Ministry for Primary Industries
Manatū Ahu Matua

A review of current international approaches to standardisation and calibration in trawl survey time series

New Zealand Fisheries Assessment Report 2015/46

N.W. Bagley
P.L. Horn
R.J. Hurst
E. Jones
S.J. Parker
P.J. Starr

ISSN 1179-5352 (online)
ISBN 978-1-77665-017-0 (online)

August 2015

New Zealand Government

Growing and Protecting New Zealand

Requests for further copies should be directed to:

Publications Logistics Officer
Ministry for Primary Industries
PO Box 2526
WELLINGTON 6140

Email: brand@mpi.govt.nz
Telephone: 0800 00 83 33
Facsimile: 04-894 0300

This publication is also available on the Ministry for Primary Industries websites at:
http://www.mpi.govt.nz/news-resources/publications.aspx
http://fs.fish.govt.nz go to Document library/Research reports

Table of Contents

EXECUTIVE SUMMARY

Bagley, N.W.; Horn, P.L.; Hurst, R.J.; Jones, E.; Parker, S.J.; Starr, P.J. (2015). A review of current international approaches to standardisation and calibration in trawl survey time series.

***New Zealand Fisheries Assessment Report 2015/46.* 54 p.**

This report describes international practice with regards to standardising trawl survey methodologies and calibration under changing vessel or gear scenarios. Literature was reviewed in an attempt to determine international best practice. A fishing unit (i.e., vessel with gear) is standardised when it adheres to clear, comprehensive and unambiguous instructions on how the vessel will operate, and how the gear will be constructed and how it will be fished. Two fishing units are calibrated when the ratio of their catchabilities is known, given that, individually, they both adhere to standardised procedures. Scientists from key areas (i.e., USA, Canada, and Europe) involved in carrying out surveys or conducting stock assessment analyses that use these survey data provided their experiences and views. In particular, perspectives were sought around the cost-effectiveness of standardisation and calibration associated with changing gears or vessels and any associated data quality or stock assessment issues.

Although standardisation is widely used to ensure that multi-vessel surveys are comparable, it is considered to be less effective than calibration, but is much cheaper and therefore more prevalent. The extent of standardisation can vary markedly among survey series, whether single or multi-vessel. It is also apparent that 'standardised' vessels can exhibit similar or markedly different catchabilities. Key elements to standardise are vessel characteristics, trawl gear from the winches to the codend, and gear deployment procedures.

Vessel calibration experiments are expensive and time consuming, and must be carefully planned and executed if they are to produce useful and precise correction factors. Carrying out simultaneous paired tows is the preferred fishing method. Several approaches to conduct these experiments have been found to produce satisfactory results. Ideally, planning of the experiment should include simulation studies using existing survey data.

Interviews with regional scientists provided insights into the rationale behind choosing survey platforms and the selected designs, as well as to calibration decisions. The approaches adopted in each region were largely a result of the historical experience with surveys, the availability of survey vessels, and the capacity to conduct robust calibration experiments. All regions considered standardisation to be critical, but some adopted much more involved standardisation procedures and actively conducted research to further understand factors influencing catchability. Others did not have the resources to standardise to the same degree. Calibration experiments were less prevalent due to cost, and where they had been carried out, the results were sometimes too variable to be used, but almost universally considered best practice in theory when using multiple units (i.e., vessels or gear) or changing from one standard unit to another. Failure to calibrate adequately is likely to result in time series being split and the need for modelling approaches to estimate ratios between fishing units.

1. INTRODUCTION

The Ministry for Primary Industries has an ongoing need for high quality data on stock abundance for the management of New Zealand fisheries. During the last 20 years, time series of trawl surveys have been developed for several inshore (usually out to a maximum of 400 m depth), middle depth (usually 200–1000 m depth), and deepwater (about 600–1500 m depth) fisheries to help estimate stock size by providing a time series of relative biomass estimates. Trawl surveys provide fishery-independent indices of stock abundance. Some inter-annual variability is incorporated into the stock assessment model by using a process error (around 20%), but a primary assumption is that individual fish in that stock have the same average probability of being caught from one survey to another. This assumption will be open to question if there is significant alteration of the survey vessel, vessel equipment, trawl gear, or deployment procedures.

In New Zealand, standardised trawl survey time series have been developed using specialised research trawlers (R.V. *Tangaroa* and *Kaharoa*) using standardised survey methodologies to minimise variability in catchability associated with vessel or gear performance and to maintain a reliable time series of abundance estimates for key species (Hurst et al. 1992, McMillan 1996, Stevenson & Hanchet 1999). Obtaining an acceptable level of precision has also been an important objective for key species and a two-phase stratified random survey approach (Francis 1984) has been used to enable this to be achieved more cost-effectively. All surveys have collected biomass data on all species caught and have maximised biological data collection opportunities on as many commercial and by-catch species as is practical for stock assessment use. There have also been a wide variety of other research objectives fulfilled by these research surveys (including data collected for trophic studies, collection of material for ageing and stock identification studies, collection of multibeam data for better habitat information, and incorporation of acoustic data collection to develop indices of mesopelagic abundance).

New Zealand middle depth surveys, unlike some of the longer-running international surveys, benefitted from starting at a time when there was reasonable gear deployment and mensuration technology available (e.g., warp tensioning systems, SCANMAR doorspread sensors). However, technology is always improving and the uptake of new systems to improve gear performance (e.g., bottom contact sensors) may be desirable and may require calibration with old systems or protocols. The development of new time series using multiple vessels (in inshore waters) has also recently been considered (Parker et al. 2009). Finally, the continuation of existing time series with alternate vessels will be an issue for the future, either when existing research vessels are decommissioned or if other vessel options are considered to be appropriate. These scenarios will likely require some level of calibration if a consistent time-series is to be maintained.

This report, funded by the New Zealand Ministry of Fisheries (now Ministry for Primary Industries) project DEE2010-01, reviews current international practices with regards to standardised trawl survey methodologies and to the methods used for calibration under changing vessel, gear, or procedural scenarios. There were two specific objectives under this project:

1. To review trawl survey methodologies currently used worldwide to estimate the relative or absolute abundance of fish for use in stock assessment.
2. To document the use of calibration methodologies to standardise the abundance indices over time from different trawl systems and different vessels (including the use of commercial fishing vessels).

Under objective 1, international practice with respect to standardised trawl survey methodologies is reviewed with a focus on standardisation of vessels, trawl gear, and gear deployment procedures. It does not include a review of survey design or catch sampling procedures, and focuses primarily on stratified random bottom trawl surveys for finfish as these are the focus of most of the existing New Zealand trawl surveys.

Changes to the survey vessel, gear, or deployment procedures (e.g., tow speed, tow duration, and changes to the trawl or rigging) are expected to have some influence on catchability of some or all target species (Dunn 2006). Some of these changes can be desirable, in that they may improve survey performance or address sampling deficiencies (e.g., the ability to catch smaller sized fish of a species of interest, or improve bottom contact). Advances in knowledge of fish behaviour may also lead to changes in the standard gear rigging (e.g., to reduce escapement of fish under the groundrope). Gear trials and tests with survey trawls (i.e. flume tank model tests) may also lead to changes in the trawl configuration where problem areas or deficiencies with survey gear are identified.

Under objective 2, international practice with respect to calibration is reviewed, recognising that it is not always possible or cost-effective to maintain exactly the same vessel, gear construction etc. When such changes occur, calibration enables surveys to be input as a single time series into a stock assessment, and also allows for multi-vessel surveys (i.e., surveys where several different vessels are used in a single year, and/or different vessels are used in different years) with non-standardised vessel/gear combinations to be used.

Internationally, calibration between vessels and gears has been carried out, particularly in areas where trawl surveys are conducted by multiple nations and vessels, or where long time series (e.g., from the 1960s) have required updating vessels, trawl gear or deployment procedures. Some of these changes will have been in response to new gear monitoring technology that has suggested improvement in gear performance or gear deployment (Brown et al. 2007).

Methods used for calibration include comparative fishing trials, often coupled with modelling approaches to develop conversion factors. There are differing challenges for single vessel combinations (e.g., when replacing an existing vessel) and across multiple vessels carrying out a simultaneous survey. The effectiveness of these calibrations for stock assessment, and issues around cost and benefits, will be explored through the literature and also through other researcher contacts.

International practice was reviewed through a literature review and also by communication (i.e., questionnaires and follow-up interviews) with key researchers involved in the main trawl survey series in the USA, Canada, and Europe. This report includes consideration of:

- The influence that changing trawl vessels and various components of the trawl system may have on abundance indices;
- The important parts of the survey methodology to standardise or calibrate;
- The range of approaches used, specifically, what has worked, what has not worked, and lessons learned;
- The strengths and limitations of standardisation and calibration;
- The costs and benefits of standardisation and calibration (data quality versus cost, efficacy and utility and validation of results).

2. METHODS

To determine international best practice for bottom trawl survey methodologies for the two project objectives, we focused our review on Europe, the USA, and Canada because many of the methodologies are well established and constantly under review in these areas. Two methods were used to conduct the review.

Firstly, a literature review (published and grey literature, where available) of time series trawl survey standardisation and calibrations methodologies was conducted. Key documents for this review were the ICES Study Group on Survey Trawl Standardisation (SGSTS) Reports (ICES 2005b, 2006, 2009a) and the Report of the Workshop on Survey Design and Data Analysis (WKSAD) (ICES 2004, 2005a). These provide updated, state-of-the-art methodologies for standardisation of survey bottom trawls (including trawl gear design, preparation, maintenance, trawl performance monitoring), gear deployment protocols and calibration.

Secondly, we identified key researchers (and one skipper) involved in bottom trawl surveys in Europe, the USA, and Canada and requested that they complete a questionnaire and take part in a follow-up interview. The purpose of this was to gain potential insights into the reasons for implementing various methodologies, problems encountered, and lessons learned that may not be explicit in the literature, or not yet published, and the cost-effectiveness of options chosen. The questionnaire sent to all participants asked:

1. How many bottom trawl surveys currently occur in your region (with aim of demersal fin-fish biomass estimates)?
2. Do any of these use commercial vessels or are they government research vessels only?
3. Are there any commercial-only bottom trawl surveys (completely staffed and outfitted by industry)? How successful has this been?
4. For commercial vessels, what is the arrangement for standardising gear, gear performance, and sampling catch? How successful has this been?
5. Are any multi-vessel surveys? If yes, how are catch rates standardised among vessels? How successful has this been? What other challenges have occurred?
6. Has your region ever had to change vessels or gear during a time series? If yes, what was done to calibrate the catch rates of the old vessel to the catch rates of the new vessel? Are there any other issues to consider?
7. Are you planning any upcoming vessel or gear changes requiring calibration and what do you anticipate the process will entail?
8. Are there any other contacts you would recommend with experience with the above?

These interviews also created the opportunities to discuss the important factors that influenced the level of standardisation chosen, and the approach adopted for vessel and gear calibration in each region.

From the two activities, we have summarised our findings under the following categories:

- When is it necessary to standardise or calibrate;
- What aspects of trawl survey protocol are important to standardise or calibrate, and what methodologies are available to do that;
- What are the practical considerations (e.g., cost-benefit, feasibility, risks) of the available methodologies.

3. RESULTS

3.1 Standardisation methodologies

3.1.1 When should a standardisation be conducted

To ensure that a bottom trawl time series can provide fishery-independent indices of stock abundance, minimising variance in catchability between survey years is essential. Standardisation of a fishing system is clearly the best way to minimise this variance. In multi-vessel surveys, standardisation is essential, as well as some level of calibration. Increased emphasis has been placed on standardisation over the past 25 years and this is well documented (Stauffer 2004, ICES 2005b, 2006, 2007, 2009a, Anonymous 2007).

The development of comprehensive standardised procedures is essential before the start of a survey series. New Zealand bottom trawl standardisation methodologies for time series bottom trawl surveys by R.V. *Tangaroa* and *Kaharoa* have been documented (Hurst et al. 1992, McMillan 1996, Stevenson & Hanchet 1999) and are not repeated in detail here.

3.1.2 What should be standardised

Key aspects to standardise with respect to vessel characteristics, trawl gear, and gear deployment procedures are described below. Other aspects such as survey design, catch sampling procedures, preparation for sea, analysis of data, and personnel training, are also important but are not included here.

<u>Vessel(s)</u>
Standardisation of a single vessel across a time series relies on the survey protocols described at the start of the series. However, two aspects can change as time goes by. First, the vessel will age and the engine in particular may become less efficient. However, this is probably of little consequence as long as the trawl net can still be deployed, fished and retrieved according to specified standards. Second, the vessel is likely to undergo maintenance or even major refits. Again, this may be of little consequence unless the sound signature of the vessel is changed markedly (Mitson 1995), or power, propulsion, or system components that influence fishing equipment are altered (e.g., hydraulics).

Standardisation of vessels that participate in a multi-vessel survey series is likely to be more critical, particularly if no calibration experiments are conducted. Different vessels towing the same gear may achieve different gear geometry and/or different or similar catches per unit of effort (CPUE) (e.g., Azarovitz 1981, Byrne et al. 1991, IPROSTS 2001). Different vessels towing different gear can produce markedly different CPUE (e.g., von Szalay & Brown 2001, Tyson et al. 2006). If it is possible to use vessels that are somewhat similar in design and power then between-vessel differences in CPUE may be minimised.

<u>Trawl winches</u>
Modern trawl winch systems can be very complex, with development and design driven by commercial imperatives to maximise catch rates. Trawl winches are known to wear and become less efficient over time and require regular maintenance. Upgrades to software and winch computers are essential as components become unreliable and computers need replacing. Adopting any of these changes may have a bearing on comparability of time series surveys.

Changes to winch systems over time have included:

- A move away from the old style traditional “locked trawl warp” using the winch brakes to hold the gear.
- Self-tensioning winches (auto-trawl) using balanced warp tensions on each winch to adjust warp length. This system minimises effects of tide, windage on the vessel, and towing on a slope by keeping the warp tension the same on each winch by adjusting the wire length. Warp is also paid out when the vessel rises on a swell and retrieved as the ship drops back between the swells. Winches and winch computer settings need to be standardised (e.g., winch pressure settings, wire counters) to ensure consistency over time.
- Trawl-mounted monitoring systems that receive data from sensors on the trawl and make adjustments to the warp length to keep trawl geometry constant. These may use water flow and cross current information from net mounted sensors to make adjustments to the warp length automatically.

Experiments comparing the three winch types on the performance of a survey trawl are given by Kotwicki et al. (2006). Results from this work showed that auto-trawl systems were more effective in decreasing the effect of environmental factors on trawl performance. By incorporating an auto-trawl system into standard survey procedures the precision of relative abundance estimates may be improved. Recommendations from Kotwicki et al. (2006) included taking a cautious approach before changing a survey from locked winches to auto-trawl systems as calibration may be required.

<u>Vessel electronics</u>

Recent advances in electronics have resulted in the availability of many sensors that can be mounted on the vessel or trawl gear to monitor performance of the fishing gear. These mensuration devices are also capable of fixing the vessel position to great accuracy as well as estimating accurate vessel speed. Sensors now available include bottom contact sensors that allow better definition of the start and end times for each tow and groundrope contact with the seafloor, speed through the water sensors (compared with speed over the ground), and sensors that monitor trawl door performance (e.g., door spread, tilt, and angle). Therefore, it is possible to ‘tune’ the trawl gear or modify vessel operations using these data, and thus make adjustments that might improve CPUE relative to previous surveys in the series. Clearly it is desirable to ensure that the gear is fishing properly (by using some of the newly available technology), but it should be kept in mind that the capacity for improved monitoring of trawl performance on a tow-by-tow basis may lead to a mean catchability that is higher (or different) than when trawl behaviour was not so well monitored. Standardised procedures are required to ensure that clear decisions are made about how to use these types of new information and to ensure continued comparability with earlier survey years.

<u>Fishing gear</u>

Along with the new or improved technology come issues associated with the replacement of original components, including trawl gear, electronics, and software. Availability of replacement parts for standard trawls can become difficult as time series become longer and net manufacturing firms change their stocks. Changes in commercial fishing gear requirements may see some equipment which was previously considered to be standard become obsolete or no longer held in stock by suppliers. Trawl gear is constantly being changed and refined by the fishing industry as they strive for better catch rates and, importantly in recent years, enhanced fuel efficiency. Trawl gear used in time series trawl surveys typically comprise:

- a standard trawl design, including cod-end mesh size;
- floats (type, number and placement);
- a standardised ground rope configuration;
- standard trawl configuration, including sweep and bridle lengths, layback;
- standard trawl doors and configuration;
- standard trawl warp construction, e.g., steel with fibre core centres, dyform or swaged (compressed) wire rope, and lay (left or right hand);
- routine monitoring and maintenance of the gear during the survey.

Changing any of these may result in a change in trawl performance, consequently changing efficiency and potentially the catchability.

Deployment procedures

Key deployment procedures to standardise to ensure a consistent approach across years include:

- towing speed;
- warp-to-depth ratios;
- monitoring of gear geometry and the use of accepted values for various measures;
- station occupation (e.g., random station position to be the mid-point of the tow, tow along the contour keeping the depth as constant as possible);
- criteria for determining invalid tows (e.g., due to foul ground, large catches, trawl performance);
- trawling hours (e.g., daylight only);
- setting and haulback procedures.

Technological advances (e.g., trawl mounted sensors) have improved the way in which some deployment procedures can be monitored. One example of this is towing speed. Variability of towing speed would be expected to have some effect on catch rates, and these may differ among species and among fish length categories within a species. Trawl speed is typically recorded as speed over the ground in lieu of recordings of the actual net speed through the water, although the important factor may vary by species (i.e., speed through the water may be relevant for a pelagic species, but speed over ground may be more relevant for a bottom-tending species). Improvements in GPS position fixing since the late-1980s have led to more accurate measurements of speed over the ground. Sensors capable of measuring speed through the water are now available that use data from measured current flows across a trawl net.

Vessel crew

The ICES (2009a) report emphasises the importance of human factors in achieving a standard approach to trawl survey time series. The “skipper effect” is acknowledged as a factor affecting fish catchability, but it is rarely discussed in detail and is difficult to estimate (Mahévas et al. 2011). It can result in variation in how a trawl path is traversed within the habitat at a station, how the gear is set on the bottom, and how the gear is lifted off the bottom; all can influence catch rates. As well, the quality and methods used for net repair can cause potential changes to gear configuration between tows. Crew effects include science and vessel staff as well as others who are involved in preparing gear for a survey (e.g., a net shed completing trawl repairs, and the wire manufacturer for the supply of trawl warp, sweeps and bridles). Standardisation targets need to be achievable, and all personnel involved with a survey should be made responsible for ensuring that the aspects they are involved with conform to standards, with training and routine monitoring being seen as key components. Vessel crew and scientific staff are difficult to standardise, consequently rigorous deployment procedures are also required to go with the standards. In concept, if a crew or skipper effect exists, it is a signal that standardisation protocols are not sufficient or not followed. The resulting variance in catch rates due to this factor can exist within a survey through multiple crews during a survey, or between surveys if vessel staff change. If a calibration study is conducted, vessel crew effects are incorporated as part of the fishing units compared.

3.1.3 Practical considerations

Standards need to be easily understood by scientists designing the surveys as well as by survey and vessel staff who conduct them. Standards need to be clear and achievable and vessel crew should be involved in their development.

Standardising a single research vessel and gear used consistently in a trawl survey time series is relatively straightforward, has few risks, and results in little added expense. It is common to define, at the start of a survey series, how the vessel will operate, and how the trawl gear will be constructed, rigged, and fished. Once the protocols have been set, it is a matter of rigorously following them in all future surveys and identifying when small changes have occurred that may influence performance.

While risks when standardising a single vessel are relatively minor, the standards need to be thought through in advance. For instance, it is possible that the material used in the construction of the original fishing system may become unavailable (e.g., trawl mesh). This can be somewhat mitigated by ensuring that the research provider holds stocks of spare material. Alternatively, but less desirable, a substitute material with characteristics as similar as possible to the original could be found. It is also possible that newly developed technology introduced to the vessel could enable it to fish ‘better’ (i.e., have an improved mean catchability), even though fishing procedures are not modified. This undesirable outcome is achieved because new technology enables easier identification of tows where the gear is not fishing properly, whereas similar malfunctioning tows may have gone unidentified in earlier surveys (Zimmermann et al. 2003).

The procedure to standardise a single commercial vessel with the gear used in a trawl survey time series should be similar. However, over time, there is considerable incentive to introduce technology to the vessel that will enable it to fish more efficiently and little incentive to ensure that it operates and fishes in exactly the same way that it did when it first started the survey series. Most often, survey scientists would have no input into these decisions, so vessel choice criteria need to be very specific regarding required characteristics when initially selecting the vessel, as well as introducing a clause which asks for stability in key vessel characteristics. Another concern when taking on a commercial vessel in a long time series is that there is a greater likelihood that it will be on-sold or unavailable in the future than would be expected for a dedicated research vessel.

Standardising the performance of several vessels that work together in a multi-vessel survey is harder to achieve, particularly where vessels are managed by different institutes. Explicit and well-written protocols are required. Even then, variation can creep in over time. For example, Wileman (1984) pointed out that the range of doors used with the standard International Bottom

Trawl Survey (IBTS) trawl could be producing 13% less to 75% more spread than the ones recommended in the manual. Even when swept area can be standardised between vessels, differing catchabilities can result from variations in sweep angles. However, even if all vessels use the same gear which has been standardised in terms of configuration and operation, there remain numerous effects which are more difficult to standardise, and which can affect catchability. Such factors include vessel sound signature, vessel behaviour in different environmental conditions, differences in vessel power influencing speeds at shooting and hauling, and skipper effects. Individual vessels are likely to have different catchabilities, even with rigorous standardisation of gear and procedures. The differences will be even greater if different gear configurations or gear operations are used.

3.1.4 Absolute abundance surveys

The issues of standardisation and vessel effects are less problematic if absolute swept area abundance estimates are being made, as long as the same gear is used, in the same way and the same metrics are measured, and an estimate of catchability is available. A recent example of such a survey is the Scottish monkfish survey of the eastern Atlantic, which aims to produce a minimum estimate of total number and biomass of anglerfish (ICES 2009b). The survey area covered is large and has necessitated the use of multiple vessels; a combination of research and commercial. The trawl gear is manufactured under strict quality control system and provided to the commercial vessels, which are required to meet a relatively specific tender specifying length, horse power, door type etc. All vessels were provided with acoustic door and wingspread sensors as well as bottom contact sensors and a means to record the raw data from these sensors. Wing and doorspread values are interpolated to provide continuous values and, along with GPS, swept area is calculated, matched to bottom contact sensor data, and truncated to measurements made only when the trawl is on the bottom. Considerable effort has been devoted to estimating factors affecting the catchability of the monkfish including herding by the sweeps (Reid et al. 2007a) and losses under the ground gear (Reid et al. 2007b). Herding corrections are based on model estimates which use observations of behaviour from underwater video. The mean estimate of fish encountering the sweeps and bridles and being herded into the path of the net was 4%. Estimates of escapes underneath the groundgear are based on bagging experiments that suggested that whilst 70% of fish in the path of the net were captured, a clear length dependency existed, with 75% of fish below 30 cm being caught in the bags. In the case of this survey, the water depth and benthic, sedentary nature of the target species means that, unlike some species (e.g., Olsen et al. 1983), there is assumed to be a minimal effect of vessel noise on fish behaviour ahead of the trawl and consequently on different noise signatures between vessels. In terms of standardisation, possible issues remaining are the sea keeping abilities in poor weather conditions of smaller participating vessels, that may reduce time available for sampling and/or affect net geometry and catching efficiency. In addition, the particular fishing styles of different commercial skippers may be hard to truly standardise.

3.2 Calibration methodologies

3.2.1 When should a calibration be conducted

The ICES (2009a) SGSTS report suggests that calibration is required when alteration to trawl survey components may change catchability. Minor changes to the trawl gear may be expected to have little influence on catchability but, when major changes, or many minor changes are made, calibration experiments may be needed. Calibration of bottom trawl surveys requires the estimation of a conversion factor that allows the catch per unit effort found by one survey vessel and gear combination to be related to another, usually on a species specific basis.

Advice on calibration of trawling gears and vessels used for standardised fish surveys is given in ICES SGSTS and WKSAD reports (ICES 2004, 2005b, 2006, 2009a). It divides changes to survey trawls and trawling procedures into three key categories and recommends when calibrations should be considered.

a. <u>Minor improvements designed to allow better compliance with the standards agreed for the survey</u>.

The following changes to the survey standards are considered minor and, if they are the only changes made, should not require calibration. This is because the correction factor estimated for such a small change in protocol would be in the expected error range of the catchability. Minor improvements include:

The replacement of minor trawl components when they become unavailable.
The supply of trawl gear is primarily aimed at the commercial sector and components can be difficult to source over time. Trawl gear is constantly evolving with new designs and technology leading to more efficient nets, with recent efforts aimed at reducing fuel consumption. Obtaining components for older trawl nets becomes increasingly difficult, with some materials having to be made specifically for research trawls. New Zealand research survey experience has shown that a trend to reduce mesh size in the fore-parts of some trawls now means that suppliers no longer hold stocks of 12 inch mesh (as used in some standard research trawl nets). The continued availability of mesh from the original manufacturers is in doubt, and this issue has been recently highlighted at NIWA because the original mesh-making machines which were formerly located in Japan have been moved to Indonesia. Therefore, net maintenance needs to be done well in advance of surveys to ensure that there are no supply issues. For some of the more scarce materials, supplies should be held in stock.

The availability of new instrumentation to monitor trawl geometry and trawl performance, and thus ensure consistent trawl geometry.
There is an increasing number of sensors that can be mounted directly onto the trawl and trawl gear to monitor trawl geometry and performance. There may be a tendency to ‘tune’ the trawl gear using new information provided by new sensors, and thus deliberately make adjustments to improve net configuration. Fitting sensors to the trawl, doors or groundrope may alter some of the standard parameters (e.g., by adding weight to the headline of the trawl, or changing water flow across the trawl doors). The use of bottom contact sensors allow tow length to be more accurately determined, and show whether the trawl had consistent bottom contact, possibly leading to the rejection of tows which would have historically been used in analysis. Improved estimates of speed over the ground and position fixing have come with improvements to satellite position fixing systems (GPS). New speed through the water sensors allow the effect of current and tide to be minimised, thus keeping the net speed constant through the water. They can also be used in conjunction with winch electronics to adjust the trawl gear to maintain trawl geometry.

Improved protocols for procurement, construction and repair of nets.
Net shed personnel and vessel crew completing trawl maintenance need to be familiar with the standardised trawl net specifications and survey requirements. However, it is possible that, over

time, techniques could be developed to improve the rigging of the trawl gear (e.g., attachment of fishing line to the groundrope or mesh to framing ropes, or changing of mesh type in high wear areas).

b. Modest changes or departures from agreed standards whose effects are individually hard to estimate.

Modest changes include those that improve net performance, such as the use of sensors to determine tow length and trawl door performance. They also include changes in standard procedures, e.g., manipulating the geometry of the trawl to fit within a specified target by changing warp to depth ratios, or changing door attachment configuration to alter doorspreads. ICES (2009a) suggest that such modest changes could be saved up and assessed in a single calibration procedure.

c. Major changes that depart significantly from agreed standards for the survey.

Major changes are considered to be:

- deliberately altered standards, e.g., to allow improved net trawl doors or footgear to be used;
- replacement of standard equipment components that are no longer available, e.g., trawl doors, or changing the entire trawl net plan;
- changing or dropping standards that are too difficult or too expensive to apply or standards that are thought to be defective or unsuitable, e.g., when flume tank studies show that the net is not fishing effectively;
- changing a survey vessel.

For such large changes, ICES (2009a) recommends that a full calibration be carried out at the time of the change, although several changes could be saved up to be covered by one calibration of entire fishing systems, as for “modest changes”. The following additional points are suggested for careful consideration before investing in a calibration exercise (ICES 2009a).

- Decide if the proposed changes to the gear, vessel, or fishing technique lead to a prior expectation of large changes in catchabilities for key species (i.e., a bias in pre- and post-change catchabilities that is greater than the general level of precision in the abundance estimates estimated from the survey series).
- Decide whether different factors are likely to be important for different species and age groups.
- Decide on the required precision of the calibration factors (and hence the required time and resources). This can be carried out, for example, by simulating stock assessments that use the survey data, and by considering the effect of a) not adjusting the survey time series, b) adjusting the survey time series with conversion factors estimated with particular levels of precision. The required precision will depend on the assessment method, the other indices that are used to tune the assessment, and the potential that the resulting data may lead to uninformative or faulty management advice.

3.2.2 What should be calibrated

3.2.2.1 Vessel(s)

Calibration of survey vessels may be needed when there is a change of vessel due to vessel replacement (Pelletier 1998) or temporary unavailability of a vessel used in time series surveys. Vessel calibration is also recommended where multiple vessels are involved in a single simultaneous survey.

A survey vessel replacement is inevitable with a long running time series. Research vessels often have expected working lives of 30–40 years, while those for commercial vessels are probably slightly shorter. As vessels age, reliability can become a factor with more down time expected for repairs and maintenance, resulting in less availability and lengthening survey times. The “life” of a vessel may also be influenced by financial or operational considerations.

Vessels have distinct characteristics such as hull design, size, and power. Different vessels have their own unique sound signatures (Mitson 1995) and the effects of these on catchability vary with fish species and depth (Godø 1994, Ona et al. 2007, de Robertis et al. 2008, Wood 2011). Larger vessels are more stable and not so prone to extremes of pitching and rolling in rougher sea conditions. Differing vessel designs and hull shapes will also have different sea handling characteristics that may influence trawl performance. Different trawl winch systems (*see* Standardisation section) will also have an effect on trawl geometry and performance and these effects may vary with warp to depth ratios.

Findings from a workshop held in Canada in 1980 (Doubleday & Rivard 1981) concluded that a large proportion of the variability in survey catches could be attributed to variability in vessel performance. They proposed dedicated studies to address the serious problem of standardisation of results among different research vessels.

Pelletier (1998) reviewed the calibration of research survey vessels in fisheries and presented details of an experiment and analysis using two French research vessels, and recommended that calibration studies should be conducted before old vessels are replaced with new ones and in multi-vessel survey programmes when vessels are changed over time. Practical limitations to this need to be considered and include ensuring that the vessel being decommissioned could function in keeping with its original trawl survey operational procedures. Depending on the experimental design, difficulties in doing this could include the cost of keeping the old vessel going while outlaying a large expenditure for a new replacement vessel, the need for additional crew for any calibration experiments, and possibly extra standardised trawl gear costs to outfit two vessels.

3.2.2.2 Calibration options

Calibration experiments can provide individual correction factors between different vessels, gear combinations and deployment procedures. Different factors may be calculated for individual species, as well as for size classes within a species (and also for regions and seasons). The three possible calibration options for trawl surveys listed below are suggested by ICES (2009a).

a. <u>Doing nothing</u>

If calibration factors are estimated with poor precision the effect of change to gear or vessel could be masked. Any bias introduced by poorly estimated correction factors may be greater than the true difference (bias) between the two gears or vessels. Correcting relative fishing power differences can remove bias due to systematic error, but may increase the variance of the mean CPUE estimate (Munro 1998). Munro's work highlights the importance of simulation studies in planning calibration experiments.

Multi-vessel research surveys off the western coasts of Canada, the United States and Europe all put emphasis on a high level of standardisation in lieu of calibration experiments (see Appendix 1).

b. <u>Comparative fishing trials</u>

Comparative fishing, using pairs of trawl tows close in space and time, can be used for estimating a calibration factor for gear and vessel effects (ICES 2009a). Options include:

- towing side by side at the same time
- towing down the same tow path with a short interval between tows
- towing along parallel tows at an interval
- tows from a single vessel while changing gear
- twin trawl from a single vessel

Experiments with vessels fishing side by side are likely to give better homogeneity of species composition, while options towing down the same trawl track require a disturbance effect to be estimated (Lewy et al. 2004). Multi-vessel surveys can incorporate paired tows when two vessels are in close proximity during a survey or as part of specific calibration experiments, depending on the sample size required.

Calibrating a gear change on a single vessel may involve completing several trawls with the same net before changing nets to repeat comparable tows with the other gear (thus minimising the number of gear changes at sea). ICES (2009a) suggested twin trawling as an option using the old and new trawls as long as the geometry can be kept the same as if the nets were towed singly. However, the vessel would need to be set up to twin trawl, and twin trawls are typically smaller than single trawls, introducing a scale factor. This approach has been used to assess differences in catchability between two different ground gears used in the North Sea International Bottom Trawl Survey (IBTS) (see Appendix 1).

Comparative fishing trials with two vessels can be carried out on a special paired-vessel cruise or by one vessel shadowing another at selected stations during a usual survey (ICES 2009a). Difficulties and disadvantages of comparative fishing trials are:

- Organising for two, fully staffed vessels to be in the same place at about the same time is costly and operationally difficult to achieve, particularly if the vessels come from different countries.
- There is a high risk of failure due to break downs or poor weather.
- There is a risk of poor precision for the estimated factors unless a relatively large number of parallel trawls can be achieved and species of interest are consistently encountered.

The fish species of interest need to occur reliably in moderate or large numbers. Zero catches are problematic as they provide no information about a conversion factor, and low catches in either haul may also provide little useful information. These effects will vary by species. If low catch numbers occur frequently an adjustment of calibration factors by species or the length of fish may not be feasible (ICES 2009a). Calibration factors will relate to the conditions at the time of the trials and these can be confounded with vessel effects, e.g., habitat type, weather, and their effect on trawl catchability. ICES (2009a) cautions that a factor estimated under one set of conditions may not be comparable with another. Ideally, comparative fishing trials should be broadened to cover the range of conditions encountered during a survey but this will inevitably increase costs substantially as more multiplicative factors are added.

c. Modelling

Calibration factors can be estimated theoretically by modelling a fish population using available survey data to estimate catches expected if conditions did not change, then by estimating a factor to align the observed catches between the two gears (or vessels) with the expected (ICES 2009). This avoids costly calibration experiments but requires a confidence in model accuracy, precision of abundance estimates, and adequate inclusion of the factors affecting catch.

3.2.2.3 Practical considerations for calibration

Catch rates in trawl survey time series are the primary source of information for tracking fish stock abundance. If it is necessary to introduce new survey vessels or to change the fishing gear, then the continuity of the time series data will be compromised. Hence, it is necessary to compare the efficiency and the selectivity of the different gears (or vessels) and to estimate conversion factors that allow the catches taken by one fishing system to be expressed in units equivalent with another system. However, such standardisations are not straightforward, as indicated from the available literature on this topic.

a. Key review by Pelletier (1998)

Pelletier (1998) produced a review of vessel calibration studies up to the late 1990s and highlighted several points that needed to be considered when conducting this type of study, i.e., the choice of sampling design, methods of gear monitoring, and approaches to data analysis. Two sampling design approaches had been used to compare vessels. In the first approach, two vessels towed independently in small areas assumed to be homogeneous with respect to fish abundance and environmental conditions (e.g., Sissenwine & Bowmann 1978). In the second approach, designs based on paired hauls considerably reduced the consequences of spatiotemporal variabilities because the two vessels towed simultaneously at the same speed and as closely as possible to one another. International workshops generally concluded that the second approach was preferable (e.g., Anonymous 1992). Pelletier (1998) also concluded that, if possible, stations should be selected so as to maximise catch for all species of interest, because low and zero catches are less informative in estimating conversion coefficients. However, if the calibration experiment must be conducted at the same time as the regular survey, then station location will be dictated by the normal survey design.

Pelletier (1998) listed a wide variety of data analysis methods that had been applied to calibration data. Differences involved data transformations, treatment of zero catches, and the actual model used to fit the data and derive conversion factors. In many of the previous experiments, the empirical distributions of the CPUE data were severely skewed with a high proportion of zero values. Zero observations were either excluded from the analysis or were retained in the analysis by adding a small number to them (an approach which may not be statistically defensible). Also, effective sample sizes often did not allow estimation of age- or length-specific conversion coefficients, or conversion coefficients could only be estimated for groups of species. Pelletier (1998) concluded that the ideal statistical model to estimate conversion factors should allow for skewed catch distributions and explicitly capture the existence of zero catches. She went on to describe and use a quasi-likelihood model to calibrate two vessels fishing the same gear, with estimates of precision derived by bootstrapping. Regardless of the type of analysis to be conducted it is clear that the estimates of conversion coefficients will be more precise when the species is present in a large number of hauls, the species catch is consistent, and the species is uniformly distributed over the sampled area. Therefore, species that exhibit many zero or small catches are likely to yield poor estimates of conversion coefficients.

In contrast with previous calibration studies, Pelletier's (1998) estimation method was defined before the experiment took place and was used in computer simulations to plan the survey. Therefore, the sampling design and statistical model were fully consistent, and the data analysis did not require any ad hoc procedure, e.g., data transformation to handle null catch. Pelletier (1998) concluded that to maximise the probability of conclusive results, particular attention should be paid to survey planning and to controlling as many sources of variability as possible during the experiment. Preliminary work based on data simulations and prior information about the surveyed resource should be used to plan this type of experiment. Sampling in areas of high abundance minimises the risk of null catch and decreases the effective sample size.

Following the review by Pelletier (1998), experiments to produce calibrated correction coefficients have generally used paired hauls. However, there are two groups of experiments: those with paired hauls that are nearby and non-overlapping (generally side-by-side), and those where exactly the same trawl track is fished twice.

b. <u>Fishing side-by-side</u>

Von Szalay & Brown (2001) analysed catch per unit of effort data from side-by-side trawl tows to calibrate two different vessels using different gear. The work aimed to enable the integration of data from separate survey series conducted by the National Marine Fisheries Service (NMFS) and Alaska Department of Fish and Game (ADF&G), thereby providing biomass estimates based on larger sample sizes particularly for walleye pollock in bays where ADF&G sampling densities were high compared with NMFS. The two vessels were of a similar size and power, but the survey gear and the standard procedures were different, e.g., a towing speed of 3 compared to 2 knots, headline height was three times greater on the NMFS gear, the groundrope had roller gear on one vessel but not on the other. A total of 33 paired side-by-side trawls were conducted, with the two vessels about 0.5 km apart. Catches were sorted by species, weighed, and length-frequency measurements taken for four key species. They estimated the fishing power differences, by species, using the estimator of Kappenman (1992). Kappenman's estimator is based on the assumption that the unknown, but typically skewed, CPUE distributions for a given species are the same for both vessels, except possibly for the values of the scale parameters of the distributions. Wildebuer et al. (1998) recommend a minimum of 50 hauls be used to produce the Kappenman's estimator, but von Szalay & Brown (2001) concluded that their sample of 33 tows was sufficient because of the relatively narrow depth range and homogenous substrate of their experimental area. The NMFS vessel was more efficient at catching the two roundfish species, with fishing power correction factors of 3.84 (pollock) and 1.72 (cod), while the ADF&G vessel was more efficient for the remaining two flatfish species (0.73 and 0.75). Length-based fishing power correction factors, intended to fine-tune the species-specific correction factors to individual size categories, were not significantly different for three of the species. However, it was concluded that separate correction factors were necessary for small (4.01) and large (2.33) length classes of walleye pollock. It was suggested that this within-species difference in fishing power was a function of headline height (three times higher for the NMFS net) and the propensity of smaller pollock to be off the bottom and therefore more likely to be caught in the NMFS trawl.

González Troncoso & Paz (2003) analysed the results of a calibration trial off Spain between an old and a new research vessel. Fishing power correction factors and their confidence intervals were calculated for the target species using six analytical methods reported in the literature, i.e., ratio of mean CPUE, linear regression model, generalised linear regression model by haul, generalised linear regression model by stratum, Kappenman's ratio of scale parameters, and a length conversion method. In total, 90 acceptable paired hauls, each 30 min long were completed with the vessels remaining as close to each other as safety considerations permitted. Trawl design differed between vessels. The different analytical models produced a relatively narrow range of conversion factor point estimates (0.28–0.39), but the sizes of the 95% confidence intervals were quite variable across analysis methods. The most precise methods were the Kappenman model and the ratio of CPUE.

Only 24 parallel hauls were carried out for the calibration between the old and new research vessel *Scotia* in Scotland. As well as the vessel effect, the tow duration was reduced from 1 hour to 30 min. Catch rates-at-length of one vessel relative to the other were modelled using loess smoothers and nonparametric curves which were then combined to estimate the average relative catch rate (Zuur et al. 2001, Fryer et al. 2003). Fixed effects were modelled by allowing the mean selection parameters to vary between hauls according to the experimental design matrix. The hypothesis was that the relative catch rate of *Scotia* III to *Scotia* II should be equal to 0.5 for all length classes. For the most abundant species (haddock), the relationship between relative catch rate and length varied widely between paired hauls, but the relative catch rate of 0.5 was not strongly refuted up to a length of 35 cm. At size classes above this, there was marginal evidence that the newer vessel caught more fish (using a bootstrap hypothesis test). The study concluded that the width of the confidence bands indicated that the mean curve was not estimated with sufficient precision to provide acceptable conversion coefficients, probably due to the small sample size.

As part of a programme to develop Namibian capacity for undertaking its own trawl surveys, the hake survey historically carried out by Norwegian research vessels was replicated in full over two years by a commercial vessel working in parallel with identical gear and rigging (Strømme & Iilende 2001). In both years, biomass and distribution patterns were similar. An estimator was calculated to express the similarity of a pair of stations by taking the ratio between the deviation from their means and the mean, referred to as a similarity index. Where no difference between vessels exists the index should be zero, and a value of +1 or -1 indicates that the catch from the second and first vessel respectively was zero. On average, the absolute deviation of the similarity index from its mean was 0.20 in both years indicating a small systematic vessel effect with the commercial vessels catching slightly more. The authors pointed out that the estimator used in this study was less sensitive to outliers than regression or covariance techniques.

Oeberst & Grygiel (2004) reported on a study to calibrate the power of new international standard trawl nets (to be used in Baltic fish surveys) relative to nets previously used by different countries. The CPUE of paired trawl stations were compared, with a pair defined as having a separation distance less than 15' of latitude and longitude, with similar depth range (± 10 m), and within 15 days of each other. The authors believed that fish densities and length distributions from trawls paired in this way would be comparable, but no other reported studies have allowed the paired tows to be separated in time by such an extent.

Two vessels (*Wilfred Templeman* and *Alfred Needler*) have been used in many of the multi-species trawl surveys conducted by the Newfoundland Region of Fisheries and Oceans. These are 'sister' vessels, similar in construction, but with the *Alfred Needler* having a larger engine size (2600 hp, relative to 2000 hp). Cadigan et al. (2006) conducted paired-trawl experiments involving these two vessels using a Campelen 1800 shrimp trawl, a standard survey trawl used by NAFO. On level bottom, the two vessels towed side-by-side at an intended distance of 0.9 km apart. Both vessels followed normal standardised survey protocols. On slope edges, where side-by-side tows were not feasible due to depth differences, one vessel towed ahead of the other, alternating the lead vessel on a tow-by-tow basis with no overlap in the area covered by the two tows. The analysis of 106 paired trawls by Cadigan et al. (2006) included a generalised linear mixed effects model with an auto-correlated random effect that was used for estimating the ratio of catchability when there is substantial local variability in stock abundance fished by each vessel. Estimates were compared with those from a fixed effects approach involving standard logistic regression. The mixed effects model fitted the data better, producing estimates of relative efficiency that were not heavily influenced by a small number of outliers, and indicated that differences in catchability were small and not statistically significant. However, the logistic regression approach produced some small but statistically significant differences. Overall, the results suggested that there were no significant differences in catchability between the two vessels and gear. Total length frequencies were similar, especially when considering the between-tow variability in catches.

Tyson et al. (2006) calibrated five different research vessels that conducted trawl surveys in Lake Erie. Correction factors were estimated for the ten most commonly collected species and age-groups for each vessel using the between-vessel ratio of CPUE. Most of the correction factors ranged from 0.5 to 2.0; differences in CPUE were most evident for vessels using different sampling gears, although significant differences also existed for vessels using the same gears. Tyson et al. (2006) concluded that standardising gear is important for multiple-vessel surveys, and that there were significant differences in catchability stemming from vessel effects. In contrast, Helser et al. (2004) estimated random effects coefficients associated with several commercial vessels and the R.V. *Miller Freeman* that were involved in a series of multi-vessel surveys. They concluded that vessels could be assumed to be drawn from a common random normal distribution of effects, and that combining data from the chartered commercial vessels and the research trawler may be appropriate to develop indices of abundance for stock assessment purposes.

The Northeast Fisheries Science Center collected data from 636 standardised paired tows by R.V. *Albatross* and R.V. *Henry B Bigelow* to enable the calibration of these two vessels. The sampling strategy compared fishing by these vessels that used different trawl nets and different operating procedures (NEFSC 2007). The vessels made paired tows at locations selected as part of the standardised bottom trawl surveys. The *Albatross* completed a tow, then the *Bigelow* towed a parallel track from the same starting location offset by 500 m. Care was taken to ensure that the tow depths remained comparable between vessels. The paired tow data were analysed using a variety of approaches (Miller et al. 2010), and it was found that the ratio estimator and the beta-binomial maximum likelihood estimator both performed well. However, it was clear that the length distribution of some species differed between vessels as a consequence of different trawl selectivities, so size-dependent calibration factors were sometimes necessary. Miller et al. (2010) suggested that calibration factors should not be estimated if there are fewer than 30 sets of paired tows where both vessels caught some fish of a given species due to the resulting estimates being unreliable. However, they recommended that whatever the estimator, it should be used with caution and also noted that despite large sample sizes, conversion factors could not be developed for all species due to low frequency of occurrence.

Comparative trials between an old (R.V. *Gadus Atlantica*) and new (R.V. *Teleost*) research vessel were conducted to maintain continuity of a NAFO groundfish trawl survey series (Warren 1997). Changes were also made in trawl design (the Engel 145 High Lift otter trawl was replaced by a Campelen 1800 shrimp trawl, directed at catching more small fish) and towing procedures (duration and speed were changed from 30 to 15 minutes and 3.5 to 3.0 knots, respectively). This experiment did not rely solely on side-by-side hauls. On flat bottom, the vessels towed side-by-side, as close as safety would allow. On sloping ground, one vessel followed the other (with the order being alternated) while ensuring that there was no overlap in the area trawled. The numbers and lengths of five major groundfish species were recorded from 285 successful paired tows, with equations developed for converting catch-at-length for five species. Criteria were developed for determining whether, during the paired tows, one vessel fished on an aggregation that was essentially missed by the other vessel; such pairs were omitted from the final analysis or down-weighted. Bootstrapping was used to estimate the precision of the conversions. The new trawl and shorter towing times were reported to be distinctly less effective at catching larger fish (and this problem was compounded because of the relative scarcity of larger fish at the time the trials were conducted). Also, problems occurred when small size classes of fish taken with the new net were not recorded in the paired tow using the old trawl. Consequently, conversion factors were limited to lengths that were recorded by both vessels.

c. <u>Fishing on the same track</u>

Lewy et al. (2004) addressed the assumption that paired trawl hauls over the exact trawl track and close together in time are more similar than paired trawl hauls along parallel tracks. They did this because, although paired trawls had often been used to calibrate two sets of gear, a number of papers had demonstrated considerable spatial variability in catch rates between nearby stations (e.g., Ehrich 1991, Kvist et al. 2001). However, fishing in immediate succession in the same track creates a new problem, as the first haul may affect the fish density and species composition available for the second haul by removing fish or by inducing changes in behaviour that alters fish distribution. This effect, which is termed the disturbance effect, therefore needs to be estimated and accounted for. A generalised linear model for paired haul catches was analytically derived and the gear conversion and disturbance parameters with their precision were obtained. Simulation studies carried out showed that the estimated conversion factors were nearly unbiased. The analyses indicated that the disturbance effect was quite significant, implying that the fish density available for the subsequent haul was reduced by 60%. The disturbance effect was not influenced by original fish density. Because of the independence of the spatial fish distribution, Lewy et al. (2004) considered this method to be preferable to the paired hauls design.

A series of calibration trials have been carried out between 2001 and 2009 during the Icelandic Groundfish survey to compare catches between both the introduced research vessel and older style commercial trawlers and also between the commercial trawlers themselves (ICES 2005b, Sólmundsson pers. comm.). Successive tows along the same track were carried out with a 1.5 hour gap in between. In addition, paired repeat tows by the same vessel were carried out to allow a disturbance factor to be calculated. A total of 50–100 hauls were carried out per experiment with an overall total of 457 pairs using different vessels and 428 pairs using one vessel. A generalised linear model with quasi log link and three different variance functions was used to evaluate the results. The noise signatures of the different vessels were also measured. In early trials using one specific research vessel, regardless of model used, the quieter research vessel had higher catch rates, whilst the catch rates of the commercial trawl vessels were not significantly different. However, new data have since been collected, adding a new research vessel and with a total of six different commercial trawlers. These additional results suggest that the catch rates of the research vessels are within the range of commercial survey trawlers and it was concluded that it was unnecessary to apply correction factors to the survey estimates (Sólmundsson pers. comm.).

Oeberst et al. (2000) compared three trawl nets by trawling twice along the same track alternating the order of the compared nets. The analysis showed that it was necessary to estimate separate conversion factors for each species. However, the authors concluded that the sample sizes in this study (10 and 18 pairs of tows) were too small to obtain precise conversion factors.

Kingsley et al. (2008) used a single vessel in an experiment to calibrate an old and a new trawl net that were used for shrimp surveys. The new trawl net design included changed groundgear and trawl doors, with the new rig expected to fish better on rougher bottom. Experiments were carried out by trawling twice along the same track using combinations of either the same trawl fished twice, or the two different gears fished in one order or the other. The principle behind the analysis was that a "disturbance effect" could be calculated from the results of fishing twice with the same gear, and used to correct the difference in catch when different gears were used, to end up with a calibration factor (i.e., a catchability ratio) between the two gears. However, unlike Lewy et al. (2004), Kingsley et al. (2008) found that the disturbance effect was influenced by initial shrimp density; the second catches of shrimp were smaller at stations with large catches in the first haul, and vice versa. Catch models were fitted to the data (both size aggregated catch weights and size specific counts) using likelihood and Bayesian methods. The new trawl was estimated to fish with about 87% of the catchability of the old trawl. The catchability ratio varied with the size of shrimp caught, but the differences were not significantly different.

d. <u>Data analysis</u>

It is apparent that a number of experimental approaches have been developed for estimating the relative fishing power differences between different vessels or gears (for method descriptions and reviews see, e.g., Gavaris 1980, Pelletier 1998, Wilderbuer et al. 1998, González Troncoso & Paz 2003, Holst & Revill 2009, Cadigan & Dowden 2010, Miller et al. 2010). Each approach has different advantages and disadvantages, and the most appropriate method for an individual situation depends on the scientific scope, resources available, and biological responses of the organisms (Lewy et al. 2004). The analysis of paired trawl data to produce correction factors is clearly a topic area that is evolving. Perhaps the most recent contributions to this debate were from Cadigan & Dowden (2010) and Cadigan & Power (2010), who concluded that a generalised linear mixed-effects model (GLMM) gave reliable results over a wide range of spatial variations in densities even if there was between-tow variation in catchability, and from Miller et al. (2010) who found that the ratio estimator and the beta-binomial maximum likelihood estimator (MLE) performed well. Miller et al. (2010) concluded that the GLMM and beta-binomial MLE approaches were analogous.

Despite the wide range of analytical options, it does appear that the preferred survey method is the paired haul design, rather than consecutive hauls on the same track (Pelletier 1998, Lewy et al. 2004). This is because, for paired hauls, the vessels tow simultaneously within perhaps 0.5 km of one another, so that the abundance of sampled organisms should be close to the same for the two vessels, with the only difference in CPUE being a function of gear performance or catchability (Pelletier 1998).

One aspect typical of most calibration studies is that the estimated confidence intervals around the correction factors are quite wide, and if the confidence interval includes 1 (i.e., equal catchability between vessels), it is difficult to justify applying the correction factor (Tyson et al. 2006).

When a correction factor is applied, an observation is changed into an estimate with associated error (Munro 1998, von Szalay & Brown 2001). Therefore, correcting for fishing power differences was found to be worthwhile only when it reduces the total error in the estimate of the mean CPUE. Given this criterion, it is quite possible to have a statistically significant difference in fishing power and not apply a correction factor, because the added variance may exceed the benefits of the reduced bias (von Szalay & Brown 2001). Consequently, Munro (1998) developed a decision rule for the application of a correction factor based on the conjecture that its application was only beneficial if it reduced the error in the estimate of the mean CPUE, i.e., if the mean squared error (MSE) of the corrected CPUE was greater than that for the uncorrected CPUE, then the correction factor should be applied. MSE is an appropriate measure to assess the effect of the application of a correction factor because MSE is the sum of the variance (between the estimated CPUE and "true" CPUE) and the square of the bias (from the application of a correction factor).

The difficulty in coming up with well estimated coefficients is reflected in the fact that, apart from those estimated for the Baltic Sea surveys, very few of the calibration coefficients estimated for vessel and gear effects within European surveys are actually used to adjust the trawl time series in the ICES database (Sparholt, pers. comm.). For many of the assessments, such as North Sea haddock, survey time series are either treated as separate tuning indices, split up into before and after major vessel or gear changes, or the indices are calculated from combined datasets which ignore such effects (ICES WGNSSK 2011).

One factor that is unlikely to be an issue in contemporaneous calibration of different vessels and/or fishing gears, but might have implications if historical survey results are calibrated and attached to a modern survey series, was outlined by Zimmermann et al. (2003). They noted that historical hauls in a long-running survey series often had unusually low CPUE, probably because the trawl failed to contact the seabed. It was suggested that technological advances in the equipment used to monitor trawl performance since the mid 1980s had increased the ability to recognise malfunctioning tows, and consequently CPUE had risen. The authors derived a minimum CPUE from the survey with the best monitoring of bottom contact of the time-series, and used it as a criterion to eliminate trawls with poor bottom contact from earlier surveys. The truncated data sets produced some significantly larger biomass indices for some species. The analysis suggested that changes in CPUE over a time-series may sometimes be related more to changing survey fishing methods than to changes in abundance.

Cotter (2001) concluded that, in favourable circumstances, statistical modelling of abundance indices can be a useful alternative to comparative trawling experiments. The model involved fitting year-class curves with added terms for intercalibration and non-linearity factors for each species separately, to produce estimates of relative abundance. Much variability of whole-survey population abundance indices was explained using only the estimated year-class strengths and a coefficient of total mortality (Z) for the species. This modelling approach was used to calibrate surveys in the North Sea between 1977 and 1997 for four commercially important species. However, an age-related factor was required to allow for apparently lower catchabilities of young fish. Also, while this method may possibly be useful to calibrate data from historical surveys where the vessel or gear no longer exist, it is reliant on the availability of accurate catch-at-age data and a precise estimate of Z by year.

e. Conclusions from the literature review

Most of the literature indicates that the best method to calibrate two fishing units is by using parallel paired hauls, rather than fishing along the same track. The number of paired tows necessary to produce statistically sound correction factors (i.e., ratios of catchability between fishing units) will vary depending on the design of the comparison experiment, the distribution and density of the species for which corrections factors must be obtained, and the required level of precision. It appears that in calibration experiments, there is a requirement for at least 30, and probably nearer 50 pairs of tows, where both sets of tows catch a reasonable amount of the target species. However, many factors are known to influence catchability (Dunn 2006), and may affect competing fishing units differently. So whilst 30–50 paired tows may be sufficient to estimate a reasonably precise catchability ratio between fishing units for a single species in a homogenous environment, many more tows will be required if the survey series targets a range of species, across conditions where catchability would be expected to vary (such as depth, bottom type, catch composition, mean catch size, and weather conditions). The number of tows required to achieve a robust calibration therefore seems likely to be at least an order of magnitude higher, as indicated by the Icelandic groundfish survey trials which have used in excess of 400 hauls.

A maximisation of the amount of useful data from paired tows will occur when they are conducted in areas where the target species density is uniform. However, if the species is distributed unevenly by size over area or depth, then it will be necessary to ensure that sufficient tows are conducted to fully sample the length range in the area to be surveyed. The areal coverage and sample size must also ensure that sufficient data are collected to enable length-based analyses of correction factors for all ‘key’ species. Careful survey planning is essential, as is controlling as many sources of variability as possible. If a specific experiment is to be conducted (rather than a standard survey where two vessels complete all the planned stations as paired tows), then data simulations using prior information about the surveyed resource is advised when designing the experiment.

It is logical to assume that small differences between the compared fishing units are likely to produce small differences in fishing power. Therefore, where there is an existing survey time series, but it has become necessary to use a new vessel (e.g., because the current one must be retired) or to modify the fishing gear (e.g., because the trawl netting material used is no longer manufactured), it would be ideal to make the new fishing configuration and procedure as similar as possible to the old one.

3.2.2.3 New Zealand calibration experience

Calibration of vessels or trawl gears used for research time series in New Zealand has been limited to a few cases where calibration tows have been attempted or modelling approaches have been used to estimate relative catchability between vessels.

Without a dedicated deepwater research vessel in New Zealand until 1991, middle depths and deepwater trawl surveys on key New Zealand species were historically conducted from domestic commercial vessels and overseas research and commercial vessels in an *ad hoc* manner. Most of the surveys done in New Zealand from the late 1970s through to the arrival of R.V. *Tangaroa* in 1991 were in this category.

The Chatham Rise orange roughy wide-area trawl surveys used two different commercial vessels (*Otago Buccaneer* and *Cordella*) during the time series. Clark (1996) reported on the change of vessel between 1987 and 1988, and noted that an increase in the mean biomass index between those two years might be due to either a change in biomass, or a change in catching power between vessels. Attempts were made in the late 2000s to use the Chatham Rise survey results more sensibly by developing an informed Bayesian prior to account for the vessel/gear differences. The priors used in the stock assessment model (P. Cordue, ISL, unpublished) were primarily based on commercial catch and effort information from the two vessels when fishing in the same area at the same time (M. Dunn, NIWA, pers. comm.). Using this information, the stock assessment model estimated that the *Cordella* had a higher catchability for orange roughy than did the *Otago Buccaneer*, yet the reverse was suggested from an evaluation of vessel characteristics, such as vessel size and power, and trawl size and configuration (M. Dunn, NIWA, pers. comm.). This result highlights the uncertainty which arises from determining the likely magnitude and direction of catchability ratios between vessels in the absence of controlled experiments.

The R.V. *Kaharoa* replaced the R.V. *Ikatere* in 1981. The *Kaharoa* completed some trials using the *Ikatere* Granton trawl in 1982, with the experimental objectives including a comparison of catch rates between the new *Kaharoa* high opening bottom trawl (HOBT) and the *Ikatere* trawl. Tows using each trawl type were done in blocks with tows repeated to ensure that stations were sampled both by nets. Fifty eight shots were completed, 22 with the *Ikatere* gear and 36 with the *Kaharoa* HOBT trawl. Results showed that the *Kaharoa* gear had much greater catch rates than the *Ikatere* trawl. The most notable difference was in catches of larger (over 25 cm FL) snapper and gurnard.

Some attempts were made to compare biomass estimates from different vessels and gear combinations in other fisheries. Horn (1991) compared the *Shinkai Maru* and *Cordella* abundance estimates for jack mackerel and other key species, based on differences in the trawl configuration. An efficiency level of 36% was suggested for the *Cordella* gear relative to the *Shinkai Maru*. However, Horn (1991) concluded that robust comparisons could not be made between surveys which were conducted in different seasons and used gear of apparently different efficiencies.

Similarly, the conclusion that reliable between-vessel comparisons could not be made strictly from an analysis of catch composition was drawn by Hurst & Bagley (1997) for the *Shinkai Maru* and *Tangaroa* in surveys off Southland, and Hurst & Schofield (1990) for the *Shinkai Maru* and *Amaltal Explorer* on hoki surveys south of New Zealand and on the Chatham Rise.

The Chatham Rise 2002 *Tangaroa* orange roughy survey did some comparative tows with the commercial vessel *Ocean Ranger* to estimate correction factors between the two ships. The survey design required abundance estimates to be made quickly so as to minimise fish movement, hence the use of two vessels. The estimated catchability for the *Ocean Ranger* was 0.9 relative to the *Tangaroa* (Ian Doonan, NIWA, pers. comm.).

A gear change was made to the northwest Chatham Rise orange roughy trawl survey time series, substituting a smaller mesh full wing trawl for the rough bottom trawl previously used. The new "rat catcher" net had a full wing trawl with 6.5 inch mesh in the fore-part, a light full-bottom-contact groundrope, and 40 mm cod-end mesh, compared to the rough bottom trawl with 12 inch mesh in the fore-part, a 600 mm high groundrope, and 100 mm cod-end mesh. A paired trawl experiment was made using the *Tangaroa* to compare the catching efficiency of the two trawls. Tows were sequentially repeated on the same grounds, with the trawled area rested for about one day between hauls. The full wing "rat catcher" trawl with smaller mesh was estimated to be about four times more efficient than the rough bottom trawl at catching orange roughy (Ian Doonan, NIWA, pers. comm.).

Parallel tows were done using R.V. *Tangaroa* and F.V. *San Waitaki* (the *San Waitaki* of 1995, which was later replaced by a new vessel also called *San Waitaki*) to see if male and female orange roughy catchability was the same for both vessels, and to compare catch rates (Tracey et al. 1997). A minimum vessel separation of 0.5 km was maintained during trawling and 28 paired tows were completed. The analysis by Francis (1996) found that the *Tangaroa* was less likely that the *San Waitaki* to catch male orange roughy.

3.3 Questionnaire and interview results

The information gathered from questionnaire answers and interviews is presented for each Institute in Appendix 1. Based on this information, the underlying reasons for differing survey approaches are discussed in the following section, along with the perceived strengths and weaknesses of basing trawl surveys on dedicated research or commercial vessels.

3.3.1 Underlying reasons for adopting specific survey approaches by institution

Each of the institution/regions selected for interviewing has adopted a specific strategy for standardising its survey gear and operations along with making decisions on how much calibration will be undertaken when survey gear or operational behaviour changes. The range of survey designs covered by our interviews included single research vessel surveys, and multi-vessel and multi-nation trawl surveys based on research vessels, commercial vessels or a combination of both vessel types. One of the common features of the institutions that we interviewed was that the survey design approach tended to be a product of the survey history of the region taking into account capacity, with a measure of pragmatism. The main components of this mix are described by the four factors in the following sections.

3.3.1.1 Local system knowledge

All survey approaches for the interviewed institutions attempt to standardise trawl performance and fishing operations to ensure constant catchability. However, different institutions do this to different standards. At one end of the spectrum, some institutions standardise trawling a minimal amount, typically specifying a net plan for the trawl gear to be used, along with the distance and speed at which to tow, and what minimal mensuration equipment is required. At the other end of the spectrum, other institutions specify in detail every aspect of the survey that may influence catchability, including monitoring capacity and procedures to hold these factors constant. This includes providing all survey gear to the vessel, specifying vessel characteristics and equipment to a narrow range, and scripting all behavioural aspects related to fishing; from choice of fishing locations, through gear deployment and retrieval, and trawl repair and maintenance. The assumption at both ends of the spectrum is that vessel alone has minimal effects on catchability, especially when compared to the overall variability in biomass estimates. Although some experimental evidence exists indicating that these factors influence catchability, the degree to which the results are species-specific, survey specific, or small relative to other factors influencing biomass estimates is uncertain, resulting in different practices among regions.

In addition to standardisation, the decision to implement calibration experiments to understand and adjust for major changes in gear or vessel configuration is usually made at the survey level. This can be as comprehensive as a system calibration, where in addition to factors that are routinely standardised, the assumption is made that vessel, skipper and/or gear effects exist and are important to calibrate, either serially between changing vessels, or in parallel when multi-vessel surveys are conducted. This calibration can be of entire fishing systems, and determine a relative catchability for any given number of metrics. These can range from an overall vessel conversion factor to species-size-area-season specific conversion factors for each vessel involved. The latter would be prohibitively expensive to conduct, though some regions have had this as their target for serial government survey vessel calibration (USA-NEFSC; CAN-Eastern DFO). The assumption here is that much of the variance in biomass estimates can be attributed to variation in catchability due to trawl or vessel performance, and that careful calibration can reduce this variance and result in more precise biomass estimates.

3.3.1.2 Economics

Obviously, the funds available for surveys are limited. Funding, and the prospects for stable future funding, are ongoing constraints on conducting long-term survey series. Even in the short term, the funding required to operate a single survey, including all aspects relating to its performance, is significant. Funding requirements for calibrating vessels and operations, although less frequent, are also extremely expensive as discussed above. Lack of funding was the primary reason cited for the low level or lack of calibration work conducted (e.g., USA-NEFSC, NWFSC, CAN-West Coast DFO, Europe). Several scientists lamented that their initial calibration designs were scaled back because funding was lacking or that the level of standardisation possible was constrained by low funding. Low funding was cited as one reason for switching from a government research vessel to a commercial platform (e.g., USA-AFSC), although this was not always the case (Iceland, Spain). Another factor that enters the economic decision to fund a survey is the value of the fisheries being managed using data from the survey.

3.3.1.3 Infrastructure

Survey support: Organisations with long-standing government survey time series usually have permanent gear technologists on staff and the infrastructure to build, maintain, and store survey equipment. These regions have the funding and infrastructure to standardise gear, test gear, conduct calibration research, maintain gear at original specifications, and store large amounts of gear (e.g., 4–6 trawls, doors, floats, warp, plus electronics). Regions with little or no history of large scale surveys do not have this capability, and building this capacity is expensive. These regions usually turned to commercial fishermen for expertise, survey platforms, and gear.

Vessel availability: There has been a trend for government research vessels to be used across more disciplines, requiring more of their days at sea for work other than surveys. Surveys are frequently long, and often occur during the main summer weather window. For many jurisdictions in recent years, the prioritisation of research vessel time for surveys has been reduced, which has influenced the number and scale of surveys, the choice of platform for future surveys, and the amount of vessel time available for calibration work.

A related factor is the availability of commercial vessels with the appropriate characteristics to conduct a survey using the desired gear. In some regions, vessels are typically small and cannot meet the criteria required for standardisation (e.g., auto-trawl winches, fishing depth constraints). In other regions, commercial vessels are the same size as research vessels and can meet required criteria. This can substitute for in-house infrastructure to some degree and permit more options for conducting surveys (e.g., USA- NWFSC).

3.3.1.4 Historical approach and political desire

The choice of approach to standardising and calibrating surveys was largely historical for each region. Regions with long time series typically started with single government research vessels (USA - NEFSC), although in some cases commercial vessels were used initially (Icelandic Groundfish survey). When it was necessary to retire these vessels, a status quo decision to replace them and calibrate to the extent possible was typical.

If research vessels were calibrated decades ago, often the procedure was essentially the demonstration of a conversion factor between two entire fishing systems. The trials between *Scotia II* and *Scotia III* included a halving in tow time, and trials undertaken in the IPROST study were between different vessels using different or modified gear. Many of the trawl performance measurements that are easily monitored today were not available at that time to standardise the performance of trawl components, and the supporting science to show the effects of different trawl components on trawl performance had not been conducted. The usual approach was to use paired fishing trials with a fairly low sample size to compare catchability of one vessel relative to another.

When survey series were started more recently, this often occurred under pressure to calibrate many more factors, standardise more factors, and respond to economic or other concerns. In the USA, more recent surveys have had political pressure to reduce costs and encourage collaborative research (e.g., AFSC, NWFSC, NEFSC). The western regions have turned to commercial survey platforms because appropriate vessels were available, standardisation was deemed more important than calibration, government vessel availability was being reduced, and using commercial platforms was seen as collaborative research. However, in the Northeast of the USA, political pressure due to survey problems being exposed led to more funding, a complete redesign of the survey methodology, and an extensive vessel calibration experiment (although the calibration experiment was still reduced because of cost). In Europe, the same pressure to capture Industry “buy-in” and criticism of “out-dated” trawl gears has been a major driver behind a number of Industry-Science collaborative survey approaches. The Scottish-Irish monkfish survey previously described ensured the input of commercial fishers into all aspects, including gear specifications and survey stratification. CEFAS and the National Fishermen’s Organisation collaborated to carry out a series of 11 spatially restricted surveys using commercial vessels in areas nominated by fishermen. These surveys revealed a spatial coherence in catches of cod that was missed by the broader scientific survey (Cotter 2004). The author suggested that such surveys could potentially be informative supplementary indices, or might be used to locate aggregations of fish that could be assigned to special strata for additional survey effort.

3.3.2 Strengths and weaknesses of research versus commercial survey platforms

The consultations highlighted some of the strengths and weaknesses of the competing survey platforms: commercial vessels or dedicated research vessels. The requirement to calibrate may become more likely and more frequent when commercial vessels are used, owing to changing vessel availability and the upgrading of individual vessels over time. Standardisation is likely to be more achievable on research vessels owing to there being no commercial incentives to improve catchability. However, the choice to adopt a commercial or dedicated research survey platform is influenced by other considerations that are rarely documented in technical survey reports.

The interviews discussed a number of perceived strengths and weaknesses of research and commercial survey platforms. Very little documentation was available to substantiate any claims, but there was clearly a dichotomy of opinions related to a number of factors. Opposing views were noted for each factor (although not all interviews discussed each factor explicitly). To be concise, the two views are stated for each factor.

- Fishing experience of skippers and crew
    - Research vessel skippers and crew do not fish often enough to be sufficiently proficient in areas where fishing can be technically challenging; therefore aspects such as “untrawlable habitat” can change, gear performance may not be as reliable, gear repair is slower, and the survey can be slower.
    - Research vessels often fish more than 100 days per year on surveys, so fishing behaviour is consistent (if not expert) and there may be a greater appreciation of the importance of standardized procedures.
- Perceived cost
    - Research vessels are more expensive per day.
    - Research vessels carry more staff, have more consistent equipment over time, collect more data for more variables, and can provide 24h survey capability.
- Staff accommodation
    - Most commercial vessels fitting the criteria for survey are large enough to accommodate the staff required, although this is not the case in all regions and fisheries. Vessel crew often assist in data collection.
    - Research vessels typically accommodate more staff, and often staff accommodation is better. Commercial vessels typically have small crews, and adding significant staff is problematic for accommodation and safety requirements.
- Catch sampling requirements
    - Commercial vessels can provide the data necessary from the survey.
    - Research vessels can provide 24 h sampling capability, are often designed for efficient catch sampling, have integrated computer systems for high data quality, can provide more scientific staff to collect more data on more species. Commercial vessels do not have the staff, storage space, or facilities for additional sampling such as histology, tissue collection, or specimen collection.
- Vessel continuity and consistency
    - Multiple commercial vessel availability provides competition and flexibility. Continuity is only important if calibrating is important. Changes in staff and equipment are not important if survey standards are specified.

    - Research vessels are more likely to be available in the future for a consistent time series. Survey staff are more likely to have input into any vessel changes. Commercial vessels can change equipment and staff at will with no input from survey scientists.
- Schedule flexibility
    - Research vessel schedules are planned months to years in advance. Small shifts in survey timing are not usually permitted. Commercial vessels are more flexible in scheduling advance, shifting schedules, and dealing with unexpected circumstances.
    - Commercial vessels operate on opportunity cost. Availability is dependent on other options, which can change.
- Relationships with fleet
    - Using commercial vessels requires interaction of scientists and fishermen, which results in better understanding of both disciplines by each group, and a better survey with more accepted results by stakeholders.
    - Conducting surveys is a government responsibility which should remain isolated from potential commercial influences.

4. DISCUSSION

The need for standardised vessel(s) properties, trawl gear and gear deployment procedures to ensure comparability of individual surveys within time series is well recognised and procedures to follow are well documented. Any change of vessel or significant changes to gear or trawl deployment procedure, without some form of calibration, will lead to uncertainty about the comparability of the results within the observations in a survey time series. The implications of any changes need to be evaluated by the relevant stock assessment working groups and fishery managers before the results of surveys conducted following a change are to be accepted. It seems likely that most quantitative stock assessment practitioners would no longer consider the time series to be intact.

One approach that could be followed would be to build a Bayesian informed prior relating the catchabilities of the two fishing units, effectively assuming that the catchabilities of the two surveys are linked. Alternatively, the stock assessment model could estimate independent catchabilities for each vessel series. A calibration study would greatly aid in the development of the informed prior, allowing the model to gain much more information from the sequential survey designs than if the model assumed the surveys were independent. This may be a helpful approach for species-size catchability conversions where sample size is limited in the calibration experiment. Information from other species and size classes could be used to provide information on what the relative catchabilities should be. However, when sequential surveys are used, a Bayesian approach provides the opportunity for the model to explain variance in the time series by varying catchability between surveys, which may mask an abundance signal. As the precision of the prior is improved through the quality of the calibration study, the overall time-series will be able to contribute more information to the assessment. As a result, calibration experiments to help determine and informed prior are recommended for valuable survey time series.

It is essential to carefully determine the best model and approach when planning any calibration work. The literature review and interviews of calibration approaches indicate that both parallel paired tows and paired tows along the same track have been used, but that parallel tows would be the preferred approach, as long as fish populations can be shown to be homogeneously distributed over the scale required. This method avoids any requirement to estimate a disturbance effect. If, following the estimation of calibration correction factors, it is found that relative abundance estimates differ vastly from what is expected, then the correction factors are likely to be brought into question. Because doubts may not arise until some years after the calibration experiment was done, the opportunity to retrospectively redo the work may be lost, e.g., in the case of a vessel change.

Calibration experiments at sea are expensive to run. A large number of comparative tows is likely to be needed to ensure a reliable result with acceptable precision, in particular where correction factors need to be applied to several species or to selected length ranges within a species. Even large scale experiments do not guarantee clear results (e.g. Icelandic Groundfish Survey). A range of calibration approaches have been undertaken, but in only a few cases have conversion factors been estimated with enough confidence to be applied in the relevant stock assessments. In most cases the logistics and costs of carrying out 50 or more hauls were just not feasible with the funding available and a gradual build-up of small numbers of paired tows over time was all that was practically possible (e.g. European North sea and Western IBTS surveys).

A consideration of which species require calibration factors is an important part of any planning. The experiment requires adequate coverage of the important survey depths where each species occurs to ensure representation. Trawl surveys are used for fish abundance estimation, but also provide key inputs for other work. Requirements for other uses also need to be considered when planning changes survey design and also when designing the calibration experiment. Long running time series provide inputs to ecosystem based modelling, information on abundance, and length or age trends for secondary species. Large numbers of individual species correction factors may be needed when redesigning a key trawl survey so that these other tasks can continue. It is always possible that ‘like’ species groups (e.g., flatfish or skates) can be combined using a single correction factor.

Trawl gear, electronics, and other aspects of fishing are constantly evolving and improving. It has been said that “fishing is like fashion — one person changes and the rest follow”. The use of commercial vessels in trawl surveys would need to be closely monitored so that these changes do not occur without some discussion of the possible effects on standardisation and catch rates. The ability to return to a standard fishing configuration for any survey work would be very important. Changes and modifications to trawling equipment can result from issues with the supply of standard components and mesh size over time, a degree of forward planning (e.g., holding spare stock) may help alleviate such issues.

Consultation with institutes in USA, Canada and Europe highlighted a range of survey designs from single research vessel surveys, to multi-vessel and multi-nation surveys using research, commercial, or both vessel types. Where commercial vessels were used in surveys, there were undoubtedly greater challenges with respect to standardisation and sometimes limitations on the data that could be gathered. In a number of cases, commercial vessels used historically were being replaced with research vessels as they became available, as these were perceived as the best option (e.g. Spain Western IBTS survey, Iceland groundfish survey, Mediterranean MEDITS survey). However, in other situations, where additional, or species-specific surveys are being initiated (e.g. monkfish and cod-focused surveys in NE Atlantic and Celtic Sea), local knowledge and “buy-in” from the fishing industry were perceived as key benefits, along with more cost-effective access to multiple vessels to cover larger areas at any one time. In almost all interviews, the key priority was seen as making sure that the trawl gear was fishing in a consistent fashion over the course of different surveys and from different vessels. This was achieved through detailed gear specifications and construction controls, and by documentation of fishing procedures. Specific vessel effects were rarely accounted for, either due to the perception that they were minor compared to other sources of variation, or, where this was clearly not the case, through the lack of sufficient funds to carry out calibration trials at the level required.

5. CONCLUSION and RECOMMENDATIONS

- If there is a need to change or replace a vessel that has carried out a trawl survey time series, then producing a calibration between the 'new' and 'old' vessel is strongly recommended.
- To be successful, a calibration experiment must be carefully and comprehensively designed and executed, with an understanding on how the data will be used in subsequent analysis; this is an expensive, time consuming exercise.
- Conducting paired side-by-side tows is the best practice for calibration experiments, with consideration also given to the likely analysis method at the experimental planning stage.
- To maintain continuity of research trawl survey series in New Zealand, calibration experiments will be required in the future for research vessels at the very least (e.g., R.V. *Kaharoa* may be only available for the next decade). Planning for this should start well in advance of the sale or decommissioning of the existing vessel.
- If a calibration is required, then this is a good time to 'modernise' the gear to be used by the new vessel to take advantage of improved technologies to catch fish and monitor and standardise gear performance.
- However, while some 'modernisation' will help to future-proof the gear, it is desirable to make as few large changes as possible to trawl design and materials. This will minimise unexpected changes in selectivity by length, and hence, reduce the likelihood that size-based correction factors will be required.
- Regardless of calibration requirements, standardisation of gear and operational procedures across years and across vessels is important, although both significant and non-significant differences in catchability for standardised vessels are known to occur.
- Standardisation is markedly less expensive than calibration.
- If standardisation is used instead of calibration (either for multi-vessel surveys, or where a vessel replacement is required), then it is clearly desirable to use standard gear and implement standard procedure as much as possible across vessels. Ideally, all fishing gear from the winches to the codend should be the same, the vessels should be of similar size and fishing power and, as a result, the trawl nets should be fishing in the same way.

6. ACKNOWLEDGEMENTS

We thank all the individuals who agreed to participate in the questionnaire and interview process that was so important and informative to this review. We also thank Matthew Dunn for a thoughtful review of the document, and Marianne Vignaux for editorial comments that clarified several sections. This work was funded by the Ministry of Fisheries under Project DEE2010-01.

7. REFERENCES

Anonymous (1992). Report of the workshop on the analysis of trawl survey data. *ICES CM 1992/D*:6.

Anonymous (2007). International bottom trawl survey in the Mediterranean (Medits) Instruction Manual – Version 5.

Azarovitz, T.R. (1981). A brief historical review of the Woods Hole Laboratory trawl survey time series. In: Doubleday, W.G.; Rivard, D. (eds.), Bottom Trawl Surveys. *Canadian Special Publication of Fisheries and Aquatic Sciences 58*: 62–67.

Brown, R.W.; Fogarty, M.; Legault, C.; Miller, T.; Nordahl, V.; Politis, P.; Rago, P. (2007). Survey transition and calibration of bottom trawl surveys along the northeastern continental shelf of the United States. *ICES CM2007/Q*:20. 25 p.

Burns, F.; Stokes, D. (2005). Comparative study between *Celtic Explorer* and *Scotia* during the Quarter 4 IBTS survey in area VIa. WD to 2005 IBTS meeting, Hamburg, 101: 118 p.

Byrne, C.J.; Forrester, J.R.S.; Hayes, D. (1991). Survey vessels and gear modifications and their possible effects on assessment analysis. In report of the 12th northeast regional stock assessment workshop. US Department of Commerce, NOAA, *Northeast Fisheries Science Center Reference Document 91-03*. 187 p.

Cadigan, N.G.; Dowden, J.J. (2010). Statistical inference about the relative efficiency of a new survey protocol, based on paired-tow survey calibration data. *Fishery Bulletin 108*: 15–29.

Cadigan, N.G.; Power, D. (2010). Vessel calibration results for redfish (*Sebastes* sp.) from comparative fishing between the CCGS Teleost research vessel and the MV Cape Beaver fishing vessel. *Canadian Science Advisory Secretariat, Research Document 2010/062*. 17 p.

Cadigan, N.G.; Walsh, S.J.; Brodie, W. (2006). Relative efficiency of the *Wilfred Templeman* and *Alfred Needler* research vessels using a Campelen 1800 shrimp trawl in NAFO Subdivision 3Ps and Divisions 3LN. *Canadian Science Advisory Secretariat, Research Document 2006/085*. 20 p.

Clark, M.R. (1996). Biomass estimation of orange roughy: a summary and evaluation of techniques for measuring stock size of a deep-water fish species in New Zealand. *Journal of Fish Biology 49 (Supplement A)*: 114–131.

Cotter, A.J.R. (2001). Intercalibration of North Sea International Bottom Trawl Surveys by fitting year-class curves. *ICES Journal of Marine Science 58*: 622–632.

Cotter, A.J.R. (2004). Can fishers teach scientists how to improve fish surveys? Selected results from spatially intense, commercial FV surveys of 9 English fisheries in 2003–4. Working Document presented to the ICES WKSAD, Aberdeen, 2004.

De Robertis, A.; Hjellvik, V.; Williamson, N.J.; Wilson, C.D. (2008). Silent ships do not always encounter more fish: comparison of acoustic backscatter recorded by a noise-reduced and a conventional research vessel. *ICES Journal of Marine Science 65*: 623–635.

Doubleday, W.G.; Rivard, D. (eds.) (1981). Bottom trawl surveys. *Canadian Special Publication of Fisheries and Aquatic Sciences 58*: 1–273.

Dunn, M.R. (2006). A review of experimental methods for determining catchability for trawl surveys. *New Zealand Fisheries Assessment Report 2006/51*. 31 p.

Ehrich, S. (1991). Comparative fishing experiments by research trawlers for cod and haddock in the North Sea. *Journal du Conseil International pour l'Exploration de la Mer 47*: 275–283.

Fiorentini, L.; Dremiere, P.Y.; Leonori, I.; Sala, A.; Palumbo, V. (1999). Efficiency of the bottom trawl used for the Mediterranean international trawl survey (MEDITS). *Aquatic Living Resources 12*: 187–205.

Francis, R.I.C.C. (1984). An adaptive strategy for stratified random trawl surveys. *New Zealand Journal of Marine and Freshwater Research 18*: 59–71.

Francis, R.I.C.C. (1996). Orange roughy sex ratios and catchrate distributions in the Chatham Rise Spawning Box. New Zealand Fisheries Assessment Research Document 96/13. 27 p. (Unpublished report held in NIWA library, Wellington.)

Fryer, R.; Zuur, A.; Graham, N. (2003). Using mixed models to combine smooth size-selection and catch-comparison curves over hauls. *Canadian Journal of Fisheries and Aquatic Sciences 60*: 448–459.

Gavaris, S. (1980). Use of a multiplicative model to estimate catch rate and effort from commercial data. *Canadian Journal of Fisheries and Aquatic Sciences 37*: 2272–2275.

Godø, O.R. (1994). Natural fish behaviour and catchability of groundfish. *ICES CM 1994/G*:14.

González Troncoso, D.G.; Paz, X. (2003). Testing methods for estimating the factor power correction obtained from the comparative fishing trial: C/V *Playa de Mendíña* versus R/V *Vizconde de Eza*. NAFO SCR Doc. 03/5. (NAFO Scientific Council Meeting, June 2003.) 8 p.

Helser, T.E.; Punt, A.E.; Methot, R.D. (2004). A generalized linear mixed model analysis of a multi-vessel fishery resource survey. *Fisheries Research 70*: 251–264.

Holst, R.; Revill, A. (2009). A simple statistical method for catch comparison studies. *Fisheries Research 95*: 254–259.

Horn, P.L. (1991). Trawl survey of jack mackerels (*Trachurus* spp.) off the central west coast, New Zealand, February-March 1990. *New Zealand Fisheries Technical Report 28*. 39 p.

Hurst, R.J.; Bagley, N.W. (1997). Trends in Southland trawl surveys of inshore and middle depth species, 1993-96. *New Zealand Fisheries Technical Report 50*. 67 p.

Hurst, R.J.; Bagley, N.W.; Chatterton, T.D.; Hanchet, S.M.; Schofield, K.A.; Vignaux, M. (1992). Standardisation of hoki/middle depth time series trawl surveys. MAF Fisheries Internal Report No. 194. 87 p. (Unpublished report held in NIWA library, Wellington.)

Hurst, R.J.; Schofield, K.A. (1990). Preliminary biomass estimates of hoki and selected species from trawl surveys of Southland/Sub-Antarctic and the Chatham Islands, Oct-Dec 1989, and comparisons with the previous *Shinkai Maru* surveys. New Zealand Fisheries Assessment Research Document 90/6. 23 p. (Unpublished report held in NIWA library, Wellington.)

ICES (2004). Report of the Workshop on Survey Design and Data Analysis (WKSAD). *ICES CM 2004/B*:07. 48 p.

ICES (2005a). Report of the Workshop on Survey Design and Data Analysis (WKSAD). *ICES CM 2005/B*:07. 170 p.

ICES (2005b). Report of the Study Group on Survey Trawl Standardisation (SGSTS). *ICES CM 2005/B*:05. 67 p.

ICES (2006). Report of the Study Group on Survey Trawl Standardisation (SGSTS). *ICES CM 2006/FTC*:05. 67 p.

ICES (2007). Manual for the International Bottom Trawl Surveys. IBTS Manual, Revision VIII. 69 p.

ICES (2009a). Report of the Study Group on Survey Trawl Standardisation (SGSTS). *ICES CM 2009/FTC*:09. 127 p.

ICES (2009b). Report of the Workshop on Anglerfish and Megrim (WKAGME). *ICES CM 2009/ACOM*:28. 112 p.

ICES WGNSSK (2011). Report of the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK). *ICES CM 2010 / ACOM*:13. 1048 p.

IPROSTS (2001). International program of standardised trawl surveys. Final report by IFREMER and MARLAB to the Commission of European Communities, Contract Reference DG XIV Study Contract 98-057. 71 p.

Kappenman, R. (1992). Robust estimation of the ratio of scale parameters for positive random variables. *Communications in Statistics, Theory and Methods 21*: 2983–2996.
Kingsley, M.C.S.; Wieland, K.; Bergström, B.; Rosing, M. (2008). Calibration of bottom trawls for northern shrimp. *ICES Journal of Marine Science 65*: 873–881.
Kotwicki, S.; Weinberg, K.L.; Somerton, D.A. (2006). The effect of autotrawl systems on the performance of a survey trawl. *Fishery Bulletin 104*: 35–45.
Kvist, T.; Gislason, H.; Thyregod, P. (2001). Sources of variation in the age composition of sandeel landings. *ICES Journal of Marine Science 58*: 842–851.
Lewy, P.; Nielsen, J.R.; Hovgård, H. (2004). Survey gear calibration independent of spatial fish distribution. *Canadian Journal of Fisheries and Aquatic Sciences 61*: 636–647
Mahévas, S.; Trenkel, V.M.; Doray, M.; Peyronnet, A. (2011). Hake catchability by the French trawler fleet in the Bay of Biscay: estimating technical and biological components. *ICES Journal of Marine Science 68*: 107–118.
McMillan, P. (Comp.) (1996). Trawl survey design and data analysis procedures for deepwater fisheries research. NIWA Internal Report No. 189. 20 p. (Unpublished report held in NIWA library, Wellington.)
Miller, T.J.; Das, C.; Politis, P.J.; Miller, A.S.; Lucey, S.M.; Legault, C.M.; Brown, R.W.; Rago, P.J. (2010). Estimation of Albatross IV to Henry B. Bigelow calibration factors. *Northeast Fisheries Science Center Reference Document 10-05*. 233 p. http://www.nefsc.noaa.gov/publications/crd/crd1005/
Mitson, R.B. (1995). Underwater noise of research vessels: Review and recommendations. ICES Cooperative Research Report 209.
Munro, P.T. (1998). A decision rule based on the mean square error for correcting relative fishing power differences in trawl survey data. *Fishery Bulletin 96*: 538–546.
NEFSC (2007). Proposed vessel calibration for NOAA ship Henry B. Bigelow. *Northeast Fisheries Science Center Reference Document 07-12*. 26 p. http://www.nefsc.noaa.gov/publications/crd/crd0712/
Oeberst, R.; Ernst, P.; Frieß, C.C. (2000). Inter-calibrations between German demersal gears HG 20/25 and TV3 520 as well as between the gears TV3 520 and TV3 930. *ICES CM 2000/K*:20. 27 p.
Oeberst, R.; Grygiel, W. (2004). Estimates of the fishing power of bottom trawls applied in Baltic fish surveys. *Bulletin of the Sea Fisheries Institute 1(161)*: 29–41.
Olsen, K.; Angell, J.; Pettersen, F.; Løvik, A. (1983). Observed fish reaction to a surveying vessel with special reference to herring, cod, capelin and polar cod. *FAO Fisheries Report 300*: 131–138.
Ona, E.; Godø, O.R.; Handegard, N.O.; Hjellvik, V.; Patel, R.; Pederson, G. (2007). Silent research vessels are not quiet. *Journal of the Acoustical Society of America Express Letters 121*: 145–150.
Parker, S.J.; Hurst, R.J.; Francis, R.I.C.C. (2009). Design options for an elephantfish, red gurnard, and flatfish complex trawl survey of the East Coast of the South Island. *New Zealand Fisheries Assessment Report 2009/59*. 22 p.
Pelletier, D. (1998). Intercalibration of research survey vessels in fisheries: a review and an application. *Canadian Journal of Fisheries and Aquatic Sciences 55*: 2672–2690.
Reid, D.G.; Allen, V.J.; Bova, D.J.; Jones, E.G.; Kynoch, R.J.; Peach, K.J.; Fernandes, P.G.; Turrell, W.R. (2007a). Anglerfish catchability for swept-area abundance estimates in a new survey trawl. *ICES Journal of Marine Science 64*: 1503–1511.
Reid, D.G.; Kynoch, R.J.; Penny, I.; Peach, K.J. (2007b). Estimation of catch efficiency in a new angler fish survey trawl. *ICES CM2007/Q*:22.

Sissenwine, M.P.; Bowman, E.W. (1978). An analysis of some factors affecting the catchability of fish by bottom trawls. *International Commission for the Northwest Atlantic Fisheries, Research Bulletin 13*: 81–87.

Stauffer, G. (2004). NOAA Protocols for groundfish bottom trawl surveys of the nation's fishery resources. NOAA Technical Memorandum NMFS-SPO-65. 211 p.

Stevenson, M.L.; Hanchet, S. (1999). Design and data analysis procedures for inshore trawl surveys. *NIWA Technical Report 53*. 18 p.

Strømme, T.; Iilende, T. (2001). Precision in systematic trawl surveys as assessed from replicate sampling by parallel trawling off Namibia. *South African Journal of Marine Science 23*: 385–396.

Tracey, D.M.; Anderson, O.F.; Clark, M.R. (1997). A two-vessel survey of orange roughy in the Chatham Rise "Spawning Box" July–August 1995. *New Zealand Fisheries Technical Report 49*. 27 p.

Tyson, J.T.; Johnson, T.B.; Knight, C.T.; Bur, M.T. (2006). Intercalibration of research survey vessels on Lake Erie. *North American Journal of Fisheries Management 26*: 559–570.

Von Szalay, P.G.; Brown, E. (2001). Trawl comparisons of fishing power differences and their applicability to National Marine Fisheries Service and Alaska Department of Fish and Game trawl survey gear. *Alaska Fishery Research Bulletin 8(2)*: 85–95.

Warren, W.G. (1997). Report on the comparative fishing trial between the *Gadus Atlantica* and *Teleost*. *NAFO Science Council Studies 29*: 81–92.

Wilderbuer, T.K.; Kappenman, R.F.; Gunderson, D.R. (1998). Analysis of fishing power correction factor estimates from a trawl comparison experiment. *North American Journal of Fisheries Management 18*: 11–18.

Wileman, D.A. (1984). Model testing of the 36/47 m GOV young fish sampling trawl. Report of the Danish Institute of Fisheries Technology, Hirtshals. 43 p.

Wood, B.R. (2011). Fish reaction to tonal vessel noise — Comparison with ICES CRR 209 methodology. *Applied Acoustics 72*: 399–411.

Zimmermann, M.; Wilkins, M.E.; Weinburg, K.L.; Lauth, R.R.; Shaw, F.R. (2003). Influence of improved performance monitoring on the consistency of a bottom trawl survey. *ICES Journal of Marine Science 60*: 818–826.

Zuur, A.F.; Fryer, R.J.; Newton, A.W. (2001). The comparative fishing trial between *Scotia II* and *Scotia III*. Fisheries Research Services (FRS) Marine Laboratory Report No. 03/01.

APPENDIX 1: Calibration and standardisation summaries from interviews and questionnaires

Three areas (the USA, Canada and Europe) where time series trawl surveys are important inputs for stock assessments were selected and interviews of researchers involved with long running time series surveys conducted. The questions asked are listed in Section 2 above. The following tables give summaries from each of the regional interviews on survey standardisation and calibration approach for bottom trawl surveys.

The individuals interviewed and / or returned questionnaires and information for this report were:

Individual consulted	Institute
Antonello Sala	ISMAR-CNR, Ancona, Italy
Brian Harley	CEFAS, England
Chris Roberts	Skipper of the Viking Storm
David Stokes	Marine Institute, Galway, Ireland
David Somerton	NOAA Alaska Fisheries Science Center, USA
Don Standsbury	Department of Fisheries and Oceans, Canada
Doug Beare	IMARES, The Netherlands
Fátima Cardador	IPIMAR, Portugal
Greg Workman	Department of Fisheries and Oceans, Canada
Haraldur Einarsson	Icelandic Marine Research Institute, Iceland
Henrik Sparsholt	ICES, Denmark
Iñaki Quincoces	AZTI - Marine Research Unit, Spain
Jim Boutillier	Department of Fisheries and Oceans, Canada
Keith Bosley	NOAA Northwest Fisheries Science Center, USA
Larwrence Hufnagle	NOAA Northwest Fisheries Science Center, USA
Norman Graham	Marine Institute, Galway, Ireland
Patricia Burke	NOAA Northwest Fisheries Science Center, USA
Paul Fernandes	Marine Scotland, Aberdeen, Scotland
Phil Politis	NOAA Northeast Fisheries Science Center, USA
Rick Stanley	Department of Fisheries and Oceans, Canada
Rob Johnston	NOAA Northeast Fisheries Science Center, USA
Robert Kynoch	Marine Scotland, Aberdeen, Scotland
Russel Brown	NOAA Northeast Fisheries Science Center, USA
Steve Walsh	Department of Fisheries and Oceans, Canada
Terry Henwood	NOAA Southeast Fisheries Science Center, USA

Region:	**NOAA, Alaska Fisheries Science Center, USA**
Number of bottom trawl surveys in region:	Three A) Bering Sea B) Aleutian Islands C) Gulf of Alaska (Plus some ooperation with State of Alaska surveys.)
Number of vessels involved:	Five (two in Bering Sea).
Commercial or Research vessel platform?	Commercial. Historical Aleutian Island surveys were conducted with government research vessel.
Historical survey platform:	Single government research vessel.
Standardisation:	
Gear provided or specified:	Warps: diameter, construction, lay, length.
	Doors: specified or provided.
	Sweeps: provided.
	Trawl: provided, custom built by NOAA specialists, measured before, during and after survey. N= 6 per vessel, one stored on deck.
Vessel equipment specified:	Autotrawl - calibrated, power, doors.
Sonar:	E-60, required to collect calibrated acoustic data.
Trawl mensuration:	Simrad acoustic.
Data acquisition system:	Provided and installed, fishing and catch sampling.
Personnel:	
Skipper:	Evaluated on experience, trained, reviewed.
Crew:	Evaluated on experience, trained, reviewed.
Science staff:	Five – provided.
Behaviours standardised:	Vessel speed from deployment to net on deck, choice of fishing location, trawl mensuration envelope, warp scope, trawl component maintenance and replacement specifications, catch handling.
Calibration	
Has survey vessel been modified over time?	Yes, but no calibration conducted.
Has survey gear been modified over time?	No
Detail:	Lack of new research vessel time allocation.
	Old vessel mechanical problems, significant maintenance required to conduct calibration.
	Large commercial vessels available with appropriate gear and power.
	Fishing experience desired.
	Unknown future vessels, high cost associated with calibrating multiple vessels.
How are vessels calibrated?	Vessel effects (including skipper) treated as a random factor drawn from a pool of vessel effects with a normal distribution. Catch rates are standardised by species for each survey.
Calibrated for all species with biomass estimates?	Modelled
Approach Summary	The Bering Sea survey has always been conducted with commercial vessels. The AFSC was not able to continue other survey time series when the government vessel was retired due to cost and survey time necessary. Their approach was to begin a new multi-vessel, commercial platform of similar vessels, and standardise every factor believed to influence catchability.
Standardise or calibrate?	Strictly standardise gear and fishing behaviour: calibration not needed. Small vessel effects removed through standardisation.

Region:	**NOAA, Northwest Fisheries Science Center, USA**
Number of bottom trawl surveys in region:	One: A) West Coast Note: other midwater/acoustic surveys are also conducted
Number of vessels involved:	Four (two vessels per pass, two passes (summer/autumn)).
Commercial or Research vessel platform?	Commercial. Historical combination of government vessel for deepwater, commercial vessels for triennial survey.
Historical survey platform:	Single government research vessel, and commercial platform for triennial. Government survey vessel for acoustic surveys.
Standardisation:	
Gear provided or specified:	Warps: diameter, construction, lay, length.
	Doors: specified or provided.
	Sweeps: provided.
	Trawl: provided, custom built by NOAA specialists, measured before, during and after survey. Multiple per vessel, one stored on deck.
Vessel equipment specified:	horsepower, doors.
Sonar:	E-60
Trawl mensuration:	Simrad acoustic.
Data acquisition system:	Provided and installed, fishing and catch sampling.
Personnel:	
Skipper:	Evaluated on experience, trained, reviewed.
Crew:	Evaluated on experience, trained, reviewed.
Science staff:	Three – provided.
Behaviours standardised:	Vessel speed from deployment to net on deck, choice of fishing location, trawl mensuration envelope, warp scope, trawl component maintenance and replacement specifications, catch handling.
Calibration	
Has survey vessel been modified over time?	Multiple vessels used, little overlap among years as some new vessels join.
Has survey gear been modified over time?	No.
Detail:	Lack of new research vessel time allocation.
	Old vessel mechanical problems, significant maintenance required to conduct calibration.
	Large commercial vessels available with appropriate gear and power (from Canada historically, US recently).
	Fishing experience desired.
	Unknown future vessels, high cost associated with calibrating multiple vessels.
How are vessels calibrated?	Vessel effects (including skipper) treated as a random factor drawn from a pool of vessel effects with a normal distribution. Catch rates are standardised by species for each survey.
Calibrated for all species with biomass estimates?	Modelled.
Approach Summary	The West Coast had a mixed survey history, with a commercial platform of large Canadian or Alaskan trawlers used for a triennial survey, and a government vessel for annual slope surveys. In 1998–2000, due to cost and time allocation and an ageing government vessel, a new annual multi-vessel survey of four commercial vessels was established.
Standardise or calibrate?	Strictly standardise gear and fishing behaviour: calibration not needed. Small vessel effects removed through standardisation.

Region:	NOAA, North East Fisheries Science Center, USA
Number of bottom trawl surveys in region:	Two: A) Spring B) Autumn (Plus some cooperation with NEMAP shallow water adjunct survey.)
Number of vessels involved:	Two.
Commercial or Research vessel platform?	Government vessels, long time series back to 1963.
Historical survey platform:	Single government research vessel, expanding to two government vessels.
Standardisation:	
Gear:	Warps: diameter, construction, lay, length.
	Doors.
	Sweeps.
	Trawl: custom built by NOAA specialists, measured before, during and after survey. N = 3–5.
Vessel equipment:	Autotrawl – calibrated, power, doors.
Sonar:	E-60 to collect calibrated acoustic data.
Trawl mensuration:	Simrad acoustic.
Data acquisition system:	Custom government, fishing and catch sampling.
Personnel:	
Skipper:	Government, consistent presence.
Crew:	Government, consistent presence.
Science staff:	14
Behaviours standardised:	Vessel speed from deployment to net on deck, choice of fishing location, trawl mensuration envelope, warp scope, trawl component maintenance and replacement specifications, catch handling.
Calibration	
Has survey vessel been modified over time?	Yes. Calibrations for specific gear changes (e.g. doors) in past, and calibrations to convert catch rates from Albatross to Delaware. Post 2003, complete fishing system redesign, with aim to calibrate to a new vessel (Bigelow) in 2008.
Has survey gear been modified over time?	Yes (some intentional and some not). Monitor important performance, calibrate if changes are detectable. All minor intentional changes documented (e.g. new float type).
Detail:	Large scale calibration of fishing systems resulted in calibration coefficients for a number of species, though not all. Calibration was originally planned as multi-year experiment, cut to only one year (e.g. lower sample sizes and no seasonal effects).Gear design started from trawl design, flume tanks, through fishing performance.
	Commercial vessels not available with appropriate specifications.
How are vessels calibrated?	Vessel effects (including skipper) treated as a fishing system. Large scale paired tow comparison (636 paired tows) was still an abbreviated version.
Calibrated for all species with biomass estimates?	Conversion factors by species for most common species.
Approach Summary	A large effort to redesign east coast trawl surveys stemmed from discrepancies and political pressure arising in 2003. The response (after allocation of more resources to the problem) was a complete redesign of the fishing system including purpose, flume performance, and vessel system calibration with a new research vessel.
Standardise or calibrate?	Standardise gear and fishing behaviour to a typical degree, but consider skipper and vessel and gear as a system, so that variation included in usual conversion factor. Desire is to calibrate for as many species as possible, over season, latitude, and complete

	fishing system. Focus on understanding gear performance and monitoring all factors that may influence that. Calibrate when monitoring shows deviation.

Region:	NOAA, South East Fisheries Science Center, USA
Number of bottom trawl surveys in region:	Two: A) Shrimp – Groundfish (shallow: 10–120 m). B) Pelagic (bottom trawl with 30 foot headrope height). Note: other longline surveys conducted and shrimp survey integrated with state waters survey.
Number of vessels involved:	Two (multi-vessel among years).
Commercial or Research vessel platform?	Government.
Historical survey platform:	Single government research vessel, but switching between government vessels.
Standardisation:	
Gear provided or specified:	Warps, doors, sweeps maintained on each vessel.
	Trawl: provided, custom built by NOAA specialists.
Vessel equipment specified:	None.
Sonar:	Only for depth, bottom suitability.
Trawl mensuration:	No.
Data acquisition system:	Catch sampling.
Personnel:	
Skipper:	Government employed, but fish surveys frequently.
Crew:	Government employed, but fish surveys frequently.
Science staff:	7
Behaviours standardised:	Vessel speed from deployment to net on deck, choice of fishing location, warp scope, catch handling.
Calibration	
Has survey vessel been modified over time?	No, although a new survey vessel is coming online (PICES).
Has survey gear been modified over time?	Gear has undergone minor changes, made slowly to minimise any effect.
	A brief comparison of paired tows showed high variability in catches between vessels. No further calibration attempted.
	Fishing experience not desired, as deemed to be immaterial if consistent.
How are vessels calibrated?	Assume no significant vessel effect. Calibration experiment is to determine that no large effect exists.
Calibrated for all species with biomass estimates?	No
Approach Summary	The groundfish survey has evolved from a shrimp survey. Government vessels were used historically. States contributed survey tows to the time series from different vessels following the same net plan. Catch rates are variable in paired tests.
Standardise or calibrate?	Standardise to extent practical, test catch rates between vessels to ensure no detectable difference. Calibration testing does occur, but not on a species level, and not aimed at determining conversion factors.

Region:	**Western Canada, Department of Fisheries and Oceans**
Number of bottom trawl surveys in region:	Six (two are shrimp surveys): A) West coast Vancouver Island B) Queen Charlotte Sound C) Hecate Strait D) West coast Haida Gwaii E) Shrimp – West coast Vancouver Island F) Shrimp – Queen Charlotte Sound
Number of vessels involved:	Mainly one Government vessel, five commercial vessels varying among years depending on survey — see below.
Commercial or Research vessel platform?	Government: West Coast Vancouver Island (last four years) Hecate Strait last two of three surveys). Commercial: West Coast Haida Gwaii (different vessel each of four surveys), Queen Charlotte (one vessel, but no longer). Shrimp: Mostly Government, but one or two years with commercial.
Historical survey platform:	Typically a single government research vessel, but with exception of West Coast Haida Gwaii which has been commercial since inception in 2006.
Standardisation:	
Gear provided or specified:	Warps and doors and a package provided by vessel per standards.
	Trawl: plan specified, two owned by DFO for government vessel, two by stakeholders.
Vessel equipment specified:	Door-spread and headline height only.
Sonar:	Only for depth, bottom suitability.
Trawl mensuration:	Door spread, headline height, bottom contact, but not on all tows
Data acquisition system:	Tow scored but not determined valid.
Personnel:	
Skipper:	Government employed for most surveys, variable on commercial vessels. Note two Government shifts so variable staff.
Crew:	Government employed for most surveys, variable on commercial vessels. Note two Government shifts so variable staff.
Science staff:	Always have DFO staff onboard, contractors or crew may assist. Aim for whole haul sort.
Behaviours standardised:	Instruction provided but concern about skipper effect, even on government vessels with more than one crew. Especially because local knowledge can introduce bias. Instructions define towing speed, scope. Shrimp survey still standardises, but believes vessel effect is small. Interview with commercial skipper indicates that specialist knowledge is required to safeguard net, or fit a tow within the trawlable habitat of the designated block. It is not known if government skippers have equivalent specialist knowledge or how area-specific this knowledge may be.
Calibration	
Has survey vessel been modified over time?	No major changes, not discussed in detail.
Has survey gear been modified over time?	Some small changes in nets during repairs noticed, but otherwise not mentioned.
How are vessels calibrated?	Have not calibrated. Calibration is good idea but very expensive. Would calibrate if vessels or gear changed radically, but rely on standardisation. Calibration therefore in special circumstances.
Calibrated for all species with biomass estimates?	No.
Approach Summary	
Standardise or calibrate?	Standardise to extent practical, though not all potential factors influencing catchability are standardised, but any bias is assumed to be constant. Calibration testing does not occur (though some

	small gear comparisons have been done in the past). Calibration is supported if cost-effective.
Region:	**East Coast Canada, Department of Fisheries and Oceans**
Number of bottom trawl surveys in region:	Spring survey – one vessel. Autumn survey – two vessels. Shrimp survey – one vessel.
Number of vessels involved:	Three
Commercial or Research vessel platform?	Groundfish surveys are research vessel platforms. Shrimp survey is commercial platform. Two historical groundfish surveys by industry only, both terminated by industry.
Historical survey platform:	Research vessels.
Standardisation:	
Gear provided or specified:	All gear is government built. Commercial survey uses commercial gear held standard.
	Warps: diameter, construction, lay, length.
	Doors: specified or provided.
	Sweeps: provided.
	Trawl: Campelen trawl.
Vessel equipment specified:	Autotrawl - calibrated, power, doors.
Sonar:	No acoustic data collected.
Trawl mensuration:	Simrad acoustic – used since first Scanmar,
Data acquisition system:	Government.
Personnel:	
Skipper:	Government long term employees, Vessels fish majority of time.
Crew:	Government employees.
Science staff:	Government employees.
Behaviours standardized:	Vessel speed from deployment to net on deck, choice of fishing location, trawl mensuration envelope, warp scope, trawl component maintenance and replacement specifications, catch handling – all in protocols. Mensuration data not specifically used in tow validation. Behavioural biases minimized through detailed standard protocols. Surveys are 24 h, 2 crews × 2 vessels × rotate 28 d.
Intercalibration	
Has survey vessel been modified over time?	Yes, calibration experiments conducted in 1980s, 1990s (intensive), 2006 as vessels change.
Has survey gear been modified over time?	Yes, but calibrated in 1980s and 90s. Now gear is maintained as standard and expected to remain in the future.
Detail:	Calibration experiments have been conducted whenever vessel and or gear have changed, generating species length-based conversion factors for main species. Trouble with changing abundance and size classes through time. Final calibration experiments with most recent vessels not conducted due to vessel availability.
	New vessels expected in next five years, planning vessel calibration experiment.
	Commercial platform or Industry-only surveys require standardisation and vessel calibration if changes occur.
	Focussed experiments conducted on any changes that may influence catchability (adding CTD).
How are vessels calibrated?	Side-by side (or in line) paired tows, sample size dependent on species and size classes present, but large sample size needed (200–300 pairs).
Calibrated for all species with biomass estimates?	No, only 6–8 main species and for main sizes (i.e., if large fish were not present in the year calibrated). For other species, the survey time series is split as two series.
Approach Summary	
Standardize or calibrate?	High priority on both. Detailed protocols exist on all aspects of survey gear construction, maintenance, deployment, fishing

	behaviour, and analysis. All gear components specified with tight tolerances. Calibration experiments needed with any change in vessel (even sister ships), or gear change. Purposefully minimize gear options within surveys such as changing sweep length with depth.

Region: Western IBTS	IPIMAR, Portugal
Number of bottom trawl surveys in region:	Q4 IBTS – Portuguese Groundfish Survey running since 1979.
Number of vessels involved:	One (with replacement research vessel when not available).
Commercial or Research vessel platform?	Research.
Historical survey platform:	Research vessel.
Standardisation:	
Gear provided or specified:	Norwegian Campelan trawl with rollers – one of three trawl types used in Western IBTS.
Vessel equipment specified:	Scanmar sensors used.
Personnel:	
Behaviours standardised:	Tow duration and speed standardised.
Details:	
Calibration	
Has survey vessel been modified over time?	Primary vessel has been unavailable due to breakdown in 1996, 1999, 2003 and 2004, so replacement was used in these years. Vessels were of similar size but differ in tonnage and engine power (1500 Hp, versus 1200 Hp).
Has survey gear been modified over time?	Using a different vessel necessitated using a different gear with bobbins and different groundrope size, bigger wing mesh size, lower vertical and wider horizontal opening.
How are vessels calibrated?	Vessels fished side by side with a minimum distance of 0.25 nm. A total of 32 paired tows. Catch ratios estimated by haul – variability higher for small and intermediate sizes. Estimated the catch conversion factor and 95% confidence limits by fitting a spline based smooth curve to the relationship between overall catch ratio and hake length – the conversion factor for hake < 10 cm was 13, decreasing to 5–6 between 10–15 cm and flattening to 1 at 30 cm.
Calibrated for all species with biomass estimates?	
Approach Summary	
Standardise or calibrate?	Standardisation within vessels, but not complete between-vessel standardisation. Some calibration was conducted to estimate conversion factors.

Region: Western IBTS Surveys	AZTI - Tecnalia / Marine Research Unit, Basque Country
Number of bottom trawl surveys in region:	Two surveys.
Number of vessels involved:	
Commercial or Research vessel platform?	One survey uses a commercial vessel, the other a research vessel. Both will use research vessels as of 2012.
Standardisation:	
Gear provided or specified:	The same vessel is chartered every year with the same crew. The trawl is provided.
Vessel equipment specified:	The vessel's own gear mensuration is used.
	The catch is sampled following IBTS protocols.
Personnel:	
Behaviours standardised:	
Calibration	
Has survey vessel been modified over time?	
Has survey gear been modified over time?	
Detail:	With a new research vessel being introduced next year, the aim is to carry out some form of calibration exercise using parallel hauls with the old (commercial) and new (research) vessel. This will be subject to budget constraints.
Has survey vessel been modified over time?	
Has survey gear been modified over time?	
Detail:	With a new research vessel being introduced next year, the aim is to carry out some form of calibration exercise using parallel hauls with the old (commercial) and new (research) vessel. This will be subject to budget constraints.
How are vessels calibrated?	
Approach Summary	
Standardise or calibrate?	Noted that abundance indices from the commercial vessel were slightly higher than those recorded from the research vessel in the same sampling zone. Surveys from Research vessels only are the preferred option. Standardisation as per ICES co-ordinated programme although the gear used is specific to Spain and not the same as that used in surveys further north. Clearly intercalibration is considered useful although extent may be limited by cost.

Region: Western IBTS Surveys and others	**Marine Institute, Galway, Ireland (two interviews)**
Number of bottom trawl surveys in region:	Q4- IBTS-IRE – Irish groundfish surveys started on the west coast in 1990 covering parts of ICES divisions VIa south and VII. In 1997, another survey covering the Irish and Celtic Sea was started. Design and coverage has changed over the years. Q3 Irish ISP monk survey VIa and VII – intermittent participation since 2006 – absolute abundance survey. Q1 Celtic Sea cod survey – since 2009, intermittent. Q3 – Porcupine (VIIc) *Nephrops* trawl survey – since 2010.
Number and type of vessels involved:	Q4 – IBTS survey forms part of multi-vessel and multi-nation Western IBTS survey carried out by RV *Celtic Explorer*. Q3 Irish ISP monk survey – commercial and research. Q1 Celtic Sea cod survey – single commercial vessel. Q3 – Porcupine (VIIc) *Nephrops* trawl survey –single commercial vessel provided by Industry.
Historical survey platform:	Q4 – IBTS **-** Prior to 2003, the survey was divided into the west coast survey and Irish sea and Celtic sea survey. The latter was carried out by the smaller RV *Celtic Voyager*, with commercial vessels covering the west coast regions that the research vessel couldn't work in. The smaller research vessel used with a scaled down version of the survey net (GOV 28.9/37.1). Commercial vessels used their own gear fitted with 20 mm liner – the same vessel was chartered each year where possible. Since 2003, RV *Explorer* has carried out the combined survey. The other three surveys are relatively new and initiated with commercial fishing Industry input and support.
Standardisation:	
Gear provided or specified:	Q4 IBTS survey – uses a GOV trawl rigged as per the IBTS Manual. As well as the standard groundgear ("A"), a modified groundgear ("D") is also used, depending on the substrate. Two different sweep lengths used for different depths. Variations to recommended rigging, sweep lengths, doors and other aspects exists between countries.
	Q3 Irish ISP monk survey – uses a commercially designed monkfish trawl supplied by the Institute. Detailed manuals produced to ensure that the gears are rigged in the same way. Significant input of gear technologists as well as Industry.
	Q1 Celtic Sea cod survey – Trawl based on commercial design, with small mesh liner.
Vessel equipment specified:	Q4 IBTS survey – uses Scanmar units record headline height and door spread. Celtic Explorer carried an EK60 (18, 38, 120, 200 kHz) and is fully compliant with ICES noise requirements recommended by ICES. Equipped with fish processing lab.
	Q3 Irish ISP monk survey – the tender document requires that vessels must be a minimum of 27 m, and carry 2000 m wire, Morgere Ovalfoil OF12, 1700 kg (5.82 m^2) doors with a specified wire rig, and a hull mounted scanmar hydrophone. Door and wingspread sensors are supplied by the Research Institute along with bottom contact sensors. Catch is processed as per standard protocols, using paper sheets, transcribed into an electronic database.

	Q1 Celtic Sea cod survey – strong emphasis and extensive dialogue on maintenance of standardised methods to maintain stable catchability.
Personnel:	For IBTS surveys onboard *Celtic Explorer*, skipper and crew employed by Institute and survey carried out by science staff, mainly biologists. For monkfish survey, skipper and vessel crew chartered with vessel and must carry three scientific staff. Same arrangement of vessel crew plus scientists for other commercial surveys.
Behaviours:	
Detail:	Q4 IBTS survey - the survey design follows the IBTS recommendation for this region with a semi-random depth stratification based on a library of known clear tows. The catchability of certain species, such as cod is perceived as being low and there are concerns about the differences in catchability between Scottish and Irish components of the survey (ICES, WGCSE Report, 2010) and internal consistency of the time series (vessel and gear change in 2003). The Celtic Sea cod survey was a response to the above concerns. VMS logbook linked data were used to stratify the survey in close consultation with the Industry. Extensive consultation required in terms of ensuring the concepts of standardised survey design, stable catchability and appropriate sampling procedures. For surveys that deviate from normal fishing procedures, greater funding and scientific staffing is definitely advised. Noted that even with highly co-ordinated surveys, door / sweep angles can deviate up to 10–15º, which are likely to be a function of different doors and rigging and that this can have an impact on spreading power.
Calibration	
Has survey vessel been modified over time?	As above – change from RV *Celtic Voyager* (31.4 m) to RV *Celtic Explorer* (65 m) in 2003. No major calibration trials. The VIIb k area haddock assessment uses a swept area raising metric to maintain the time series.
Has survey gear been modified over time?	As above, change from scaled-down GOV to standard GOV for use with larger vessel and use of modified ground gear (D).
How are vessels calibrated? Detail:	Q4 IBTS survey – Pre-2003, the *Celtic Voyager* participated in comparative fishing trials with *Thalassa* and *Scotia* as part of IPROST project. Hauls were carried out side-by-side, no more than 0.5 nm apart. Locations of tows were relatively spread out. A multivariate approach was taken, using principal component analysis (PCA) biplots, to assess interactions between species and assess differences in species composition between hauls. For species with sufficient numbers, individual haul catch rates from one vessel were modelled relative to the other vessel and the fitted curves combined and bootstrapped confidence intervals estimates. There was no strong evidence that catch rates were different.
	The survey was not used in assessments and no calibration trials were carried out when *Explorer* replaced the smaller vessel. The current survey series is used in selected assessments. A retrospective conversion factor for *Voyager* v *Explorer* is based on indices standardised by swept area and using SURBA analysis to look at cohort tracking.
	ICES has recommended that different countries build a database of comparative hauls where time allows. This occurs on an ad-hoc basis where vessels happen to be in the same area. For *Celtic*

	Explorer, a relatively small database exists (10–30 total) between Irish vessel and Scottish, Spanish and French.
	Monkfish survey – no inter-calibration carried out.
Calibrated for all species with biomass estimates?	IBTS – for selected species that the survey series is used as a tuning index for, e.g. cod and haddock.
Approach Summary	A number of different surveys conducted with different approaches. Historical survey used commercial vessels where the research vessel was not capable of operating. With new, larger research vessel this has been the preferred option for this time series - produces relative abundance indices, along with considerable additional biological information and emphasis is on striving for better standardisation with ad-hoc, small scale calibration through paired tows where feasible to build up a data set over a number of years. The Western IBTS is not fully standardised and surveys are used selectively for particular stocks and often not in full. Where there is a feeling that survey has changed from the standard, suggests the value of swept area index (as opposed to number per hour as is traditional for IBTS) and monitoring age-based indices for year effects. Newer surveys make more use of commercial fishing experience and vessel availability. The monkfish survey is unique in aiming to produce absolute abundance estimates which reduce influence of vessel effects as long as gear and protocols are well standardised.
Standardise or calibrate?	Standardise gear, use vessels of similar size and power, scientific staffing / management and move to swept area absolute abundance indices.

Region: North Sea and Western IBTS Surveys	Marine Scotland (two interviews)
Number of bottom trawl surveys in region:	4 x IBTS – SCO (Q1 and 3 North Sea, Q1 and Q4 Ground Fish Surveys in Western division). Q3 Rockall haddock survey. Q3 West of Scotland (VIa) deepwater survey. Q4 ISP monkfish survey.
Number and type of vessels involved:	IBTS – multi-vessel surveys using research vessels (RV *Scotia* III). Rockall and deepwater surveys – single vessel research surveys (RV *Scotia* III). Monkfish survey carried out with combination of *Scotia* and chartered commercial vessels.
Historical survey platform:	R.V. *Scotia* III replaced *Scotia* II in 1998.
Standardisation:	
Gear:	IBTS – Scotia uses the standard 36/47 GOV for all surveys. In the North Sea the standard ground gear "A" (200 mm rubber discs) and a modified ground gear "B" (305 mm rubber bobbins) in northerly areas with harder ground. On west coast surveys, a third type of ground gear, "C", with 533 mm bobbins in the bosom section is used. Remainder of gear rigged as per IBTS manual apart from that the use of 110 m sweeps in depths over 70 m is not done. Q3 Rockall haddock survey uses GOV rigged with ground gear C. West of Scotland deepwater survey uses a commercial net with 533 mm rockhopper ground gear. In 2008, changes were made to the survey gear to bring it into line with the recently commenced Irish deepwater survey using 400 mm rock hopper discs. In 2009 tow time was reduced from 2 hr to 1hr This breaks the time series, but has improved the catching efficiency for many species. Monkfish survey – commercial design net is made to strict specifications and supplied to all vessels involved to ensure consistency. A quality control system is in place for repairs.
Vessel equipment:	IBTS surveys use scanmar sensors to monitor headline height, trawl speed, wing and door spread along with a bottom contact sensor. Use Hyper terminal to collect raw data strings. EK60 (18, 38, 120, 200 kHz). Catch sampling protocol as per IBTS manual.
	Q3 Rockall haddock survey – as above. West of Scotland Deepwater survey – use headline height, wing and door spread sensors down to 500 m, but only door spread at greater depth.
	Monkfish survey – detailed tender document as per Irish Appendix. Scanmar and bottom contact sensors supplied to the vessel.
Personnel:	IBTS – Vessel crew employed by commercial shipping company through competitive tender. Many are ex-fishermen with considerable experience. Survey carried out by scientific staff, biological background, many with extensive survey experience. In recent years, a gear technologist has been on each survey for the purpose of monitoring gear deployment, net geometry, and overseeing any repairs.
Details	IBTS North Sea – stratified by statistical square with fixed stations (from a library of clear tows).

	IBTS West coast survey – as above until 2011. Survey design changed to a stratified random design with stratification based on community structure from multivariate analysis of species composition. Station allocation based on variance in gadoid abundance. Monkfish survey design stratified primarily by depth, but also substrate and bottom water temperature, with station allocation based on fishermen's perception of density. Research vessel initially designated the area of lowest abundance (east of Scotland), but as a large area, was found to contain the highest total abundance. In following surveys, the RV was used instead to survey the Rockall strata; this is the most remote area with the poorest weather conditions, and also has a number of closed areas to protect cold water corals. *Scotia* is a larger vessel than the participating commercial vessels and also equipped with technology to be able to inspect survey tows to avoid trawling on coral. This survey aims for an absolute abundance estimate. Effort has gone into estimating the efficiency of the trawl gear for the target species and ensuring that the calculation of swept area is accurate. Have used video observations of behavioural reactions to parameterise an individual based model used to estimate herding efficiency of sweeps – only 4% of fish are herded into path of trawl (Reid et al 2007a). Losses under the fishing line have been estimated using catch bags – only 25% of fish under 30 cm are caught in the net. Swept area is calculated using GPS, scanmar readings (wings and doors) and bottom contact sensor data is used to adjust for periods when ground contact is lost (Pout, pers. comm.) Deepwater survey – depth stratified with fixed stations from library of known clear tows.
Behaviours standardised:	IBTS surveys – towing speed, warp lengths, gear mensuration, catch processing all as per IBTS manual. Deepwater survey – as above. Monkfish survey – survey sites pre-allocated and given to scientist in charge. Tow direction decided by the skipper, 1 hr duration and specified speed. Standard protocols followed for catch sampling carried out by scientists onboard with the remainder of the catch processed by the crew for landing.
Calibration	
Has survey vessel been modified over time?	There has been a change in vessel for the IBTS surveys when *Scotia* III replaced *Scotia* II. Deepwater and monkfish surveys all initiated after this vessel change. Monkfish surveys use three commercial vessels each year, which are not the same from year to year.
Has survey gear been modified over time?	IBTS – Pre-1991 used a Dutch herring trawl; 1992–1997 used a "Aberdeen 48 ft" trawl. From 1998 used GOV. Trawl and towing time (1 hr reduced to 30 min) changed with new vessel. In 2011, the ground gear and sweep lengths used on the west coast survey have been changed from a "C" to "D" to come into line with Irish surveys, with chain toggles attaching fishing line to ground gear replaced with steel rings and 110 m sweep lengths used in deeper water.
Detail:	Prior to the change in ground gear, comparative tows have taken place – one small scale experiment with just eight hauls (Burns &

	Stokes 2005), and another larger exercise with over 100 tows in one particular area (Rockall). Results not currently available. Have also undertaken twin trawl and alternate haul trials of the different ground gears used in the North Sea, found that both ground gears caught similar length ranges for gadoids and flatfish, but greater numbers of flatfish were caught with ground gear "A" (unpublished results).
How are vessels calibrated?	An inter-calibration exercise carried out between old vessel / gear and towing time and new vessel, gear and towing time. 24 paired tows side by side. Used a mixed-model approach to combine smooth size-selection and catch comparison curves over hauls with pointwise standard errors or confidence bands. Fixed effects included by allowing the mean selection parameters to vary between hauls according to some design matrix (Zuur et al. 2001, Fryer et al. 2003). Concluded that the null hypothesis of a constant relative catch rate of 0.5 was not strongly refuted for haddock, although there was some evidence that the new vessel with shorter tows caught more large haddock. The outcome was to apply a factor of 0.5 to calibrate abundances between the two vessels. As per recommendations from ICES, ad-hoc vessel comparisons are carried out where feasible – trials have been carried out with the German vessel – 14 paired tows (random positions within a limited defined area and within two days of each other). There were variations in net geometry achieved and *Scotia* caught more juvenile gadoids. For the monkfish survey there is no inter-calibration – an absolute swept area abundance index calculated.
Calibrated for all species with biomass estimates?	No, insufficient data for all species.
Approach Summary	For the IBTS surveys – accept that there are differences between vessels and between the variations in gear rigging. Emphasis placed on improving the application of standardisation protocols and assuming that the effects do not bias the results to a significant extent given the large variation in catch rates among hauls anyway. Recent changes to survey gear and design have been made to bring into line with IBTS recommendations and / or improve the data, at the potential cost of breaking the time series. For the monkfish survey, the emphasis is on same gear, fished in the same way (under scientific supervision) and using the same metrics to produce swept area abundance that is corrected for variations in gear geometry.
Standardise or calibrate?	Combination of both. Standardise with build-up of comparative fishing trials over time. Move to swept area absolute abundance where feasible.

Region: North Sea	**IMARES, Netherlands**
Number of bottom trawl surveys in region:	Q1 IBTS North Sea – cover western, central and southern North Sea. Targets gadoids, mackerel, herring and sprat. Q3 SNS – Solenet survey (coastal waters of south-eastern North Sea) provides abundance indices for juvenile North sea plaice and sole. DFS – Demersal Fish Survey – international inshore fish survey targeting demersal fish. BTS – international Beam trawl survey, western, central and southern North Sea. Target species North sea plaice and sole.
Number and type of vessels involved:	IBTS – RV *Tridens.* SNS – single research vessel (*Tridens* / *Isis*). DFS – three research vessels from Netherlands (*Isis, Stern, Schollevaar*) + vessels from three other countries. BTS – RV *Tridens* and *Isis* along with vessels from three other countries.
Historical survey platform:	IBTS – same research vessel. SNS – 1969 – 1995 used RV *Tridens*. Since 1995 used RV *Isis.* DFS – same three vessels used since 1969. BT – initiated in 1985 using RV *Isis*. Survey area enlarged in 1996 and *Tridens* included.
Standardisation:	
Gear:	IBTS – GOV trawl with ground gear "A". DFS – vessels use either 3 m or 6 m beam trawls. BTS – Netherlands uses 8 m beam trawl, other countries use 4 m or 7 m beam trawl.
Vessel equipment:	
Sonar:	
Trawl mensuration:	Scanmar headline and door spread used on IBTS surveys.
Data acquisition system:	
Personnel:	
Behaviours standardised:	IBTS – follow manuals. SNS – Standard tow speed (3.5–4 knots) and duration (15 min). DFS – standardised tow tie, speed and mesh sizes.
Detail	SNS – 10 transects running perpendicular or along the coast. DFS – stratified by geographic area and depth with fixed trawl positions. Use to produce plaice and sole indices. Number per 1000 m^2 and calculate a weighted average.
Calibration	
Has survey vessel been modified over time?	Attempts have been made to calibrate the catches of the *Tridens* and the *Isis* in the BTS but not successful. The two data sets are used as separate tuning indices along with the SNS survey (which combines data from the two vessels).
Has survey gear been modified over time?	
Detail:	
How are vessels calibrated?	
Calibrated for all species with biomass estimates?	
Approach Summary	Standardisation, ignore the vessel effects or use as separate tuning indices.
Standardise or calibrate?	

Region: Iceland	Icelandic Groundfish Survey
Number of bottom trawl surveys in region:	2 – a Spring (formally since 1985) and Autumn (since 1996) Groundfish survey that covers the Icelandic continental Shelf down to 500 m.
Number of vessels involved:	Over the history of the survey, 13 commercial trawlers and 2 research vessels have been used. In any one year, 4–5 vessels are used to cover the area
Commercial or Research vessel platform?	Started with only commercial, now a combination of both commercial and research vessels.
Historical survey platform:	At start of Spring time series, five identical commercial trawlers (Japanese) were rented by the Institute. With the acquisition of new research vessels in 1996, these have been also been used.
Detail:	The survey was initiated by commercial fishermen who worked with scientists to design a survey to assess cod stocks. Since then the Institute has become more involved. One of the key aims of the survey was to improve the relationship between the Institute and fishing Industry. Stratification is by statistical rectangle with squares grouped into 10 strata depending on cod density. Allocation of stations between strata was in proportion to area and density. Within each strata, stations allocated semi-randomly; half picked at random and half chosen by vessel skippers as known areas of high cod abundance. A total of 550–600 tows. These stations are now fixed and stratified by depth and water temperature. The survey area has been reduced since 1996 to reduce costs, but as more species added to be assessed, more stations have been added to cover their distribution.
Standardisation:	
Gear provided or specified:	Initially the commercial trawlers used their own gear, (Granton - a common commercial net). This has been standardised in cooperation with the skippers (in 1986?) and has remained relatively unchanged since then. Where feasible, all the nets used are made at the same factory because previous experience showed that different manufacturers produced nets from the same plans that fished differently. Also carry out manual checks in the gear workshops and ask cruise leaders to check that it is rigged according to the manual. Includes specified trawl doors and sweep lengths at different depths.
	The ground gear is particularly heavy (iron bobbins), and repair costs can be high, which is unpopular with the skippers, and headline height is low compared to modern trawls.
Vessel equipment specified:	Use scanmar to monitor door angle and spread and headline height and also trawl-mounted hydrographic sensors (water temperature). Use Hyper-terminal to collect raw data strings from scanmar units.
	Commercial vessels don't generally have calibrated scientific echosounders so only collect acoustic data on the research vessels.
	Institute has developed a computerised catch sampling system which automatically flags which fish should be biologically sampled. For surveys on the commercial vessel, the scientists build a working area with bench space etc in the fish processing factory.
Personnel:	Combination of commercial crew and scientists on the commercial vessels and Institute-employed crew and scientific staff on the research vessels.
Behaviours standardised:	Towing speed, distance and the tow path, but not direction, trawl geometry all standardised. Initially protocol was a fixed amount of warp out, but it was apparent that some skippers were sometimes using autotrawl to minimise risk of gear damage – surveys were done for a fixed cost so time spent repairing nets was not

	desirable. Hence use of autotrawl is now part of the standard protocol.
Calibration	
Has survey vessel been modified over time?	The same style of commercial trawlers have been used throughout the survey history, but they have been variously modified over time in terms of length and engine power. Research vessels have been included latterly.
Has survey gear been modified over time?	Very small changes have been made to the trawl gear. However, door weight has increased from around 1750 kg to 1900 kg.
	Changes have occurred in design as described above.
How are vessels calibrated?	A series of calibration trials were carried out over four years to compare catches between the research vessel and the commercial trawlers and also between the commercial trawlers. Paired hauls with successive tows along the same track, with a 1.5 hour gap. A disturbance factor calculated by allowing the same vessel to carry out repeated tows. 50–111 hauls per experiment. A GLM with quasi log link and three different variance functions to evaluate results. Also measured the noise signature of the different vessels Regardless of model used, the research vessel had higher catch rates whilst the catch rates of the commercial trawlers was not significantly different. Do not use a conversion factor.
Calibrated for all species with biomass estimates?	Key target species; cod, haddock, saithe, redfish, plaice and catfish.
Approach Summary	Focus is on standardisation, ignoring the vessel effect although there is concern that as the old commercial trawlers are replaced or phased out, newer vessels will need to be found and this will present issues in the future.
Standardise or calibrate?	

Region: Mediterranean	**Institute of Marine Sciences of the National Research Council (ISMAR-CNR), Ancona**
Number of bottom trawl surveys in region:	MEDITS survey – Mediterranean International Trawl Survey covering continental shelf and upper slope from 10–800 m. Total of 1100 hauls.
Number of vessels involved:	8–11 vessels from multiple countries. Initially France, Spain, Italy and Greece, but area has expanded to include other countries.
Commercial or Research vessel platform?	Both research and commercial.
Historical survey platform:	Before 1994, an Italian trawl survey was carried out using commercial vessels based locally in seven different regions with their own trawl gear (GRUND). Now Italian part of MEDITS uses one research vessel and six commercial vessels.
Standardisation:	
Gear provided or specified:	With the start of MEDITS, the trawl gear was changed to a standard French design (GOC). Gear technologists are employed to ensure the trawl is properly rigged on each vessel. Trawl rigging detailed in manual, bridle length specified for shallower and deeper than 200 m, door type, warp diameter and depth ratio specified.
Vessel equipment specified:	Scanmar recommended, but only used on some vessels. Vemco mini temperature logger used on all trawl headlines since 1998.
Personnel:	Vessel crew chartered with commercial boats, scientific staff carry out the catch sampling.
	Full catch sampling usually achieved on research vessels, but, depending on facilities, on some commercial vessels, the catch is preserved and biological analysis carried out on return to the laboratory.
Behaviours standardised:	Towing speed and duration (30 min for less than 200 m, 60 min for hauls greater than 200 m). Detailed description of shooting and hauling operations.
Details:	Stratified by depth with stations allocated randomly within strata, which are now fixed.
Calibration	
Has survey vessel been modified over time?	The Manual recommends using the same commercial vessels if feasible over time. This is the case in Italy, but a new research vessel has been introduced.
Has survey gear been modified over time?	Has been modified slightly to improve bottom contact at depths over 200 m – floatation reduced, links between footrope and net shortened and tickler chain removed in1996.
How are vessels calibrated?	No calibration between vessels has been carried out. Alternate haul trials were carried out in 1996 comparing the standard MEDITS trawl to a typical traditional trawl used by local fishermen (Fiorentini et al. 1999). Four fishing trips, each carrying out between 9–20 hauls. Different nets were used on alternate days. Catch data were converted into abundance per swept area and geometric means used as estimators to compare efficiency. Found higher variance among hauls than among vessels, but concluded that the MEDITS trawl was less efficient for a number of target species than the traditional commercial trawl. These conversion factors have been applied to compare MEDITS with historical survey.
Calibrated for all species with biomass estimates?	
Approach Summary	
Standardise or calibrate?	Standardise and put effort into ensuring that the trawl is fishing properly on all vessels involved by carrying out gear trials. Also suggests overlap of survey area where feasible.