevent wasteage of resources on inappropriate data collection. Consultation with ICES EGs & SSGs, STECF, RCMs/RCGs; stakeholder Advisory Committees, European Commission and other RFMOs. Establish workshops to develop / test methods.		Year 1–3	Consultation reports Documented guidelines
e	Plan workshops and studies focused on specific methodological development needs	Workshops and studies are effective for attracting people with specific skills.			Workshop reports

Summary of the Work Plan

Year 1	Consolidate 3–year workplan; establish membership & skills needed; consultation within SSGIEOM on broader QAF implementation (e.g. surveys); establish links and working procedures with ICES EGs, ICES DataCentre, external bodies, external experts; develop draft QAF guidelines for benchmarks; work with test case benchmark in autumn 2015 (Irish Sea); first PG plenary meeting summer 2015. Specific ToRs for the plenary meeting will be to: A) Review all or a representative selection of previous ICES benchmark and associated data compilation and evaluation meetings to determine how these were implemented, focusing particularly on how (if at all) data quality was evaluated, how this information was utilized at the benchmark assessment meeting, how proposals for new work or data collection were arrived at and prioritized, and where there were shortfalls that need to be addressed through establishing a clearer framework for each type of benchmarking process. B) Review the responses to the data-quality questionnaires for discards estimates included in the 2015 data call for stock assessment EGs, and how the information was used by the EGs. C) Using the planned benchmark meeting for the Irish Sea (WKIrish) as a test case, work with the assessment team to identify the data needed, and use this as a test case to develop an initial draft framework and guidelines for compilation and evaluation of relevant data for benchmark assessments, including provision of time-series of data quality indicators (bias and precision) that can be incorporated directly in assessment models or used as supporting information. D) Clearly define the scope and working practices of PGDATA and identify the working relationships that PGDATA should establish within ICES (e.g. ICES SCICOM/ACOM Steering Groups; survey and other data collection EGs; assessment EGs; ICES DataCentre) and with external bodies. E) Review and adapt the work programme for the next two years of PGDATA, and develop the ToRs for the 2016 meeting. F) Consider the need for specific workshops prior to the 2016 core-group meeting, or study proposals to address PGDATA goals. G) Identify what data on recreational fishery that should be collected, with focus on the spatial and temporal resolution that is needed to support the fisheries management advice. H) InterCatch and role of ICES Data Group in PGDATA
Year 2	Planning and workshop to develop MSE-type tools for evaluating contribution of data quality to variance of assessment estimates and quality of advice, and evaluating relative impacts of data improvements; guidelines to other SSGIEOM EGs on QAF implementation; Further development and testing of QAF procedures in benchmarks; consultations with end-users; 2nd PG meeting
Year 3	Review of progress / results in implementing QAF; further implementation in benchmarks; Methodological Workshop-developing and testing criteria for evaluating data needs and requests; consultations with end-users on data needs; 3rd PG meeting; evaluate future PGDATA workplans.

Supporting information

Priority	This PG has high priority for improving the effectiveness of the ICE benchmarking process and the quality of ICES advice, and for ensuring th best use of available resources for data collection. An objective of the PG is t help ICES to develop advice using the most appropriate assessments give the quality of the data, and to be able to explain uncertainties in th assessments due to aspects of data quality and how these are reflected in th advice. This objective addresses single species, mixed fishery an multispecies assessments carried out by ACOM and SCICOM EGs, wi particular focus on regional benchmarking. A further goal is to develo objective procedures to identify where data quality improvements will hav greatest impact on quality of advice, and to ensure that proposals to colle new data or amend existing data collection schemes can be made in a informed way taking account of factors such as feasibility, methods fo collection and use of the data, impact on advice, costs of data collectio relative to precision, implications for regional sampling schemes or survey and how the quality of the data can be evaluated. ToR g) is of high priority since the work will be used as basis of the advice t the EU request data collection on recreational fisheries. PGDATA shou considered the work done by WGRFS 2015 to address the request.
Resource requirements	The national science programmes which provide the main input to this grou are already underway, and will need to commit resources to suppo participation of staff in the PG. Due to relevance of the PG to fisher management under the CFP and to the DC-MAP, use of national EMFF fun to co-finance involvement in the PG should be agreed as eligible.
Participants	The core PG participation required to address annual work plans and plenar meetings will require experts in statistics, sampling design, survey modelling, stock assessment, management strategy evaluation methods an other modelling approaches needed, DC-MAP implementation; and RCG and efforts are needed to ensure participation of experts directly involved i specific work areas, such as regional benchmark processes, which are bein addressed. Other experts, including external experts from USA an elsewhere will be invited when required. EC DG-MARE involvement will b beneficial. A broader pool of experts and other national scientists will b identified for participation in meetings and workshops and to facilitate two way communication between PGDATA and national institutes.
Secretariat facilities	Support needed from Secretariat involved in setting up benchmark meetings
Financial	No financial implications.
Linkages to ACOM and groups under ACOM	This is a joint ACOM-SCICOM Expert group. There will be strong and dire linkages with ACOM and with assessment EGs involved in region benchmarks targeted for case studies.
Linkages to other committees or groups	There will be a very close working relationship with all the groups SSGIEOM and with ACOM benchmarking groups.
Linkages to other organizations	There will be linkages with STECF, RCMs/RCGs; stakeholder Advisor Committees, European Commission and other RFMOs

Annex 2: PGDATA agenda

PGDATA Agenda

ICES Planning Group on Data Needs for Assessments and Advice

Swedish University of Agriculture Sciences (SLU Aqua) in Lysekil. 30 June to 3 July 2015

Start time: 09h.30 , 30 June 2015 **End time**: 12h.30 3 July 2015

Purpose of the meeting: A core purpose of this year's meeting is to analyse how well the benchmark assessment process at ICES is working in terms of evaluating and acting upon the quality of available data sets, and PGDATA will establish more comprehensive generic and detailed guidelines for this process using an Irish Sea case study (WKIRISH). Further details see TORs.

Use of the results of the meeting: Guidelines for data quality in benchmark processing

Pre-meeting preparation by PG members: Read a selection of ICES benchmark assessment reports that will be located on the PGDATA Sharepoint site, referring also to a questionnaire we developed (also on Sharepoint) to see how the reports deliver against existing ICES guidelines for benchmarks, focussing on the data compilation and evaluation aspects and how data quality issues were communicated to assessment scientists and acted upon.

Agenda item:	Process and outcome:	Topic leaders:	Duration
Tuesday June 30th	**START: 9.30 am**		
Welcome	Welcoming words and house keeping	Katja Ringdahl	**10 mins**
Introduction	Introduction of participants. General background to PGDATA – what is our role and how we link with other ICES expert groups and Steering Groups to deliver the ICES Science and Advisory Plans. ToRs for the meeting Working procedures and expected outcome of the meeting	Mike Armstrong and Marie Storr-Paulsen	**1 hr**
Adjusting agenda	Go through agenda, add / amend if needed	Marie Storr-Paulsen	**10 mins**

InterCatch/ RDB and role of ICES Data Group in PGDATA	Presentation of development and future plans with ICES databases IC and RDB. relevance for PGDATA	Henrik Kjems-Nielsen	**20 mins**
Data call: discards data quality tables	**ToR B**: Present the responses to the data-quality questionnaires for discards estimates included in the 2015 data call for stock assessment EGs, and how the information was used by the EGs. What did we learn? How could they be adapted for future use and extended to longer time series and other types of data as part of benchmarking and data quality monitoring?	Mike Armstrong / Marie Storr Paulsen / Cristina Morgado	**1 hr**

Lunch 1 hour starting around 12:30.

Afternoon session – ICES benchmark process and guidelines with Irish Sea WKIRISH as test case for development (**ToR A and C**)

Benchmark processes	Introduction on the benchmark processes and how ICES sees the development across single species assessments, mixed fishery assessments and regional ecosystem assessments	Jörn Schmidt	**20 mins**
Current ICES guidelines	Present what ICES currently expects from stock assessment benchmarks	Marie Storr-Paulsen	**10 mins**
Review of previous ICES benchmarks	**ToR A.** How data quality was evaluated and then used in the assessments; how proposals for new work or data collection were arrived at and prioritised, and where there were shortfalls that need to be addressed through establishing a clearer framework for each type of benchmarking process.	Jörn Schmidt / Mike Armstrong / Cristina Morgado	**30 mins**

Developing detailed guidelines for the benchmark data compilation and evaluation process using an Irish Sea stock (whiting) for forthcoming WKIRISH benchmark as case study (over 3 days)	**ToR C**: Using the planned benchmark meeting for the Irish Sea (WKIRISH) as a test case, identify the data needed, and use this as a test case to develop an initial draft framework and guidelines for compilation and evaluation of relevant data for benchmark assessments, including provision of time series of data quality indicators (bias and precision) that can be incorporated directly in assessment models or used as supporting information. We propose this is done in plenary, breaking it into data categories that map into the Stock Annexes: - Stock identity and mixing - Biological parameters (M; maturity; growth and changes over time) - Fishery descriptions and changes over time - Landings series - Discards - Length/age compositions and weights - RV surveys - Fishery-dependent indices To be agreed: What needs to be compiled, evaluated and documented through the benchmark data compilation process? How can stock assessors be informed on data quality issues in a way that helps define which models are appropriate, how they should be configured and how each data set should be weighted (where appropriate) to reflect the relative quality of different data used? Outcomes: generic and detailed guidelines for data compilation and evaluation procedures in benchmarks; Guidance on specific aspects of the WKIRISH benchmark data compilation.

Wednesday 1 July. Start 9.00 am. Finish before 17:00 in time for social event. Break for 1-h lunch around 12:30.			
Developing detailed guidelines for the benchmark data compilation and evaluation process using Irish Sea stocks for forthcoming WKIRISH benchmark as case study	Continued work on **ToR C**. May develop some break-out groups to deal with specific topics	Mike Armstrong, Marie Storr-Paulsen; Pieter-Jan Schön; Jörn Schmidt	**All day**
16:00 - Summing up	Summing up of day's outcomes including any breakout groups	Chairs and breakout group rapporteurs	
17:00h Social event. 2-3 hour boat ride (it is an old tugboat run by a museum) along the beautiful Swedish coast, returning to Lysekil around 20.00. Organisers: Katja Ringdahl & Francesca Vitale			

Thursday 2 July. Start 9.00 am. Finish 18:00. Break for 1-h lunch around 12:30.			
Developing detailed guidelines for the benchmark data compilation and evaluation process using Irish Sea stocks for forthcoming WKIRISH benchmark as case study	Continued work on **ToR C**: Developing and reviewing text on the detailed guidelines.	Mike Armstrong, Marie Storr-Paulsen; Pieter-Jan Schön; Jörn Schmidt	
12:00 Summing up	Summing up of day's outcomes including any breakout groups	Chairs and breakout group rapporteurs	
15:00. Future PGDATA work – evaluating data needs for end users	Evaluating impacts of data quality on assessments, and cost-effectiveness of data collection. Presentation by Jon Helge Vølstad on Norwegian work using simulation studies Work-shop proposal and a study proposal	Mike Armstrong, Marie Storr-Paulsen, Jon Helge Vølstad	
1Recreational fishery data needs	Respond to additional ToR on request from European Commission on needs for recreational fishery data, frequency of collection and use in stock assessments – mainly to add / respond to WGRFS response to same ToR.	Mike Armstrong	**30 mins**

Friday 3 July. Start 9.00 am. Finish 12:30pm			
16:00 – what do end-users want from PGDATA	Chairs of RCMS, WGCATCH, WGBIOP and other expert groups and steering groups at meeting to give their view on the most useful outputs that PGDATA could provide for their groups.	Group chairs	**5 mins each**
Scope of PGDATA	Define the scope and working practices of PGDATA and identify relationships within ICES (e.g. ICES SCICOM/ACOM Steering Groups; survey and other data collection EGs; assessment Egs; ICES Data Centre) and with external bodies.	All	**1 hour**
Future work in PGDATA – refining the three-year work plan; and identifying workshops or study proposals	Review and adapt the work programme for the next two years of PGDATA, including intersessional work, and develop the ToRs for the 2016 meeting. Consider the need for specific workshops prior to the 2016 core-group meeting, or EU study proposals to address PGDATA goals.	Mike Armstrong and Marie Storr-Paulsen	**1 hour**
Summing up/ Discussion	Report deadlines and responsibilities; work plan for 2015/16.	Mike Armstrong and Marie Storr-Paulsen	**30 mins**
Any other business			**30 mins**
Ending at 12.30 am			

Annex 3 Participants list

Name	Address	Phone/Fax	E-mail
Mike Armstrong chair	Cefas Pakefield Road Lowestoft Suffolk NR33 0HT United Kingdom	+44 1502 524362	mike.armstrong@cefas.co.uk
Lotte Worsøe Clausen	DTU Aqua Jægersborg Allé 1 2920 Charlottenlund Denmark	+45 21362804	law@aqua.dtu.dk
Laurent Dubroca	Ifremer Port-en-Bessin Station PO Box 32 Avenue du Général De Gaulle 14520 Port-en-Bessin France		laurent.dubroca@ifremer.fr
Henrik Kjems-Nielsen ICES	ICES H.C. Andersens Boulevard 44-46 1553 Copenhagen Denmark		henrikkn@ices.dk
Uwe Krumme	Baltic Sea Fisheries Alter Hafen Süd 2 18069 Rostock Germany	+49 381 8116-148	uwe.krumme@ti.bund.de
Cristina Morgado ICES Professional secretary	ICES H.C. Andersens Boulevard 44-46 1553 Copenhagen Denmark	+45 33 38 67 21 +45 33 63 42 15	cristina@ices.dk

Name	Address	Phone/Fax	E-mail
Kjell Nedreaas	Institute of Marine Research PO Box 1870 Nordnes 5817 Bergen Norway	+47 99538549	Kjell.nedreaas@imr.no
José Castro	Insituto Español de Oceanografía Avda. Beiramar, 37 36202 Vigo Spain	+34 986 213107	jose.castro@vi.ieo.es
Nuno Prista By correspondence	IPMA Avenida de Brasilia 1449-006 Lisbon Portugal		nmprista@ipma.pt
Katja Ringdahl	Institute of Marine Research Turistgatan 5 453 30 Lysekil Sweden	(+46) 10 478 4043	katja.ringdahl@slu.se
Jörn Schmidt	University of Kiel Dept. of Economics Wilhelm-Seelig-Platz 1 24118 Kiel Germany	+49 431 880 5632	jschmidt@economics.uni-kiel.de
Pieter-Jan Schön	Agri-food and Biosciences Institute AFBI HQ 18a Newforge Lane Belfast BT9 5PX UK	+44 28 90255015 fax +44 28 90255004	pieter-jan.schon@afbini.gov.uk
Marie Storr-Paulsen chair	DTU Aqua Jægersborg Allé 1 2920 Charlottenlund Denmark	+45 35 88 34 41	msp@aqua.dtu.dk
Christoph Stransky	Institute of Sea Fisheries Palmaille 9 22767 Hamburg Germany	+49 4038905228	christoph.stransky@ti.bund.de

Name	Address	Phone/Fax	E-mail
Els Torreele	ILVO Ankerstraat 1 8400 Oostende Belgium	+32 59569833	els.torreele@ilvo.vlaanderen.be
Sieto Verver	Centre for fisheries research 1970 AB IJmuiden The Netherlands	+31 317 4890045	sieto.verver@wur.nl
Francesca Vitale	Institute of Marine Research Turistgatan 5 453 30 Lysekil Sweden	+46 10 478 4052	francesca.vitale@slu.se
Jon Helge Vølstad	Institute of Marine Research PO Box 1870 Nordnes 5817 Bergen Norway	+47 46929039	jonhelge@imr.no
Lucia Zarauz	AZTI Sukarrieta Txatxarramendi ugartea z/g E-48395 Sukarrieta (Bizkaia) Spain	+34-946029400	lzarauz@azti.es

Annex 4. Draft guidelines for ICES benchmark data evaluation process

Introduction

These guidelines were developed to provide ICES benchmark data evaluation teams with comprehensive advice on tasks that should be completed prior to and during the benchmark data evaluation meeting. The guidelines cover all types of data and biological parameters commonly used in stock assessments. For some benchmark assessments, only some of the data types and parameters may require full evaluation depending on the issues list for the benchmark, or if previous benchmark data evaluation workshops have carried out a full evaluation which only requires an update with more recent data.

The guidelines include a description of tasks required for each data or parameter type, which will also relate to a specific Term of Reference for the benchmark data evaluation.

The compilation and evaluation of data, and their analysis as part of a stock assessment, are the endpoint of a process that commences with a stock assessment Expert Group (EG) identifying the need to benchmark a new assessment procedure for one or more stocks. Key stages in the process, which is expected to take place over two years, are illustrated in Figure 1.

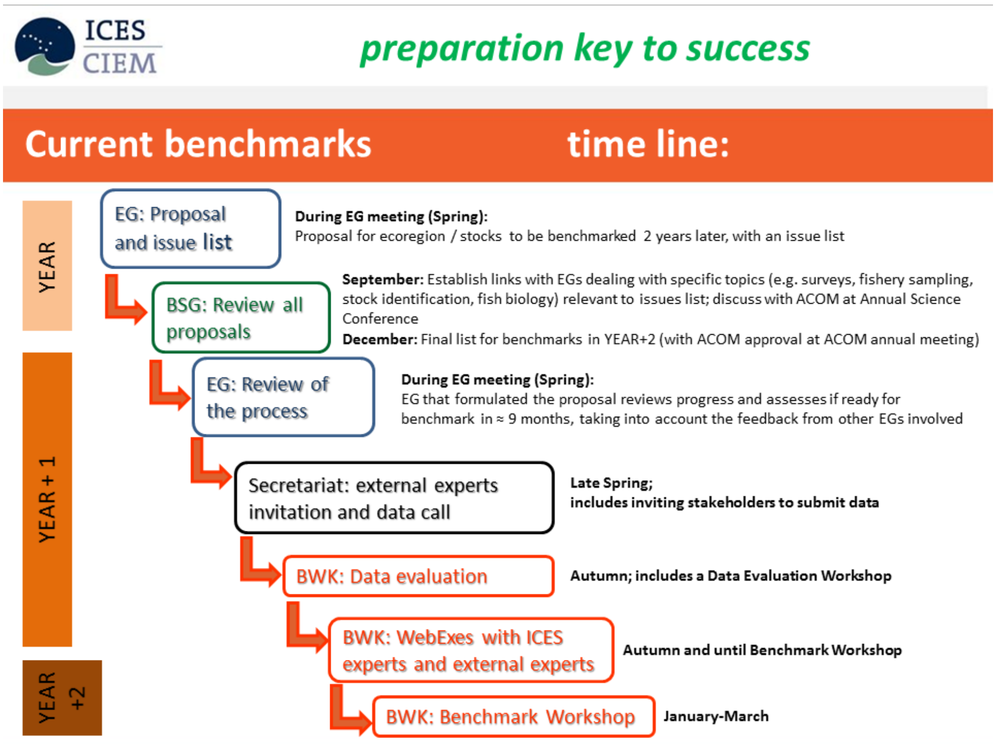


Figure 1 Example schema for a typical ICES benchmark assessment process. BSG = ICES Benchmark Steering Group; BWK = benchmark workshop; ACOM = ICES Advisory Committee on Management. The periods of the year are examples and will vary by stock and timing of the parent assessment EG.

It cannot be emphasized too strongly that the benchmark data evaluation workshop, and the benchmark stock assessment workshop, are the endpoints of programmes of work carried out by team members over a sufficient period of months prior to the workshops. For the data evaluation process, it is expected that working documents and proposed datasets will be compiled for each term of reference for the data evaluation process, and reviewed at the data evaluation workshop. The data evaluation workshop will therefore be confined to:

- Reviewing working documents and other material addressing each term of reference, prepared by the appropriate specialists in advance of the meeting;
- Agreeing on data inputs, biological parameters, quality indicators and alternative data scenarios where needed.
- Providing a spreadsheet containing all the proposed data inputs and parameters.
- Recording the work completed and recommendations in a separate data evaluation workshop report.

Additional data analysis may be required at the workshop if it is needed and is possible, but the process will start to fail if compilation and evaluation of data is left too late and is being done at the workshop itself, which lasts only a few days. Information from stakeholders must be sought well in advance of the workshop so that it can also be evaluated for presentation at the workshop.

The data evaluation can only succeed if people with the required skills are available and have the necessary time allocated to complete the pre-workshop tasks, attend the workshop, prepare the workshop report, and clarify any issues raised by the benchmark stock assessment team after they have studied the report and the recommended data inputs. This includes input from people from data Expert Groups, such as those providing survey data or dealing with sampling design and analysis. To minimize such issues arising, the primary stock assessors and stock coordinator should be involved throughout as it is essential that data are evaluated in the context of the existing or potential new assessment approaches.

If it is not possible to develop a data evaluation team with the required skills and time allocation to do the necessary work for all the stocks covered, the feasibility of carrying out the benchmark assessment for all or some of the stocks covered should be reconsidered by ICES.

Depending on requirements, the data evaluation workshop will:

- Provide a full and detailed description and evaluation of the data with quality indicators if possible and candidate parameters if there is not a recent ICES benchmark data evaluation for a given data category that remains valid.
- If there is a recent ICES benchmark data evaluation that remains valid, summarize the findings, provide links to the data evaluation report, and update any time-series of data and quality indicators since the previous evaluation.
- Fully document all the datasets and parameters recommended for use in the benchmark assessment and not only the new or changed datasets.

The following sections describe tasks recommended to be completed for each generic Term of Reference for a full benchmark assessment data evaluation process. A particular benchmark may only require some of these to be addressed in full. A supporting information section (Appendix 1) contains more detailed explanatory text for some of the data types.

Guidelines to the data evaluation team, by data type

1. Stock Identification

Explain the basis for existing assumptions on stock structure and mixing rates between stock areas, or proposed new assumptions which form the basis for spatial aggregation of fishery and survey data and/or adjustments to datasets to account for stock mixing. If no changes are to be made to stock boundaries, or to any procedures to separate fishery or survey catches into stocks within a stock mixing area, provide a brief summary of the current definition in a document for the data evaluation workshop report. This should include a map showing the existing stock boundaries along with brief text explaining the basis for the stock assumption and any methods to quantify stock mixing rates. Provide links to the Stock Annex, previous benchmark data evaluation reports or any other documentation explaining the basis for these assumptions and methods.

Prior to the data compilation workshop: If the assessment expert group (EG) believes that current stock areas or mixing rates may need revision, the EG will in the first instance seek expert advice from the ICES Stock Identification Methods Working Group (SIMWG; ICES 2015a and earlier), or, where spatial patterns in biological parameters such as growth or maturity are being used as evidence, the ICES Working Group on Biological Parameters (WGBIOP; ICES 2015b). Appendix 1.1 provides examples of how the EG may develop an initial issues list and revise this following expert advice. Prior to any data compilation and evaluation for new stock areas, it must first be determined if the evidence is sufficiently robust and the work to create new datasets and parameters is feasible. Carry out the following tasks, in liaison with the other expert groups being consulted:

- Conduct an initial review and summarize this in a Working Document containing: i) a full explanation of the reasons for reviewing the stock structure or mixing rates; ii) an evaluation of the robustness of the evidence – e.g. the quality and comparability of data on growth, maturity, recruitment patterns, genetic structure, tagging results, and meristics or other population characteristics used as evidence of stock structure; and iii) an evaluation of the feasibility of aggregating or disaggregating catch and survey data and revise biological parameters to reflect the new stock definitions in time for the proposed benchmark data evaluation meeting.
- If the evidence is not sufficient to warrant a revision of stock structure, or if it is not possible to develop datasets for revised stock definitions in time for the benchmark assessment, the assessment EG should consult with the ICES benchmark steering group to decide if the benchmark should continue using the existing stock definitions or be postponed.
- If it is decided to revise the stock boundaries and it is possible to complete such work in time for the benchmark data evaluation meeting, then provide revised historical landings, catch composition data, abundance indices and biological parameters required according to the new stock boundaries. If the assessment EG intends to account for mixing rates between stocks by adjusting input data (e.g. plaice stocks in VIId & e), thoroughly review the evidence of mixing rates and provide a plausible range of uncertainty to allow the sensitivity of the assessment and forecast to this to be evaluated. This is also required if mixing rates are to be estimated within a multistock statistical assessment model, to help develop prior distributions of input values.

- If the data can be compiled in time, provide the disaggregated or aggregated datasets for the revised stock definitions together with an evaluation of quality of the datasets. Depending on sampling and survey coverage, expanding or splitting the stock area could lead to changes in data quality at the stock level depending on the quality of data from the different sampling areas included in the range of the stock or could imply that the dataseries has to be truncated.

2. Review and recommend life-history parameters

Life-history parameters (e.g. growth parameters, maturity ogives, fecundity, natural mortality), for use in assessments should be analysed. Where applicable, provide appropriate models to describe growth, maturation, and fecundity by age, sex, or length (see Appendix 1.1 for how the EG may develop an initial issues list and revise this following expert advice):

Growth, maturity and fecundity

The data evaluation process will address the life-history parameters and specific issues identified by the parent assessment EG for the benchmark data evaluation process. Summarize the findings in a Working Document. If previous benchmark data evaluations of these parameters remain valid, provide references and links together with a summary of the parameter values and their precision if calculated.

- The sources of the samples (e.g. which trawl surveys or fishery sampling schemes), and the laboratories involved in data collection over time.
- Coverage of the sampling in terms of geographic areas and seasons, relative to the known distribution of the population at different life stages, and an evaluation of possible effects of any spatio-temporal mismatch with the stock biology.
- Results of latest age validation studies, calibration studies and exchanges to document bias and precision of age estimates, highlighting any persistent differences between laboratories and changes in interpretation of age material over time.

If a full evaluation is required, then a further documentation is required of design, interpretation and analysis:

- Selectivity characteristics of gears providing samples, where these may lead to skewed distributions of size at age.
- Numbers of independent primary sampling units such as survey trawl hauls or commercial fishing trips with samples for the species, and total numbers of individuals sampled, by year. Gaps in sampling coverage that will affect quality of estimates should be identified.
- How the sampling units were selected (e.g. opportunistic or using a design-based random sampling scheme).
- Methods and criteria for identifying mature fish in samples, with reference to maturity keys, sampling protocols and calibration workshops or studies.
- Description of fecundity estimation methods, if applicable.
- Description of analysis methods including use of statistical models to estimate growth, maturity and fecundity parameters by age, length or sex as appropriate.

- Derived parameter estimates, with diagnostics and evaluation of quality and evidence of stability or trends in parameter values over time.
- Recommended parameters for use in assessments. For statistical age and length based assessment models, specific statistics such as standard deviation of length-at-age, CV of ageing errors or age-error matrices, may be required and should be requested by the assessment working group.

Natural mortality

Within ICES, decisions on appropriate values for M generally rely on: i) results of multispecies models such as the Stochastic Multispecies Model SMS for the North Sea updated at intervals by the ICES Working Group on Multispecies Assessment Methods (WGSAM), or integrating single-species models with other forms of multispecies or ecosystem models; ii) methods that predict average or age-based natural mortality from life-history parameters such as growth and maturity parameters and maximum observed age; iii) an assumption such as 0.2 used in the absence of other information, or by comparison with similar species. For some data-rich stocks assessed using statistical models, it may be possible to derive some inference on plausible rates of natural mortality based on likelihood profiles over a range of M, or from tagging results included in the model.

Depending on information available, carry out the following tasks:

- If an estimator such as SMS, or another approach using multispecies models, is used by the assessment EG and it is intended to continue with this approach, or if it is proposed to start using M estimates from such a model, provide a reference and link to the latest model update and the values of M by age and year for the stock. Summarize information on the quality of the estimates given in the multispecies assessment report. It is important to consult the expert group providing the multispecies model estimates when preparing the data evaluation, to ensure the correct information is provided for the benchmark stock assessment.
- If life-history methods to infer M are to be proposed, provide the results of a range of plausible models from the literature, proposing a baseline method together with alternatives that could be used for sensitivity testing.
- If estimates of M have previously been derived from an assessment model including tagging, or inferences have been made from likelihood profiles or other modelling approaches, summarize the findings of the relevant EG report including any information provided by the EG on the quality of the estimates or inferences.
- If there is no existing information, derive a range of plausible values for M for species with generally similar life histories and give supporting arguments.

3. Describe the history of fishery management regulations

Information is needed on management regulations and actions that are expected to have caused changes in the quality of fishery catch data or the selectivity patterns of fisheries that are of relevance for the scientific assessment of the stocks and provision of advice.

If there is an existing Stock Annex, this should already provide a history of management measures relevant to the assessment and advice. If this is not sufficiently complete and adequate it must be reviewed and updated. Carry out the following tasks where appropriate (much of this will be generic to many stocks within a region; See Appendix 1.2 for supporting information):

- Provide a chronological description of management regulations and actions applied to fleets (rather than those specific to stocks), and the known or expected impacts on data quality and fishery selectivity in general. Include information such as: spatio-temporal closures; gear regulations (mesh size, selective devices, length of nets); change in minimum landing size (reference size); direct regulation of fishing effort; decommissioning schemes (including how much of the targeted fleets are removed and the impact on overall fleet capacity) and any other measures having a significant impact on the amount of fishing and selectivity of fishing fleets.
- Provide a chronological description of management regulations and actions that are specific to the stock being benchmarked. This could include: TACs; individual boat limits; minimum conservation reference size (MCRS); implementation of the landing obligation, etc.
- Provide a chronological description of management regulations or actions that affect the compliance with management measures and the completeness and quality of fishery data supplied to assessment working groups. This may include changes in catch reporting systems such as national Buyers and Sellers regulations, introduction of landing obligation with *de minimiz* rules, changes in the quota system and in vessel monitoring and control.
- For stocks where an understanding of changes in fishery selectivity is needed for the assessment model, document any management regulations or actions that are expected to cause a change in selectivity for the stock being benchmarked, and evaluate the known or likely outcomes.
- For stocks where commercial CPUE or LPUE is to be evaluated for providing abundance indices, identify management regulations or actions that are expected to cause a change in catchability or selectivity of the relevant fleets for the stock being benchmarked.
- Where possible make use of graphical or tabulated summaries to give a clearer overview of changes over time. Some examples are given in the supporting information section (Appendix 1.2).

4. Develop time-series of catch estimates with bias and precision indicators.

The guidelines in this section relate to total retained or discarded fishery catch for all types of fishing. Separate data evaluations are needed for catches that are recorded exhaustively (e.g. landings logbooks), and for those estimated through sampling schemes (e.g. discards and recreational catches). See Appendix 1.4 for additional information.

For exhaustively collected data:

- Provide full documentation of the derivation of the catch figures for the time-series available for assessment, and any adjustments made to official statistics. Such adjustments might have been made to allocate landings to the correct fishing ground, adjust for stock mixing, to disaggregate mixed-species landings records using sample data, or make other corrections for misreporting or underreporting. Explain how the adjustments are made.
- Document and explain differences between the landings figures recommended by the data evaluation team and the official statistics
- Evaluate the reliability of catch estimates in terms of historical biases and trends in bias, where evidence of such biases exist.
- Propose catch dataseries which are appropriate to use in a stock assessment. If there are historic data of poor quality, for example due to known or suspected inaccuracies in reporting, provide (if possible) different plausible catch histories that could be used for sensitivity analyses in the benchmark assessment. Consult with stakeholders in drawing up such scenarios.

For data collected non-exhaustively through sampling schemes, the description of the surveys and evaluation of data quality can be complex, requiring detailed examination of survey design and sampling achievement down to the level of sampling strata. Seek assistance from ICES Expert Groups dealing with such surveys well in advance of the data meeting (e.g. ICES Working Group on Recreational Fishery Surveys; ICES Working Group on Commercial Catches). Contact the ICES Secretariat and the chairs of these EGs to determine a process by which the sampling survey experts may contribute to the documentation and evaluation of catch data from surveys of recreational fisheries or commercial discards and landings. This may require ToRs to be added to the next meeting of these EGs so this needs to be considered well in advance of the benchmark. The following data evaluation tasks will be required:

- Provide an overview of the survey methods adopted, with links or references to detailed scheme descriptions. This covers the design of the schemes, including: definition of the population being sampled; sampling frames and their coverage; primary and lower level sampling units and how they are selected; stratification of the sampling units and reasons for this; other relevant data collected such as recording of non-responses or refusals; and how the data are analysed to provide estimates of total catches.
- Document historical changes in sampling schemes that may indicate changes in data quality (bias and precision) over time. The tables for documenting quality of discards estimates, included in the ICES 2015 data call, provide one format for documenting changes in sampling schemes (see Appendix 1.3) though other formats are possible.
- Evaluate the reliability of catch estimates in terms of historical biases and trends in bias, and in terms of precision. Where standard errors or CVs of

estimates are provided, document these. Also provide simpler quality indicators such as numbers of primary sampling units sampled. See additional notes in the supporting information section for further details.

- Propose catch dataseries which are appropriate to use in a stock assessment together with data quality indicators to help the stock assessment team to decide which data to use, to weight different dataseries if necessary, and to interpret the diagnostics of assessment models.

5. Estimate the length and age distributions of fishery landings and discards, with bias and precision indicators.

As with estimation of catches by surveys, the description and evaluation of additional sampling schemes to estimate length and age composition, and evaluation of data quality, can be complex. Include this aspect of data collection with catch estimation when seeking assistance from ICES Expert Groups dealing with such sampling schemes. Assistance from the ICES Working Group on Biological Parameters (WGBIOP) should be sought in relation to quality of age estimates (see guidelines for biological parameters). See Appendix 1.5 for more information.

With input from the relevant EGs as described above, where appropriate, document the derivation and quality of existing length and age composition data for fisheries, and of any new datasets that have been made available, as follows:

- Using ICES reports on age validation and calibration studies for the stock, or any other documentation on precision and bias in age readings: (i) evaluate if age readings are reliable enough for use in an assessment – i.e. sufficiently low bias, and (ii) provide metrics of precision such as CV or an age error matrix that can be incorporated into a statistical stock assessment model. Compare the age structure of catches across countries. Identify any systematic differences in interpretation of otoliths, scales or other material between laboratories, and any drift over time in age interpretation by national laboratories, where information is available. Seek guidance from the stock assessors on the metrics of bias or precision needed for incorporation in assessment models.
- Provide a summary of the historical design of national shore-based and at-sea sampling schemes or any other schemes to estimate length and age compositions, and the methods of raising data to give compositions at the national scale. Describe how total catches-at-age are derived from combination of length and age sampling (ALKs), or from age sampling on its own.
- Tabulate achieved annual sampling rates (numbers of primary sampling units such as numbers of fishing trips sampled for length and age, with supporting information on numbers of fish measured or aged). This should ideally be done by country and sampling stratum in each year together with the estimated annual landings or discards for each stratum. Use these data to identify deficiencies and gaps in sampling.
- Describe how length and age compositions are raised and aggregated within and across countries to give international estimates (e.g. be métier or métier group through InterCatch). Identify if the methods are statistically sound and the sample sizes are sufficient in each stratum to support the degree of resolution being applied, or if there is a substantial amount of subjective “borrowing” of estimates from other countries and métiers especially

if done without reference to the quality of borrowed data. Consult with experts from the ICES Working Group on Mixed Fisheries Advice (WGMIXFISH) on their information needs. If necessary rework the raising and aggregation using more statistically robust methods for comparison with InterCatch results.

- Describe how individual live weights are derived (e.g. direct measurement or from length-weight relationships) and evaluate known or potential errors introduced by this.
- Provide a recommended dataset of length and age compositions (CANUM; LANUM) for landings, discards (and recreational catches where appropriate), and associated weights at age (WECA). If possible, provide estimates of precision (e.g. relative standard error or CV) for the raised international landings and discards at age. Consult the stock assessment team on whether numbers or weights at age should be sums-of-products (SOP) corrected so that the sum of numbers-at-age and weights at age is equivalent to the total catch weight figure input to the assessment.
- Use the information on sampling design, sampling achievements, coverage, precision over time and ageing errors to provide advice to the stock assessment team on changes in overall data quality (bias and precision) that will allow an objective decision to be made on whether the data can be used for all or some years, or weighted in an assessment model.
- Evaluate the internal consistency of proposed catch-at-age datasets in terms of consistent tracking of year classes, and identify the most likely sources of poor year class tracking based on the data quality information available. This will help identify further research or additional sampling needed to improve data quality. Unless otherwise instructed, provide age compositions out to the oldest true age to allow flexibility in setting a plus group. Information on numbers of fish sampled at age each year can be useful statistics to help determine the most appropriate plus group for the assessment.

6. Develop recommendations for addressing fishery selectivity (pattern of catchability at length or age) in the assessment model.

Most age-based or length-based stock assessment models require some assumptions about selectivity i.e. how catchability varies with size or age in fisheries. Selectivity in this context is a combination of the selectivity properties of fishing gears of different design, and factors influencing the probability of fishing operations encountering fish of different sizes and ages, for example related to distribution of fishing or behavior patterns of the fish.

Statistical assessment models may involve fitting selectivity patterns of varying complexity (e.g. asymptotic or various types of domed curves) separately to individual fleets or groups of fleets. To help the assessment team decide on appropriate selectivity patterns and any changes over time, carry out the following tasks:

- Examine the spatio-temporal distribution of fisheries relative to the known distribution of fish of different sizes or ages, for example from trawl surveys.

- Review any available information on how the behaviour of different sizes of fish affect their likelihood of interacting with fishing gear at any location.
- Review existing information on selectivity characteristics of the main types of fishing gears used for the assessed stock, based on gear selectivity studies or other published studies.
- Refer to the guidelines for documenting changes in management regulations to identify expected changes in selectivity, and consider how changes over time in the contribution of catches by fleets with different selectivity characteristics may have altered the overall selectivity pattern for the combined fisheries.
- If data allow, use a fleet-aggregated model such as VPA to derive partial F's at age by fleet/gear using the separate age compositions for the component fleets, and hence provide inferences on relative selectivity of the different fleets and any historical changes.
- If an assessment is to be explored in which domed selectivity is to be assumed for some fleets, it can be helpful to have one fleet for which selectivity is most likely to be asymptotic and where the catches and input length or age data are sufficient to allow a good fit. Advise on which fleets (national or international), if any, are most likely to fulfil this criterion, based on the tasks given above.

7. Recommend values for discard mortality rates, where appropriate, and indicate the range of uncertainty in values.

ICES assessment EGs have, for most assessed stocks, assumed that all discards die. The potential for dispensations from the EU landings obligation for species with high discard survival has resulted in a range of studies on the mortality rates of fish and shellfish discarded or released alive from fishing operations. Many recreationally caught fish are also released alive after capture and have variable survival rate depending on a range of factors such as deep hooking, bleeding and water temperature. There are numerous published studies on post release survival of marine species, though relatively few are from Europe. Increasingly, ICES assessment EGs will need estimates or inferences of mortality of discarded or live-released fish caused by the fishing operation.

Carry out the following tasks to provide information on estimated or potential discard mortality:

- Review existing information on discard mortality for the assessed stock, or for similar species in similar fisheries and conditions, following the guidelines provided by the ICES Workshop on Methods for Estimating Discard Survival (WKMEDS: ICES 2015 and previous).
- Where supported by data or comparisons with similar stocks studies elsewhere, recommend discard mortality rates and range of uncertainty. Include thorough rationale for recommended discard mortality rates.
- Provided justification for any recommendations that deviate from the range of discard mortality provided in available research and published literature.

8. Review all available and relevant fishery-independent and dependent data sources on fish abundance, and recommend which series are considered adequate and reliable for use in stock assessments

Fishery-independent data

Assessment EGs make extensive use of research surveys to provide absolute estimates of abundance, or more commonly, relative abundance indices, for tuning length or age based stock assessments. In many data-limited assessments, surveys provide the main source of information on stock trends. Survey data may be used as size/age-aggregated indices or as length or age based indices. Some assessment models require the parameters of the selectivity pattern of a survey at length or age to be fixed or estimated, and for indicators of data quality such as CVs or effective sample sizes to be input to the model separately for the total abundance indices and the length or age compositions.

As with survey estimates of fishery catches and catch compositions, the evaluation of fishery-independent survey data can be complex and will require outputs and support from expert groups dealing with design and implementation (e.g. International Bottom Trawl Survey Working Group, IBTSWG) and those dealing with interpretation and end-use of survey data (e.g. Working Group on Improving Survey Data for Assessment and Advice, WGISDAA). To facilitate this process, EGs that coordinate and monitor specific surveys such as the IBTS should become responsible for providing the specific types of information on historical design, data quality and abundance indices required for the evaluation and use of survey data, with guidance from groups such as WGISDAA on what is needed. Where indices are required by age or maturity, the Working Group on Biological Parameters, WGBIOP, can be consulted for advice on data quality and use where necessary. The appropriate survey EGs must be consulted at the initial stages of the benchmark process (See Figure 1) to identify their tasks in providing advice or carrying out the evaluation work needed for the benchmark.

In collaboration with the survey EGs tasked with providing the necessary information in their reports, or other experts responsible for the surveys under consideration, carry out the following tasks to evaluate each fishery-independent dataseries:

- Document main objectives, timing, frequency, spatial coverage, survey sampling design including definition of sampling units, sampling gear, sampling intensity, stratification and methods for allocation of sampling effort to strata, subsampling procedures, and other relevant characteristics. Provide maps of survey coverage in relation to expected species/stock area of occupancy.
- Evaluate the suitability of the survey for providing abundance indices for the species/stocks being assessed given known aspects of fish behavior, habitat preferences and vertical- horizontal distribution.
- Document changes in survey design, coverage, vessels and gears over time. Evaluate the potential for bias caused by systematic or step-changes in catchability over time due to such changes. Document any calibration factors applied following vessel or gear changes, and any estimates of uncertainty around these.
- Refer to guidelines for biological parameters to evaluate if age or maturity readings are of sufficient quality to derive abundance indices by age and maturity, including any changes in age interpretation or maturity criteria that

would compromise integrity of time-series (liaise with WGBIOP where required).

- Describe the analytical methods used for deriving indices of abundance including any disaggregation by sex, maturity, length or age class. Describe any selection methods used in the analysis of the survey data to provider assessment inputs – for example restricting the analysis to spatial subareas (domains) or time of day of observations, or use of any modelling approaches such as GLMs or GAMs.
- Describe the methods for deriving estimates of precision and provide the estimates for each year over the time-series – see Appendix 1.6 for further details and caveats.
- Review any evidence that may help identify the shape of the selectivity pattern by length or age for the survey, if needed for the assessment. This is a complex function of gear selectivity, distribution of fish of different sizes relative to the survey coverage, and aspects of fish behavior at a trawl station that affect the probability of fish of different sizes or ages interacting with the gear.
- Tabulate the recommended survey indices and quality indicators for use by the assessment EG.
- Tabulate all other survey data provided and evaluated, but not considered suitable for the assessment.

Fishery dependent data

Fishery dependent abundance indices continue to be used for some species, with or without fishery-independent data, and may be the only information available on stock trends for some data-limited stocks. Assessment and advisory groups need to understand the limits imposed by the quality and resolution of such data. See Appendix 1.6 for more details on the limitations of such data. Advice from the ICES Working Group on Fishing Technology and Fish Behaviour (WGFTFB) should be sought in evaluating the suitability of a fleet for providing abundance indices and for evaluating issues such as technology creep.

If fishery-dependent data are to be evaluated, consult the background documents listed in Annex 1 and carry out the following tasks, collaborating where needed with ICES WGFTFB:

- Document all fishery CPUE/LPUE series evaluated, addressing target species, fleet sectors, fishing gears, coverage, and regulatory measures affecting fleet behaviour. Evaluate the suitability of each CPUE/LPUE fleet for the species being assessed, in terms of known aspects of the fisheries and fish behaviour in relation to gear design and fleet coverage.
- If developing a CPUE index including discards, evaluate the quality of the discards data for each year in the series, following the approaches outlined above for developing time-series of fishery discards and landings.
- Define and describe the available effort metrics (e.g. hours, days, trips, number of hooks or nets, horsepower, soaking time, search time or any combinations of these), and evaluate which, if any, of the metrics are appropriate, and why.

- Describe the methods for data selection (e.g. subsetting of fishery trips according to vessel size, time, area, gear or species composition). Provide maps of coverage of the selected vessels in relation to the entire selected fishery (e.g. VMS).
- Develop fishery CPUE/LPUE indices by appropriate strata (e.g. area, and fishery) and include measures of precision and assessment of bias; rank indices with regard to their suitability for use in assessment modelling. Describe methods of analysis of CPUE/LPUE data including any statistical modelling carried out.
- Evaluate the potential for changes in catchability over time due to changes in vessels, fishing gear and methods, or spatio-temporal activities. Document the methods and rationale for any factors used to correct for changes in fishing efficiency, and feasible ranges for time-trends in efficiency.
- For age-based CPUE/LPUE indices, evaluate the internal consistency of age compositions and correlations between fishery dependent CPUE/LPUE series and surveys. Indicate if CPUE/LPUE fleets with age compositions provide a large fraction of the total international catches.
- Where needed for exploring assessment models, evaluate the length or age selectivity of the CPUE/LPUE fleet as described above for fishery length and age compositions. Indicate the extent to which components of the age composition are mainly observed in the fishery dependent CPUE/LPUE and not in the scientific surveys.
- Recommend and tabulate fishery dependent datasets that are appropriate to use in the assessment, together with any quality indicators such as precision estimates or plausible alternative scenarios for catchability trends.

9. Longer term or episodic/transient changes in environmental drivers known to influence distribution, growth, recruitment, natural mortality or other aspects of productivity and which are relevant to assessments and forecasts.

There are potential circumstances where the data inputs to an assessment model, or the assumptions in the model, need to take into account environmental drivers. These may be episodic or transient phenomena such as mortality or changes in fish distribution caused by low-oxygen water or lethal temperature events, or longer term trends in environmental conditions. The data evaluation team should source and review existing information and make recommendations on how this information should be used by the assessment team, as described below.

Longer term environmental drivers

Regional integrated ecosystem assessment groups, ecosystem overviews or scoping workshops may have identified environmental time-series that are relevant to an assessment or forecast – for example trends in environmental variables that affect recruitment and could be included as covariates in an assessment or used to modify decisions on recruitment for short-term or medium-term forecasts. Environmental variables may also be related to changes in growth and distribution, or catchability in surveys. Compile any such datasets supplied by regional integrated ecosystem assessment groups etc. and make available to the assessment team together with any specific comments on quality of those data (taking advice from ICES Data Information Group where needed)

Episodic / transient events

Identify any episodic / transient environmental events that have been shown to affect abundance or population dynamics of the stock being assessed, where these need to be accounted for in the assessment model and any associated predictions and advice. Data that could be used by the assessment team for this purpose should be developed if not already supplied by other expert groups (e.g. low oxygen or salinity events, exceptional warm or cold periods)

10. Review progress on existing recommendations for research to develop and improve the input data and parameters for assessments, and develop and prioritize new proposals.

Provide a review of existing recommendations for research to develop and improve the input data for the assessment, and what has been achieved. If work is still ongoing, describe progress, problems encountered, how these will be resolved and expected finalization of the work. If this cannot be progressed, consider a recommendation that the work should be stopped.

During the data evaluation workshop, proposals for changes to data collection or needs for new data or studies may be identified. The workshop must identify the relative priorities of the recommendations and expected impact on the quality of the assessment, and take into account feasibility.

11 Develop a spreadsheet of assessment model input data that reflects the decisions and recommendations of the Data Workshop.

Develop a spreadsheet of assessment model input data and parameters that reflects the decisions and recommendations of the data evaluation workshop, covering all aspects of data and parameter estimates covered in 1 – 9 above. This will include quality indicators such as age-error matrices and time-series of CVs or sample sizes that are needed for input to the assessment model, in addition to plausible ranges of parameters such as M, and alternative catch histories where needed. Also document any data that were evaluated by the data evaluation team but not recommended for use.

This is a key output of the data evaluation process. The benchmark assessment workshop will use this table to indicate which data were used, and explain why any of the data are not used or are modified.

12 Prepare the data evaluation workshop report providing complete documentation of workshop actions, decisions, list of working documents, other information used by the workshop, and a list of any additional tasks to be completed following the workshop with dates and responsibilities for completion.

Finalize and agree the report of the data evaluation workshop, and the spreadsheet of recommended assessment input data, within two weeks of the end of the workshop. This is to allow the stock assessment team time to evaluate the recommendations, seek any clarification from the data evaluation team, or conduct any of their own analyses if they disagree with the findings of the data evaluation workshop.

The data evaluation workshop report and Excel table of recommended inputs should stand as separate documents alongside the assessment workshop report with both being available from the same ICES web page.

Appendix 1 ICES benchmark data evaluation process: Supporting Information

Appendix 1.1 Example Issue lists for biological parameters and stock structure

The first table below presents possible ways to handle issues related to stock structure and life-history parameters (described here as the "first draft issues list") and it would be compiled by the assessment EGs. This table includes the issues that potentially need to be handled. In many cases, advice is needed from specialist EGs to develop the work needed, and this could be included in the ToRs for their next meeting or for intersessional input to the benchmark data evaluation process. The issues contained in this table and addressed to the specialist EGs have to be specific in order for them to be able to tackle the issues and provide a valid response back to the assessment EG chair, stock coordinator and stock assessor. It is the assessment EGs responsibility to ensure that advice on the issues is obtained from other ICES EGs with the necessary expertise.

The second table is a template for an adjusted issue list to be completed by the assessment EG (chair, stock coordinator and stock assessor) based on initial exploratory work and feedback received from the other EGs consulted. This table will define the work needed to be done by the benchmark data evaluation team to address each issue. The table will specify which data to compile and evaluate in terms of quality when answering issues related to stock structure and life-history parameters. A similar approach can be adopted for any data types where the ability to carry out the data evaluation for the benchmark assessment, or the specific tasks, depends upon the outcomes of an initial exploratory exercise and expert consultation.

Appendix 1.1 Table 1– First draft issue list for parameters related to stock structure and life history to be completed by assessment EG (chair, stock coordinator or stock assessor)

The table below is meant as a rough-guide for the assessment working groups to identify necessary actions to be taken if problems with either stock structure or estimated life-history parameters are suspected. This is not input to the data evaluation workshop but is communicated to the relevant experts, ICES EGs, etc. for advice, leading to the development of an adjusted issues list given in Table 2 depending on the outcomes of the consultations with experts.

Issue topic	Problem / Aim	Work needed / possible direction of solution	Data needed	ICES groups to contact (as first step)
Stock structure	Possibility that two or more currently defined biological stocks belong to a single, larger stock unit and should be aggregated.	Literature search for existing evidence e.g. genetics, tagging, meristics, biological parameters	Literature; reports…	SIMWG
		Analysis of current datasets on biological parameters such as size at age, weight-at-length, maturity pattern between areas.	Available biological data from survey and catch	SIMWG, WGBIOP
		Analyse recruitment synchrony within current area and with surrounding stock areas.	Time-series of survey indices by age, by area within existing stock area, and recruitment series from current and adjoining stock areas.,	SIMWG
		Evaluate feasibility of aggregating catch data and revising survey indices and biological parameters to cover new expanded stock area	Fishery and survey data.	SIMWG; WGCATCH; WGISDAA; WGBIOP
	Possibility that two or more biological stock units are being assessed as one stock and should be split.	Literature search for existing evidence e.g. genetics, tagging, meristics, biological parameters, recruitment synchrony within existing area etc.	Literature; reports…	SIMWG

Issue topic		Problem / Aim	Work needed / possible direction of solution	Data needed	ICES groups to contact (as first step)
			Review available evidence (if any) for extent of mixing of stocks across the revised stock boundaries and effect on fishery and survey data from each area.	Check with SIMWG	SIMWG
			Source evidence of temporal stability of the stock boundaries and mixing rates, for example spatial expansion / contraction of range related to stock size and/or environment.	Spatial data on environmental variables related to stock distribution, and on stock structure parameters, over time	SIMWG
			Evaluate feasibility of splitting catch data and revising survey indices and biological parameters to cover new expanded stock area. Need some understanding of the relative size of the stocks, and the effect of splitting the data on the quality (bias and precision) of the assessment data for the new stocks	Fishery and survey data with high spatial and temporal resolution	SIMWG; WGCATCH; WGISDAA; WGBIOP
Life-history Parameters gg	Age and Growth	Very variable / inconsistent age readings within and between laboratories	Review calibration studies/exchanges results and any existing validation studies. Source other evidence such as tagging results, daily growth analysis, modal progressions.	Document repository/data from calibration sets and the CRR on age validations. Literature and Survey data	WGBIOP
		Unexplained differences in weight-at-length or age between countries and across time	1) Analyse consistency of methods used in each country. 2) Explore potential for real spatio-temporal variations in fish condition and size-at-age affecting national data	National data on age, length and weight, by area and time, and sampling protocols	WGBIOP

Issue topic	Problem / Aim	Work needed / possible direction of solution	Data needed	ICES groups to contact (as first step)
	Biases in growth parameter estimates due to gear selectivity	Evaluate how gear selectivity may be skewing length-at-age distributions.	Length-at-age distributions from the different data sources in relation to gear design or other relevant parameters	
Reproduction	Maturity keys not applied consistently with International standards and guidelines	Consult maturity workshop results. Investigate if correction factors can be estimated from histological/macroscopical comparisons to apply in SSB calculations	Document repository/data from calibration sets.	WGBIOP
	Timing and sampling scheme for maturity ogives may not be appropriate	Examine the biology of the stock in question and the suitability of the sampling for maturity ogives	Literature search on spawning times and spawning migration patterns; sampling dates and sampling design for maturity data	WGBIOP
	Fecundity estimates needed e.g. for computing spawning potential	Check if fecundity parameters or data and methods descriptions on fecundity exist for the stock or neighbouring stocks of the species, and if the methods used are consistent with current standards and protocols. If data are for a different area and stock, evaluate the appropriateness of the data or parameters.	Papers, reports, data files.	WGBIOP
Natural mortality	Need update from existing multispecies models	Consult the current multispecies model updates.	Most recent model outputs	WGSAM or other relevant EG

Issue topic	Problem / Aim	Work needed / possible direction of solution	Data needed	ICES groups to contact (as first step)
	Factors other than predation may be significant cause of M	Search literature/reports for evidence of mortality on stock due to parasites, starvation, spawning stress, environmental conditions or other processes that are known or suspected to be operating.	Papers, reports	WGBIOP
	Link with life-history parameters needed	Review suitable life-history based methods predicting overall M or M at age	Literature for methods; growth and maturity parameters	WGBIOP

Appendix 1.1 Table 2. Adjusted issue list for Data Evaluation process related to Stock structures and Life-history parameters, following initial exploration and advice from other Expert Groups consulted

The table provides an example of an updated issues list for the benchmark data evaluation incorporating advice from other experts or EGs consulted on issues related to stock structure and life-history parameters. The work needed by the data evaluation team, the data for this and the desired output for the Benchmark workshop are outlined. The same approach may be adopted for other data types.

Issue	Reply from experts	Work needed / possible direction of solution	Data needed	Desired output to the benchmark WK
Stock structure	There is no need to continue a split substock structure as currently done. Lumping is recommended	Aggregate existing fishery dataseries (landings, discards, catch-at-age). Derive new abundance indices from surveys that have previously been split into areas covering the old stock structure; derive biological parameters appropriate to larger stock. Estimate the uncertainty in the merged dataseries and compare with the existing split dataseries	All data by substock	Merged datasets and parameters with associated quality indicators
Stock structure	There are more stocks than currently assessed as one stock. Splitting into substocks is recommended and should be attempted.	Disaggregate existing fishery dataseries (landings, discards, catch-at-age). Derive new abundance indices from subsets of survey strata and stations where the survey covers a larger area; derive biological parameters appropriate to the revised smaller stocks. Estimate the uncertainty in the split dataseries and compare with the existing collated dataseries. Discuss relevant sampling protocols for collating split data keys for future assessment	High resolution fishery and survey data, and biological sampling	If it proves possible to carry out the splitting of data: Disaggregated datasets and parameters, with quality indicators If it proves impossible to carry out the splitting of data: Retain existing stock boundaries; carry out further work to identify data needs for splitting stocks in future.

Issue	Reply from experts		Work needed / possible direction of solution	Data needed	Desired output to the benchmark WK
Stock structure	There is no need to continue a split substock structure as currently done. Lumping is recommended		Aggregate existing fishery dataseries (landings, discards, catch-at-age). Derive new abundance indices from surveys that have previously been split into areas covering the old stock structure; derive biological parameters appropriate to larger stock. Estimate the uncertainty in the merged dataseries and compare with the existing split dataseries	All data by substock	Merged datasets and parameters with associated quality indicators
Life-history Parameters s	Age and Growth	(Very variable / inconsistent age readings within and between laboratories) (a) Yes – ageing problem exists	Provide bias and precision indicators to determine the adequacy of age related parameters, based on outcomes of ICES QA workshops and exchanges.	Results of ICES QA workshops and exchanges and further investigations.	Age data with recommendations on reliability and use of the data, including evidence of the problems, specific quality indicators needed, and further work needed to resolve problems.
		(b) No ageing problems exist	Provide bias and precision indicators needed for assessment, based on outcomes of ICES QA workshops and exchanges.	Results of ICES QA workshops and exchanges	Age data with specific quality indicators needed for assessment
		Unexplained differences in weight-at-length or age between countries and across time: a) Are most likely due to inconsistent approaches between countries	Document the differences in procedures for deriving individual fish weights by laboratories and evaluate the biases and additional variability induced in assessment data by these, both in terms of final weights at age and any biases propagated through sample raising procedures using estimated sample and catch weights.	National data on age, length and weight, by area and time, and sampling protocols.	Weight-at-age data with supporting evaluation of bias induced by inconsistent or inappropriate methods, including changes in bias over time.

Issue		Reply from experts	Work needed / possible direction of solution	Data needed	Desired output to the benchmark WK
Stock structure	There is no need to continue a split substock structure as currently done. Lumping is recommended		Aggregate existing fishery dataseries (landings, discards, catch-at-age). Derive new abundance indices from surveys that have previously been split into areas covering the old stock structure; derive biological parameters appropriate to larger stock. Estimate the uncertainty in the merged dataseries and compare with the existing split dataseries	All data by substock	Merged datasets and parameters with associated quality indicators
		b) Are most likely due to real differences in fish condition	No additional work needed; estimate weights at age using agreed methods	National data on age, length and weight, by area and time, as required by sample raising procedures	Mean weights at age as required for assessment.
		Biases in growth parameter estimates due to gear selectivity: Corrections needed for selectivity at younger ages	Identify analytical methods to correct for this where possible.	Length-at-age data	Parameter estimates corrected for selectivity.
	Reproduction	Maturity keys not applied consistently with International standards and guidelines: a) Yes: there is likely to be bias associated with maturity staging	Document inconsistencies and their impact on maturity ogives, including changes in interpretation affecting time-series. Develop correction factors if possible and document the methods	Information allowing standardization across national maturity keys. Maturity data and supporting data for raising to population level.	Annual maturity ogives, standardized as far as possible, or recommendation that maturity data cannot be standardized historically causing bias that may vary with time.

Issue	Reply from experts	Work needed / possible direction of solution	Data needed	Desired output to the benchmark WK
Stock structure	There is no need to continue a split substock structure as currently done. Lumping is recommended	Aggregate existing fishery dataseries (landings, discards, catch-at-age). Derive new abundance indices from surveys that have previously been split into areas covering the old stock structure; derive biological parameters appropriate to larger stock. Estimate the uncertainty in the merged dataseries and compare with the existing split dataseries	All data by substock	Merged datasets and parameters with associated quality indicators
	b) Low potential for bias – process data using standard methods; Consider annual maturity ogives or smoothed trends	Calculate yearly maturity ogives for the stock; Investigate statistical models to develop smoothed trends over time.	Available quality assured maturity-at-age data from appropriate survey/sampling time	Variable maturity ogives for the time-series, with quality indicators.
	Timing and sampling scheme for maturity ogives may not be appropriate: a) Sampling scheme appears biased by over-representation of samples from spawning grounds at spawning time.	Document the sampling schemes and identify the potential for bias in maturity ogives by the current and historical schemes. Identify other sources of data that are from more representative sampling even if sample sizes are smaller, and provide estimates and quality indicators.	Time-series of maturity data and sampling protocols	Maturity ogives with advice on likely biases if used in assessment, or maturity ogives based on less biased data. Recommendations for more appropriate sampling design.
	b) Sampling design does not appear biased	No additional work needed; estimate maturity parameters using agreed methods	Time-series of maturity data	Maturity ogives with quality indicators

Issue	Reply from experts	Work needed / possible direction of solution	Data needed	Desired output to the benchmark WK
Stock structure	There is no need to continue a split substock structure as currently done. Lumping is recommended	Aggregate existing fishery dataseries (landings, discards, catch-at-age). Derive new abundance indices from surveys that have previously been split into areas covering the old stock structure; derive biological parameters appropriate to larger stock. Estimate the uncertainty in the merged dataseries and compare with the existing split dataseries	All data by substock	Merged datasets and parameters with associated quality indicators
	Fecundity estimates needed for computing spawning potential: a) Data / parameters are available for the stock	Document parameters, data and descriptions of methods; evaluate consistency of different studies if more than one; provide estimates of fecundity at size or age with quality indicators	Fecundity data or published parameters	Fecundity at size or age as required for assessment, plus quality indicators.
	Natural mortality s Need update from existing multispecies models Yes – new natural mortality values are available	Acquire the updated time-series of natural mortality-at-age. Discuss the quality of this new time-series and explain changes from previous values. If no update, provide the relevant information from the previous model, and provide information from report on quality / reliability of estimates.	Model outputs	Updated time-series of M plus
	Factors other than predation may be significant cause of M: Yes – there is some evidence of this	Document the evidence. Provide information, if available, on possible range of additional mortality being generated, and the life-history stages and years affected.	Results of specific studies	Range of possible additional mortality and years/ages affected.

Issue	Reply from experts	Work needed / possible direction of solution	Data needed	Desired output to the benchmark WK
Stock structure	There is no need to continue a split substock structure as currently done. Lumping is recommended	Aggregate existing fishery dataseries (landings, discards, catch-at-age). Derive new abundance indices from surveys that have previously been split into areas covering the old stock structure; derive biological parameters appropriate to larger stock. Estimate the uncertainty in the merged dataseries and compare with the existing split dataseries	All data by substock	Merged datasets and parameters with associated quality indicators
	Link with life-history parameters needed Yes: most appropriate method	Document life-history parameters, methods of inferring M from these, and the estimates from the preferred methods	Biological parameters	Range of plausible M at age based on life history

Appendix 1.2 Notes on fishery management regulations and actions that are expected to have caused changes in the quality of fishery catch data or the selectivity patterns of fisheries.

Fishery management regulations can induce large changes in the quality of catch statistics, for example due to changes in compliance, discard patterns or catch reporting. Technical conservation measures related to gear design or spatio-temporal closures can cause changes in fishery selection patterns that need to be reflected in stock assessment models where fleet selectivity patterns are estimated.

If there is an existing Stock Annex, this should already provide a history of management measures relevant to the assessment and advice although it may not be sufficiently complete and adequate and must be reviewed and updated. Ideally, ICES assessment working groups would have access to a detailed regional chronological summary of fishery management measures that affect fleets, such as effort limits, gear restrictions, spatio-temporal closures and decommissioning schemes, and specific measures for stocks such as TACs, boat landings limits, minimum conservation reference sizes and implementation of the landings obligation. However there is (currently) no consolidated and complete description of these for any region to refer to. STECF reports contain useful information and should be consulted for historical perspectives and evidence of impacts of management measures, for example STECF 12/20:

http://stecf.jrc.ec.europa.eu/documents/43805/432011/2012-11_STECF-12-20+Defining+selectivity+under+TM+regulation_JRC76897.pdf

For the stocks being benchmarked, the data evaluation team should source information that is available to identify chronological changes in management measures that are known to (or are likely to) have affected data quality or fishery selectivity to an extent that will affect the choice, configuration, diagnostics of the assessment model. Where possible make use of graphical or tabulated summaries to give a clearer overview of changes over time. Some examples are given in Figure 1 and Table 1 below:

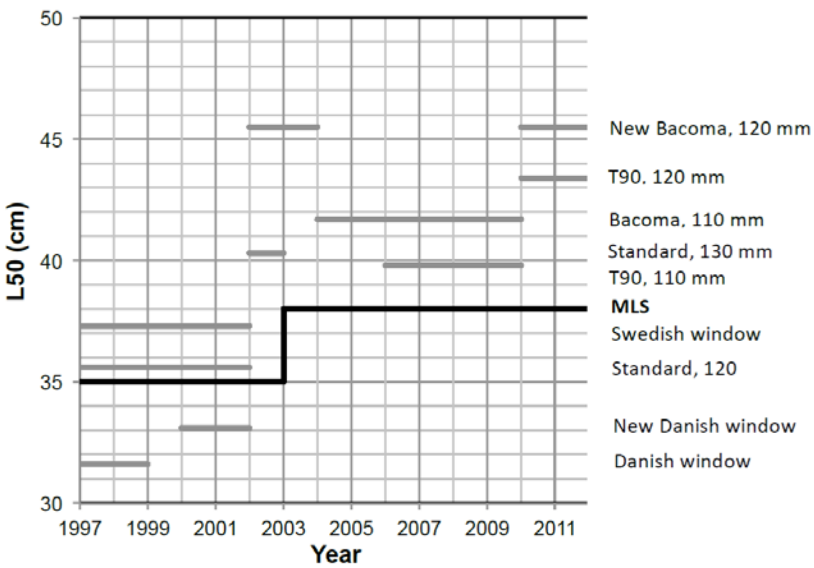


Figure 1. Schematic presentation of regulation in changes over time and the effect on the selectivity (L50). The figure is modified from Feekings *et al.* 2013.

Table 1 Example of a format for summarizing a time-series of catch management measures and their outcomes.

Year	International management			National management					
	Management system	TAC (t)	TAC uptake %	Country 1		Country 2		Country 3	
				Management system	Quota uptake	Management system	Quota uptake	Management system	Quota uptake
1995	None	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1996	None	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1997	None	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1998	None	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1999	TAC	10,000	50	Weekly limits	67	Monthly limits	33	None	50
2000	TAC	10,000	60	Weekly limits	67	Monthly limits	67	None	50
2001	TAC	10,000	60	Weekly limits	67	Monthly limits	73	None	45
2002	TAC	10,000	55	Weekly limits	67	Monthly limits	67	None	38
2003	TAC	10,000	62	Weekly limits	60	Monthly limits	67	None	60
2004	TAC	10,000	65	Weekly limits	73	Monthly limits	73	None	53
2005	TAC	10,000	65	Weekly limits	67	Monthly limits	63	None	65
2006	TAC	10,000	70	Weekly limits	80	ITQ	73	None	60
2007	TAC	10,000	70	Weekly limits	80	ITQ	70	Weekly limits	63
2008	TAC	10,000	70	Weekly limits	83	ITQ	73	Weekly limits	58
2009	TAC	10,000	65	Weekly limits	73	ITQ	67	Weekly limits	58
2010	TAC	10,000	70	Weekly limits	73	ITQ	77	Weekly limits	63
2011	TAC	10,000	80	ITQ	87	ITQ	80	Weekly limits	75
2012	TAC	10,000	80	ITQ	87	ITQ	77	Weekly limits	78
2013	TAC	12,000	70	ITQ	69	ITQ	69	Weekly limits	71
2014	TAC	8,000	100	ITQ	100	ITQ	100	Weekly limits	100

Appendix 1.3 ICES 2015 Assessment Expert Groups summary of national discards sampling programme design – designed by PGDATA

Country:……… **ICES Assessment Working Group:…….**

Sampling programme name:……..

Stocks covered by programme : …….

	Sampling frame		Survey design and sample selection		Estimation procedures			Self evaluation of potential for bias (1-3 where 1 is the best)[2]
Years/ time periods	Vessel sizes and gears covered	ICES Divisions covered	Describe Survey design and vessel selection	Stratification	Raising procedure for stratum estimates for a stock	Methods to impute missing stratum estimates "borrowing procedures"	Variance estimates	
2012	All national registered vessels >10m using towed demersal otter trawls, beam trawls and seines;	IVa,b,c	Non-random selection of vessels on opportunistic basis to meet sampling quotas by stratum.	4 x quarter 3 x area (divisions) 3 x gear (otter; beam; seine)	Trip-raised estimates summed for sampled vessels in stratum, and then raised to total fleet using reported total fleet landings of stock and reported landings of stock by sampled vessels.	Discard rates for "nearest neighbour" sampled strata applied to reported landings of stock for missing strata. Decision on which strata to impute from is based on sample sizes and expert judgment.	Analytical	3
2013	As above	As above	As above	As above	As above	As above	As above	3
2014	Vessels 7 – 9.9 m included; vessels using fixed/drift nets included.	IV a,b,c and VIId	Stratified random selection of vessels, with evenly spread sampling effort across year.	2 x area	Trip-raised estimates summed for sampled vessels in stratum, and then raised to total fleet using sampling probabilities (total number of trips reported for fleet divided by numbers of sampled trips).	Stratum estimates for sampled strata are combined and raised to all reported landings including missing strata.	Bootstrap	1

[2] See guidelines below for completion of the tables.

Name of person completing table:		Date:	

Appendix 1.3 continued: Excel file for ICES 2015 Assessment Expert Groups summary of national discards data quality, 2012–14

ICES 2015 Assessment Expert Groups summary of national discards sampling quality by stock

(some example entries are included and should be deleted)

Country:

Sampling programme name: *(link to Word table)*

Species: *(e.g. grey gurnard)*

Stock area: *(e.g. Subarea VI and Divisions VIIa–c and e–k)*

Stock code: *(e.g. gug-celt)*

Overall assessment by national scientists of reliability of the survey for this stock
Data for 2014 considered good quality due to improved design and coverage; data for 2012 and 2013 are much less reliable and should be used with caution

Year	Total No. fishing trips sampled in strata covering stock area *	No. sampled trips where fish of this stock was discarded	No. of trips where discards of this stock were measured for length	No. of trips where discards of this stock were aged	Raised discards estimate (total weight , tonnes)	RSE of raised estimate **	Refusal rate (%)***	Self evaluation of potential design bias (from final column in Word table)	Additional comments (e.g. problems with missing stratum data)****
2012	50	25	10	0	120	0.7	15	3	No beam trawl samples in Q3 & Q4
2013	60	28	20	0	250	0.6	10	3	
2014	120	55	50	0	100	0.5	5	1	

* Only include trips using gears that could catch this species, and where all or part of the trip was in the stock area (e.g. for sprats, exclude longline trips!)

**RSE (relative standard error) of raised total discards estimate $\hat{y}$, i.e. RSE = SE ($\hat{y}$)/$\hat{y}$

*** Refusal rate is a number indicating the percentage of the vessels approached that declined to have observers on board

**** If one or more important strata for the species were intended to be sampled but were not sampled for any reason, this could lead to bias. Highlight years with such problems.

Appendix 1.3 continued Instructions for completing the Word table and Excel file describing national discards sampling programmes

There are two files for each country to complete, to give each ICES stock assessment Working Group some basic information to help evaluate the quality of new (and existing) discards estimates that are included in this year's ICES assessment WG data call.

- The first is a Word table (above) which provides text describing the sampling programme and how it has changed over the last three years for which data are requested, highlighting any issues with design that could lead to large (and potentially varying) bias in time-series of discards estimates. This table should refer to a sampling programme that is used for a defined set of stocks covered by the data call. If an individual country has more than one sampling programme applicable to different stocks and areas covered by the assessment WG (e.g. observer sampling for some stocks and self-sampling for others) the word table has to be filled out for every programme. Remember for each one to write the name of the programme and which stocks are covered by it.
- The second file is an Excel file to capture, on a stock-by-stock basis, some basic statistics about the amount of sampling that has taken place, and includes some simple data quality indicators. Please enter the name of the sampling programme, ensuring it matches the one used on the corresponding Word table. Use a separate worksheet for each stock.

It is emphasized that this is a short-term exercise for Assessment WGs in 2015. It does not represent an in-depth evaluation of data quality as would be expected for a benchmark data compilation and evaluation, and which would require more detailed scrutiny of sampling design against best practice for the full time-series of data, identifying quality issues arising at the implementation and analysis stages, scrutinizing information at the scale of individual survey strata, and viewing a range of diagnostics to evaluate how representative the sampling was. ICES expert groups on fishery sampling (SGPIDS1-3; WKPICS1-3; PGCCDBS) have explored detailed quality assurance reporting and these reports should be consulted for further guidance on data quality evaluation.

To help the Assessment WGs evaluate the quality of your discards estimates, we are therefore also asking you to give a self-assessment of the potential for bias in your sampling scheme and resultant estimates, using the following scores derived from a scoring table shown on the next page:

1) Least potential for bias – Assessment WG can be confident in using the discard data (provided there are sufficient samples for a given stock)
2) Some issues with potential bias – WG should use discard data with some caution, and comment on the quality issues with the data.
3) Large potential for bias –WG should consider carefully before presenting and using the discard data, or exclude the data for all or part of the time-series from assessments, and comment on the quality issues with the data.

Ideally this scoring would be done using the detailed quality assurance reporting procedures mentioned above, but in the absence of this, a simpler approach is proposed.

Guidance for completion of the final column in the Word table ("Self evaluation of potential for bias") is given on the next page

In addition, the Excel table contains a box at the top right, where you can enter any other comments to the assessment working group concerning your views on the quality of the discards estimates for the stock covered by the table. The assessment WGs may carry out additional checks on the internationally aggregated discards data, for example internal consistency of age compositions and residual patterns around model fits to data.

Guidance for completion of final column in Word Table "Self evaluation of potential for bias". *(See glossary below for terminology and guidelines)*

Design of survey	National sampling frame coverage	How representative are the sampled trips in each stratum?	Score
Probability-based sampling design with selection of vessels that is random or close to random.	Frame covers most vessels in total national population of active vessels that discard the stocks covered by the assessment WG.	Diagnostics show that sampled trips are representative of the overall national population of vessels, or data can be post-stratified and re-weighted to be representative.	1
		Diagnostics show that sampled trips are consistently and markedly different from the population, and this cannot be corrected by re-weighting, or no diagnostics have been scrutinised.	3
	Relatively large numbers of vessels are excluded from the frame (e.g. small vessels; vessels in remote ports).	Diagnostics show that sampled trips are representative of the overall national population of vessels, or data can be post-stratified and re-weighted to be representative.	2
		Diagnostics show that sampled trips are consistently and markedly different from the population, and this cannot be corrected by re-weighting, or no diagnostics have been scrutinised.	3
Ad-hoc sampling design where vessels are selected in a non-random, opportunistic or subjective way	Frame covers most vessels in total national population of active vessels that discard the stocks covered by the assessment WG.	Diagnostics show that sampled trips are representative of the overall national population of vessels, or data can be post-stratified and re-weighted to be representative.	2
		Diagnostics show that sampled trips are consistently and markedly different from the population, and this cannot be corrected by re-weighting, or no diagnostics have been scrutinised.	3
	Relatively large numbers of vessels are excluded from the frame (e.g. small vessels).	Diagnostics show that sampled trips are representative of the overall national population of vessels, or data can be post-stratified and re-weighted to be representative.	2
		Diagnostics show that sampled trips are consistently and markedly different from the population, and this cannot be corrected by re-weighting, or no diagnostics have been scrutinised.	3
Other designs	Special case of very high observer coverage or proven reliable self sampling coverage of most of fleet		1
	A relatively small subset of vessels is selected as a reference fleet and each one is sampled at intervals throughout year.	Diagnostics show that sampled trips are representative of the overall national population of vessels, or data can be post-stratified and re-weighted to be representative.	2
		Diagnostics show that sampled trips are consistently and markedly different from the population, and this cannot be corrected by re-weighting, or no diagnostics have been scrutinised.	3

Appendix 1.3 continued: Glossary

Probability-based sampling:

Vessels are selected in a way that controls the probability of individual vessels being sampled (e.g. a selection of vessels is made from all the vessels in the frame, and the target number of trips to sample in a stratum is set to achieve a desired proportion - e.g. 1% - of the total number of trips, which represents the probability of an individual vessel being selected in the stratum). The probabilities (sampling fractions) are used for raising the estimates for the sampled vessels to all the vessels in a stratum within the sampling frame (other auxiliary variables such as landings weight or fishing effort may be included in the raising procedure). It is important to note that if a subset of vessels is deliberately excluded from the frame (e.g. very small vessels, or those predominantly using a particular gear) these have a predetermined selection probability of zero, and this is part of a probability-based scheme. This differs from sampling where vessels could be sampled but are rejected on a subjective, ad-hoc basis by an observer in favour of another vessel – in this case there is no predetermined list of vessels that are excluded from the frame and the selection probabilities are therefore not controlled.

Random vessel selection:

To be a true probability-based random selection, all vessels in the frame must have a chance to be selected, with a given probability of selection in each stratum. This can be done using randomized draw lists. In some cases not all vessels need to have the same probability to be selected but it has to be a known probability. The probability could differ in a controlled way between vessels according to an auxiliary variable such as vessel size, trip numbers or trip duration which is positively correlated with quantities discarded. Many countries are moving towards greater randomization of sample selection but may currently adopt vessel selection procedures that are not strictly random but follow a protocol that spreads sampling across vessels in a way that tries to achieve representative coverage and minimizes the freedom of observers to make ad-hoc decisions. This can be considered closer to "probability based" than to "ad hoc" sampling, and this aspect of the sample selection procedure should be clearly stated in the "survey design and sample selection" column of the Word table. It is important that diagnostics are developed to evaluate how representative the sampling has been of the total population of vessels and their trips.

Ad hoc sampling:

This term is used here to define any vessel selection procedure where vessels are **not selected at random** and observers are given the freedom to choose which vessels to sample from within a gear group or area for example to meet a quota of X trips. There can be many reasons for an observer to select a given vessel; for example, if large vessels are chosen in preference because they are more comfortable to work on, or small vessels with day trips are chosen preferentially to help meet sampling targets, or vessels working only from the nearest ports are repeatedly chosen in preference to ones from more distant ports, to minimize travelling. If the observer can chose who to contact rather than make a random selection or follow some other procedure to spread sampling across the fleet in a representative way, this is to be considered as an ad hoc sampling.

Reference fleet:

A fixed set of vessels that is selected at the start of the year, and where each one is sampled throughout the year either by full observer or self-sampling coverage, or by random or systematic random sampling of trips. The reference fleet may be partially or fully re-selected each year or be fixed for a longer period. (Example is the Norwegian coastal reference fleet where types of vessels are selected in proportion to activities of similar vessels and gears in the overall fleet.)

Refusal rate:

The percentage of the vessels approached for sampling where the skipper declined to have observers on board. Large refusal rates indicate a potential for bias if most refusals are by captains of vessels that have, for example, persistently above-average discard rates. Potential for bias due to this may also be apparent from logbook data showing how representative the refused trips are, compared to the population of vessels as a whole in a stratum (see below). Similar biases may result from an **observer effect**, where captains alter their discarding patterns with an observer on board.

Representative sampling and diagnostics to evaluate it.

On average, over many annual implementations of a fully randomized sampling scheme, the occurrence of fishing trips by gear or area (e.g. ICES rectangle) should be more or less the same in the collection of samples as in the population. If not, the sampling has been biased by:

- Bias towards selecting vessels with particular gears and fishing patterns (a design problem, possibly due to ad-hoc sample selection)
- High refusal rates by skippers of vessels with different discarding patterns from the rest of the population;
- Different behavior of vessels with observers on board compared with non-observed vessels (an observer effect).

Achievement of representative sampling is most likely when using a randomized, probability-based approach with extensive coverage of the frame, but this does not guarantee that captains will not alter their behavior with observers on board, or refuse access to avoid observation of high discard rates, leading to biases in fleet-raised estimates of discards.

The use of a variety of diagnostic tools to understand if sampled trips are representative, on average, of the non-sampled population, is an essential component of good practice. Even the best designed, probabilistic survey may give very inaccurate estimates if there are strong observer effects or high refusal rates at the implementation stage. This may not happen, but without diagnostics there is a potential for bias and the bias cannot be quantified. ICES Expert groups such as PGCCDBS, SGPIDS and WKPICS have given advice on diagnostic tools, and graphics tools for this have also been developed in the COST project.

It should be noted that in any year, a random sample of trips may, by chance, differ noticeably in composition or spatio-temporal fishing patterns from the logbook records held for the rest of the population of vessels and trips in a stratum. Data therefore need to be scrutinized over several years. Where there is evidence that this a chance occurrence and not a reflection of known or suspected observer effects or biases due to refusals, there is a potential to improve accuracy by post-stratification and re-weighting of sampled trips in each stratum using characteristics such as trip duration or gear group that is accurately and consistently recorded for the sampled and unsampled trips. This can only be done if there are sufficient samples for each of the post-strata.

Completion of the third column in the scoring table "*How representative are the sampled trips in each stratum*" should therefore use diagnostics for several years, if these are available and have been scrutinised. Use the score for "*Diagnostics show that sampled trips are consistently and markedly different from the population and this cannot be corrected by re-weighting*" if there is a persistent, marked difference between the samples and the population in a stratum and it is not possible or valid to correct for any resultant bias using post-stratification and re-weighting of samples based on the exhaustive trip data from EU or other logbooks.

Sampling frame

The total collection of vessels and trips from which selections are made for sampling. This may or may not cover the complete population of vessels or trips – for example if vessels below a certain size, or from certain ports, are excluded. The frame coverage must be documented, and any differences between vessels and activities inside and outside the frame investigated and documented, in order to understand the potential for bias.

Survey design:

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

Stratification:

Strata are non-overlapping groups of vessels or trips within a sampling frame that may have different sampling rates – e.g. stratification by vessel characteristics, quarter or region. A sampled trip can occur in only one stratum.

More information:

If you are looking for more information on discard sampling programmes and designs, and evaluation of data quality, ICES have hosted a series of planning groups, workshops and working groups on these topics (WKACCU, WKPRECISE, WKMERGE, WKPICS 1-3, SGPIDS 1–3, WGCATCH) – all reports can be found at the ICES homepage for each group or in the ICES library.

Appendix 1.4 Estimating commercial and recreational fishery catches and associated measures or indicators of bias and precision.

(a) Fishery landings databased on exhaustive data collection

These data may include:

- Official commercial landings figures from logbooks and sales slips, without any adjustment
- Official commercial landings adjusted in some way, for example to allocate landings to the correct fishing ground, adjust for stock mixing, to disaggregate mixed-species landings records using sample data, or make other corrections for misreporting or underreporting. Explain how the adjustments are made.

(b) Fishery catch estimates from sampling schemes

Commercial fishery landings, discards or recreational catch data may be derived from sampling surveys such as:

- Survey-based estimates of commercial landings for example through data provided by random samples of vessels from a list frame of vessels (mainly for small-scale fisheries with no EU logbook requirements).
- Surveys to estimate components of the landings of vessels that are not recorded exhaustively or accurately from an existing exhaustive logbook or sales slips scheme, including to estimate species compositions of landings recorded in mixed species categories.
- Surveys to estimate commercial fishery discards from at-sea sampling or self-sampling schemes
- Surveys to estimate recreational fishery catches, typically from a range of off-site or on-site surveys (see reports of ICES Working Group on Recreational Fisheries Surveys).

(c) Evaluating the reliability of catch data: i) historical biases and trends in bias

Without information on potential magnitude of biases in catch data and how they may have altered over time, it is impossible to know if historical population size can be reconstructed from the catches in an assessment for all or even a recent period with sufficient accuracy to support fishery management decisions. Knowledge of bias can also allow construction of plausible alternative scenarios for time-series of fishery catches, which can be used for investigating the sensitivity of management advice to this.

For exhaustive data collection schemes (e.g. EU logbooks) an evaluation should be provided of known or suspected biases in official landing statistics due to the following causes, or other causes, and a full explanation must be provided of any methods applied in an attempt to correct for these biases:

- Misreporting by area or species
- Incorrect allocation of landings records to fishing grounds by the national authorities

- Under- or over-reporting of landings within an area, including removals which are not taken into account by official statistics, such as reporting exemptions for small catches by under-10m vessels without EU logbooks
- Inaccurate species identification (e.g. reporting as a mixed-species category)
- Changes in conversion factors

For sampling schemes to estimate quantities of discards, recreational catches or small-scale commercial fishery catches, an evaluation should be provided of known or suspected biases in estimates due to the following causes, and a full explanation should be provided of any methods applied in an attempt to correct for these biases:

4) **Bias related to survey design**: Sampling schemes where sampling is ad-hoc and opportunistic and does not follow any statistically sound design, will almost certainly be biased to some extent, and the bias may vary systematically over time if the schemes or personnel are changed. The table developed by PGDATA for the ICES data call for discards data in 2015, to document changes in sampling designs over time (Appendix 1.3), could be used as a template. If there are historical periods where details of the sampling schemes are no longer fully understood, this should be indicated. Particular data quality issues should be explored in greater depth to determine if or how the data can be used.

5) **Bias that arises during implementation of the survey**: If sampling becomes non-representative of the population being surveyed, estimates will probably be biased. Examples include: high refusal rates by skippers who have above-average discard rates; "observer effects" where there is a different fishing behavior with observers on board; or "recall bias" where respondents to a recreational fishery survey tend to overestimate or underestimate catches recalled for a previous period. Major gaps in sampling, for example strata with no or inadequate samples, may cause bias depending on how this is dealt with during analysis.

6) **Biases that arise at the data analysis stage**: There are often large raising factors from sampled fishing trips to the total fishery, and these must reflect the stratification and sampling probabilities at each stage in sampling. If samples are aggregated and raised in a different way, the raised figures may be biased. In some cases auxiliary variables such as effort or landed weight (for the stock or group of stocks) are used to improve the accuracy of discards estimates, but may themselves be biased and an evaluation of the appropriateness of methods used is needed.

Some bias diagnostics

Some benchmark assessments have attempted to use the bias "scorecard" and traffic lights system developed by the ICES Workshop on Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment (WKACCU; ICES 2008). This is a useful tool for national laboratories to check for bias in their data, but has generally proved too detailed for use in benchmark assessments. The data evaluation team should focus on documenting biases that are sufficiently large to cause concern over their impact on the quality of an assessment.

In addition to potential for bias inherent in the data collection methods, a range of diagnostics can be developed. Some examples are given below of information that can be used to assess these sources of bias, and how it can be presented in the report for the benchmark data evaluation meeting:

- Comparison between official landings for individual trips and the landings observed during port sampling programs or from observer data collected at sea. This is only valid for landings on shore (including composition of size categories) if it is known that the complete landing is available at the sampling site. An example for an observed trip at sea is given below:

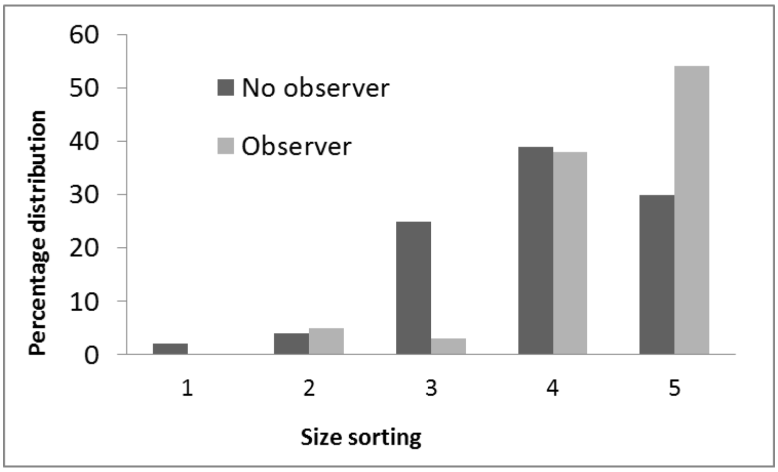


Figure 2. Size sorting distribution of cod caught by OTB_MCD_90-119_0_0, in Kattegat 3AS.

- Exploration of other sources of information on landings, for example information from vessel inspections by the control agency (if it was collected randomly)
- A comparison between spatial distribution of the catches as reported in logbooks, with maps of effort based on VMS data
- Compilation of conversion factors used by different countries; methods used to derive the factors, and changes over time
- Information about how the catch is split by species when landed together (mixed species landings)
- For mixed-stock pelagic fisheries, information on how the split between stocks is made in catch statistics, in particular if there are industrial fisheries involved
- For survey-based catch estimates, scrutiny of sampling coverage to identify magnitudes of gaps in coverage such as missing stratum data, evaluation of systematic observer effects by comparison of fishing patterns and landed catch composition of observed and non-observed vessels using same gear in the same area and time period.
- Checking if appropriate auxiliary variables are used in raising procedures for bycatch and discards surveys. Auxiliary variables such as effort, landings of the stock or a group of stocks, should be correlated with the amount of bycatch or discard. If there is no correlation that is persistent over the time-series, these variables will not improve the estimates and could cause a bias that might also be changing systematically over time.

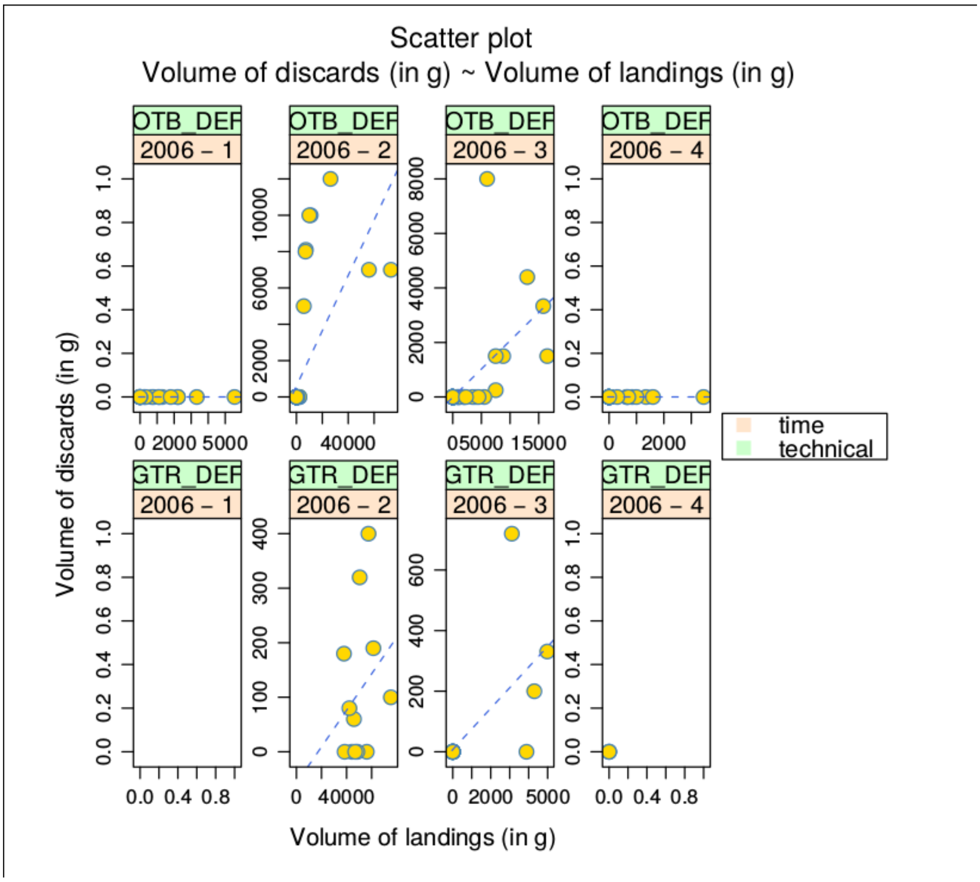


Figure 3. Example of correlation plots between discard quantities and the landed weight of the stock, where landings are being examined as the auxiliary variable in a ratio estimator to estimate total discards by fleet. (See ICES Workshop on Discard Raising Procedures, WKDRP, ICES 2008, for more information).

(d) Evaluating the reliability of catch data: ii) precision of catch estimates from sampling schemes

Almost all commercial discards and recreational fishery catch estimates are from sampling surveys. If these are carried out using statistically sound methods with sufficient samples per stratum, and the estimation methods correctly reflect the hierarchical cluster sampling that is typical, it is possible to provide a reliable value of standard error for the estimate of total annual discards or recreational catch for a stock or for any post-strata such as region or fleet component.

If sampling rates are very low, or are zero, for some strata, or if there are many sampled trips but only a very small number have discards of the stock being benchmarked, the standard error estimates may be unreliable. If discarding of a species is a rare event, total fleet discards may be so low in relation to landings that the reliability of the estimates is not an important issue, but this would need to be evaluated. Information on the achieved sampling, particularly the annual number of primary sampling units (e.g. vessel trips) for each independent survey, and the numbers with discards of that species, is important for evaluating the precision estimates or may on its own provide a proxy for precision.

Estimates or indicators of precision for each year in a time-series may be provided in the following ways:

- Standard errors or relative standard errors (SE/mean)

- Effective sample sizes ESS (for a random, or stratified random, cluster sampling scheme – the number of random, non-stratified independent samples that would give the same precision)
- Number or primary sampling units (as a proxy for effective sample size).

ESS and numbers of PSUs are useful mainly for crude comparisons of the likely magnitude of differences in precision between estimates for different fleets or strata (e.g. 3vs.30 PSUs) or if there are enough data to allow meaningful estimates.

There is no standard way to present this information in the most useful way to all benchmark assessment processes. ICES PGCCDBS (ICES, 2011c), WKPICS (ICES, 2013b) and the Excel table format that countries were requested to complete in 2015 as part of the ICES data call for stock assessment data (see Appendix 1.3) all provide examples of how information could be presented. The most useful analysis of sampling coverage may require scrutiny of sampling levels for individual strata and countries in relation to landings or discards quantities. In future, development of the Regional Databases and codes for interrogating the databases and estimating discards, which will be tools to support the work of the planned EU Regional Coordination Groups, will provide greater flexibility to explore the quality of national data and combined international estimates based on fishery sampling to estimate discards or biological variables based on sampling schemes.

Appendix 1.5. Notes on estimating the length and age distributions of fishery landings and discards, with indicators of bias and precision.

Length compositions of fishery catches are usually derived from multistage sampling programmes, where for example a port is selected, then vessel landings are selected, then boxes of fish within each landing, then individual fish are measured and/or sampled for age determination. Such schemes should be based on a sound statistical design and implementation to ensure representative sampling of ports, vessels or trips within strata, and the analysis should properly take account of the sampling probabilities at each stage and adopt statistically defensible approaches for other estimation procedures such as use of ratio estimators or model-based estimators. Detailed accounts of sampling designs and estimators, and evaluation of data quality, are given in the reports of ICES WKDRP (ICES 2008), WKPRECISE (ICES 2009a), WKMERGE (ICES 2010b), WKPICS (2011–2013), SGPIDS (ICES 2011–2013) and WGCATCH (ICES 2014).

Given the many steps involved in generating an international raised length or age composition, there is considerable potential for errors at each stage that propagate through to affect the quality of the final data inputs to an assessment model. These include:

- Biases associated with sampling designs that are ad-hoc, opportunistic and depart substantially from a stratified random design.
- Sample sizes too small in any or all strata to provide usable information.
- Data analysis methods that do not properly account for sampling design or use inappropriate ratio estimators or modelling approaches.
- "Borrowing" of age length keys from areas /fisheries where age frequencies within length classes differ from the fisheries the data are being applied to.
- Errors in ageing – both random and systematic, including systematic differences in interpretation of otoliths, scales or other material between laboratories, and drift over time in age interpretation by national laboratories.

The current implementation of InterCatch at ICES expects uploading of national data by métier or métier group to support mixed-fishery models, and this requires an additional estimation step of post-stratifying national sampling data to derive fleet-raised estimates of length and age compositions by métier or métier group before uploading, which can lead to some métiers having low or zero sampling in a year. Stock coordinators are then required to "borrow" estimates from other countries and métiers, which could be done subjectively and without reference to the quality of borrowed data. The data evaluation team should advise on the appropriateness of this approach for the stock being assessed, and if necessary rework the raising and aggregation using more statistically robust methods for comparison with InterCatch results.

To some extent the quality of age composition data can be evaluated through inspection of how accurately the data appear to track year classes over time. However in many cases there is considerable variability, and without a detailed investigation of how data were collected and interpreted, or any estimates or indicators of precision and bias, it is impossible to understand what are the main contributors to this uncertainty, what is needed to improve the data, and the costs of this.

Appendix 1.6 Notes on fishery–independent and dependent data sources on fish abundance

(a) Estimating precision for fishery-independent surveys

The interpretation of standard errors or CVs will depend on the design of the survey, for example randomized or fixed station designs and the coverage of the stock area – obtain advice from WGISDAA on this. Outliers or other factors could result in unusual changes in apparent survey precision, and should be investigated and documented. Consult the benchmark stock assessment team for the type of quality indicators needed – for example CVs (if appropriate) of the overall indices by age class each year, or CVs of the total age-aggregated numbers with separate precision indicators for length or age compositions such as effective sample sizes. There may be additional "year effects" in surveys where some factor such as weather or tides cause changes in catchability over most or all the stations, and result in underestimation of the true variance of the indices if only the between-station variability is used. This may need to be accounted for in assessment models where input CVs are used for providing prior weights to survey data, to avoid over-weighting individual surveys. Consult WGISDAA and stock assessment team for advice. For age-based indices, evaluating the internal consistency of age compositions can be useful for highlighting year effects common to all or most ages, although these may also be induced by changes in interpretation of age material.

(b) Issues with fishery dependent data

In principle, fishery dependent indices such as commercial CPUE data can provide useful indices of population trends in abundance provided that changes in CPUE are proportional to changes in stock size. This issue has previously been addressed by the ICES Workshop on the Utility of Commercial CPUE and VMS Data in Assessments (WKCPUEEFFORT; ICES 2011d), and useful guidelines and recommendations are found here. The supposition of linearity or proportionality between CPUE and stock status is based on the assumption that catchability of the fleet remains constant over time and that recorded or apparent effort is stable and reflective of actual or effective effort.

In practice, these assumptions are violated due to changes in catchability associated with technological creep, resulting in improvements in gear efficiency and the ability of fisheries to maintain catch rates even when the overall abundance declines, by targeting 'hot spots'. Furthermore, in many fisheries, it is often the landings rather than the catch that is actually monitored, more correctly we should use the term landings per unit of effort (LPUE). Where discarding contributes a significant source of mortality and, more critically, discarding profiles vary, LPUE estimates may suffer a degree of bias.

An EU JRC Workshop on Transversal Variables linking economic and biological effort data (EU 2015) showed that calculation of days at sea and fishing days in the EU Member States is carried out using several different methods, and recommended a workshop on ways to estimate fishing days for different types of vessel, gears and reporting schemes and to recommend how to harmonize data collection across Member States. Definition of effort and how it is recorded is therefore an important task.

Appendix 1.7. References

EU 2015. Report on the Workshop on Transversal Variables (Linking economic and biological effort data (call) design). Luxembourg: Publications Office of the European Union. EUR – Scientific and Technical Research series. 80 pp. ISSN 1831-9424 (online), ISSN 1018–5593 (print). ISBN 978–92–79–46593-2. doi: 10.2788/385049

ICES 2007a. Report of the Workshop on Sexual Maturity 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 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 Discard Raising Procedures. ICES ACOM

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

ICES 2009a. Report of the Workshop on methods to evaluate and estimate the precision of fisheries data used for assessment (WKPRECISE), ICES CM 2009/ACOM:40, 43 pp.

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

ICES 2010b. Report of the Workshop on methods for merging métiers for fishery based sampling (WKMERGE), ICES CM 2010/ACOM:40, 94 pp.

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, 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 Commercial Catches, Discards and Biological Sampling (PGCCDBS), ICES CM 2011/ACOM:40.

ICES. 2011d. Report of the Workshop on the utility of commercial CPUE and VMS data in assessment (WKCPUEFFORT), 5–7 April 2011, Copenhagen, Denmark. ICES CM 2011/ACOM:49. 71 pp.

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

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

ICES 2012c. Report of the second Workshop on Practical Implementation of Statistical Sound Catch Sampling Programmes. ICES CM 2012/ACOM:54

ICES 2012d. Report of the Planning Group on Commercial Catches, Discards and Biological Sampling; PGCCDBS 2012. ICES CM 2012/ACOM:50.

ICES 2013a. Report of the Working Group on Recreational Fisheries. ICES CM 2013/ACOM:23.

ICES 2013b. Report of the Study Group on Practical Implementation of Discard Sampling Plans (SGPIDS). ICES 2013/ACOM:56

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

ICES. 2015. Report of the Working Group on Commercial Catches (WGCATCH), 10–

14 November 2014, ICES HQ, Copenhagen, Denmark. ICES CM 2014\ACOM:36. 295 pp.

ICES 2015a. Interim Report of the Stock Identification Methods Working Group (SIMWG). ICES CM 2015/SSGEPI:13

ICES. 2015b. First Interim Report of the Working Group on Working Group on Biologi-cal Parameters (WGBIOP), 7–11 September 2015, Malaga, Spain. ICES CM 2015/SSGIEOM:08. 67 pp.

ICES 2015c. Report of the Workshop on Methods for Estimating Discard Survival 3 (WKMEDS 3). ICES CM 2015 / ACOM:39.

ICES. 2015d. Report of the Working Group on Data Needs for Assessments and Advice (PGDATA), 30 June–3 July 2015, Lysekil, Sweden. ICES CM 2015/SSGIEOM:06.