Guidelines for the Quality Assessment of Estimates Derived from Catch Monitoring Programs in Canada

Jacques Allard and Hugues P. Benoît

Fisheries and Oceans Canada
Maurice Lamontagne Institute
Mont Joli, QC
G5H 3Z4

2021

Canadian Technical Report of Fisheries and Aquatic Sciences 3414

Fisheries and Oceans Canada Pêches et Océans Canada

Canada

Canadian Technical Report of Fisheries and Aquatic Sciences

Rapport technique canadien des sciences halieutiques et aquatiques

Canadian Technical Report of
Fisheries and Aquatic Sciences 3414

2021

GUIDELINES FOR THE QUALITY ASSESSMENT OF ESTIMATES DERIVED FROM CATCH MONITORING PROGRAMS IN CANADA

by

Jacques Allard[2] and Hugues P. Benoît[1]

[1]Fisheries and Oceans Canada
Maurice Lamontagne Institute
Mont Joli, QC
G5H 3Z4

[2]Atlantic Statistical Analysis Inc.
Dieppe, NB
E1A 4Z7

Cat. No. Fs97-6/3414E-PDF ISBN 978-0-660-36891-7 ISSN 1488-5379

Correct citation for this publication:

Allard, J. and Benoît, H.P. 2021. Guidelines for the quality assessment of estimates derived from catch monitoring programs in Canada. Can. Tech. Rep. Fish. Aquat. Sci. 3414: v + 53 p.

TABLE OF CONTENTS

ABSTRACT

Allard, J. and Benoît, H.P. 2020. Guidelines for the quality assessment of estimates derived from catch monitoring programs in Canada. Can. Tech. Rep. Fish. Aquat. Sci. 3414: v + 53 p.

The Department of Fisheries and Oceans Canada (DFO) adopted a National Fishery Monitoring policy in 2019. To implement that policy, DFO has created a framework for the Quality Assessment (QA) of estimates derived from fishery monitoring programs, the most common estimates being those of total catch. This framework defines a comprehensive list of statistical and operational characteristics of estimation processes that must be examined to assess the quality of estimates. This report offers guidance on applying the QA framework. We review the quality assessment framework and present practical suggestions on efficiently carrying out a QA. Suggestions are given on the elicitation of expert judgements, which are important in many assessments. We present considerations on examining the statistical and operational characteristics defined by the framework, beginning with considerations applicable to most monitoring programs followed by considerations specific to seven major types of monitoring programs used by DFO. Finally, we suggest a method to assess the impact of the absence of monitoring, a step which can be essential to the assessment of the quality of an estimate of total catch if an unmonitored component of the catch is important.

RÉSUMÉ

Allard, J. and Benoît, H.P. 2021. Guidelines for the quality assessment of estimates derived from catch monitoring programs in Canada. Can. Tech. Rep. Fish. Aquat. Sci. 3414: v + 53 p.

Le ministère des Pêches et Océans du Canada (MPO) a adopté en 2019 une politique nationale de surveillance des pêches. Pour mettre en œuvre cette politique, le MPO a créé un cadre pour l'évaluation de la qualité (ÉQ) des estimations dérivées des programmes de surveillance des pêches, les estimations les plus courantes étant celles des prises totales. Ce cadre définit une liste complète des caractéristiques statistiques et opérationnelles des processus d'estimation qui doivent être examinées pour évaluer la qualité des estimations. Ce rapport offre des conseils sur l'application du cadre d'assurance qualité. Nous passons en revue le cadre d'évaluation de la qualité et présentons des suggestions pratiques pour mener efficacement une ÉQ. Des suggestions sont données sur la sollicitation de jugements d'experts, ceux-ci étant importants dans de nombreuses évaluations. Nous présentons des réflexions sur l'examen des caractéristiques statistiques et opérationnelles définies par le cadre, en commençant par des réflexions applicables à la plupart des programmes de surveillance et en suivant par des réflexions propres à 7 grands types de programmes de surveillance utilisés par le MPO. Enfin, nous suggérons une méthode pour évaluer l'impact de l'absence de surveillance, étape qui peut être essentielle à l'évaluation de la qualité d'une estimation de la capture totale si une composante non surveillée de la capture est importante.

1 INTRODUCTION

Understanding and managing the impacts of fisheries on fish populations is essential to the conservation of fish populations and to the sustainability of the fisheries. For many fish populations, the mortality in the kept and discarded catches is the most important anthropogenic impact and generally the most accessible information. Fishery monitoring programs are used worldwide to estimate catch and other parameters related to the fisheries, including fishing effort. Fisheries and Oceans Canada (DFO) has developed a unified framework to assess the quality of estimates obtained from fishery monitoring programs it manages (DFO 2019a, 2020). The purpose of this report is to provide additional guidance to promote an efficient, informed, objective and consistent application of the quality assessment (QA) framework. We discuss overarching considerations that may apply to all types of monitoring programs, and considerations specific to the most common types of fishery monitoring programs. In particular, we review the manner in which the design and implementation of monitoring programs can affect quality of an estimation process.

DFO'S NATIONAL FISHERY MONITORING POLICY

Fisheries and Oceans Canada has recently adopted a national fishery monitoring policy to ensure that it has dependable, timely and accessible information to manage fisheries sustainably and to minimize harm to non-harvested incidentally captured taxa (DFO, 2019b). An important goal for the policy is to implement a fair and consistent approach for setting the type and degree of monitoring employed across fisheries managed nationwide by DFO under the *Fisheries Act*, taking into account the conservation risk level encountered by each fish population. In particular, removals by fisheries posing heightened conservation risk to one or more valued resources should be monitored in such a way as to provide high quality estimates. To reach this goal, DFO is using a two-prong approach. On the one hand, DFO has developed a framework to assess the risk posed by fisheries to the fished population and the corresponding level of quality required of estimates of catches and of other parameters relevant to the fish population (DFO, 2020). On the other hand, it has developed a framework to assess the quality of estimates obtained from existing fishery monitoring programs, which is the subject of the present document (DFO, 2019).

The adequacy of estimates to support the correct decision on management actions, or their fitness-for-purpose, is termed ***dependability***. The results from the risk assessment and from the quality assessment are used to establish dependability, based on the principles outlined in DFO (2020) and Benoît and Allard (2020).

FISHERY MONITORING PROGRAMS

Numerous approaches exist to collect the data required to produce estimates of catch and other parameters relevant to fishery management. Common catch monitoring approaches include fishery logbooks, dockside monitoring of landings, monitoring by third-party at-sea observers, creel surveys and questionnaires. Monitoring approaches have very diverse characteristics: (near) real-time reporting or periodic reporting, resource user self-reporting or independent monitoring, paper vs electronic recording, etc. Furthermore, monitoring in Canada and worldwide occurs in a large diversity of fisheries ranging from small recreational or subsistence fisheries to large industrial fisheries (Beauchamp et al., 2019).

Ensuring that adequate data are available to support decision making for the management of fisheries in light of this diversity is a major challenge.

THE QUALITY ASSESSMENT FRAMEWORK

In estimating total catch of a fish population, quality is defined as a measure of the validity of an estimate, or how close to the true value the estimate is likely to be, e.g., how close the estimate of total catch of a stock is to the true total. The quality of an estimate depends on the variability and bias of an estimation resulting from sampling randomness (for sampling surveys) and from the implementation of the monitoring and estimation protocol (Babcock et al., 2003; ICES, 2013), a combination referred to as a ***parameter estimation process*** or, simply, an ***estimation process***. DFO Science Branch has developed a ***Quality Assessment*** (QA) ***Framework*** to assess the quality of an estimation process, including a generalized and unified approach to quantifying it (Allard and Benoît, 2019; DFO 2019b, DFO 2020).

The approach has been implemented in a Microsoft Excel – Visual Basic application commonly referred to as the quality assessment tool (QAT), which is applicable to many situations. The QAT records the necessary information and computes the two main indicators of quality, bias, and variability.

REPORT SCOPE

While the risk screening is carried out at the fish population level, the quality assessment of the estimation process must be first carried out at the monitoring program level, since each monitoring program will have different characteristics. A fish population can be impacted by a single fishery covered by a single monitoring program, in which case a single quality assessment will be required to assess, for example, the estimation of total catches from that fish population. However, more complex cases frequently occur.

This report is focused on estimates obtained in one of the following ways: an estimate obtained from a single monitoring program (e.g., a snow crab fishery); an estimate that is the sum of estimates obtained from two or more monitoring programs (e.g., the catch obtained from a dock-side monitoring program of a trawl fishery and that obtained from a logbook monitoring program of a gillnet fishery); and, an estimate that is obtained by multiplying estimates obtained from two monitoring programs (e.g., the effort obtained from an aerial survey and the catch per unit of effort obtained from a creel survey).

We present considerations that apply specifically to the QA of the estimation processes based on data from the following seven common types of monitoring programs:

- Mandatory resource user reports with an auditing mechanism allowing for 100% verification. These include independent monitoring that allows for an audit of the fisher reports, including remote sensing vessel monitoring systems (VMS) and 100% video monitoring (§6).
- Mandatory resource user reports without an auditing mechanism allowing for 100% verification. These include catch or effort logbook reports, verbal hails, and others (§7).
- Mandatory dock-side monitoring (100% or <100% coverage) (§8).
- Mandatory purchase slip monitoring (§9).
- At-sea observer programs (§10).
- Creel surveys (§11).

- Post-season questionnaires (§12).

Additionally, we treat the case where there is an absence of monitoring (§13). In some situations, absence of monitoring may be inconsequential to the overall dependability of a catch estimation process, notably if the anticipated catches are relatively small, nonetheless an assessment of the contribution to quality may be necessary to draw this conclusion and is important for initiating a discussion of the appropriateness of the absence of monitoring.

In sampling surveys, assessing departures from the sampling protocol is an important aspect of a QA. The present report focuses mostly on surveys based on simple random sampling or systematic sampling, except for monitoring by at-sea observer where cluster sampling is often employed and is partially addressed.

The present document may be helpful for many estimation processes not included in the above list, especially for the assessment of characteristics unrelated to the sampling protocol.

Some estimates are obtained by a combination of monitoring programs and complex mathematical operations. For example, the estimate of the total catch of Chinook salmon from the Pacific Coast recreational fishery combines information from several monitoring programs including dockside and aerial surveys to estimate effort and creel surveys to estimate catch per unit of effort. Complex estimation processes are outside the scope of the present report.

QA CONCEPTS VS COMPUTATIONAL TOOL INPUTS

Any tools used to apply the quality assessment framework, including the QAT mentioned above, will require inputs to compute the assessment. Differences between the presentation of the concepts in this guidance document and the inputs required by a tool may be subtle. Tool users should carefully follow tool's user guide to avoid errors.

CONTENTS

Following this introduction, Section 2 reviews the concept of quality assessment and the quality assessment process as developed in Allard and Benoît (2019; consult that document for the details). Section 3 presents general guidance on how to approach a quality assessment. Section 4 presents suggestions on the elicitation of expert judgements, which are a key input to the QA process. Section 5 presents general considerations concerning the statistical and operational characteristics that must be examined to complete a quality assessment. Sections 6 to 12 present considerations specific to each type of catch monitoring covered by this report. The last section is a brief discussion.

The reader encountering the QA process for the first time should read Sections 2, 3 and 4. When preparing for a specific QA, we recommend reading the section specific to the type of monitoring before looking at the appropriate subsections of §5.

2 OVERVIEW OF QUALITY ASSESSMENT

2.1 The object of a Quality Assessment

In this report, the term fish population is used to designate the group of individuals from the same (sub-) species of fish or other aquatic life form to which a collective management plan is in place or could be implemented. The definition therefore includes the concept of stock and of designable unit in the context of the Committee on the Status of Endangered Wildlife in Canada.

Most QAs concern the estimation of one of the following parameters: the landed catch, the retained but not landed catch (e.g., fish used as bait), the discarded catch, the total of these catches, the total catch of a certain fish population component (e.g., undersized fish), or the total effort. Other types of parameters estimated include a mean, a proportion (e.g., proportion of soft-shell crab in the snow crab catch), a ratio (e.g., the ratio of catch of species Y to catch of species X, or the catch per unit of effort). This report focuses on these types of parameters with an emphasis on the first group.

The type of the parameter has an impact on the quality assessment. For example, the sampling ratio (sampling size/population size) is required when estimating total catch from a sampling survey but not when estimating a catch ratio.

The quality assessment is concerned with the typical application of the estimation process, as opposed to, say, its last annual application. Therefore, it entails consideration of typical population sizes, sample sizes, and systematic departures from the protocol, i.e., departures recurring year after year. Isolated occurrences (e.g., missing data due to a rare labour dispute or a rare database loss) should not be considered in the assessment.

The quality assessment methodology can also be applied to one-off situations including pilot projects.

2.2 Measures of quality of an estimation process

Estimation processes based on catch monitoring have one or both of the following objectives: estimation and compliance applications.

In estimation applications, the objective is to obtain an estimate of a parameter (e.g., total catch) for management or research purposes (e.g., stock assessment). In these applications, the bias (opposite of accuracy) and the variability (opposite of precision) of the estimation process are the measures of the (lack of) quality. Bias describes a tendency for the estimate to be below (negative bias) or above (positive bias) the true value. A bias close to zero is desirable. If conservation is the policy goal, only a negative bias is relevant to, say, an estimation of total catch (a positive bias can only lead to further fishing restrictions). Variability is positive (similar to the standard deviation) and a value closer to zero is desirable. Variability describes how estimates would vary if the estimation process was carried out simultaneously several times. To obtain standardized measures, the bias and variability are expressed as a proportion (%) of the anticipated true value of the parameter (Allard and Benoît, 2019).

In compliance applications, the objective is to determine whether a regulatory limit or cap has been respected. These limits are almost exclusively upper limits on catch, a proportion (e.g., percentage of undersized fish) or effort. In these applications, the quality of the estimation process is defined as the probability of avoiding a decision detrimental to conservation i.e., typically, the probability of declaring that a cap has been breached when it is the case. This probability corresponds to the concept of “sensitivity” used in describing the quality of medical tests and elsewhere. It is a value between 0 and 1 and values closer to 1 are desirable. Under the QA framework, this probability is computed under the assumption that the estimate is a normal random variable with the expected value and the standard deviation derived from the anticipated true value, the bias and the variability determined by the QA.

Regulatory limits are most often fixed control points based on general considerations, such as bycatch limits which can be set at somewhat arbitrary low levels and quotas set relative to past catches believed to be sustainable. In some cases, limits to ensure sustainability are based on modelling or analysis and can be expressed with uncertainty. For example, the uncertainty and possibly bias, of a conservation limit can be quantified within a single model (e.g., using parameter uncertainty), with respect to several competing plausible models (e.g., by examining the range of regulatory limits obtained) or in closed-loop simulations. In these cases, uncertainty and bias on the limit can be included as part of the quality assessment (details in Benoît and Allard, 2020).

2.3 Structure of a quality assessment

The quality of the estimation depends on the design and implementation of the monitoring program(s) that is(are) used to carry out the estimation.

In many cases, the estimation is obtained from a single monitoring program. For example, total snow crab landings in a snow crab fishing area may be obtained by a dock-side monitoring program. In some cases, components of a single fishery are covered by different monitoring programs. For example, the 4RS3Pn cod fishery is partitioned into a commercial fishery for which landed catch in three separate sectors is monitored by a different program (each implemented as a census for that sector; dockside monitoring, harvester selfreporting via a hail, and purchase slips) and one recreational fishery that is not subjected to structured monitoring. Similarly, several fisheries may contribute to the total catch of a single fish stock. In all these cases, a quality assessment should be completed for each component (excluding perhaps those contributing trivially to catch) and the result of these assessments combined to obtain an assessment of the total, as described in section §2.5.

In other cases, an estimate can be the product of two estimates, for example independent catch-per-unit-effort and total effort estimates. In these cases, a quality assessment is completed for each component and the result of these assessments combined to obtain an assessment of the product, as described in section §2.6.

In some fisheries, multiple monitoring programs record data pertinent to the parameter of interest in parallel (e.g., at-sea observers and dockside monitoring). However, in most of these cases, data from a single monitoring program are used for official reporting and decision making. The estimates obtained from this monitoring program should be the focus of the quality assessment. Data from the other programs may be used to validate data from the main program, to impute missing values or to correct erroneous values. These other programs can contribute to improving the quality of the estimation process, a role that should be documented in the quality assessment.

2.4 Assessment of an estimation based on a single monitoring program

This section applies to the quality assessment of estimation based on a single monitoring program.

2.4.1 Statistical characteristics

Each monitoring program is based on a specific protocol. The protocol specifies an estimator, i.e., a mathematical formula applied to the observations. The most common estimators in fishery management are based on means or totals. For example, in the case of a monitoring program by at-sea observers that randomly samples trips, the estimator of total discards of a species may be the mean of the discards for observed trips multiplied by the total number of trips. The statistical characteristics of an estimator are its statistical bias and its standard deviation. They contribute respectively to the estimation process bias and variability. The statistical bias and the standard deviation of an estimator are computed using theoretical or numerical (e.g., bootstrapping) statistical methods. They are usually reported along with estimates.

STATISTICAL BIAS

For a sampling survey, the characteristics of the sampling protocol and the mathematical properties of the estimator can create bias. The usual estimators of the mean and of the total, under simple random sampling, are not statistically biased while some other estimators are statistically biased (e.g., ratio estimators). It is sometimes possible to correct for estimator bias using statistical methods. In the QA, the statistical bias should be reported and assessed unless a reliable correction has been applied, in which case it should be reported as 0. Allard and Benoît (2019) provide examples of statistical biases. For a census, the statistical bias is 0.

STANDARD ERROR

For a sampling survey, the statistical standard error (SE) of an estimate measures the variability of the estimates due to the randomness included in the sampling protocol. The computation of the standard error depends on the sampling protocol presumed to have been used and the estimator. For example, computations required for simple random sampling and for stratified sampling are different. In nearly all cases, the SE can be estimated by bootstrapping. For all estimation processes based on sampling surveys, the SE is required for the assessment of the variability. Although a properly estimated SE should be employed for the final quality assessment, interim proxy values can be used temporarily in a preliminary assessment. For example, for an estimation of a mean based on random sampling, the interim estimate of the standard error can be a small multiple (e.g., 2 × or 3 ×) of the 0.25 × [largest observation – smallest observation]/[sample size]^{½}. Further details on estimating the SE or its proxy are available in Allard and Benoît (2019). For a census, the standard error is 0.

2.4.2 Operational characteristics

In practice, a protocol is rarely implemented perfectly. The QA process divides areas of concern with the implementation of a protocol into 15 classes, referred to as “operational characteristics” or “OCs”. An important goal of the QA is to assess each operational characteristic for departures from the protocol and other errors, and to quantify the contribution of these departures and errors to the bias and variability of the estimation process.

The 15 operational characteristics (OC01 to OC15) can be grouped in the following classes: errors in the determination of the statistical population (under- or over-coverage), missing values; various data errors including measuring errors, adjustment, imputation, and modelling errors; and, for survey sampling only, departures from the sampling protocol and observer effects. Section 5 provide details on the OCs.

The impact of an operational characteristic on the quality of the estimate depends on the proportion of the population or of the observations to which the characteristic applies, which we call the ***extent*** of the impact. For example, data handling errors are likely to apply only to a small proportion of the observations.

Assessing the operational characteristics depends primarily on expert judgements. To be justifiable and relevant, expert judgements must be obtained and reported in a rigorous and consistent way (see §4).

2.4.3 Computation of the measures of quality

The bias and variability of the estimation process are obtained by applying standard probability bias and variance formulae to the contribution of the statistical and operational characteristics to the bias and variability.

For bias, the contributions of the operational characteristics are added up arithmetically, with the following exceptions. For OC01-Undercoverage and OC02-Overcoverage, when estimating a total, the excluded units and incorrectly included units further contribute to the bias through the correction that must be applied to the population size.

For variability, the contribution of OCs related to data errors (OC07 to OC15) are added quadratically according to the formula for the variance of a sum of independent random variables. For sampling surveys, the contribution of OCs related to departures from the sampling protocol (OC03 to OC05) and to observer effects (OC06) are applied multiplicatively since changes in sampling protocol typically impact the variance of an estimate multiplicatively (e.g., comparison of the variance of an estimate based on stratified sampling vs that of an estimate based on simple random sampling).

When relevant, the contributions are weighted by the extent of occurrence. For example, the contribution to bias of logbook underreporting of a bycatch will be less if it is assessed to occur in 20% of the logbook reports than in 80% of the logbook reports.

2.5 Assessment of estimation obtained by an addition of estimates

Some estimates are obtained as the total of several estimates. This is the case, for example, when a fish population is intercepted by several fisheries or when components of a fishery are monitored by different programs. In such cases, a QA must be performed for each estimate and the results of these QAs must be combined to assess the quality of the estimate of the total.

The measures of quality of a total are the same as those of an estimate obtained from a single monitoring program.

Computationally, the bias and the variability of the estimate of the total are obtained by applying the mathematical formulae for the bias and the standard error of the sum of random variables, weighted by the contribution of each component to the total. See Allard and Benoît (2019) and DFO (2019) for further details.

2.6 Assessment of estimates obtained by multiplication of two estimates

Some estimates are obtained as the product of two estimates. A common example of such estimates is the estimation of total catch as the product of estimates of the catch per unit effort (CPUE) and of total effort. This approach is common in recreational fisheries, where the CPUE may be obtained from creel surveys or harvester questionnaires, and effort may be obtained from overflights or monitoring of access points. A second example is estimation of discards as the product of estimates of a discard ratio (kg discarded/kg landed catch) and of the total landed catch. The discard ratio may be obtained by an at-sea observer program and the total landed catch from a purchase slip return program. In these cases, the quality of each estimation process, as well as the quality of the estimation of the product, must be assessed.

The assessment of the quality of an estimate that is a product of two estimates depends on the assessments of these two estimates in a way that can amplify errors in these assessments. Therefore, the QA of the two input estimation processes should be carried out with extra care.

The measures of quality of a product are the same as those of an estimate obtained from a single monitoring program. Computationally, the bias and variability of a product of estimates are obtained from the bias and variability of each estimate according to the formulae for the product of independent random variables.

3 GUIDANCE FOR AN EFFICIENT QUALITY ASSESSMENT PROCESS

The QA framework promotes a rigorous (semi-) quantitative evaluation of quality. The framework also attempts to cover all issues potentially impacting the extremely diverse estimation processes used in Canadian fishery science and management.

In some circumstances a highly detailed review involving inputs and calculations of high precision could be required. This is possible if there is conservation concern and resource users insist that less rigorous (and less expensive) monitoring is adequate. Such a level of scrutiny may not be required in all cases and may not be sustainable given the policy requirement to assess the quality of estimation processes for a large number of fish populations subjected to many fisheries and monitoring programs. Below we briefly review elements that could contribute to an efficient quality assessment process.

FOCUS ON THE MONITORING PROGRAMS SUPPLYING THE DATA USED FOR RESEARCH AND DECISION MAKING

In many commercial fisheries, several monitoring programs may provide information on a given parameter. The QA should focus on the primary monitoring programs, i.e., those are used for decision making, research or reporting. The QA should recognize the existence of the secondary program(s) and report their contribution to the quality of the estimation process. If the secondary program is thought to provide a higher quality estimation, it could be assessed and a comparison the quality of the two programs could be reported.

RECOGNIZE THE STRUCTURE OF THE ESTIMATION PROCESS AND ALLOCATE THE ASSESSMENT EFFORT ACCORDINGLY

From a conservation point of view, assessing the quality of the estimation of the total catch is the overarching goal of the QA. When a fish population is intercepted by only one fishery that lands its entire catch, that goal is often reached by a single QA involving a single monitoring program. However, some fisheries are subject to several monitoring programs and some fish populations are intercepted by multiple fisheries. In these cases, several QAs must be carried out and their results combined in a QA of the overall estimation process.

A single fishery may be covered by multiple monitoring programs monitoring different components of catch (e.g., landed catch and discarded catch). Commercial fisheries may involve both 100% dockside monitoring of landed catch by independent observers and mandatory logbook self-reporting by fishers of their landed and discarded catch. The estimate of the total catch will be the total of the estimate of the landed catch from the dockside monitoring and the estimate of the discarded catch from the logbooks. In other fisheries, the landed catch may be monitored dockside in major ports and by self-reporting in remote ports.

For such fisheries, we suggest that the assessment effort be proportional to the contribution of the components to total catch. If the objective is to estimate total catch, and a fishery lands 10,000 t of a fish population while discarding 10 t, the QA should focus on the estimation of the landings while the QA of the estimation of the discards could be a cursory QA or place holder QA. A similar evaluation is recommended for a fish population intercepted by several fisheries. For example, if a fish population is targeted by a 10,000 t commercial fishery and a subsistence fishery roughly estimated at 100 t and subject to discarding by a third fishery roughly estimated at 10 t, the QA of the estimate of the subsistence fishery would be cursory and that of the discards limited to a place holder QA.

By a *cursory QA*, we mean assuming a ‘best estimate’ or a ‘worse-case’ values for the bias and variability of the estimation process. By a *place-holder QA*, we mean stating a relative bias of 0% and a relative variability of, say, 50% of the estimated catch (reflecting an unbiased but very imprecise expert judgement). Including place-holder QAs for small fisheries, whether or not they are monitored, ensures a transparent and thorough QA of the estimation process, while reducing the burden of undertaking a detailed QA for components that contribute to only a small portion of removals from the population. For a large unmonitored fishery, see §13 below.

These suggestions do not apply to a fish population that is intercepted by numerous equally small fisheries. Unless there is good a priori evidence that fishing mortality has negligible population-level impacts, full QAs may be required to avoid the negative consequences of low-quality monitoring leading to a “death by a thousand cuts”.

ILLEGAL, UNREPORTED, AND UNREGULATED (IUU) FISHERIES AND OTHER UNMONITORED FISHERIES

IUU fisheries are, by their nature, not covered by catch monitoring programs. Their contribution to errors in catches can generally not be reduced by improving catch monitoring, only by improving enforcement actions or increasing incentives against IUU. For this reason, IUU fisheries have been excluded from the program of quality assessment of estimation processes since its inception (DFO, 2019). This is different from unmonitored regulated fisheries, where catch monitoring could be put in place, and where accounting for the absence of monitoring is therefore important for QA.

Section 13 suggests an approach to quantifying the impact of the absence of a monitoring program of a fishery. While it is intended for regulated unmonitored fisheries, the approach could be employed to quantify the impacts of IUU on quality if this were of interest for purposes external to the design and implementation of catch monitoring.

FOCUS EFFORTS ON THE OCS CONTRIBUTING MOST TO BIAS AND/OR VARIABILITY

The Quality Assessment methodology provides a structured framework to review the potential contributors (the OCs) to the bias and variability of the estimation process in order to ensure that analysts undertake as complete an analysis as possible. This report highlights, for each type of monitoring program considered, the OCs that are not applicable, unlikely to be pertinent, or that require special scrutiny. The analysts should further limit their effort when analysing OCs that have only a minor impact.

FOR SAMPLING SURVEYS, VERIFY THE APPROPRIATENESS OF THE STATISTICAL METHODOLOGY

For survey sampling, the estimator should use appropriate formulae and statistical bias and the standard error should be computed according to the sampling protocol. It is important to verify that this is the case, even in simple situations (e.g., when estimating total catch using data from a sampling survey, verify that the scaling to the population is applied correctly). If methodological errors are found, these should be corrected going forward. In the uncommon situation where the computation is done according to methods agreed up by negotiation and as such rectifying the calculations is not feasible in the short term, the consequences of these errors should be described and expressed by amending the statistical bias and the standard error reported by the QA.

ONLY RECURRING OR EMERGING ISSUES SHOULD BE CONSIDERED IN THE QA

Only systematic and recurring, or emerging issues should be included in the analysis. The impact of a hurricane on an at-sea observer program is an example of a non-recurring event. Recurring discrepancies between logbook data from observed and non-observed trips are a strong indication of systematic (and possibly intentional) errors in logbooks. An increase in observer data errors following a reduction in observer training is an example of an emerging issue that should be reported.

THE DESCRIPTION AND DOCUMENTATION OF ISSUES IS MORE IMPORTANT THAN THE QUANTIFICATION OF THEIR IMPACT

The QA should include a clear description of issues identified and present the information that supports their identification. Quantification of the impact is required to obtain standardized measures of quality. However, this quantification can be coarse. For example, stating that the proportion of the observations impacted by an issue is 20% vs 30% should be inconsequential to the post-assessment decision making.

CONTRIBUTIONS TO BIAS ARE MORE IMPORTANT THAN CONTRIBUTIONS TO VARIABILITY

From a conservation point of view, the contribution of OCs to the bias of the estimation process are more important than contributions to its variability. A biased estimation process may lead to *long-term* incorrect decision-making detrimental either to conservation or to the fishery. A high variability estimation process will more likely lead to incorrect recurring (e.g., annual) decisions in the *short term*, alternating between decisions detrimental and favorable to either conservation or the fishery.

Important issues should get the most attention

Some specific issues affecting quality are likely to be important in many fisheries. These include, for example, logbook errors, observer effects and observer errors. The analysts should focus their effort on identifying, describing, and documenting these issues and, especially, their contribution to the bias of the estimation process. For example, issues impacting 1% of the observations are unlikely to meaningfully impact the overall quality assessment. Similarly, if three experts assess the contribution of an OC to bias to be around 3%, 5%, and 9%, respectively, obtaining a consensus value is not necessary. If the three expert assessments are 25%, 50% and 90%, obtaining a consensus value is desirable.

Minor issues should be acknowledged

Issues likely to occur systematically but rarely, or to have a minor impact, should be acknowledged but analysed only briefly. Acknowledging minor issues helps ensure transparency and completeness of the QA process. For example, in programs involving a single transcription of data from a paper record to a computerized database, data handling errors (OC12) will be recurring, but relatively rare given quality assurance and control procedures, and will not contribute to bias since the errors will be random. These errors should be acknowledged by inputting, say, an extent of occurrence of 0.1% (1 error per 1000 transcriptions), a bias of 0% and a variability of 50%. The suggested variability is high but corresponds to a transcription interchanging two digits, i.e., inputting 817 instead of 187. Due to the rare occurrence, the impact of this issue on the overall assessment will still be negligible.

Common issues and best practices

Some contributions to the estimation process bias and variability are likely to be common to many monitoring programs or classes of monitoring programs. These include errors in data reported by independent observers, and errors resulting from equipment, data handling, processing errors, and adjustments. As experience undertaking QAs is gained nationally, common values or ranges of values for these errors will be identified and using them will both streamline the QA process and promote consistency. This is likewise true for best practices related to the identification, description, and documentation of common issues.

4 ELICITATION OF EXPERT JUDGEMENTS

Some information necessary to carry out the Quality Assessment of a parameter estimation process is typically not available in the form of “hard data”. Therefore, the QA must rely in part on “soft data” and, especially, the judgement of “experts”, where an expert is a person who has relevant knowledge about a subject matter. In this section, we discuss how to obtain and report expert judgements.

In a typical QA, few Operational Characteristics will be relevant and, among those, even fewer will require a detailed assessment. For example, in some estimation process based on sampling survey, observer effects may be the largest source of bias and warrant a detailed assessment while in most estimation processes, measurement scale error will be a negligible source of bias or variability. The procedure described in this section should be applied only for OCs requiring a detailed assessment.

4.1 The elicitation of expert judgements

Expert judgements play an important role in fishery management. For example, the set of models considered to estimate a reference value is rarely, if ever, based on hard data. Similarly, the criteria used to select of the "best" model often include expert judgement, implicitly or explicitly.

Expert judgements are necessary in the quality assessment of estimation process due to limited data availability. For example, the accuracy and precision of visual estimations of catch weights by fishers or at-sea observers may not have been studied for all fisheries; underreporting of some bycatch in logbooks may be known to exist from comparisons between trips with and without at-sea observers but may not have been quantified; the impact of the presence of an at-sea observer on fishers' choice of fishing grounds and, indirectly, on bycatch can be very difficult to estimate based on data alone.

Expert judgements are also required when information from a similar situation is used to guide the assessment. For example, results from an American research program on compliance in the Chinook fishery may be used in the QA of a Canadian estimate of catch, with an expert assessing the dissimilarity and adjusting the results accordingly.

Outside of fishery management, the ubiquitous importance of expert judgements is easy to recognize. Whether in the selection of cancer treatments or in political decisions on national economic policies, expert judgements must at least complement the often-limited hard data available. This recognition has led to the development of formal scientific methodologies to use expert judgements. The title of Dias et al. (2018) gives an accurate description of this field: "Elicitation –The Science and Art of Structuring Judgement". While applying formal elicitation methods is far beyond the needs of QAs, most suggestions presented in this section are inspired by elements of the theory of elicitation.

4.2 Obtaining expert judgements

EXPERTS RELEVANT TO QAS

Depending on the characteristics, relevant experts will include current and recent monitoring program managers, conservation and protection agents, fishery scientists, people familiar with the sampling protocol, observers, and resource users.

NUMBER OF EXPERTS

In a famous experiment, Galton (1907) showed that the median of the judgements of many people could provide a very good estimate of a true value (within ±1% in the example given, with 787 judgements). This observation has developed into a large body of scientific work about the “Wisdom of Crowds”.

The practical consequence of this observation is that more experts is better than fewer. Consequently, when carrying out a QA, it is preferable to obtain the independent judgement of several experts on items that are important to the assessment. For example, in estimating bias in logbook reporting, the analyst could obtain the judgement of the current and of the previous two managers of the fishery, or of the fishery manager and of two conservation and protection officers. In general, the experts should be as independent from one another as possible (e.g., the judgements of fishery officers working independently is preferable to that of officers usually working in the same team).

WHY AND HOW TO OBTAIN QUANTIFIED JUDGEMENTS

The statement “The bycatch of cod by the lobster fishery is not well estimated” may be useful to highlight the need for research on cod bycatch by the lobster fishery but is not usable in the quality assessment of the estimation of the total weight of cod caught by all fisheries in NAFO division 4T. To be useful, a quantitative statement is required, e.g., “While cod bycatch by the lobster fishery is not currently accounted for, it is estimated to be between 10 t and 25 t with around 20 t being most likely”. Since the currently used estimate is 0 t (“not currently accounted for”), the contribution to the bias of the estimation process for total cod catch can be assessed as –20 t and to its variability as 3.75 t [(25 – 10)/4].

In general, each expert should be asked to contribute his/her estimate of the lowest plausible value, his/her estimate of the highest plausible value, and his/her best estimate. Obtaining and reporting the expert’s lowest and highest plausible values allows the QA to describe the uncertainty inherent to the expert judgement.

A typical question might be as follows.

Considering logbooks in fishery X, what is your estimate of the bias in reporting of the weight of cod discards as a percentage of the true weight? Underreporting should be expressed as a negative value. Realistically, what do you think is the LOWEST PLAUSIBLE bias? _________% Realistically, what do you think is the HIGHEST PLAUSIBLE bias? _________% What is your best estimate of the bias? _________%

Elicitation literature describes the three estimates as follows. The analyst would like to obtain 5th, 95th and 50th quantiles of the expert’s subjective probability distribution of his estimates. One can think of the estimates requested as follows, in terms of betting odds. The expert should be ready to:

- Make a bet with 1 to 20 odds that the true value is higher than his/her lowest plausible estimate: for a $1 bet, if the true value is above his lowest plausible estimate, he/she makes 0.05$ (plus his/her $1 bet); otherwise, he/she loses the bet.

- Make a bet with 20 to 1 odds that the true value is higher than his/her highest plausible estimate: if the true value is above his highest plausible estimate, he/she gets $20 (plus his/her $1 bet); otherwise, he/she loses the bet.
- Make an ordinary bet with 1 to 1 odds that the true value is above his/her best estimate: if the true value is above his best judgement estimate, he/she makes $1 (plus his/her $1 bet); otherwise, he/she loses the bet.

Technical note: The 5th and 95th quantiles of the of the expert's subjective probability distribution are like the limits of a *90%* confidence interval. If the estimator has a normal distribution, the standard error of the estimate would be computed from the confidence interval formula as [95th quantiles - 5th quantiles]/3.28. In this document, we have supposed that the lowest and highest plausible values are like the limits of the *95%* confidence interval and we have therefore assessed the standard error of the estimate as [97.5th quantiles – 2.5th quantiles]/4. In the context of elicitation of expert opinions, we consider the difference to be immaterial.

COMBINING EXPERT JUDGEMENTS

Once the judgement of each expert has been obtained, a discussion between the experts may be held to obtain a consensus assessment with corresponding lowest and highest plausible values. If obtaining a consensus assessment is not feasible because holding a discussion between the experts is not feasible or because the experts cannot come to a consensus, for simplicity, the median of the expert's best estimates could be used in the QA.

REPORTING EXPERT JUDGEMENTS

Ideally, and to ensure transparency, the 3 values provided by each expert should be reported (e.g., for logbook bias, lowest: -30%, best: -22%, highest: -15%) together with the rationale given by the expert (e.g., "Data from unannounced vessel boarding during the last 3 years indicate logbook underreporting averaging 22%; experts lowest and highest plausible values took into account the small number of boarding and informal conversations with fishers"). If a consensus assessment has been obtained and it is quite different from the initial expert estimates, the discussion leading to the consensus should be summarized.

4.3 Appropriate balance

Fully implemented, the formal methods of elicitation include a calibration of the experts using seed questions for which the true value is known to the analyst but not to the expert or will be known soon. The result of the calibration is used to weight the judgements of the experts when they are combined. Other methodologies have been developed to improve the performance of elicitation of expert judgements and the measurement if the inherent uncertainty. Hemming et al. (2018) offers an easily accessible and instructive case study in elicitation in the context of natural resource management. We consider that the proposed minimal elicitation procedure is sufficient for the objectives of the QA and we do not promote the full implementation of elicitation methods in QAs. However, we suggest that minimal awareness of these methods is useful.

5 CONSIDERATIONS COMMON TO MOST CATCH MONITORING PROGRAMS

General considerations associated with the contributions of operational characteristics to bias and variability of estimation processes are described below. In several cases, these considerations differ between censuses (100% monitoring) and sampling surveys (<100% monitoring).

5.1 Statistical characteristics (general considerations)

CENSUS VS SAMPLING SURVEY

In most cases, monitoring programs are either based on a census, where the protocol demands that the whole statistical population be observed, or a sampling survey, where a subset (a sample) of the statistical population is observed. Examples of statistical populations include the set of all fishing trips in a fishery, the set of all recreational fishers during a season, and the set of all possible daily effort counts in a fishery. Logbook monitoring programs are most often censuses. At-sea observer programs can be censuses (100% coverage of trips) or sampling surveys (<100% coverage of trips). The distinction between censuses and sampling surveys is an important consideration for quality assessment as will be evident throughout this document.

POPULATION SIZE AND SAMPLE SIZE

The following terms and concepts are used:

In this report, hereafter, “population” refers to a statistical population while “fish population” refers to *a group of individuals of the same (sub-)species* as defined above.

The population size is the typical number of units in the sampling/census frame, i.e., the number of units that are known to the monitoring program and from which the sample is drawn or that are covered by the census.

For a census, knowing the typical population size accurately is only consequential for small population sizes (e.g., < 50); otherwise, a coarse estimate is sufficient.

For a sampling survey, the sample size is the number of units that will be typically observed. Ideally, it will be the number of units mandated by the sampling protocol, but this may not always be the case due to operational limitations. The sampling ratio is the quotient of the sample size and the population size. For a sampling survey, estimating a total (e.g., total catch) requires knowing at least two of the following: the typical population size, the typical sample size, and the typical sampling ratio.

PARAMETER TYPE

The parameter types considered here are total, mean, ratio and proportion. When computing the bias and variability of the parameter estimation process, the last three types are treated similarly.

TYPICAL OBSERVED PARAMETER VALUE

The typical observed parameter value can be taken as the median for the values observed in the last several years if the value has been stable, or some recent observed value if the value is increasing or decreasing rapidly. In some cases, the parameter may show two or more distinct modes (e.g., the estimate of the catch of age 1 recruits in a fish population that produces periodic high recruitment years separated by several years of low recruitment). In such cases, the QA may have to be carried out for each mode.

STATISTICAL BIAS

In most sampling surveys considered in this report, the statistical bias is 0. Note that statistical bias is possible when an estimation process involves a multiplication of a ratiobased estimator, however this bias is captured as a modelling error (see section 5.17).

STATISTICAL VARIABILITY: THE STANDARD ERROR

In sampling surveys, the typical standard error of the estimator is required to complete the assessment. In some cases, it will not have been computed, a fact that should be clearly noted in the assessment. If the sampling protocol is similar to simple random sampling, we suggest that the following approximations of the standard error be use as interim values:

For estimation of a total:

$$[population\ size] \times \frac{[largest\ typically\ observed\ value] - [smallest\ typically\ observed\ value]}{4\sqrt{[sample\ size]}}$$

For other parameter types:

$$\frac{[largest\ typically\ observed\ value] - [smallest\ typically\ observed\ value]}{4\sqrt{[sample\ size]}}$$

5.2 OC01 Undercoverage (general considerations)

OC01 and OC02 concern the determination of the statistical population.

For the purposes of catch monitoring, the primary sampling unit is typically a fishing trip, a fishing haul, or a licence holder (e.g., for a post-season survey). The target population is the set of units for which estimates are desired, usually according to a time period (e.g., a fishing season) and a geographical area (e.g., a designated fishing area). The frame is the list of all units that are available to be observed by a monitoring program. By frame coverage, we mean the relationship between the frame and the target population. In a census, all units in the frame are observed. In a sample survey, the sample is drawn from the frame. Ideally, and in many monitoring instances, the frame and the target population are identical.

Undercoverage occurs when a subset of the target population is not included in the frame. Units in this subset are not known to exist when the monitoring takes place. By contrast, missing values occur when data from a unit known to exist are not available (see OC07 and OC08 below).

Examples of undercoverage include an incomplete list of recreational fishers targeted for a post-season interview, vessels neglecting to hail-out prior to departure, causing trips to be excluded from an at-sea observer program and vessels neglecting to hail-in prior to arrival, causing trips to be excluded from a dockside observer program. In such a case, the existence and extent of undercoverage may be assessed from the results of aerial surveys, for example.

Illegal, Unreported, and Unregulated (IUU) fisheries are not part of undercoverage as defined here since they are by definition not subjected to a monitoring program. However, unreported events (e.g., when a mandatory hail is not sent out) in a monitored fishery are part of undercoverage as defined here.

Undercoverage is not expected when:

- The resource users are subject to a licencing process, their number is relatively small, and the reporting is mandated by the licence or a condition of licence renewal;
- There is independent monitoring of fishing activities such that the population can be defined independently of the monitoring program, such as by mandatory vessel tracking (VMS) or 100% video monitoring; or,
- There is independent surveying of fishing trips for follow-up enforcement of mandatory reporting.

Contribution to bias: For estimation of a mean, a proportion or a ratio, undercoverage will contribute to bias if there is a difference between the excluded units and the population (e.g., recreational fishers not hailing in when their catch exceeds the daily limit):

- Excluded subset associated with large values: the undercoverage contributes negatively to the bias of the estimation process;
- Excluded subset associated with small values: the undercoverage contributes positively to the bias of the estimation process;
- Excluded subset associated equally with small and large values, or with mean or middle values: no (further) contribution to bias.

Undercoverage will further contribute negatively to the bias for the estimation of a total (e.g., total catch): since a part of the fishing effort will not be accounted for. In a census, values for the units excluded from the frame will not be added to the total; in a sampling survey, they will not be accounted for in the computation, which, in many cases, requires multiplying the observed mean by the quotient [frame size]/[sample size].

Identifying the likely source or motivation for anticipated undercoverage is key to evaluating the likely contribution to bias. If the exclusion of units is random with respect to the quantity being measured, the contribution to bias should be zero. Units associated with small values might be excluded because the catches are deemed by resource users too small to hail for follow-up monitoring or are ignored by creel surveyors that focus their sampling efforts only on productive fishing grounds. Such exclusions would be inconsequential for an estimation of total catch but would contribute positively to bias in the estimation of a CPUE. On the other hand, the exclusion of units associated with large values appears more likely to result from deliberate self-exclusion on the part of resource users, such as vessels prone to illegal discarding avoiding timely pre-departure hail-outs, causing trips with large illegal discard to be excluded from an at-sea observer program.

Contribution to variability: No contribution.

5.3 OC02 Overcoverage (general considerations)

OC01 and OC02 concern the determination of the statistical population.

Overcoverage occurs when units outside the target population are incorrectly included in the frame.

For example, if some lobster fishers systematically misreport their trap location, their activity may be incorrectly assigned to a lobster fishing area covered by another monitoring program, creating overcoverage in that fishery (and undercoverage in their true fishing area). However, to be relevant for overcoverage, such a discrepancy needs to be recurring and systematic. It is reasonable to expect that, if this had been known to occur, corrective action would have been taken. Consequently, overcoverage is expected to be rare in DFO's fishery monitoring programs, or simply unknown. Furthermore, overcoverage is not expected in:

- all third-party monitoring programs (dockside, video, at-sea observer), which are paid for by the resource users and therefore unlikely to include additional units;
- all monitoring programs that require reporting by resource users as a condition of licence (logbooks, purchase slips), because that licence is specifically tied to a fishery; or,
- creel and in situ surveys as these will be tied to a specific time, location and therefore fishery.

Contribution to bias: Overcoverage will contribute to bias if there is a difference between the incorrectly included units and the target population:

- Incorrectly included subset associated with large values: the overcoverage will contribute positively to bias;
- Incorrectly included subset associated with small values: the overcoverage will contribute negatively to bias; or,
- Incorrectly included subset associated equally with small and large values, or with mean or middle values: no (further) contribution to bias.

Overcoverage will contribute positively to the bias for estimation of a total (e.g., total catch): since additional activity will be incorrectly accounted for.

Contribution to variability: No contribution.

5.4 OC03 Unaccounted-for clustering of samples (general considerations)

This OC does not apply to censuses. OC03 to OC06 concern departures from the sampling protocol and observer effects.

In a sample survey, cluster sampling consists of partitioning the population into subsets called clusters, taking a random sample of clusters, and then, taking a census from each selected cluster (one-stage cluster sampling) or a sample from each selected cluster (twostage cluster sampling). Cluster sampling may require a larger sample size to achieve a desired standard error or may lead to a larger standard error for a given sample size, compared to simple random sampling. However, cluster sampling may be a cost-effective way to reduce the standard error of an estimate.

Cluster sampling is unaccounted-for if it occurs, but the appropriate mathematical formulae (Lohr, 2009) are not used to compute the estimate. If computations for simple random sampling are used, unaccounted-for cluster sampling may lead to an underestimation of the variability of the estimation process.

Unaccounted-for cluster sampling is not expected in a well implemented monitoring program if the sample units are identified and appropriately selected ahead of time or when the logistical constraints and costs of sampling is equal among units. In contrast, unaccounted-for cluster sampling is possible when the logistical constraints, the cost or the convenience of sampling (e.g., travel, time of day, day of week), is considerably different between units or when the marginal costs of sampling within clusters is small.

Contribution to bias: No contribution anticipated.

Contribution to variability: If cluster sampling is occurring while formulae for simple random sampling are used, and clusters are more homogeneous than the population, the computed standard error will be smaller than the true standard error, and the variability of the estimation process will be underestimated. Therefore, when unaccounted-for clustering is present, comparing clusters to the general population is a suggested first step: if the clusters are similar to the general population (or, equivalently, to each other), unaccountedfor clustering will have a negligible contribution to variability.

The impact of unaccounted-for clustering on variability may be assessed by mathematically comparing estimates assuming the planned sampling protocol with those obtained by assuming alternative realized protocol.

The QAT (§1) provides facilities to assess very roughly the contribution of unaccounted-for clustering the variability of the estimation process. Details on those facilities are provided in Allard and Benoît (2019). A better assessment of the underestimation requires a good understanding of the clustering, which often will not be available to the analyst, and advanced statistical computations. Therefore, unaccounted-for clustering should be acknowledged but its contribution to variability will likely be quantified only very roughly in many cases. In others, where it is critical to assess quality precisely, an in-depth assessment of the realized sampling scheme may be required.

5.5 OC04 Unaccounted-for stratification of samples (general considerations)

This OC does not apply to censuses.

In a sample survey, stratified sampling occurs when the target population has been partitioned into subsets (strata) and a sample is drawn separately from each stratum. Strata differ from clusters: in stratified sampling, a sample must be drawn from each stratum; in cluster sampling, a sample or a census is taken from a sample of clusters.

Using appropriate computations, stratified sampling may require a smaller sample size to achieve a desired standard error or may lead to a smaller standard error for a given sample size, compared to simple random sampling. Stratified sampling may be a cost-effective way to reduce the standard error of an estimate.

In fishery monitoring, stratified sampling is common. Stratification could be temporal (e.g., equal number of observations each week of the season or a number of observations proportional to the weekly number of units) or spatial (e.g., one observer assigned to each dock). Such stratifications are used implicitly because they are convenient; they avoid the unevenness of sampling resource requirements that may otherwise exist under simple random sampling. Stratified sampling may also be implicitly required as the result of a regulatory obligation of equal coverage among units in monitoring programs with fleet-level target coverage levels, as occurs in some at-sea observer surveys.

Stratified sampling is unaccounted-for if it occurs, but the appropriate computations (Lohr, 2009) are not used. If computations for simple random sampling are used, unaccounted-for stratified sampling may lead to an overestimation of the variability of the estimation process.

Contribution to bias: Impact may result if strata are more homogeneous than the population and the sample allocation is not proportional to stratum size, as could occur in a river salmon recreational fishery if there is separate allocation of the sampling effort to each river pool in which the pools result in different catch characteristics. Estimating the bias requires some knowledge of the mean characteristics of strata and their sampling fractions.

Contribution to variability: If stratified sampling is occurring unintentionally while formulae for simple random sampling are used and strata are more homogeneous than the population, the computed standard error will be larger than the true standard error, and the variability of the estimation process will be overestimated. Therefore, when unaccounted-for stratification is present, comparing strata to the general population is a suggested first step: if the strata are similar to the general population (or, equivalently, to each other), unaccounted-for stratification will have a negligible contribution to variability.

The impact of these departures from random sampling on variability will be precautionary: ignoring unaccounted-for stratification leads to overestimating variability. However, estimating this impact may be relatively easy and could be carried out if the standard error of the estimate contributes to excessive variability relative to the requirements established at risk screening.

The impact of unaccounted-for stratification on bias and variability may be assessed mathematically by comparing estimates assuming the planned sampling protocol with those obtained by assuming the alternative realized protocol.

The QAT (§1) provides facilities to assess very roughly the contribution of unaccounted-for stratification the variability of the estimation process. Details on those calculations are provided in Allard and Benoît (2019). A better assessment of the overestimation requires a good understanding of the stratification and advanced statistical computations. Therefore, unaccounted-for stratification should be acknowledged but its contribution to variability should only be quantified very roughly unless a reduction in sampling effort is deemed desirable.

5.6 OC05 Other irregular sampling probabilities (general considerations)

This OC and its special cases do not apply to censuses.

The quality assessment process considers three classes of situations, other than unintended cluster or stratified sampling, that may cause the sample selection probabilities to be incongruent with those determined by the sampling protocol and used in analysis. Two special cases, unanticipated exclusions (OC05A) and forced inclusions (OC05B) warrant a specific treatment, while other cases or irregular sampling probabilities overall can be assessed in OC05C.

5.6.1 OC05A Other irregular sampling probabilities: Unanticipated exclusions (general considerations)

Unanticipated exclusions occur when some sampling units are unexpectedly excluded from the sampling process. Like undercoverage, unanticipated exclusions remove units from the sample frame. Unlike undercoverage, the excluded units are known to the monitoring program. For example, recurring failures to deploy monitors due to shortages, equipment failures (e.g., VMS), or cancelled overflights due to weather may create unanticipated exclusions. To be relevant the exclusions should be a recurring feature, year after year.

Contribution to bias: If there is no association between the exclusions and the property being measured (e.g., unplanned lack of monitors during a random portion of the fishing season), no contribution to bias is expected. Otherwise (e.g., lack of monitors during a productive portion of the fishing season), there may be a contribution to bias.

Contribution to variability: The computation of the SE accounts for the reduced sample size. However, unanticipated exclusions may further contribute to variability if the unanticipated exclusions are associated with extreme values (positive contribution – i.e., SE is underestimated) or non-extreme, typical values (negative contribution; SE is overestimated).

5.6.2 OC05B Other irregular sampling probabilities: Forced inclusions (general considerations)

Forced inclusions occur when some units from the statistical population not selected within the sampling protocol are included in the sample for external reasons; their probability of inclusion in the sample is 1. The most prevalent forced inclusion in catch monitoring programs results from targeted sampling. Targeted sampling of suspected non-compliant fishers is commonly used in at-sea observer surveys for enforcement and compliance purposes. Cases of targeting are typically not revealed to the catch monitoring data analyst for reasons of privacy and investigation confidentiality and cannot be accommodated properly in analyses.

Contribution to bias: If the inclusions are not associated with the property being measured (e.g., vessels targeted for safety violations), no contribution to bias is expected. Otherwise (e.g., vessels targeted for illegal discards), there may be a contribution to bias.

Contribution to variability: In most cases, the computation of the SE will underestimate the variability because the forcibly included units do not represent units other than themselves. Given appropriate information, the impact of forced inclusions on the variability in a quality assessment can be approximated using the approach described in a technical note (p. 21) of Allard and Benoît (2019) and implemented in the QAT (§1).

5.6.3 OC05C Other (or all) irregular selection probabilities (general considerations)

The probability of a unit being included in the sample may be different from the probability determined by the sampling protocol. For example, this is possible if at-sea observers prefer particular vessels because of comfort or attitude of the crew or if there are insufficient monitors or observers to address a glut of fishing activity. OC03, OC04, OC05A and OC05B are special types of irregular selection probabilities. Their contributions to bias and variability can be assessed separately when the information allows it. Alternatively, any of these special cases, as well as other types, can be included in OC05C. This is useful, for example, when a simulation was used to assess the overall departure from sample selection protocol (e.g., Benoît and Allard, 2009).

Contribution to bias and variability: These contributions should be assessed according to the irregularities in the sample selection process. In some cases, a mathematical approach can be used. For example, supposing a simple random sampling protocol, and an implementation where 20% of the trips from the most accessible wharfs and 10% of trips from the other wharfs are observed, historical data and mathematical formulae can be used to compare the biases and the standard deviations of the protocol and of the implementation. In more complex cases, the bias and the standard deviation of the implementation may be best obtained by numerical methods (e.g., simulation or bootstrap).

5.7 OC06 Observer effect (general considerations)

This OC does not apply to censuses.

An observer effect occurs when the presence, or anticipated presence, of a human observer or of a technological surveillance tool causes a change in the fishing activity such that it is different from unobserved fishing activities. Observer effects are well known to occur in at-sea observer surveys (Benoît and Allard, 2009; Faunce and Barbeaux, 2011). They may also occur in some self-reporting programs if fishers are aware of surveillance or quality assurance monitoring activities and change their activities accordingly (e.g., logbook entries when there are known compliance over-flights or at-sea boardings in the area, or avoiding fishing or transiting in areas where a monitor is suspected to be). In the case of logbook entries, the observer effect is to eliminate biased reporting of resource user data (see OC8 and OC9 below) for the monitored trips.

The potential for observer effects to exist can be assessed by comparing certain characteristics between observed and unobserved trips such as amounts of catch and effort (Allard and Benoît, 2009; Faunce and Barbeaux, 2011), the size composition of catches (an indicator of potential high-grading; Allard and Chouinard, 1997) and fishing locations. Important differences in catch or effort amounts or fishing locations may be such that observed trips are not representative of the fishery and the bias could be considered potentially very large. We recommend that intense scrutiny be given to observer effects if fishers have the latitude to alter their fishing behaviour or patterns and the incentives to do so are large. The bias contributed by observer effects has the potential to be large and could by itself render quality sufficiently low as to not meet dependability thresholds.

Contribution to bias: Observer effects are likely to contribute to bias. The magnitude of bias will depend on the degree to which the alteration in the fishing or self-reporting activity results in a change in the property being measured (e.g., catch amount) and on the extent of observer effects.

Contribution to variability: None or inconsequential contribution.

5.8 OC07 Missing values due to unintentional factors (general considerations)

A missing value is the observation from a selected unit (for a census, all units; for a sampling survey, unit in the sample) that is not obtained. A missing value is unintentional when the observation is not obtained due to events outside the control of the people (fishers, plant personnel, observers) or technology involved in the monitoring (e.g., accidental loss of a logbook).

If imputation is used to fill-in all missing values for a given fishery, there are no errors caused by the omission. Instead, errors associated with the imputation (OC14, below) should be assessed.

Contribution to bias: In most types of monitoring programs missing values due to unintentional factors will be generated randomly and should therefore not contribute to bias. However, in some programs missing values could be larger or smaller, on average, than the typical population values and therefore contribute to bias. This is possible, for example, in post-season interviews if the most productive fishers have a higher probability of not being reached because they are out fishing.

Contribution to variability: In a sampling survey, the computation of the standard error will take into account the missing values through the reduced sample size. In a census, unintentional missing values will create variability. The variability will be inconsequential if the proportion of unintentionally missing values is small. If the proportion tends to be large (e.g., larger than 25%), the parameter estimation process should be assessed as a sampling survey and the statistical standard error of the estimate computed (according to formulae for simple random sampling if the missing values are considered to occur randomly).

5.9 OC08 Missing values due to intentional factors (general considerations)

A missing value due to an intentional factor occurs when an observation is not made due to an intentional action, most often a refusal to provide information. Missing values due to intentional factors can occur in all forms of self-reported monitoring if there are financial, regulatory or societal (e.g., negative perception associated with discards) incentives to not report certain activities. They are not expected when there is 100% third-party monitoring or verification, except in cases of illegal coercion or collusion. Partial third-party monitoring or verification may also reduce their extent.

If imputation is used to fill in all missing values for a given fishery, there are no errors caused by the missing values. Instead, errors associated with the imputation (OC14) are important.

Contribution to bias: Missing values due to intentional factors are likely to contribute to bias. The degree of the contribution to bias will depend on the extent of missing values and the extent to which their catch characteristics differ from those of non-missing values. In turn, these will depend on the opportunities and incentives to not report.

Contribution to variability: No contribution.

5.10 **OC09-OC14 Errors in data, adjustments, and imputations**

Operational characteristics 9 to 14 address errors in data reported by resource users, by independent monitors and resulting from measurement, adjustments, and imputations. In practice distinguishing between reporting and measurement errors can be difficult. The QA summary computes the contribution of OC09 to OC14 to variability and bias of the estimation process identically. Therefore, the overall assessment will not be impacted by incorrect attributions among these six OCs. Grouping the impact of several of these OCs may be necessary if it is difficult to assess them individually. However, attention should be given to not report an issue in more than one OC since this would lead to multiplying the impact of the issue in the overall assessment. Correct attributions will indicate where errors are occurring and may help improve future estimation processes or help to improve quality.

Contribution to variability: For the estimation of a mean or a total, the contribution of each of these errors to the variability of the estimation process will be inconsequential if the relative imprecision

$$[standard\ deviation\ of\ the\ errors]/[typical\ observed\ value]$$

is small or the population size is large.

Technical note: For the estimation of a mean or a total, the contribution of each of these errors to the relative variability of the estimation process unaccounted for in the standard error will be

$$\frac{[standard\ deviation\ of\ the\ errors]}{[typical\ observed\ value]} \times \frac{1}{\sqrt{[population\ size]}}.$$

This statement applies to sampling surveys and to censuses, where the standard error is 0. The contribution is added quadratically.

For example, if the relative imprecision of visual weight estimates is 50% (an extremely large value) and the population size is 100, the contribution of the imprecision of the visual weight estimates to the estimation process would be only 5%.

5.11 **OC09 Errors in data reported by resource users (general considerations)**

OC09 considers recurring errors related to the implementation of the program including, for example unintentional errors due to lack of training, carelessness, etc. and intentional errors aiming to mislead fishery regulators, revenue agencies or others (e.g., environmental or consumer groups).

Intentional errors are not expected when there is 100% third-party monitoring or verification, except in cases of illegal coercion or collusion. Partial third-party monitoring or verification may also reduce their extent.

Contribution to bias: Errors will typically contribute to bias when they are intentional. Underreporting is a common form of biased error, particularly when there are quotas, limits (e.g., bycatch limits), prohibitions (e.g., protected species catch) or economic and societal pressures (e.g., by consumer groups). Over-reporting has also occurred in cases in which harvesters anticipated the imposition of future catch shares based on present catch amounts. If incentives are strong and independent monitoring or enforcement activities are insufficient, the biases may be very large. If the contribution to bias is assessed to be very large, even using a lowest-plausible value can cause the QA to render low the quality of the entire estimation process.

Unintentional errors may also contribute to bias, particularly when self-reporting is based on recall. Recall of distant events (i.e., fishing trips) is typically less accurate than that of recent events. Better recall of exceptional catch events may bias reporting.

Contribution to variability: See §5.10 above.

5.12 OC10 Errors in data reported by independent observers (general considerations)

Independent observers are typically third-party staff. Errors in data reported by independent observers are likely to belong to one of two types: unintentional errors due lack of training, inattention, errors in protocols or standards, etc., and intentional errors due to collusion with or coercion by resource users.

Contribution to bias: Intentional errors due to collusion or coercion are likely to contribute to bias. Unintentional errors due to inattention are not expected to contribute to bias. Errors due to lack of qualification or training may contribute to bias if they result in a systematic difference between the true quantity and that which is reported (e.g., frequently recording weights without subtracting the container weight). Likewise, unintentional errors in protocols or standards may lead to systematic reporting errors that contribute to bias (e.g., recreational fishing effort measured using aerial counts may unintentionally misclassify pleasure vessels as recreational fishing vessels resulting in an overestimate of fishing effort). However, one can presume that systematic and recurring unintentional errors contributing to bias would have been addressed by appropriate modifications to the monitoring programs.

Contribution to variability: See §5.10 above.

5.13 OC11 Equipment error (general considerations)

Measuring tools include scales, automated fish counting devices, and the human eye as used for visual estimations of mass or volume. This OC is concerned with the inaccuracy and imprecision of these measuring tools.

Contribution to bias: Expected to be rare.

Contribution to variability: See §5.10 above.

5.14 OC12 Data handling and processing error (general considerations)

Data handling refers to all data manipulation steps occurring after the initial recording of the observations. A data handling error occurs when a data manipulation creates an error.

All monitoring programs are susceptible to data handling errors. Errors are expected to propagate with each data handling step. Quality assurance and control procedures can help attenuate data handling errors.

Contribution to bias: Data handling errors are unlikely to contribute to bias. Biased errors could occur if an error is systematic (e.g., neglecting to convert pounds to kilograms will create a positive bias in the estimation of total catch in kilograms). However, one can presume that systematic and recurring data handling and processing errors contributing to bias would have been addressed by appropriate modifications to the monitoring programs.

Contribution to variability: Under proper quality assurance and control procedures, data handling and processing errors may be small relative to other sources of error. Unless they are believed to be further reducible or of sufficient magnitude to impact the quality of estimation processes significantly, assuming negligible contributions to variability in the quality assessment could be justified. Previous quality assessment within DFO processes have assumed contributions to variability on the order of 0-5% for all observed values. Also see §5.10 above.

5.15 OC13 Adjustment error (general considerations)

Adjustments are required when observations are obtained for measured properties that are not the same of those subject to estimation. Common examples of adjustments are conversions applied to measured weights of processed fish (e.g., gutted, head-off) to produce round-weight equivalents. An adjustment error occurs when an adjustment yields incorrect values.

This OC does not apply to monitoring programs that do not use adjustments.

Presently DFO uses a variety of adjustment equations, many of which have not been updated for a long time. Given the possibility of changes in biological properties of some stocks and of changes in industry methods, these adjustment equations may need to be reassessed.

Contribution to bias: Adjustment errors should not contribute to bias when the adjustment equations have been properly established and updated in a timely manner and applied correctly. However, if the adjustment equations are incorrect, they will likely lead to bias as the equation may consistently underestimate or overestimate the true value.

Contribution to variability: See §5.10 above.

5.16 OC14 Imputation error (general considerations)

An imputation is the replacement of a missing value by a value obtained (imputed) from other information available, including the values for spatially and temporally adjacent or otherwise similar observed units. Imputation depends on the availability of auxiliary data or information and on a statistical model to compute the imputed values. The imputation error is the difference between the imputed value and the true but unavailable value. In some cases, the quality of a statistical imputation method (accuracy and precision) can be assessed using statistical methodology, such as by simulation or cross-validation.

For estimation of a total, accounting for missing values by multiplying the observed total by [number of units]/[number of observed units] is equivalent to imputing the mean of the observed units to the missing values.

Contribution to bias: Imputation can generate bias in the estimation process if imputed values are, on average, smaller/greater than the true values that would otherwise have been observed. For example, in an estimation of bycatch of a certain fish population, if missing logbooks or hail-ins are associated with high bycatches, imputing the mean of observed values will contribute negatively to the bias.

Contribution to variability: See §5.10 above.

5.17 OC15 Modelling error (general considerations)

Some statistical estimators are based on a statistical model. A modelling error occurs when the statistical model does not fit the data perfectly.

Modelling is used to scale-up parameter estimates from samples to populations or when parameters of interest are estimated indirectly, such as in the estimation of catch from estimates of CPUE and effort. Modelling is particularly relevant for the analysis of at-sea observer surveys, where various types of design and model-based approaches are used to make inferences at the population level (Rochet and Trenkel, 2005), and of questionnaires, interviews, and creel surveys, where auxiliary variables such as effort are typically used (Pollock et al, 1994).

Contribution to bias and variability: The contributions depend on the nature of the model and how well it represents the link between the data and the parameter estimated, and is therefore case specific. For example, estimating the CPUE by the ratio estimator (i.e., [total sample catch]/[total sample effort]) assumes that the catch is proportional to effort (i.e., the model is catch = CPUE × effort + residual). If the CPUE depends on effort (e.g., if very active recreational fishers are more skilled and have higher individual CPUE), the ratio estimator will give a biased estimate due to the model lack-of-fit.

In all cases where a model is used, the contribution of the uncertainty on any parameter present in the model to the variability of the estimation process should be assessed.

6 MANDATORY RESOURCE USER REPORTS WITH AUDITING MECHANISM

This section concerns the quality assessment of estimation processes based on mandatory 100% reporting by users supplemented by one or more monitoring methods allowing for independent verification of the reports. These monitoring tools include logbooks supplemented by 100% video monitoring from which a random subset is audited, and hail effort reports supplemented by 100% GPS tracking and fishing gear winch rotation sensors. The supplemental monitoring methods allow DFO to verify the users' reports randomly or completely. These supplemented monitoring methods are a condition of licence and are implemented as censuses in that information is collected on all events of all trips (e.g., 100% video), even though data may be extracted for only an audited subset (e.g., 15% video review) (Beauchamp et al., 2019). Parameters estimated by these monitoring programs include effort (i.e., number of fishing events), location and time of fishing, and catch.

User reports that can be verified tend to give exceptionally reliable catch estimates (retained and released) of species that are part of the audit, particularly if errors or non-compliance identified by the secondary monitoring leads to extra cost to the user. For example, the extent of review of video monitoring in the Pacific groundfish fishery, a direct cost to the industry, is tied to the accuracy of the mandatory logbook reports (Stanley et al., 2011, 2015).

6.1 Statistical characteristics (User reports with auditing mechanism)

Pertinence: These monitoring programs are censuses: the statistical bias and the standard error are 0.

6.2 OC01 Undercoverage (User reports with auditing mechanism)

Pertinence: This type of monitoring is often used to estimate total catch. In this case, undercoverage will contribute negatively to the bias of the estimate. Undercoverage is unlikely in fisheries with 100% verifiable fisher reporting.

6.3 OC02 Overcoverage (User reports with auditing mechanism)

Pertinence: Overcoverage is unlikely to be pertinent.

6.4 OC03 to OC06 Sampling related OCs (User reports with auditing mechanism)

Pertinence: These monitoring programs are censuses: OC03 to OC06 do not apply.

6.5 OC07 Missing values due to unintentional factors (User reports with auditing mechanism)

Pertinence: Missing values are unlikely to occur because there is a second source of information from which these can be gleaned. If however there is no process to obtain missing values from the second source, only the systematic occurrence of unintentional missing values is pertinent to the quality assessment.

Contribution to bias: Unlikely since missing values due to unintentional factors are expected to be randomly generated and uncorrelated with the values.

Contribution to variability: Likely to be inconsequential.

6.6 OC08 Missing values due to intentional factors (User reports with auditing mechanism)

Pertinence: Due to the supplemental monitoring tools, missing values due to intentional factors are likely to be rare since the users withholding information must also ensure that the supplemental monitoring is also inactive or ineffective.

Contribution to bias: In the case of systematic missing values, the negative contribution to bias is very likely because catch (bycatch) limits and stigma associated with incidental capture (e.g., capture of protected, depleted or charismatic fauna) all create an incentive to not report.

Contribution to variability: None.

6.7 OC09 Errors in data reported by resource users (User reports with auditing mechanism)

Pertinence: Due to the supplemental monitoring tools, systematic intentional errors are likely to be rare since users supplying erroneous data incur the risk of a verification and penalties.

Contribution to bias and variability: As per the general considerations. If systematic unintentional errors are suspected, their contribution to bias and variability can be assessed by examining the results of audits.

6.8 OC10 Errors in data reported by independent observers (User reports with auditing mechanism)

Pertinence: Not applicable. Independent observers are not involved in the main monitoring program supplying data for the estimation process.

6.9 OC11 Equipment error (User reports with auditing mechanism)

Pertinence: Most fisher reports of catch amounts are based on visual estimation and subject to errors. Catch amount reported by large at-sea factory ship are based on scale weights.

Contribution to bias: Anticipated to be low given the possible verification.

Contribution to variability: Could be high for visual estimation of the estimation of weight but low for the estimation of small counts of the number of fish (e.g., large pelagic fish fisheries). See §5.10 above.

6.10 OC12 Data handling and processing error (User reports with auditing mechanism)

Pertinence: Pertinence: Likely to be pertinent if the reporting is on paper because of transcription errors and because this adds a data handling step. Less likely for electronic logbooks and other direct data entry methods.

Contribution to bias: Not anticipated.

Contribution to variability: See §5.10 above.

6.11 OC13 Adjustment error (User reports with auditing mechanism)

Pertinence: Unlikely to be pertinent if catch reporting in fisher reports records whole fish. In monitoring programs where the logbooks are recording processed fish, an adjustment may be required, in which case an assessment of adjustment error is pertinent.

Contribution to bias and variability: As per the general considerations.

6.12 OC14 Imputation error (User reports with auditing mechanism)

Pertinence: Pertinent only if imputation is systematically used. Imputation is rare under 100% mandatory reporting.

Contribution to bias and variability: As per the general considerations.

6.13 OC15 Modelling error (User reports with auditing mechanism)

Pertinence: Modelling error is unlikely to be pertinent as catch properties of interest, means and total, are estimated directly.

Contribution to bias and variability: As per the general considerations.

7 MANDATORY RESOURCE USER REPORTS WITHOUT AUDITING MECHANISM

This section concerns the quality assessment of estimation processes based on mandatory 100% reporting by fishers that cannot be verified by one or more secondary monitoring methods (e.g., 100% video monitoring). These monitoring tools include many logbooks and hail monitoring programs and are a principal means of collecting information on effort, retained catch and discarded catch. Logbooks, for example, are the most common monitoring tool across Canadian commercial fisheries and are a requirement in most (Beauchamp et al., 2019). These monitoring tools result from a condition of licence and are implemented as censuses (Beauchamp et al., 2019).

7.1 Statistical characteristics (User reports without auditing mechanism)

Pertinence: Monitoring programs of this type are censuses. The statistical bias and the standard error are 0.

7.2 OC01 Undercoverage (User reports without auditing mechanism)

Pertinence: This type of monitoring is often used to estimate total catch. In this case, undercoverage will contribute negatively to the bias of the estimation process.

Undercoverage may be pertinent if the fishery has many participants and/or fishing events, historical compliance with related regulations (e.g., mandatory hails) is low or there is limited verification. Undercoverage is likely to be pertinent in fisheries with many participants, including some recreational fisheries, if follow-ups to ensure mandatory reporting are not in place.

Undercoverage is unlikely to be pertinent if the fishery involves only a few large participants or fishing events, historical compliance is high or there is independent monitoring of fishing activities such that the population can be defined independently of the resource users, such as by mandatory GPS vessel tracking, independent surveying of fishing trips, etc. for followup enforcement of mandatory reporting.

7.3 OC02 Overcoverage (User reports without auditing mechanism)

Pertinence: Overcoverage is unlikely to be pertinent because a licence is specifically tied to a fishery. It will be pertinent only if some fishers systematically misreport the location of their catch.

7.4 OC03 to OC06 – OCs related to sampling (User reports without auditing mechanism)

Pertinence: These monitoring programs are censuses: OC03 to OC06 do not apply.

See OC08 and OC09 for the impact of observers on missing values and/or data errors, as per general conditions.

7.5 OC07 Missing values due to unintentional factors (User reports without auditing mechanism)

Pertinence: Unintentional factors include unreadable logbook entries, radio malfunction preventing a hail, etc. Only the systematic occurrence of unintentional missing values is pertinent to the quality assessment.

Contribution to bias: Unlikely since missing values due to unintentional factors are expected to be randomly generated and uncorrelated with the values.

Contribution to variability: Likely to be small.

7.6 OC08 Missing values due to intentional factors (User reports without auditing mechanism)

Pertinence: Catch (bycatch) limits and stigma associated with incidental capture (e.g., capture of protected, depleted, or charismatic fauna) all create an incentive to not report (e.g., logbooks not submitted). Absence of rigorous enforcement or of other follow-ups can be a contributing factor.

Contribution to bias: A negative and, possibly, large contribution to bias should be considered likely.

Contribution to variability: Not anticipated.

7.7 OC09 Errors in data reported by resource users (User reports without auditing mechanism)

Pertinence: Errors in data reported by resource users can be intentional (e.g., underreporting discards) or unintentional (e.g., data entry mistakes).

Reports in logbooks and other forms of mandatory self reporting are susceptible to underreporting of catch amounts, overreporting of the frequency of zero catches and reporting a smaller diversity of species (e.g., Allen et al. 2002; Walsh et al. 2002; Bremner et al. 2009). There may be regulatory or economic incentives to misreport, including catch limits, prohibitions on catching certain species, and capture of charismatic species, which could otherwise lead to short-term or longer-term fishery closures (Faunce 2011) or reduced income (e.g., difficulties with eco-certification). There may also be a desire to hide prime fishing locations or avoid enforcement actions (e.g., Stanley 1992; Metuzals et al. 2005; Rijnsdorp et al. 2007).

Contribution to bias: Considering intentional errors, a negative contribution to bias should be considered likely. Unintentional errors are unlikely to contribute to bias.

Contribution to variability: Considering intentional errors, not anticipated. For unintentional errors, See §5.10 above.

If there is also monitoring by at-sea observers of a sample of trips, comparing logbook data for trips without and with at-sea observers can provide information on this type of errors, assuming that logbook entries on trips monitored by observers will be fully compliant.

7.8 OC10 Errors in data reported by independent observers (User reports without auditing mechanism)

Pertinence: Not applicable. Independent observers are not involved.

7.9 OC11 Equipment error (User reports without auditing mechanism)

Pertinence: A fisher's estimates of catches are typically based on a visual assessment and relies on the fisher's ability to correctly identify taxa in the catch. Therefore, there may be a heightened degree of misidentification for uncommon or cryptic species.

Contribution to bias: Unlikely since error will tend to go in either direction.

Contribution to variability: Variability contributed by visual estimation of mass could be high. Variability contributed by visual estimation of numbers could be low in fisheries that catch relatively few individual fish (e.g., large pelagic fish fisheries).

For measurements of effort, the variability will depend in part on the resolution with which effort is measured and reported (e.g., days vs hours, number of nets vs total dimensions).

See **§5.10 above**.

7.10 OC12 Data handling and processing error (User reports without auditing mechanism)

Pertinence: As per general considerations.

7.11 OC13 Adjustment error (User reports without auditing mechanism)

Pertinence: Pertinence: Pertinent only if adjustments are made. For example, an adjustment in user reported data for round weight to net weight, or inversely, may be necessary when the fish is processed at sea.

Contribution to bias and variability: As per the general considerations.

7.12 OC14 Imputation error (User reports without auditing mechanism)

Pertinence: Pertinent only if imputation is systematically used, a rare occurrence for 100% mandatory user-reported data.

Contribution to bias and variability: As per the general considerations.

7.13 OC15 Modelling error (User reports without auditing mechanism)

Pertinence: Only if modelling is used. Modelling is rarely used for 100% mandatory userreported data since both means and totals can be calculated directly from the data.

Contribution to bias and variability: As per the general considerations.

8 MANDATORY DOCK-SIDE MONITORING

This chapter covers estimates based on data from dock-side monitoring program when used as the primary source of information. Dock-side monitoring programs (DMP) can be used to estimate the total landed catch, if all landing sites are monitored or, otherwise, a part of the landed catch. Dock-side monitoring cannot be used to estimate non-landed catch (i.e., catch sold, used, or discarded at sea).

8.1 Statistical characteristics (Dock-side monitoring)

Pertinence: Dock-side monitoring used as a primary source of information is often implemented as a census, although in some fisheries it is implemented as a sampling survey with sample selection based on mandatory hail-in (e.g., fall herring fishery on the Gaspé peninsula).

For census, the statistical bias and the standard error are 0. For a sampling survey, see general considerations.

8.2 OC01 Undercoverage (Dock-side monitoring)

Pertinence: Undercoverage may be pertinent if the frame is determined by a self-reporting mechanism (e.g., incoming hails where harvesters may fail to hail), but is unlikely to be pertinent if the frame is determined independently (e.g., by mandatory vessel tracking).

8.3 OC02 Overcoverage (Dock-side monitoring)

Pertinence: Third-party monitoring programs are paid for by the resource users and therefore likely to avoid including incorrect trips. In fisheries by a large fleet of small vessels, dockside observers may not be able to identify vessels that have fished outside the fishing area associated to their dock.

8.4 OC03 Unaccounted-for clustering of samples (Dock-side monitoring)

Pertinence for a census: Not applicable.

Pertinence for a sampling survey: For a sampling survey, unaccounted-for clustering will occur if, for example, dock-side observers are assigned to docks for time periods. Such an assignment creates a temporal and geographical clustering, each time period-dock combination being a cluster.

Contribution to bias and variability: As per the general considerations.

8.5 OC04 Unaccounted-for stratification of samples (Dock-side monitoring)

Pertinence for a census: Not applicable.

Pertinence for a sampling survey: For a sampling survey, unaccounted-for stratification will occur if, for example, dock-side observers are deployed evenly through a season.

Contribution to bias and variability: As per the general considerations.

8.6 OC05 Other irregular sampling probabilities (Dock-side monitoring)

8.6.1 OC05A Other irregular sampling probabilities: Unanticipated exclusions (Dockside monitoring)

Pertinence for a census: Not applicable.

Pertinence for a sampling survey: For a sampling survey, instances such as observer shortages resulting from a glut of activities or observer absences (sick leave, vacation) and delayed hail-ins that result in an incapacity of observers to receive a trip, *if systematic*, can lead to unanticipated exclusions.

Contribution to bias: An observer shortage during the most productive part of the season would lead to a negative contribution to the bias. Similarly, if delayed hail-ins are associated to trips with high proportions of bycatch.

Contribution to variability: As per the general considerations.

8.6.2 OC05B Other irregular sampling probabilities: Forced inclusions (Dock-side monitoring)

Pertinence for a census: Not applicable.

Pertinence for a sampling survey: For a sampling survey, the impact of forced inclusions should be assessed if there is targeted monitoring, typically due to enforcement actions.

Contribution to bias and variability: As per the general considerations.

8.6.3 OC05C Other (or all) irregular selection probabilities (Dock-side monitoring)

Pertinence for a census: Not applicable.

Pertinence for a sampling survey: For a sampling survey, systematic departures from the DMP's sampling protocol not included in OC03, OC04, OC05A or OC05B should be reported, although no causes for such departures were identified for dockside monitoring for this report.

Contribution to bias and variability: As per the general considerations.

8.7 **OC06 Observer effect (Dock-side monitoring)** *Pertinence*

for a census: Not applicable.

Pertinence for a sampling survey: An observer effect could result if harvesters change their destination, shorten their trips or discard part of their catch if they become aware that an observer is at their destination or scheduled to monitor their catch.

Contribution to bias: Observer effect may contribute to bias. The contribution is anticipated to be negative for a catch estimate.

Contribution to variability: None or inconsequential contribution.

8.8 **OC07 Missing values due to unintentional factors (Dock-side monitoring)**

Pertinence: The frequency of missing values – landings that were sampled but for which there are no recorded values – should be very small to nil as there are typically several procedures in place in third-party monitoring to avoid such instances.

Contribution to bias: Impact unlikely since the unintentional factors (e.g., lost records) are expected to be independent from the value observed (e.g., catch).

Contribution to variability: Impact likely inconsequential given the small frequency of occurrence. Also see **§5.10 above**.

8.9 **OC08 Missing values due to intentional factors (Dock-side monitoring)**

Pertinence: Only pertinent if coercion or collusion is reported or suspected.

Contribution to bias: The contribution to bias of missing values attributed to coercion or collusion should be assessed if it is believed to be pertinent.

Contribution to variability: None.

8.10 OC09 Errors in data reported by resource users (Dock-side monitoring)

Pertinence: Not pertinent. The data are provided by independent observers.

8.11 OC10 Errors in data reported by independent observers (Dock-side monitoring)

Pertinence: Unintentional errors by independent observers are likely, although the magnitude may not be large. Intentional errors, contributing to bias should be assessed if coercion or collusion is frequently reported or suspected.

Contribution to bias: Intentional errors are likely to contribute to bias. Unintentional errors are unlikely to contribute to bias.

Contribution to variability: Impact likely to be inconsequential.

8.12 OC11 Equipment error (Dock-side monitoring)

Pertinence: Dockside monitors typically use calibrated scales to obtain catch weights. Although measurement errors are expected, their magnitude is likely small.

Contribution to bias: Unlikely.

Contribution to variability: Impact likely to be inconsequential. See §5.10 above.

8.13 OC12 Data handling and processing error (Dock-side monitoring)

Pertinence: Pertinent except when fully automated data capture and transmission is in place.

Contribution to bias and variability: As per the general considerations.

8.14 OC13 Adjustment error (Dock-side monitoring)

Pertinence: Pertinent if adjustments are used, including when catches are landed on ice or have been subject to some processing at sea (e.g., dressing) or when the weight of the containers varies and must be estimated.

Contribution to bias and variability: As per the general considerations.

8.15 OC14 Imputation error (Dock-side monitoring)

Pertinence: Pertinent if imputation is used.

Contribution to bias and variability: As per the general considerations.

8.16 OC15 Modelling error (Dock-side monitoring)

Pertinence: For a census, typically no modelling involved. For sampling surveys, the models used to extrapolate to total catch should be assessed for modelling errors impact on bias and variability.

Contribution to bias and variability: As per the general considerations.

9 MANDATORY PURCHASE SLIP MONITORING

Purchase slips (commercial sales slips) report only fish that is landed and sold. They are inappropriate for monitoring discards. Total catch reported by purchase slips will underestimate true total catch when some fraction of landings is not sold in official channels (e.g., fish for personal use, or sold privately). Reporting of fish sales is compulsory, therefore purchase slips constitute a census of official sales. They are a relatively common tool with most DFO regions (Central and Arctic, Newfoundland and Labrador, Pacific, Quebec, and Gulf regions) still using sales slips and requiring them to be submitted as part of their licence conditions (Beauchamp et al., 2019).

The QA of an estimation process based on mandatory purchase slip monitoring is anticipated to be simple since there are few possible departures from the protocol.

9.1 Statistical characteristics (Purchase slips)

Pertinence: These monitoring programs are censuses: the statistical bias and the standard error are 0.

9.2 OC01 Undercoverage (Purchase slips)

Pertinence: For purchase slips, undercoverage refers to a purchase that is not reported and, therefore, unknown to DFO. There will be no undercoverage if the data are used to estimate landings of total official sales or total landings in a case in which all landings are officially sold.

For estimation of total landings, undercoverage will be pertinent if a non-negligible quantity of landed fish is not officially sold.

9.3 OC02 Overcoverage (Purchase slips)

Pertinence: Very unlikely to be pertinent. Purchases slips are tied to fishing licences and therefore fisheries. Purchasers are unlikely to intentionally add purchase slips.

9.4 OC03 to OC06 – OCs related to sampling (Purchase slips)

Pertinence: These monitoring programs are censuses: OC03 to OC06 do not apply.

9.5 OC07 Missing values due to unintentional factors (Purchase slips)

Pertinence: Unlikely to be pertinent. Missing mandatory purchase slips are likely to be identified by normal quality assurance verifications and to be solicited.

9.6 OC08 Missing values due to intentional factors (Purchase slips)

Pertinence: Unlikely to be pertinent. Missing mandatory purchase slips are likely to be identified by normal quality assurance verifications and to be solicited.

9.7 OC09 Errors in data reported by resource users (Purchase slips)

Pertinence: Unintentional errors in purchase slips are expected to be rare since the purchase slips are part of a financial transaction. Intentional errors may occur if, for example, there are incentives to underreport catch or revenue. *Contribution to bias and variability*: As per the general considerations.

9.8 OC10 Errors in data reported by independent observers (Purchase slips)

Pertinence: Not applicable. Independent observers are not involved.

9.9 OC11 Equipment error

Pertinence: Contribution to bias is unlikely and contribution to variability small.

Contribution to variability: Expected to be small since the purchase slips are part of a financial transaction and both the sellers and buyers have interest in using measurements with high precision. Also see §5.10 above.

9.10 OC12 Data handling and processing error (Purchase slips)

Pertinence: Pertinent only if the data acquisition and transmission are not fully electronic.

Contribution to bias and variability: As per the general considerations.

9.11 OC13 Adjustment error (Purchase slips)

Pertinence: Adjustments are used when catches have been subject to some processing at sea (e.g., dressing or where ice has not been tared before weighing the catch).

Contribution to bias and variability: As per the general considerations.

9.12 OC14 Imputation error (Purchase slips)

Pertinence: Imputation is not anticipated to occur systematically. Pertinent if imputation is used.

9.13 OC15 Modelling error (Purchase slips)

Pertinence: Not pertinent: modelling is not involved in estimating sold landings using data from purchase slip monitoring programs.

10 AT-SEA OBSERVER PROGRAMS

This section covers estimates based on data from at-sea observer programs.

For at-sea monitoring, the selection of observations is hierarchical: trips are selected, and, within trips, events (e.g., fishing hauls, observer's working time) are selected. Therefore, the survey type should be established at both levels. At the trip level, the program can be implemented as a census, with observers assigned to all trips, or as a sampling survey, with observers assigned to a sample of trips, where the trips are typically pre-announced (e.g., mandatory hail-outs). Similarly, at the event level, observers may observe all events, providing a census, or may observe a sample of events. The coverage level for trips, the percentage of trips to be monitored, is generally fishery or fleet specific and is set by DFO to meet the needs of conservation, enforcement, and program complexity (e.g., the use of multi-species individual transferable quotas). Sampling of events is generally dictated by logistics and observer capacity. For example, it is typically not possible for a single observer to monitor all hauls when fishing operations occur continuously over a 24-hour period.

The quality assessment of estimates derived from at-sea observer programs depends on the nature of sampling at the trip and event level. In addition, there may be sub-sampling of event-level catches to determine catch composition (size, sex, species, etc.). If sampling occurs at more than one level, a rigorous assessment of the impact of sampling (OC03 to OC06) would require complex simulations. In the considerations below, we suggest approximations that we consider sufficient for the purpose of the quality assessment.

10.1 Statistical characteristics (At-sea observers)

Pertinence: For censuses at both levels, the statistical bias and the standard error are 0. Atsea monitoring is often a sampling survey, at least at the event level. Therefore, statistical bias and the standard error will usually be pertinent.

10.2 OC01 Undercoverage (At-sea observers)

Pertinence at the trip level: Undercoverage may be pertinent if the statistical population is determined by a self-reporting mechanism (e.g., fishers failing to send mandatory outgoing hails) but is unlikely to be pertinent if the population is determined independently (e.g., mandatory vessel tracking).

Pertinence at the event level: Undercoverage is possible if events can be concealed from the observer, such as hauls that occur unbeknownst to the observer while they are off watch. Concealed events cannot be observed in a census or included in the sampling frame in a sampling survey. Such events would presumably contravene licence conditions and should be rare.

If undercoverage is considered systematic and consequential, we suggest that the extent of the undercoverage be established as a proportion of all events. For example, suppose that the trip-level frame consists of 80 trips while the true population size is thought to be 100 trips and the event-level frames average 6 events per trip while the true value is thought to be 8 events per trip. Then 80 × 6 = 480 events are in the overall event frame compared to 100 × 8 = 800 thought to be in the true event population. Then, the undercoverage is calculated as (800 − 480)/480 = 66.7%.

10.3 OC02 Overcoverage (At-sea observers)

Pertinence at the trip level: Unlikely to be pertinent; third-party monitoring programs are paid for by the resource users and therefore unlikely to include incorrect trips.

Pertinence at the event level: Not pertinent as it is very unlikely that events would be included incorrectly in the monitoring program.

10.4 OC03 Unaccounted-for clustering of samples (At-sea observers)

In a sampling survey, cluster sampling is unaccounted-for if the computation of the estimate does not use the mathematical formulae appropriate for cluster sampling.

Pertinence at the trip level: Not pertinent for a trip-level census. For a sampling survey, unaccounted-for clustering will occur if, for example, at-sea observers are deployed on wharf and/or time-period basis. The wharf-time-period combinations creates clusters.

Pertinence at the event level: Not pertinent for an event-level census. For an event-level sampling survey, unaccounted-for clustering is likely to occur because sampling will cluster according to the observer's working shift (temporal clustering). Notably, the clusters may be selected systematically (e.g., one 8-hour shift followed by 16 hours of rest).

If both levels are sampling surveys, to simplify the assessment, we suggest that the assessment be based on a comparison of combined trip- and event-level clusters, i.e., considering the two sampling levels as a single level. For example, consider the estimation of the number of dolphins caught incidentally in a trawl fishery and suppose a trip-level sampling from clusters of departures defined by wharfs and event-level sampling from clusters defined by days (i.e., the observer will observe the tows from some randomly chosen days). Combining sampling levels means considering each fishing day on any trip as a single cluster (i.e., ignoring the fact that the wharfs was selected first). If the values for the fishing days are typically similar, unaccounted-for clustering will not have an impact on variability.

10.5 **OC04 Unaccounted-for stratification of samples (At-sea observers)**

In a sampling survey, sampling stratification is unaccounted-for if the computation of estimate does not use the mathematical formulae appropriate for stratified sampling.

Pertinence at the trip level: Not pertinent for a census. For a sampling survey, stratification will occur if, for example, at-sea observers are deployed evenly through a season (temporal stratification) or by wharf (spatial stratification).

Pertinence at the event level: Not pertinent for an event-level census. For an event-level sampling survey, unaccounted-for stratification is possible if, for example, observers choose one event during each shift.

If both levels are sampling surveys, to simplify the assessment, we suggest that the assessment be based on a comparison of combined trip- and event-level strata, i.e., considering the two sampling levels as a single level. For example, given a trip-level stratification based on 5 locations and event-level stratification based on time periods, one would consider a single stratification based on the locations-time-periods combinations. If the values for the locations-time-periods combinations are similar, unaccounted-for stratification will not have an impact on variability.

10.6 **OC05 Other irregular sampling probabilities (At-sea observers)**

10.6.1 OC05A Other irregular sampling probabilities: Unanticipated exclusions (At-sea observers)

Pertinence at the trip level: Not pertinent for a census. For a sampling survey, instances such as observer shortages resulting from a glut of activities or observer absences (sick leave, vacation) and delayed hail-outs that result in an incapacity of observers to join a trip, if systematic, can lead to unanticipated exclusions.

Pertinence at the event level: Not pertinent for an event-level census. For an event-level sampling survey, unanticipated exclusions are unlikely.

10.6.2 OC05B Other irregular sampling probabilities: Forced inclusions (At-sea observers)

Pertinence at the trip level: Not pertinent for a census. For a sampling survey, the possibility of forced inclusion, typically due to enforcement actions, should be considered. Forced inclusions are a well-documented component of at-sea observer surveys in Canada and are often used to fulfil enforcement or deterrence objectives (Benoît and Allard, 2009).

Pertinence at the event level: Not pertinent for an event-level census. For an event-level sampling survey, forced inclusions at the event level are unlikely.

10.6.3 OC05C Other (or all) irregular selection probabilities (At-sea observers)

Pertinence at the trip level: Not pertinent for a census. For a sampling survey, systematic departures from the sampling selection protocol could include the differential selection of trips based on the accessibility of docks, or observer preferences for times of day or vessels.

Pertinence at the event level: Not pertinent for an event-level census. For an event-level sampling survey, irregular selection probabilities are most likely to be associated with either clustering or stratification and captured under OC3 and OC4, respectively.

10.7 **OC06 Observer effect (At-sea observers)**

Pertinence at the trip level: Not pertinent for a census. For a sampling survey, fishers can change fishing grounds, or alter their fishing gear or fishing and discarding methods when an observer is on-board.

Pertinence at the event level: Not pertinent for an event-level census. For an event-level sampling survey, could be pertinent if fishers can change fishing and discarding methods when the observer is active versus inactive.

Contribution to bias: At both levels, observer effect may contribute to the bias of the estimation process. The contribution is anticipated to be negative for a catch estimate.

Contribution to variability: None or inconsequential contribution.

10.8 **OC07 Missing values due to unintentional factors (At-sea observers)**

Pertinence at the trip level and event-level: The systematic presence of unintentional missing values – trips or events that should have been observed but for which there are no reported values – should be very small to nil as there are typically several procedures in place in third-party monitoring to avoid such instances.

Contribution to bias: Impact unlikely since the unintentional factors (e.g., lost records) are expected to be independent from the value observed (e.g., catch).

Contribution to variability: Impact likely inconsequential given the small frequency of occurrence.

10.9 **OC08 Missing values due to intentional factors (At-sea observers)** *Pertinence*:

Possible, in particular if coercion or collusion has been reported.

Contribution to bias: The contribution to bias of missing values attributed to coercion or collusion should be assessed.

Contribution to variability: Mostly accounted-for in the computation of the standard error.

10.10 OC09 Errors in data reported by resource users (At-sea observers)

Pertinence: Not pertinent. The data are provided by independent observers.

10.11 OC10 Errors in data reported by independent observers (At-sea observers)

Pertinence: Unintentional errors by independent observers are likely, although the magnitude may not be large. Intentional errors are possible due to coercion or collusion.

Contribution to bias: None for accidental errors. For deliberate errors, contributions to bias are likely.

Contribution to variability: Mostly accounted-for in the computation of the standard error.

10.12 OC11 Equipment error (At-sea observers)

Pertinence: Probable. Visual estimation of catch estimates may be erroneous. Misidentification of species may contribute to the errors.

Contribution to bias: Systematic equipment errors are unlikely to contribute to bias except where observer training or protocols are incorrect or inadequate, although one would presume that this would be corrected, were it to be known.

Contribution to variability: Contribution to variability of errors on the visual estimation of mass could be high. The contribution to variability could be low in fisheries that catch relatively few individual fish (e.g., large pelagic fish fisheries).

Effort reporting tends to be very precise.

10.13 OC12 Data handling and processing error (At-sea observers) *Pertinence*:

Pertinent in all cases.

Contribution to bias and variability: As per the general considerations.

10.14 OC13 Adjustment error (At-sea observers) Pertinent

if adjustments are used.

Contribution to bias and variability: As per the general considerations.

10.15 OC14 Imputation error (At-sea observers) *Pertinence*:

Pertinent if imputation is used.

Contribution to bias and variability: As per the general considerations.

10.16 OC15 Modelling error (At-sea observers)

Pertinence: Unlikely to be pertinent if the monitoring program is a census at the trip AND at the event-level.

Possibly pertinent for a trip-level sampling survey and/or event-level sampling survey, where population-level estimates are derived using design-based estimation (estimating using the statistical formula appropriate for the assumed sampling design) or model-based estimation (estimation involving covariates or spatio-temporal modelling). Model-based estimation is frequently used because the true sampling design is not known or due to its perceived simplicity in some applications (Rochet and Trenkel, 2005). The appropriateness of the model should be examined. Consequences of inappropriateness of the model, if any, on bias and variability should be assessed.

Contribution to bias and variability: As per the general considerations.

11 ESTIMATIONS BASED ON CREEL SURVEYS

Creel surveys involve interviews with resource users in the field to obtain information on catch amounts, composition, and effort. Consequently, in this section, we use the term “interviewer” instead of “observer”. They are commonly employed for recreational fisheries to obtain the data required to estimate total catch. Sampling is often undertaken on the water body or at access points. Typically, the creel surveys are used to estimate the catch per unit effort (CPUE). The CPUE is then multiplied by an estimate of total effort to estimate the total catch.

11.1 Statistical characteristics (Creel surveys)

Pertinence: The statistical population often consists of individual harvester-days or trips.

In nearly all applications, creel surveys are sampling surveys (Pollock et al., 1994). In some small tightly managed systems (e.g., lakes in a park or reserve requiring mandatory reporting), creel surveys may be implemented as censuses.

The licence conditions may or may not require that these events be reported, depending on the fishery. Therefore, the population size is often unknown but large, in which cases the sampling ratio is often not known. In some cases, a separate monitoring program is used to obtain the sampling ratio (e.g., In the Pacific Coast recreational marine fishery aerial survey of effort).

The parameters estimated include the CPUE, a ratio, and the harvester effort.

11.2 OC01 Undercoverage (Creel surveys)

Pertinence: Undercoverage may be pertinent if the fishery takes place in locations not known or from access points not easily accessible to the sampling program (e.g., remote private fishing lodges and marinas), or if harvesters avoid making their activity known to the monitoring program. Under- (or overcoverage) may also be caused by the fisher not knowing and therefore not reporting their fishing location precisely.

Contribution to bias: For example, when the creel survey is used to estimate a CPUE, the CPUE may be higher in fishing areas unknown to the monitoring program due to less fishing pressure. Similarly, CPUE may be higher for professionally guided recreational fishing departing from remote fishing lodges. In these cases, the undercoverage will lead to a negative bias. On the contrary, if remote locations with below average catches per unit of effort are used for other reasons than maximizing catches, a positive bias would result.

Contribution to variability: None.

11.3 OC02 Overcoverage (Creel surveys)

Pertinence: Overcoverage is unlikely to be pertinent because the sampling is specifically tied to a fishery and carried out by interviewers.

11.4 **OC03 Unaccounted-for clustering of samples (Creel surveys)**

Pertinence: Creel surveys have a long history of employing clustered sampling designs due to the challenges of sampling numerous waterbodies or harvest access points (Pollock et al., 1994). Unaccounted-for clustering will occur if the estimates are not computed using the equations provided by cluster sampling theory.

Contribution to bias and variability: As per the general considerations.

11.5 **OC04 Unaccounted-for stratification of samples (Creel surveys)**

Pertinence: Unaccounted-for stratification will occur if, for example, interviewers are deployed evenly through a season (temporal stratification).

Contribution to bias and variability: As per the general considerations.

11.6 **OC05 Other irregular sampling probabilities (Creel surveys)**

11.6.1 OC05A Other irregular sampling probabilities: Unanticipated exclusions (Creel surveys)

Pertinence: Systematic shortages of interviewers resulting from gluts of activities or systematic interviewer absences (e.g., for vacation), can lead to unanticipated exclusions.

Contribution to bias and variability: As per the general considerations.

11.6.2 OC05B Other irregular sampling probabilities: Forced inclusions (Creel surveys)

Pertinence: Forced inclusions are not expected in creel surveys based on interviews to obtain information on catches and fishing effort, unless the interviewers seek out specific fishers. Creel surveys involving interviewers tasked also with an enforcement role may involve forced inclusions.

Contribution to bias and variability: As per the general considerations.

11.6.3 OC05C Other (or all) irregular selection probabilities (Creel surveys)

Pertinence: Irregular selection probabilities are possible, for example if interviewers choose their monitoring sites based on perceived fishing success, receptivity of resource users or ease of access to the sites while the sampling protocol does not call for such a selection process.

If the protocol specifies geographical and/or temporal cluster sampling, and cluster selection is not programmed, interviewers may be more likely to sample certain locations than otherwise planned, such as those that are more accessible, or differentially sample at more desirable times of the day, week, or season.

Contribution to bias: If sites are chosen based on factors that could reasonably be associated with, for example, CPUE, a bias could result, e.g., selection based on fishing success or remoteness (which can affect fishing pressure and therefore CPUE). Otherwise, no bias is expected.

Contribution to variability: Irregular selection probabilities impact to variability may be assessed by simulations of the sampling protocol and the actual realized scheme.

11.7 OC06 Observer effect (Creel surveys)

Interviewer effects are possible in creel surveys if harvesters alter their fishing practices, presumably to ensure compliance with regulations, when interviewers are known to be present (e.g., fishers informed before returning to wharf that an interviewer is present). For example, harvesters may release fish that would otherwise have been illegally retained or may reduce overall effort.

Contribution to bias: For a catch estimate, a negative contribution to bias should be anticipated. Quantifying a contribution to bias requires identifying the nature of the interviewer effect and its extent. *Contribution to variability*: Not expected.

11.8 OC07 Missing values due to unintentional factors (Creel surveys)

Pertinence: The frequency of missing values – harvesters that were sampled but for which there are no reported values – should be very small or nil since creel surveys are carried out by independent interviewers and there are typically procedures in place to avoid such instances.

Contribution to bias: Impact unlikely since the unintentional factors (e.g., lost records) are expected to be independent from the value observed (e.g., catch).

Contribution to variability: Impact likely inconsequential given the small frequency of occurrence.

11.9 OC08 Missing values due to intentional factors (Creel surveys)

Pertinence: Only pertinent if coercion or collusion is reported or suspected; these effects would only be relevant if interviewers have an enforcement role.

Contribution to bias: missing values attributed to coercion or collusion should be assessed if it is believed to be pertinent.

Contribution to variability: For sampling surveys, the impact will be reflected in the standard error through the reduction of the sample size.

11.10 OC09 Errors in data reported by resource users (Creel surveys)

Note: In creel survey, errors in data reported by resource users to interviewers can be assigned to either (but not to both) OC09 or OC10 as long as the source of the errors is described accurately.

Pertinence: Pertinent if the monitoring is based on a verbal account by resource users as opposed to a direct observation by the interviewer. Potential resource users' errors include the following:

- Underreporting their catch and/or effort. Not pertinent if retained catch is directly observed by the interviewer and effort is measured on a coarse scale.
- Misidentification of the species.
- Misidentification of the fishing area (e.g., special stratum).
- Incorrect reporting of the effort (e.g., beginning and ending of the fishing period).

Creel surveys occur during or soon after fishing, consequently unintentional errors resulting from recall should be infrequent and small. On the other hand, intentional misreporting is possible.

Contribution to bias: Intentional errors are likely to create bias. Some non-intentional errors can create bias: for example, less informed fishers may call any flatfish halibut, contributing positively to the bias for halibut catch.

Contribution to variability: Impact likely to be inconsequential and mostly accounted-for in the computation of the standard error.

11.11 OC10 Errors in data reported by independent interviewers (Creel surveys)

Note: In creel survey, errors in data reported by resource users to interviewers can be assigned to either (not to both) OC09 or OC10 as long as the errors are described accurately.

Pertinence: Unintentional errors by independent interviewers are likely, although the magnitude may not be large. Intentional errors, contributing to bias should be assessed if coercion or collusion is frequently reported or suspected.

Contribution to bias: Only intentional errors are likely to create bias.

Contribution to variability: Impact likely to be inconsequential.

11.12 OC11 Equipment error (Creel surveys)

Pertinence: Creel surveys reporting counts by interviewers or based on harvester reports are likely to be very precise if catch amounts, in numbers, are small. Creel surveys reporting weights are likely to be less precise. (Intentional misreporting by harvesters covered in OC09).

Contribution to bias: Unlikely.

Contribution to variability: Impact likely to be small or inconsequential.

11.13 OC12 Data handling and processing error (Creel surveys) *Pertinence*:

Pertinent. See general guidelines.

11.14 OC13 Adjustment error (Creel surveys)

Pertinence: Likely not pertinent, as whole fish are reported.

11.15 OC14 Imputation error (Creel surveys) *Pertinence*:

Pertinent if imputation is used.

11.16 OC15 Modelling error (Creel surveys)

Pertinence: Likely to be pertinent if catch is estimated using, for example, a ratio estimator without correcting for estimator bias.

Contribution to bias: Ratio estimators such as these can be biased, for example if the true intercept is not zero or if catch does not scale proportionately with effort.

Contribution to variability: Variability in the estimates resulting from the model can be estimated based on the error of the model parameters.

12 ESTIMATIONS BASED ON POST-SEASON QUESTIONNAIRES

Post-season questionnaires are administered to resource users to obtain information on catch amounts, composition, and effort. They can be administered orally, in person or by telephone, on paper forms or electronically. They are commonly employed for recreational fisheries to obtain the data required to estimate total catch. They are used in some subsistence fisheries and marine mammal hunts. Finally, they are also used in some commercial fisheries, in particular to get information on fishing effort in the absence of logbooks. Often questionnaires are used to estimate the catch per unit effort (CPUE) which is then multiplied by an estimate of effort to estimate the total catch. Participation in questionnaires is typically voluntary (not a condition of licence).

12.1 Statistical characteristics (Questionnaires)

Pertinence: The statistical units are typically individual harvesters.

In nearly all applications, questionnaires are implemented as sampling surveys. In many applications the list of licence holders constitutes the frame. The parameters estimated often include the CPUE but could alternatively be total catch and effort.

12.2 OC01 Undercoverage (Questionnaires)

Pertinence: Undercoverage may be pertinent if there are harvesters that can legally fish without a licence, e.g., youth accompanying a licence holder. The contribution of undercoverage to estimation bias will follow the general considerations.

12.3 OC02 Overcoverage (Questionnaires)

Pertinence: Overcoverage could occur if the questionnaire is specific to a particular water body or species and, year after year, there are respondents who inadvertently provide answers that are relevant to another water body or species.

12.4 OC03 Unaccounted-for clustering of samples (Questionnaires)

Pertinence: In principle, it should be possible to administer a questionnaire according to a random sampling protocol. However, sampling may be clustered, for example if the questionnaire is distributed to harvester associations that in turn distribute questionnaires to their members. The associations form clusters. Failure to account for this in the analysis would result in unaccounted for clustering.

Contribution to bias and variability: As per the general considerations.

12.5 OC04 Unaccounted-for stratification of samples (Questionnaires)

Pertinence: A stratified sampling design may be used (e.g., with strata based on location and harvester age) and, if so, accounted-for in the computations. Unaccounted-for stratification is not anticipated for monitoring by questionnaires. *Contribution to bias and variability*: As per the general considerations.

12.6 **Other irregular sampling probabilities (Questionnaires)**

12.6.1 OC05A Other irregular sampling probabilities: Unanticipated exclusions (Questionnaires)

Pertinence: Unanticipated exclusions can occur if some participants systematically cannot be contacted (e.g., seasonal residents who have left their temporary residence) or if some respondents are unable to respond (e.g., respondents with limited English literacy skills for a written questionnaire; technologically-challenged participants for a web-based application). If questionnaires are distributed through harvester associations, an association not forwarding the questionnaires would cause unanticipated exclusions.

Contribution to bias and variability: As per the general considerations.

12.6.2 OC05B Other irregular sampling probabilities: Forced inclusions (Questionnaires)

Pertinence: Forced inclusions could be pertinent if some respondents are targeted, for example because they have been reliable respondents in the past or because they are known to be particularly active in the fishery.

Contribution to bias and variability: As per the general considerations.

12.6.3 OC05C Other (or all) irregular selection probabilities (Questionnaires) Pertinence:

No guidance specific to monitoring by questionnaires.

12.7 **OC06 Observer effect (Questionnaires)**

Not pertinent since the fishing activities take place before the reporting.

12.8 **OC07 Missing values due to unintentional factors (Questionnaires)**

Pertinence: Participation in questionnaires is typically voluntary. In such cases, missing values due to unintentional factors would describe unintentional non-participation, an unlikely systematic situation.

12.9 **OC08 Missing values due to intentional factors (Questionnaires)**

Pertinence: Participation in questionnaires is typically voluntary. In submitted responses, leaving information out on purpose creates intentional missing values.

Contribution to bias: If the response refusal is unrelated to the catch-effort characteristics (e.g., due simply to respondent irritability) no contribution to bias is expected; otherwise, a contribution to bias is possible.

Contribution to variability: As these are sampling surveys, included in the computation of the standard error.

12.10 **OC09 Errors in data reported by resource users (Questionnaires)**

Pertinence: User responses may contain errors resulting from inaccurate recall or from an intent to deliberately mislead (e.g., overstate catches in the hopes of increasing quotas the next season; hide illegal catches).

Contribution to bias: Recall biased toward unusual events such as large catches and deliberate misreporting will contribute to bias.

Contribution to variability: Impact likely to be inconsequential or mostly accounted-for in the computation of the standard error.

12.11 OC10 Errors in data reported by independent observers (Questionnaires)

Pertinence: Not pertinent

12.12 OC11 Equipment error (Questionnaires)

Pertinence: In many applications of questionnaires, catch is counted in numbers and is likely to be very precise particularly if catch amounts are small.

Contribution to bias: Unlikely.

Contribution to variability: Impact likely to be small or inconsequential.

12.13 OC12 Data handling and processing error (Questionnaires) *Pertinence*:

Pertinent.

Contribution to bias and variability: As per the general considerations.

12.14 OC13 Adjustment error (Questionnaires)

Pertinence: Unlikely to be pertinent, as counts of whole fish are usually reported.

12.15 OC14 Imputation error (Questionnaires) *Pertinence*:

Pertinent if imputation is used.

Contribution to bias and variability: As per the general considerations.

12.16 OC15 Modelling error (Questionnaires) *Pertinence*:

Pertinent if modelling is used.

Contribution to bias and variability: As per the general considerations.

13 ABSENCE OF A MONITORING PROGRAM

This chapter covers estimations in the absence of monitoring programs. In some cases, the unmonitored fishery will be the only fishery intercepting a fish population. In many cases, an unmonitored fishery will be one of several fisheries interacting with the fish population.

Given that the goal of the Fishery Monitoring Policy is fishery sustainability, the impact of the absence of monitoring of a catch component on the quality of the estimation of total catch should be assessed if the component is anticipated to have a non-negligible contribution to the total catch and similarly for the collective impact of multiple nonmonitored components. Furthermore, doing so helps ensure transparency and completeness of the QA process.

When estimating total catch or effort, fisheries that are judged to have a very small contribution to the total can be acknowledged by a cursory QA or a place-holder QA as described in §3.

We present an approach for such an assessment. This approach has two main goals, expressed here for the case of an estimation of total catch. First, if fishery management ignores the fishery in computing a total catch, an estimate of the negative bias will be derived in the assessment. Second, the uncertainty due to the lack of monitoring will be translated into an estimate of the variability of the estimation process.

The approach is the following. The estimate used by fishery management is considered to be an imputation. The quality of the estimation process is the quality of this imputation.

13.1 **Statistical information (Absence of monitoring)**

Since the estimate used is a single observation, consider that the estimate is obtained from a census of a population of size 1. The typical parameter value is the typical value used in decision making. If the fishery is ignored in the decision making, the typical parameter value is 0.

13.2 **OC14 Imputation error (Absence of monitoring)**

The imputation error will apply to 100% of the size-one population.

Using available sources of information, including expert judgements, similar situations, etc., the best estimate of the true value and the smallest and largest plausible true values of the parameter should be determined.

The contribution of the imputation error to the absolute bias will be:

$$[typical\ value\ in\ reporting\ ] - [best\ estimate\ of\ the\ true\ value].$$

The contribution of the imputation error to the absolute variability will be approximately:

$$\frac{[largest\ plausible\ true\ value] - [smallest\ plausible\ true\ value]}{4}.$$

The last expression corresponds to the fact that, for normally distributed estimators, the length of the confidence interval is approximately 4 times the standard error.

The measures of quality are the bias and variability relative to the anticipated true value.

13.3 **Examples (Absence of monitoring)**

The Quality assessment of an estimation process in absence of a monitoring program is an important but special situation. Consequently, we give some illustrations.

Suppose that the total catch, in tons, by an unmonitored subsistence fishery is estimated. The following table shows four examples of data entry with the corresponding quality assessment for an estimation application.

Table 1 Examples of assessment of the absence of monitoring.

EXAMPLE	A	B	C	D
INPUTS				
Typical value used in decision making, from historical documents	400	400	0	600
Best estimate of the true value, from expert knowledge	400	400	400	200

Smallest plausible true values, from expert knowledge	300	0	100	50
Largest plausible true values, from expert knowledge	500	800	700	300
QUALITY ASSESSMENT OF THE ESTIMATION PROCESS				
Absolute bias	0	0	-400	400
Absolute variability	50	200	150	62.5
Relative bias	0%	0%	-100%	200%
Relative variability	12.5%	50.0%	37.5%	31.3%

In Example A, the managers have been using the best estimate of the total catch obtained from well supported expert knowledge, i.e., 400 t with a plausible range from 300 t to 500 t. The estimation process is considered unbiased and the relative variability is a small 0.25 × (500 - 300)/400 = 12.5%.

Example B is similar, but the expert knowledge is much less solid: the plausible range for the total catch is from 0 t to 800 t. While the estimation process is again unbiased, the relative variability is 0.25 × (800 - 0)/400 = 50.0%.

In Example C, the managers use a total catch of 0 t, ignoring the subsistence catch, based on the absence of monitoring. Expert knowledge is somewhat better than in Example B. The relative bias is now (0 - 400)/400 = -100% while the relative variability is 0.25 × (700 - 100)/400 = 37.5%.

In Example D, the managers use a total catch of 600 t, the value established in a 20-yearold survey. However, current expert knowledge suggest that this fishery has declined, with a likely catch of 200 t, and a plausible range from 50 t to 300 t. The relative bias of the estimation process is (600 - 200)/200 = +200% and the relative variability is 0.25 × (300 - 50)/200 = 31.25%.

14 DISCUSSION

Analysts need to keep in mind that uncertainty around conservation catch limits used by DFO is either not considered at all or, when considered, does not account for all sources of uncertainty. Likewise, the thresholds used for assessing dependability of measurements with respect to bias and variability, and for limits (DFO 2020) were carefully set in a peer review process but still reflect some arbitrariness. While limits and thresholds are necessary administrative tools, dependability should be considered broadly in terms of the classes defined by the thresholds as opposed to a strict adherence to these thresholds. The analysts should therefore avoid celebrating when the quality falls just over the threshold or lamenting when it falls just under.

The quality assessment process will almost certainly identify fish populations for which catch monitoring is inadequate. Following the principles above, the analysts should search for the largest contributors to the inadequacies. If such a fish population is intercepted by several fisheries, the analysts should identify the fisheries contributing the most to the bias and the variability of the estimation process and focus on these fisheries. For each fishery identified, the OCs having an important contribution to the bias and the variability of the estimation process should be identified.

When dependability is found to be insufficient given the conservation risk for a fish population, it is likely that the above principles will lead to addressing a small number of important deficiencies. Selecting the most important deficiencies will prevent attempting solutions based on tweaking monitoring programs. Instead, the importance of these deficiencies should motivate decision makers to look for broad changes to the selected monitoring programs or their implementation, to change the type of monitoring program or even to create new monitoring types of programs. The principles laid out in this report and in Beauchamp et al. (2019) can help inform these changes.

15 ACKNOWLEDGEMENTS

We wish to acknowledge the important insights brought by several QAs carried out since the framework was adopted and we thank their authors. This report benefited considerably from discussions with and reviews by Heather Bowlby, Nick Duprey, Tom Fowler, Rob Houtman, Rob Tadey and Margaret Treble. We thank them for their well considered comments and ideas.

16 LITERATURE CITED

Allard, J., and Benoît, H.P. 2019. Unified framework for the statistical assessment of fishery monitoring programs. DFO Can. Sci. Advis. Sec. Res. Doc. 2019/013. vi + 61 p.

Allard, J., and Chouinard, G.A. 1997. A strategy to detect fish discarding by combining onboard and onshore sampling. Can. J. Fish. Aquat. Sci. 54: 2955-2963.

Allen, M., Kilpatrick, D., Armstrong, M., Briggs, R., Course, G., and Pérez, N. 2002. Multistage cluster sampling design and optimal sample sizes for estimation of fish discards from commercial trawlers. Fish. Res. 55: 11-24.

Babcock, E.A., Pikitch, E.K., and Hudson, C.G. 2003. How much observer coverage is enough to adequately estimate bycatch? Oceana, Washington, D.C. (accessed October 13, 2020).

Beauchamp, B., Benoît, H., and Duprey, N. 2019. Review of catch monitoring tools used in Canadian fisheries. DFO Can. Sci. Advis. Sec. Res. Doc. 2019/010. vi + 49 p.

Benoît, H.P., and Allard, J. 2020. The dependability of fishery monitoring programs: Harmonising the quality of estimates with the risks to the conservation of aquatic populations. DFO Can. Sci. Advis. Sec. Res. Doc. 2020/15. vi + 54 p.

Benoît, H.P. and Allard, J. 2009. Can the data from at-sea observer surveys be used to make general inferences about catch composition and discards? Can. J. Fish. Aquat. Sci. 66:2025-2039.

Bremner, G., Johnstone, P., Bateson, T., and Clarke, P. 2009. Unreported bycatch in the New Zealand West Coast South Island hoki fishery. Mar. Policy, 33: 504–512.

DFO. 2019a. Framework for the qualitative assessment of the dependability of catch data from existing fisheries monitoring tools. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2019/004

DFO. 2019b. Fishery monitoring policy. https://www.dfo-mpo.gc.ca/reports-rapports/regs/sffcpd/fishery-monitoring-surveillance-des-peches-eng.htm#toc6-1. Retrieved on 2020/10/10

DFO. 2020. Assessment of the dependability of fishery catch monitoring programs. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2020/022.

Dias, L. C., Morton, A., and Quigley, J. (Eds.) 2018. Elicitation: The Science and Art of Structuring Judgement. Springer.

Faunce, C.H., and Barbeaux, S.J. 2011. The frequency and quantity of Alaskan groundfish catcher-vessel landings made with and without an observer. ICES J. Mar. Sci. 68: 1757-1763.

Galton, F. 1907. Vox Populi . Nature 75, 450–451. https://doi.org/10.1038/075450a0 Retrieved on 2020/10/29.

Hemming, V., Walshe, T.V., Hanea A.M., Fidler F., and Burgman M.A. 2018. Eliciting improved quantitative judgements using the IDEA protocol: A case study in natural resource management. PLoS ONE 13(6): e0198468. doi.org/10.1371/journal.pone.0198468

ICES. 2013. Report of the ICES Working Group on Recreational Fisheries Surveys 2013 (WGRFS), 22-26 April 2013, Esporles, Spain. ICES CM 2013/ACOM:23. 49 pp.

Lohr, S. 2009. Sampling: Design and Analysis, 2nd edition, Duxbury Press, 600 pp. ISBN10: 0495105279.

Metuzals, K.I., Wernerheim, C.M., Haedrich, R.L., Copes, P., and Murrin, A. 2005. Data fouling in marine fisheries: findings and a model for Newfoundland. In: Sumaila UR, Marsden AD, editors. North American Association of Fisheries Economists Forum proceedings, Fisheries Centre Reports 14–1, 2005. Fisheries Centre, The University of British Columbia, Vancouver, Canada. p. 87–104.

Pollock, K.H., Jones, C.M., and Brown, T.L. 1994. Angler survey methods and their applications in fisheries management. Am. Fish. Soc. Spec. Publ. 25, 371 pp.

Rijnsdorp, A.D., Daan, N., Dekker, W., Poos, J.J., and van Densen, W.L.T. 2007. Sustainable use of flatfish resources: addressing the credibility crisis in mixed fisheries management. J. Sea Res. 57: 114–125.

Rochet, M.-J., and Trenkel, V.M. 2005. Factors for the variability of discards: assumptions and field evidence. Can. J. Fish. Aquat. Sci. 62: 224-235.

Stanley, R.D. 1992. Bootstrap calculation of catch-per-unit-effort variance from trawl logbooks: do fisheries generate enough observations for stock assessments. N. Am. J Fish. Manage. 12: 19–27.

Stanley, R.D., McElderry, H., Mawani, T., and Koolman, J. 2011. The advantages of an audit over a census approach to the review of video imagery in fishery monitoring. ICES J. Mar. Sci 68: 1621-1627.

Stanley, R.D., Karim, T., Koolman, J., and McElderry, H. 2015. Design and implementation of electronic monitoring in the British Columbia groundfish hook and line fishery: a retrospective view of the ingredients of success. ICES J. Mar. Sci. Science 72: 12301236.
Wallis, K. F. 2014. Revisiting Francis Galton's Forecasting Competition. Statist. Sci. 29, no. 3, 420--424. doi:10.1214/14-STS468.

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