# Data analysis methods for blue cod potting surveys

R. I. C. C. Francis

NIWA
Private Bag 14901
Wellington 6241

**New Zealand Fisheries Assessment Report 2011/27**
**June 2011**

**Published by Ministry of Fisheries**
**Wellington**
**2011**

**ISSN 1175-1584 (print)**
**ISSN 1179-5352 (online)**

**©**
**Ministry of Fisheries**
**2011**

Francis, R.I.C.C. (2011).
Data analysis methods for blue cod potting surveys.
*New Zealand Fisheries Assessment Report 2011/27.*

This series continues the informal
New Zealand Fisheries Assessment Research Document series
which ceased at the end of 1999.

## EXECUTIVE SUMMARY


**Francis, R. I. C. C. (2011). Data analysis methods for blue cod potting surveys.**

***New Zealand Fisheries Assessment Report 2011/27.***

This report describes work carried out under project IPA200918, whose scope was *To review historical [blue cod potting] survey data and provided recommendations on standard forms of data analysis to determine trends in population abundance and estimates of* Z *[total instantaneous mortality]*. The main conclusions and recommendations from this work are as follows.

1. The current method of calculating relative abundance indices from these surveys is appropriate, but the units for these indices should be standardised, probably to kg/pot (or kg/lift).

2. The c.v.s for these relative abundance indices should be calculated using set-based equations, rather than the current pot-based equations.

3. The general approach used since 2008 to estimate *Z* is appropriate, but errors have been made in its implementation.

4. Some small modifications are suggested to the simulation approach used to estimate the precision of *Z* estimates.

5. No evidence was found of substantial and consistent effects of station, depth, or environmental data on estimates of abundance, length/age frequency, or sex ratio.

6. For some areas, some evidence was found of between-subarea variation in age-length keys but it is doubtful whether it would be worthwhile to collect more otoliths in order to calculate separate age-length keys for each subarea. It is possible to structure the otolith sample to reduce any bias arising from spatial heterogeneity in age-length keys.

# 1. INTRODUCTION

Recreational blue cod stock status in the South Island is currently monitored using potting surveys, repeated about every three to four years, in areas corresponding to the key fisheries. This report, which was funded by MFish under project IPA200918, provides recommendations concerning the analysis of data from these surveys.

The scope of this project was

> To review historical survey data and provide recommendations on standard forms of data analysis to determine trends in population abundance and estimates of *Z* [total instantaneous mortality].

and the specific outcomes requested were

> 1. To determine the most appropriate analyses of relative abundance and associated precision, and estimates of Z, and
>
> 2. Using GLM (or other methods) to determine if the abundance estimates, length/age frequency, and sex ratio are affected by station, depth, or other environmental variables (where we have data), and
>
> 3. To determine whether there are sufficient age data collected to develop a separate age-length key for each stratum.

This project was done in conjunction with another MFish project (IPA200910), whose scope was *To develop a blue cod potting survey manual documenting standardised survey design, gear specifications and report outputs*.

## 1.1 Terminology

To avoid the confusion that has sometimes occurred in the reporting and discussion of previous surveys, a set of agreed terminology was devised, in conjunction with project IPA200910. The full set of terminology will be provided in the report from IPA200910; Table 1 includes just those terms needed for the present project.

The key point to notice about Table 1 is the difference between the three distinct concepts of 'site', 'set', and 'station'. The surveys analysed in this report used only fixed sites (though random sites will be used in future surveys). These sites were defined before the first survey in each area, and were given labels unique to that area (usually, the sites for stratum 2, say, were labelled 2A, 2B, 2C, etc). In contrast, sets were numbered sequentially within each survey (so set numbers are unique within each survey). The distinction between 'site' and 'set' is not so important in the context of a single survey, but becomes crucial when we are discussing results from multiple surveys in the same area. For example, the two Kaikoura surveys collectively had 50 sets, but used only 35 sites. Each set involved deploying a pot at each of either nine (for Marlborough Sounds) or six (for other areas) stations located close to the selected site. One source of confusion in previous reports is that the term 'station' has been variously used to mean what is here labelled 'site', 'set', or 'station'.

**Table 1: Some terminology used in this report (these definitions are an abbreviated subset of a more complete set that will be given in the report from project IPA200910).**

| Term | Definition |
|---|---|
| Site | A geographical location near which sampling may take place during a survey. It may be specified as a latitude and longitude or a section of coastline. |
| Fixed site | A site, defined before the first survey in a given area, that has a fixed known location (single latitude and longitude or length of coastline) and is available to be used repeatedly on subsequent surveys in that area. Which fixed sites are used in a particular survey is determined by random selection from the set of available fixed sites in each stratum. Sometimes referred to as an index site or fisher-defined site. |
| Random site | A site that can have any location (single latitude and longitude) generated randomly from within a stratum, given the constraints of proximity to other selected sites, for a specific survey. |
| Set | A group or cluster of pots deployed in the vicinity of a selected site in a specific survey. |
| Station | The position (latitude and longitude) at which a single pot (or other fishing gear) is deployed at a site during a survey. |

## 2. DATA

Data from 13 surveys in seven areas were available for this project (Table 2). The two earliest Marlborough Sounds surveys (in 1995 and 1996) were not included because there was some doubt as to whether their data were comparable to those from later surveys. In particular, soak times were highly variable in these surveys (they exceeded 2 h at more than a third of stations, and sometimes exceeded 12 h), whereas they were almost always close to 1 h in more recent surveys. Two recently completed surveys (gol0801 in Dusky Sound and nim0901 in North Otago) were not included because data from these surveys had still not been loaded into the appropriate databases.

**Table 2: The 13 surveys from which data were analysed in this project, grouped by survey area. The environmental data include air and sea conditions and temperatures (see text for details).**

| Area | Year | Trip code | Environmental data? | Area code | Reference |
|---|---|---|---|---|---|
| Banks Peninsula | 2002 | chj0201 | | BNKS | Beentjes & Carbines (2003) |
| | 2005 | chj0501 | | BNKS | Beentjes & Carbines (2006) |
| | 2008 | chj0801 | Yes[1] | BNKS | Beentjes & Carbines (2009) |
| Kaikoura | 2004 | mys0401 | | KAIK | Carbines & Beentjes (2006a) |
| | 2007 | mys0701 | Yes[1] | KAIK | Carbines & Beentjes (2009) |
| Motunau | 2005 | nav0501 | | MOTN | Carbines & Beentjes (2006a) |
| | 2008 | leg0801 | Yes[1] | MOTN | Carbines & Beentjes (2009) |
| Dusky Sound | 2002 | slt0201 | | DUSK | Carbines & Beentjes (2003) |
| North Otago | 2005 | suz0501 | | OTAG | Carbines & Beentjes (2006b) |
| Paterson Inlet | 2006 | gol0601 | Yes[1] | PATE | Carbines (2007) |
| Marlborough Sounds | 2001 | lhr0101 | | SNDS | Blackwell (2002) |
| | 2004 | lhr0401 | Partial[2] | SNDS | Blackwell (2005) |
| | 2007 | lhr0701 | Partial[2] | SNDS | Blackwell (2008) |

[1]All 15 variables recorded for all sets; [2]Only some variables recorded

## 2.1 Data extraction

The following data were extracted for each survey in Table 2:

– all stratum and station data in tables t_stratum, t_station of the *trawl* database (excluding stations where methods other than potting were used);
– all blue cod catch weight, and lengths by sex, from tables t_catch and t_lgth of the *trawl* database; and
– all blue cod ages (and associated data) from the *age* database.

Stratum data for trips with no data in table t_stratum were either inferred from other trips in the same area (for chj0501 mys0401 nav0501) or read from the relevant FAR (for suz0501). Where stratum areas were inconsistent with those in the survey reports (FARs or FRRs), the latter were used (this occurred for trips lhr0101 & slt0201).

Stations not used in the survey reports for calculating potting catch rates were ignored. As a check, the number of sets (and pots) per stratum, and the mean catch rates (kg per pot) by stratum, were compared to tabulated values in the survey reports (e.g., see table 3 in Beentjes & Carbines 2003). The numbers of sets and pots were always exactly as tabulated, and the mean catch rates were always the same, or close to, the tabulated values (some slight differences are to be expected if catch weights used in the reports were recalculated from length frequencies and length-weight coefficients).

## 2.2 Data preparation

In preparation for the analyses below a table of data by set was prepared. This included mean catch rates for all blue cod, and also all 'recruited' blue cod (those with length greater than or equal to 30 cm); mean length and age; proportion male; and a suite of environmental variables (where available).

All catch rates were calculated as kg/pot (i.e., without adjusting for soak time) as has been the practice in past surveys (see Section 3.2.1 for more discussion of this). Catch rates for recruited fish were calculated as $C\left(\sum_i^{\text{rec}} aL_i^b\right)\Big/\left(\sum_i aL_i^b\right)$, where $C$ is the mean catch rate for all fish, $L_i$ is the length of the $i$th measured fish, $a$ (= 0.01224) and $b$ (= 3.0746) are the default blue cod length-weight parameters in the *rdb* database, and the first summation was limited to fish with $L_i \geq 30$ cm.

For both length and age, mean values were calculated as the simple average of sex-specific means to avoid these values being affected by variations in sex ratio. The mean ages were calculated from age frequencies calculated by applying survey- and sex-specific age-length keys to the sex-specific length frequencies for each set. In the following analyses, mean lengths and ages were deemed unreliable (and thus ignored) unless at least 10 fish of each sex were measured (which occurred in only 243 of the 499 sets); for proportion male, the minimum sample size (for both sexes combined) was set at 12 (this condition, which was chosen to avoid proportions male of 0 or 1, and thus allow the use of a logistic transformation, was met in 359 sets). Unsexed fish, which were ignored in these calculations, were rare in the length frequency samples (1% or less) in most trips, the exceptions being the last two Marlborough Sounds surveys (where about 9% were unsexed) and the Dusky Sound survey (where 86% were unsexed, because most fish were released alive). For these latter surveys, a visual comparison suggested that the length distributions of sexed and unsexed fish were broadly similar, so ignoring unsexed fish should not matter.

The environmental variables are those included in table t_station in the *trawl* database (Table 3). These were all available for only four surveys (see Table 2) and were recorded by set, rather than by station, so there was no need to average across pots in a set. Where sensible, these were treated as continuous variables (e.g., the 10 levels for sea condition, and the 6 for bottom contour, both describe increasing degrees of roughness); otherwise they were treated as categorical.

**Table 3: Environmental variables recorded by set for some surveys, their units, and their treatment in the following analyses.**

| Variable | Units | Treatment |
|---|---|---|
| Cloud cover | Eighths (0 to 8) | Continuous |
| Air temperature | C | Continuous |
| Air pressure | Millibars | Continuous |
| Wind direction | Degrees true[1] | Categorical[2] |
| Wind force | Beaufort scale | Continuous |
| Wind speed | $m\ s^{-1}$ | Continuous |
| Sea condition | Code (from 0 to 9) | Continuous |
| Sea colour | Code (from 01 to 08) | Categorical |
| Swell height | Code (from 1 to 3) | Categorical |
| Swell direction | Degrees true[1] | Categorical[2] |
| Secchi depth | m | Continuous |
| Surface temperature | C | Continuous |
| Bottom temperature | C | Continuous |
| Bottom type | Code (from 0 to 12)[3] | Continuous |
| Bottom contour | Code (from 0 to 5)[3] | Continuous |

[1]A value of 999 implies no discernable direction
[2] Level 0 corresponds to a value of 999; levels 1-4 correspond to the cardinal directions E, S, W, & N
[3] The value of 0, signifying 'unknown', was never recorded in the data used here

For this report, all non-numeric stratum names were replaced by single-digit numbers as given in Table 4.

**Table 4: Stratum numbers used in this report for those surveys using alphabetic stratum names (the numbers given for the SNDS surveys are actually used in the databases – except that 10 is used instead of 0 – but not in the survey reports).**

| Marlborough Sounds surveys | | Dusky Sound survey | |
|---|---|---|---|
| Stratum name | Stratum number | Stratum name | Stratum number |
| SEPR | 0 | inner | 1 |
| IQCH | 1 | mid | 2 |
| OQCH | 2 | outer | 3 |
| EQCH | 3 | extreme outer | 4 |
| EOPE | 4 | open coast | 5 |
| OPEL | 5 | | |
| DURE (or DURV) | 6 | | |
| IPEL | 7 | | |
| MPEL | 8 | | |
| DURW | 9 | | |

# 3. ANALYSES AND CONCLUSIONS

## 3.1 The effect of ancillary variables

The first set of analyses addressed Outcome 2

> Using GLM (or other methods) to determine if the abundance estimates, length/age frequency, and sex ratio are affected by station, depth, or other environmental variables (where we have data).

which concerns the relationship between three types of *ancillary variables* (station, depth, and environmental) and three types of *outputs* (abundance estimates, length/age frequency, and sex ratio) [note that 'ancillary' is used here in its normal, non-technical, sense, meaning auxiliary or supplementary]. In this context I interpreted 'station' to mean site (as defined above), and the environmental variables to be those given in Table 3. Two abundance estimates were considered (mean catch rate for all fish, and for 'recruited' fish) and proportion male was taken to represent sex ratio. The only effects on length and age frequencies that were considered were those on mean length and mean age. Thus I considered the effects of 17 ancillary variables (site, depth, and 15 environmental variables) on each of 5 outputs (the two catch rates, mean length and age, and proportion male).

To address Outcome 2 we need a simple measure of how much any given ancillary variable(s) 'affects' an output (such as catch rate). The measure used below is percent variation explained (PVE), which was calculated as

$$\mathrm{PVE} = 100\frac{\sum_i \left(Y_i^{\mathrm{null}} - Y_i\right)^2 - \sum_i \left(Y_i^{\mathrm{full}} - Y_i\right)^2}{\sum_i \left(Y_i^{\mathrm{null}} - Y_i\right)^2}$$

where $Y_i$ is the (possibly transformed) value of the output at the *i*th site, and $Y_i^{\mathrm{null}}$ and $Y_i^{\mathrm{full}}$ are estimates of this value from two models which differ only in that the former does not use the ancillary variable(s), but the latter does. The null model used only survey and stratum as predictors, so it estimated the value of an output at a given site in a given survey as the mean of the values of that output at all other sites in the same stratum and survey (in the usual linear model notation this model could be written as $Y \sim$ survey x stratum). The catch rates were log transformed after zero catches were replaced by a small number (0.01 kg/pot), and proportion male was transformed with a logit function. These transformations ensure that the models can't predict nonsensical values (e.g., negative catch rates or proportions greater than 1). Since depth is confounded with stratum, particularly in Banks Peninsula and Kaikoura, all set depths were divided by the mean set depth for that stratum and survey, so the predictor to be evaluated was actually relative depth (Figure 1). [I also evaluated the possibility of calculating relative depth by subtracting, rather than dividing by, the mean depth, but found that the resulting PVEs were broadly similar, and so have not presented them here].

In exploratory studies like this, which consider many potential predictors, there is a danger that chance correlations will produce misleading results (Francis 2006). Two steps were taken to avoid this. First, PVE was calculated using a 'leave-one-out' cross-validation procedure. That is, $Y_i^{\mathrm{null}}$ and $Y_i^{\mathrm{full}}$ were each estimated using models whose parameters were estimated using a data set from which the *i*th set was excluded. Second, the environmental predictors were treated as a group, and a forward stepwise predictor-screening procedure was used to decide which (if any) of these predictors would be used to estimate $Y_i^{\mathrm{full}}$ (this decision would vary depending on *i* and the output being estimated). The R function, *step*, was used for this predictor screening (this function uses the Akaike Information Criterion (AIC) to determine which predictors were used in the full model). Note that when calculated using cross-validation, PVE can be negative, which suggests that the predictor being evaluated is worse than useless (i.e., estimates of the output from the full model will, on average, be *worse* than those from the null model, which does not use the predictor).

Since the relationship between output and predictor (e.g., catch rate and depth) may vary between areas, and possibly between surveys, it was important to repeat the calculation of PVEs for a range of subsets of the data defined by area and survey.

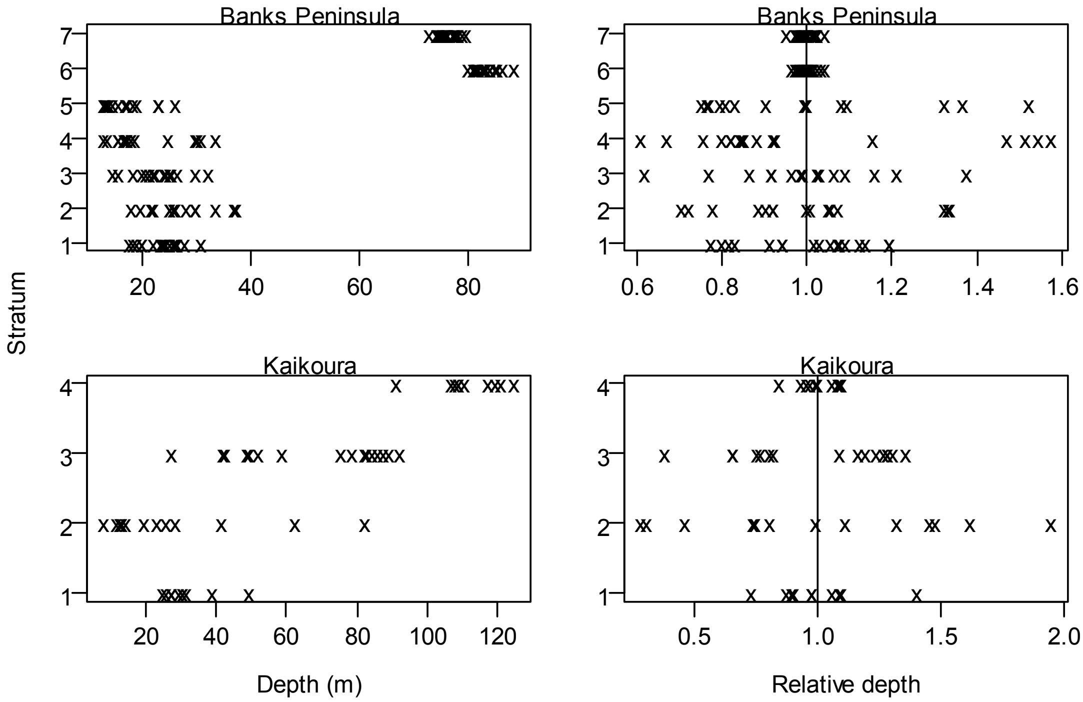


**Figure 1: Depths (left panels) and relative depths (right panels) of sets by stratum in each of two survey areas – Banks Peninsula and Kaikoura (each plotted point corresponds to a set in a survey in the given area).**

### 3.1.1 Results for site and depth

PVEs for site as a predictor varied widely from -230 to 93, but were mostly negative, and always negative when calculated over all areas (Table 5A). There are several points to consider in interpreting these results. First, if we restrict attention to the two areas where there have been three surveys (BNKS and SNDS), we sometimes find a positive PVE for a pair of surveys (e.g., 62 for proportion male in 2002 & 2008 in BNKS; and 93 for mean age in 2001 & 2007 in SNDS), but never for all three surveys in the area. This suggests we should be cautious in interpreting the two positive PVE values in MOTN, where we have only two surveys (because the apparent ability to predict site means may disappear when further surveys have been done). Second, the tabulated PVEs are only estimates, which will be particularly poor when sample sizes are low (note that the three largest PVE values – 93, 62, and 44 – are each based on only 8 sets). Sample sizes were much smaller for mean length & age (and to a lesser extent for proportion male) because sets with small numbers of measured fish were ignored (see above). To illustrate the importance of cross-validation, note that the PVE calculated for mean catch rate using all three Marlborough Sounds surveys was 66, without cross-validation, compared to -23 with cross-validation.

**Table 5: Percent variance explained (PVE) by A, site, and B, (relative) depth, in estimating each of 5 outputs (mean catch rates of all fish and of recruited fish; proportion male; and mean length and age) for specified combinations of area and surveys. Also given, for each PVE value, is the number of sets (Nset) on which it was based.**

**A, Predictor site**

| | | mncatch.all | | mncatch.rec | | pmale | | mnlen | | mnage | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Area | Surveys | Nset | PVE | Nset | PVE | Nset | PVE | Nset | PVE | Nset | PVE |
| BNKS | 2002, 2005 | 48 | -123 | 42 | -101 | 18 | -195 | 4 | -230 | 4 | -13 |
| | 2002, 2008 | 44 | -89 | 44 | -72 | 8 | 62 | 0 | – | 0 | – |
| | 2005, 2008 | 44 | -76 | 36 | -65 | 14 | 5 | 4 | -181 | 4 | -167 |
| | 2002, 2005, 2008 | 100 | -73 | 93 | -111 | 40 | -102 | 8 | -181 | 8 | -167 |
| KAIK | 2004, 2007 | 30 | -138 | 30 | -167 | 20 | -162 | 10 | -17 | 10 | -20 |
| MOTN | 2005, 2008 | 18 | 18 | 18 | 34 | 14 | -59 | 12 | -56 | 12 | -69 |
| SNDS | 2001, 2004 | 52 | 1 | 52 | 23 | 42 | -133 | 18 | -10 | 18 | -6 |
| | 2001, 2007 | 52 | -22 | 52 | -69 | 30 | -109 | 8 | 44 | 8 | 93 |
| | 2004, 2007 | 94 | -46 | 94 | -91 | 64 | -91 | 30 | -127 | 30 | -116 |
| | 2001, 2004, 2007 | 150 | -23 | 150 | -42 | 100 | -87 | 44 | -92 | 44 | -70 |
| All | All | 298 | -45 | 291 | -72 | 174 | -100 | 74 | -84 | 74 | -121 |

**B, Predictor (relative) depth**

| | | mncatch.all | | mncatch.rec | | pmale | | mnlen | | mnage | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Area | Surveys | Nset | PVE | Nset | PVE | Nset | PVE | Nset | PVE | Nset | PVE |
| BNKS | 2002, 2005 | 80 | 1 | 76 | 3 | 50 | 3 | 27 | -9 | 27 | -4 |
| | 2002, 2008 | 80 | 2 | 80 | 3 | 45 | -1 | 19 | -2 | 19 | -1 |
| | 2005, 2008 | 80 | -1 | 76 | -1 | 55 | -4 | 24 | -29 | 24 | -7 |
| | 2002, 2005, 2008 | 120 | 2 | 116 | 2 | 75 | 2 | 35 | -5 | 35 | -2 |
| KAIK | 2004, 2007 | 50 | -8 | 50 | -19 | 45 | -17 | 33 | 43 | 33 | 39 |
| MOTN | 2005, 2008 | 39 | 3 | 39 | -3 | 37 | 40 | 33 | 3 | 33 | 5 |
| SNDS | 2001, 2004 | 107 | 9 | 107 | 7 | 79 | 0 | 56 | 2 | 56 | 4 |
| | 2001, 2007 | 113 | 3 | 113 | 5 | 80 | -1 | 42 | -3 | 42 | 0 |
| | 2004, 2007 | 136 | 13 | 136 | 12 | 95 | -2 | 60 | -2 | 60 | -1 |
| | 2001, 2004, 2007 | 178 | 10 | 178 | 9 | 127 | 0 | 79 | 1 | 79 | 2 |
| All | All | 499 | 5 | 495 | 4 | 354 | -1 | 232 | 3 | 232 | 1 |

Sample sizes were usually much larger when calculating PVEs for depth as a predictor (cf values of Nset in Table 5A and 5B). This is because all sets at sites which occurred in only one survey must be omitted when calculating PVEs for site, but can be retained for those for depth.

PVEs for depth as a predictor covered a much narrower range (from –29 to 43) and were more often positive than negative (Table 5B). In BNKS and SNDS, 8 of the 10 PVEs calculated for all surveys in the area were positive, though always small (≤ 10). In these areas the most promising results were for the two catch rate outputs in SNDS, where PVEs were positive (though small) for all combinations of trips. However, an exploratory plot suggested that these results may be driven by just a few points from stratum 0 (where catch rates from the deepest sets were high) (Figure 2). This was confirmed when PVEs for this area were recalculated without stratum 0, producing values of 4 (2001 & 2004), –3 (2001 & 2007), 1 (2004 & 2007), and 0 (all years). A plot of the most promising results in the other two areas suggested that the high PVE for mean length in Kaikoura was mostly due to stratum 3 (Figure

3A) (the analogous plot for mean age in this area looked similar), and that the result for proportion male in Motunau may be driven by a few outliers (Figure 3B).

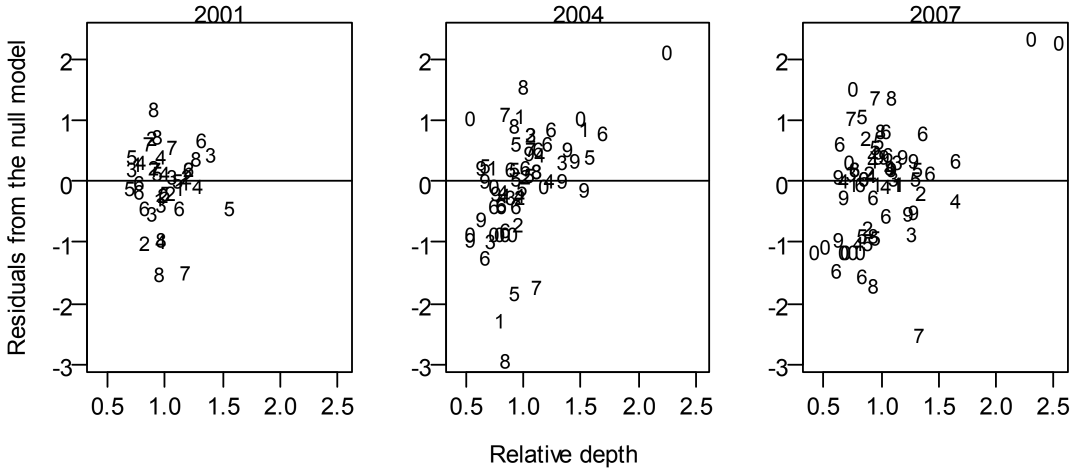


**Figure 2: Illustration of the utility of relative depth (plotted on the x-axis) as a predictor of mean catch rate (for all fish) for surveys in area SNDS (Marlborough Sounds). Each panel plots data from one survey; the y-axis shows residuals from the null model (which predicts catch rate from year and stratum) fitted to all surveys, and the plotting symbol indicates the stratum.**

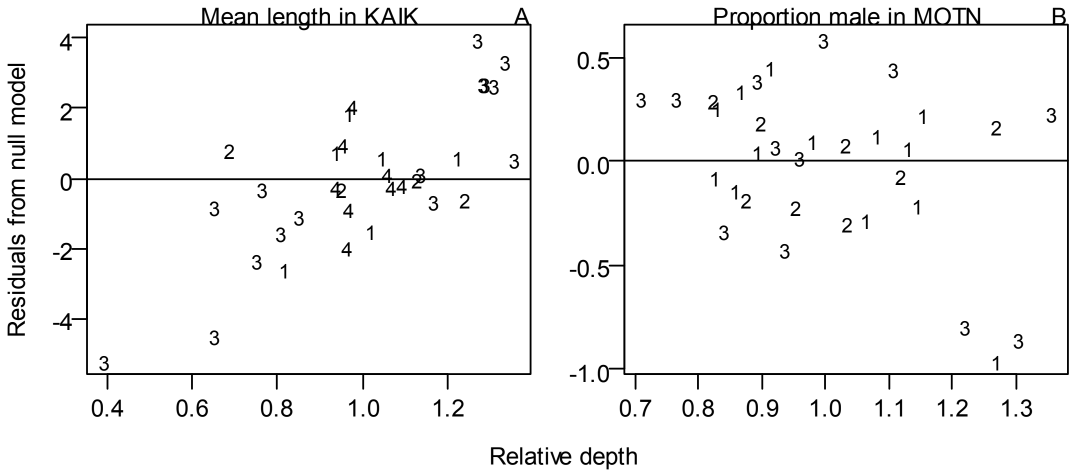


**Figure 3: Illustration of the utility of relative depth (plotted on the x-axis) as a predictor of A, mean length in Kaikoura (KAIK), and B, proportion male in Motunau (MOTN). In each panel the y-axis shows residuals from the null model (which uses year and stratum as predictors), and the plotting symbol indicates the stratum. Residuals for mean length are in cm, but those for proportion male are on a logistic scale.**

In calculating the PVEs for Table 5, the predictor (site or depth) was always included in the full model. An alternative approach would be to do as was done with the environmental predictors and use AIC to determine whether the predictor was included in the full model (with this decision being made separately for each output, and for each iteration of the cross-validation procedure). When this alternative approach was used, most of the PVEs changed, but the general conclusions given above did not. There were only three cases where the estimated PVE changed sign (all in area SNDS, with predictor depth, and outputs mean length or age) and the change was always from positive to negative.

### 3.1.2 Results for environmental predictors

PVEs for the environmental predictors were all negative, ranging from -181 to -5 (Table 6). No separate PVEs were calculated by area because this would restrict the available data to single trips, which means that the number of sets was sometimes less than, and never much more than, the maximum number of parameters estimated in the full model (recall that a categorical predictor with *n* levels requires an additional *n*–1 parameters).

**Table 6: Percent variance explained (PVE) by environmental variables in estimating each of 5 outputs (mean catch rates of all fish and of recruited fish; proportion male; and mean length and age). Also given, for each PVE value, is the number of sets (Nset) on which it was based.**

| mncatch.all | | mncatch.rec | | pmale | | mnlen | | mnage | |
|---|---|---|---|---|---|---|---|---|---|
| Nset | PVE | Nset | PVE | Nset | PVE | Nset | PVE | Nset | PVE |
| 117 | -18 | 117 | -17 | 65 | -5 | 40 | -181 | 40 | -73 |

### 3.1.3 Discussion of results for Outcome 2

Overall, the results for Outcome 2 were discouraging. No substantial and consistent relationships were found. This does not mean that none of the ancillary variables considered affect any of the 5 survey outputs. It simply means that we found no clear evidence of such effects. This is perhaps not surprising considering the limited data that were available and the approach adopted.

For example, consider the effect of site on mean catch rates. Loosely speaking, we were seeking evidence that there are some sites where catch rates are typically above average (for their stratum), and others where they are below average. This seems likely to be true, but given the day-to-day variation in catch rates at any given site, we may require data from several surveys before we can clearly see which are the above- and below-average sites. Note also that there was no area with more than 3 surveys in the data available for this study. Also, there were relatively few sites that occurred in all 3 surveys.

Another point to note is that whether we find an effect depends on how we choose to measure that effect. Consider an alternative approach using rank correlation. If we rank the sites within each stratum and survey by mean catch rate (as illustrated in Figure 4), we can calculate, for each pair of surveys, the correlation between the sets of ranks. For the three panels in Figure 4 these correlations were 0.64, 0.28, and 0.67, respectively, suggesting that site does affect catch rate in the Marlborough Sounds surveys. A randomisation test (which involved recalculating the correlations many times after randomly re-assigning the observed catches to different stations) showed that the larger two of these correlations were statistically significant (*P* values were 0.028, 0.466, and 0.003, respectively). [The same analyses applied to other areas found one significant correlation in Banks Peninsula, but none in either Kaikoura or Motunau.]

Two problems limited the chance of detecting any effect of environmental variables. First, there is a potentially large number of parameters to estimate, because there are many environmental variables, and some are categorical. For example, a full model to predict mean catch using all the environmental variables requires 45 parameters, and there are only 117 data points (i.e., sets) available for this model. This high ratio of parameters to data points means that it is difficult to distinguish between real correlations and those that might occur by chance in such a small data set. The second problem is that the effects of some environmental variables are likely to vary between areas (e.g., a wind, or swell, direction that produces low catch rates in one area may not do so in another). There were not sufficient data to allow the environmental analyses to be restricted to single areas.

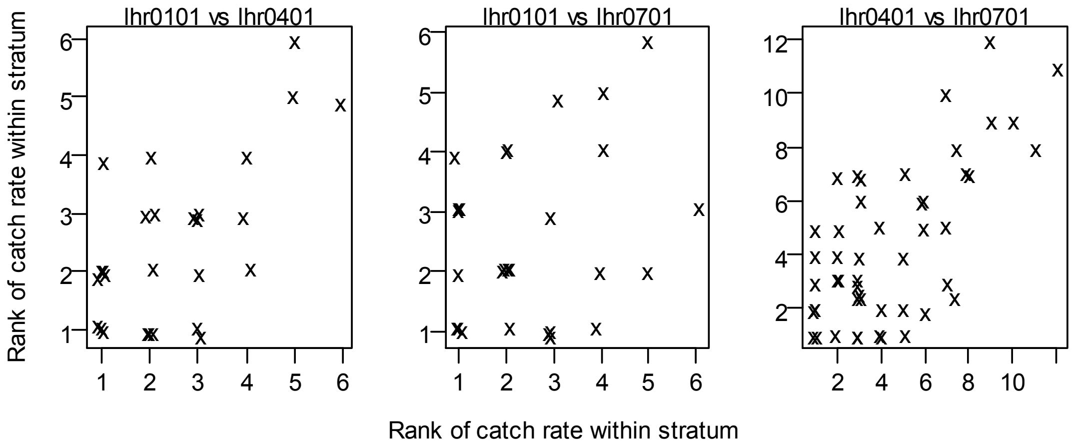


**Figure 4: An alternative approach to assessing the effect of site on mean catch rate. Each panel compares two surveys; each plotted point (jittered, to separate coincident points) corresponds to a site which occurred in both surveys; the x-value shows the rank of the catch rate at that site amongst all the catch rates in the same stratum in the first survey (a rank of 1 indicates the lowest value); and the y-value shows the corresponding rank in the second survey.**

## 3.2 The calculation of relative abundance and its precision

The data used to calculate relative abundance (and its precision) in potting surveys are $C_{pst}$, the catch (kg) from the *p*th pot in the *s*th set in stratum $t$; $A_t$ and $n_t$, the area (or coastline length) and number of sets in stratum $t$; and $m$, the number of pots per set. The calculations are the same whether the catch is all blue cod, or just those that are recruited to the fishery (i.e., those that exceed a given size limit).

### 3.2.1 Calculating relative abundance

In all surveys to date, relative abundance has been calculated as mean catch rates by set ($\overline{C}_{st}$), stratum ($\overline{C}_t$), and survey ($\overline{C}$) (all with units kg/pot) using the following equations.

$$\overline{C}_{st} = \left(\sum\nolimits_p C_{pst}\right)\Big/ m \qquad (1)$$

$$\overline{C}_t = \left(\sum\nolimits_s \overline{C}_{st}\right)\Big/ n_t \qquad (2)$$

$$\overline{C} = \left(\sum\nolimits_t A_t \overline{C}_t\right)\Big/\left(\sum\nolimits_t A_t\right) \qquad (3)$$

These are standard design-based methods of calculation. An alternative approach would be to use model-based methods (e.g., using GLMs), but this is not currently possible given that the above analyses (in Section 3.1) failed to find useful predictors of catch rate.

It is important to note that in calculating catch rates, no adjustment has been made for soak time (despite the use in some survey reports of units such as ‘kg per pot per hour’, ‘kg/hour’, ‘kg/pot/hour’, and ‘kg/pot-hour’, which suggest that such an adjustment has been made). Although the target soak time was 1 h for all surveys analysed here, the actual soak times (where these have been recorded) have occasionally deviated significantly from this target (e.g., they exceeded 1.5 h in 29 of 378 pots in lhr0101, and two pots soaked for 5 h in slt0201). The reason for not adjusting catch rates for soak times appears to be the finding, by Cole et al. (2004), that ‘the number of blue cod contained in pots changed little after 30 min’.

### 3.2.2 Four methods of calculating standard errors for relative abundance

I evaluated four alternative methods of calculating standard errors (s.e.s) for stratum and survey catch rates. The first method, which has been applied for all surveys to date, uses the following equations.

$$\text{s.e.}\left(\bar{C}_t\right) = \left[\frac{\sum_{ps}\left(C_{pst} - \bar{C}_t\right)^2}{\left(mn_t - 1\right)mn_t}\right]^{0.5} \tag{4}$$

$$\text{c.v.}\left(\bar{C}\right) = \left[\sum_t A_t^2 \,\text{s.e.}\left(\bar{C}_t\right)^2 \Big/ \left(\sum_t A_t\right)^2\right]^{0.5} \Big/ \bar{C} \tag{5}$$

This method might be called *pot-based*, because it assumes that the catches in individual pots within a stratum are independent of each other. If this assumption is not true (e.g., because catches from two pots are typically more similar when they are from the same set) then it could be better to use a *set-based* approach, in which the stratum s.e. is calculated using

$$\text{s.e.}\left(\bar{C}_t\right) = \left[\frac{\sum_s\left(\bar{C}_{st} - \bar{C}_t\right)^2}{\left(n_t - 1\right)n_t}\right]^{0.5} \tag{6}$$

but the survey s.e. is still calculated with Equation (5).

Another approach to estimating s.e.s is called *bootstrapping*. This involves creating a large number of simulated survey data sets – each representing what might have occurred in the survey – by resampling the real survey data. For each simulated data set, mean catch rates are calculated using Equations (1)–(3), then s.e.s are calculated as the standard deviations of these catch rates. For example, if $\bar{C}_{\text{boot},i}$ is the survey mean catch rate calculated from the $i$th of $N_{\text{boot}}$ simulated data sets, then the bootstrap estimate of the survey s.e. is given by

$$\text{s.e.}\left(\bar{C}\right) = \left[\sum_i\left\{\bar{C}_{\text{boot},i} - \left(\sum_j \bar{C}_{\text{boot},j} \Big/ N_{\text{boot}}\right)\right\}^2 \Big/ \left(N_{\text{boot}} - 1\right)\right]^{0.5} \tag{7}$$

There are two variants of the bootstrap technique that may be appropriate for the potting surveys. In a *conventional bootstrap*, the method of simulating one data set is as follows.

1. In stratum $t$, randomly select $n_t$ sets, with replacement, from the survey data.
2. From each selected set, randomly select $m$ catch rates, with replacement, from the recorded catch rates for that set.
3. Repeat steps 1 and 2 for all strata.

The *finite bootstrap* is a modification of the conventional bootstrap that is intended to take into account the fact that when the real survey was designed, the sites chosen for stratum $t$ were randomly selected, without replacement, from a finite set of potential sites. Let $N_t$ be the number of potential sites in stratum $t$, and $n_{\text{rep},t}$ be the integer part of $N_t/n_t$. Then the finite bootstrap uses the same procedure as the conventional bootstrap, except that step 1 is replaced by the following.

1A. In stratum $t$, construct a set of $N_t$ potential sites by taking $n_{\text{rep}}$ replicates of the $n_t$ sites that were used during the real survey and adding $N_t - n_{\text{rep}}n_t$ sites selected at random, without replacement, from the $n_t$ sites.

1B. Randomly select $n_t$ sites from this set of $N_t$ potential sites, *without replacement*.

It is important to note that the set of potential sites constructed at step 1A must be constructed anew for each of the simulated data sets.

[Technical note. The procedure used at step 1A to construct a set of $N_t$ potential sites may seem counter-intuitive, but it appears to be standard (Chao & Lo 1985, Booth et al. 1994). I briefly investigated an alternative, and to me more intuitive, procedure in which the set of potential sites was generated as the union of the $n_t$ sites that were used during the real survey and a set of $N_t - n_t$ sites selected at random, with replacement, from these sites. This produced slightly higher c.v.s than the above procedure but did not perform as well in a simulation experiment I conducted to compare the two alternative finite bootstraps.]

### 3.2.3 Calculating coefficients of variation for relative abundance

The precision of a relative abundance estimate has commonly been presented as a coefficient or variation (c.v.), calculated in the conventional way from the associated s.e. For example, when it is expressed as a percentage, the c.v. of the survey mean catch rate is calculated as

$$\text{c.v.}\left(\bar{C}\right) = 100\,\text{s.e.}\left(\bar{C}\right)\big/\bar{C} \tag{8}$$

(if the c.v. is expressed as a number, rather than a percentage, the factor 100 is omitted from this equation).

### 3.2.4 Evaluation of the four alternative methods

For each of the 13 survey data sets, I calculated the c.v.s of the survey mean catch rates, using all four of the above alternative methods, where this was possible. The finite bootstrap method could be applied only in those three areas where the numbers of potential sites (i.e., the $N_t$) were available (Table 7).

The pot-based c.v.s were always the same as, or close to, the values published in survey reports, but the set-based c.v.s were always markedly higher, though never large (Table 8). This suggests that within-set catch variability is typically smaller than that between sets in the same stratum, which contradicts the assumption underlying the pot-based method. This suggestion is supported by the fact that both bootstrap methods also produced c.v.s higher than those from the pot-based method. Compared to the set-based c.v.s, those from the conventional bootstrap were always the same or slightly higher, whereas those from the finite bootstrap were always lower (but higher than the pot-based c.v.s).

Neither of the bootstrap methods is satisfactory. The conventional bootstrap produces c.v.s that are too high because it ignores that fact that there is only a finite number of potential sites in each stratum. The finite bootstrap deals with this problem, but produces c.v.s that are too small because they do not include all sources of uncertainty in the stratum mean catch rates. There is uncertainty concerning catch rates at both the selected and unselected sites. Steps 1A and 1B of the bootstrap procedure deal with the latter component of uncertainty. As to the former, this uncertainty is in two parts: spatial and temporal. That is, the observed mean catch rate at a site will depend on exactly where the pots are placed around the site, and also on the exact day and time during the survey that the site is sampled. The spatial component is intended to be incorporated in step 2 of the bootstrap procedure, but the temporal component is not included.

**Table 7: The number of potential sites, by stratum and area (where available), and also, for comparison, the maximum number of sets occupied in any one survey in that stratum and area. The numbers of potential sites were taken from the lhr0701 Voyage Program, for Marlborough Sounds (SNDS), and from information provided by Glenn Carbines, for the other areas.**

| Area | Stratum | Number of potential sites | Max. number of sets per stratum |
|---|---|---|---|
| DUSK | 1 | 25 | 8 |
| DUSK | 2 | 32 | 8 |
| DUSK | 3 | 15 | 14 |
| DUSK | 4 | 16 | 8 |
| DUSK | 5 | 26 | 6 |
| OTAG | 1 | 10 | 5 |
| OTAG | 2 | 12 | 10 |
| OTAG | 3 | 10 | 7 |
| OTAG | 4 | 10 | 5 |
| OTAG | 5 | 10 | 7 |
| SNDS | 1 | 8 | 4 |
| SNDS | 2 | 10 | 7 |
| SNDS | 3 | 8 | 5 |
| SNDS | 4 | 9 | 8 |
| SNDS | 5 | 9 | 8 |
| SNDS | 6 | 13 | 13 |
| SNDS | 7 | 8 | 4 |
| SNDS | 8 | 8 | 5 |
| SNDS | 9 | 12 | 10 |
| SNDS | 10 | 12 | 11 |

**Table 8: Four alternative estimates of c.v.s of survey mean catch rates (%): pot-based, set-based, and conventional and finite bootstrap (the last method could be applied only where the number of potential sites per stratum was available – see Table 7).**

| | | | | Bootstrap | |
|---|---|---|---|---|---|
| Area | Trip | Pot-based | Set-based | Conventional | Finite |
| BANKS | chj0201 | 10.8 | 15.6 | 17.0 | – |
| | chj0501 | 5.7 | 7.5 | 8.7 | – |
| | chj0801 | 7.7 | 11.4 | 12.2 | – |
| KAIK | mys0401[1] | 8.6 | 11.8 | 13.1 | – |
| | mys0701[1] | 8.2 | 12.6 | 13.5 | – |
| MOTN | nav0501 | 7.3 | 11.4 | 11.4 | – |
| | leg0801 | 8.8 | 16.1 | 16.4 | – |
| DUSK | slt0201 | 6.6 | 10.7 | 11.5 | 7.9 |
| OTAG | suz0501 | 5.2 | 8.0 | 8.7 | 6.4 |
| PATE | gol0601 | 8.4 | 10.6 | 11.5 | – |
| SNDS | lhr0101 | 6.5 | 10.8 | 11.5 | 9.6 |
| | lhr0401 | 5.0 | 8.6 | 8.8 | 6.3 |
| | lhr0701 | 5.9 | 6.9 | 8.6 | 6.1 |

[1]To allow comparison with the survey reports (Carbines & Beentjes 2006a, 2009), for these calculations the area of stratum 1 in this survey area was taken as 2.9 $km^2$ for the first survey, and 9.6 $km^2$ for the second (the latter is understood to be the correct value).

### 3.2.5 Conclusions concerning relative abundance

I find no reason to change the existing practice of calculating relative abundance as mean catch rates using Equations (1)–(3). However, I have two small related concerns. The first is that it is potentially confusing to use units for these catch rates that suggest an adjustment has been made for soak time (e.g., ‘kg per pot per hour’, ‘kg/hour’). I suggest using kg/pot (or possibly kg/lift). My second concern relates to those surveys in which some actual soak times have been very different from the target value of 1 h (see above). However, it’s not clear what can be done about these surveys, except to say that there must be some doubt about the estimates of relative abundance.

As to estimating the precision of relative abundance estimates, I recommend the set-based method (Equations (6), (5), and (8)) on the grounds that it is simple and that all the other methods considered have been shown to be flawed.

### 3.3 The calculation of total mortality and its precision

Two approaches have been used to estimate blue cod total mortality, *Z*, using potting survey data. Carbines (2007) estimated *Z* for Paterson Inlet, using data from the 2006 survey of this area, but provided no estimate of precision. Carbines et al. (2008) used a slightly different approach to estimated *Z* for 14 area-time combinations, each of which was based on data from a set of strata from one survey (e.g., two *Z* estimates were calculated from the 2002 Banks Peninsula survey – one for the inshore strata, and one for those offshore). They also estimated 95% confidence intervals for *Z*. Subsequent estimates of *Z* (and confidence intervals) for blue cod (Blackwell 2008, Beentjes & Carbines 2009, Carbines & Beentjes 2009) appear to have used the same approach as that of Carbines et al. (2008).

For both approaches, the first step was to construct an overall age frequency (AF), for both sexes combined, for the all strata concerned.

#### 3.3.1 The construction of an overall age frequency

The data needed to calculate this AF are the length frequencies for each set in the survey (let $f_{lkst}$ be the number of fish of length *l* and sex *k* in set *s* in stratum *t*); an age-length key (ALK) by sex (let $K_{lak}$ be the proportion of fish of length *l* and sex *k* that are of age *a* [so $\sum_a K_{lak} = 1$ for each value of *k*]); and the area ($A_t$) and number of sets ($n_t$) for each stratum.

The calculation of the AF, which has been done using NIWA's catch-at-age software (Bull & Dunn 2002), is in principle quite simple. First, the overall length frequency (LF) for each sex is calculated as

$$f_{lk} = \sum_t \left[ \left( A_t / n_t \right) \sum_s f_{lkst} \right] \qquad (9)$$

then these LFs are converted to AFs using the ALKs

$$f_{ak} = \sum_l f_{lk} K_{lak} \qquad (10)$$

and finally the sex-specific AFs are summed to calculate the overall AF

$$f_a = \sum_k f_{ak} \qquad (11)$$

[Technical notes. For clarity, I have simplified the presentation of Equation (9) by omitting some details of the calculation which involve scaling the set LFs to the catch weight, using length-weight parameters, and allowing for catches in which fish were not measured. Full details are given in Beentjes & Francis (2011).]

Unfortunately, Carbines et al. (2008) mistakenly inputted their data to the catch-at-age software using a format ('new') that is appropriate for catch-sampling data, rather than survey data (the 'survey' format should have been used). The effect of this error was that the stratum weighting that occurs in Equation (9) was done wrongly, with the term $n_t$ being replaced by the total catch weight for stratum $t$. The effect of this error will vary from survey to survey, but it was substantial in at least one survey (Figure 5).

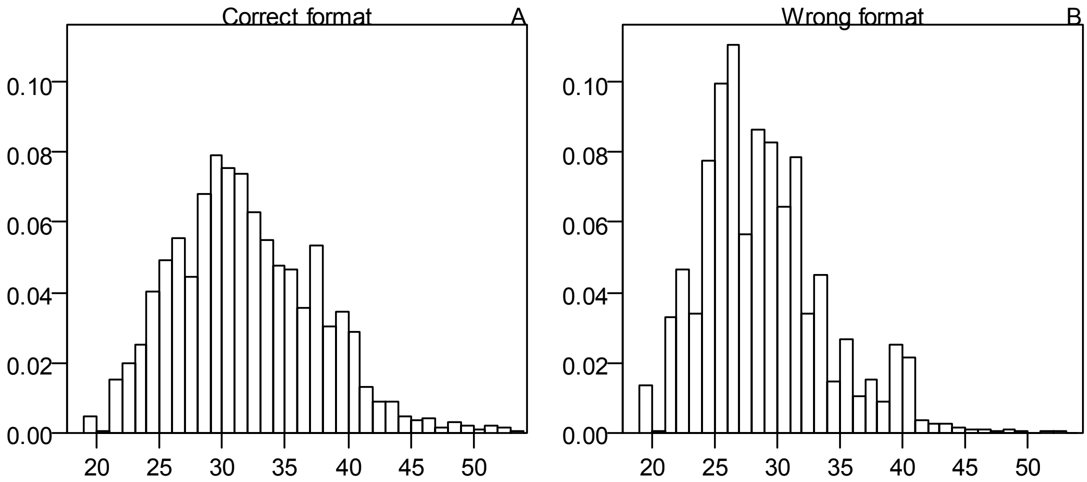


**Figure 5: Overall length frequencies for the 2004 Kaikoura survey calculated by the catch-at-age software using two different formats for the length data: A, the correct ('survey') format; and B, the wrong ('new') format.**

Another potential problem with the calculations made by Carbines et al. (2008) (and possibly Carbines 2007) is that they used a simpler form of the above equations which ignores sex (the subscript $k$ was dropped from Equations (9) and (10), and Equation (11) was not used). This is not always a bad thing to do. How harmful it might be depends on (a) whether ALKs differ markedly by sex (as they did in some, though not all, surveys – Figure 6); and (b) whether otolith samples were structured by sex. It appears these samples were sex-structured in most, if not all, surveys (this is stated in the reports from all the Marlborough Sounds surveys, and from all recent (post-2005) surveys in other areas, and is evident in the data from at least some earlier surveys – see Table 9A). When this is done, and ALKs differ by sex, the unsexed ALK, and thus the final AF, will be seriously biased. [I note in passing an odd and unexplained pattern in the aged sample for one trip. Because blue cod change sex from female to male, we expect the proportion male in any sample to increase with increasing length. In trip nav0501, the proportion was low (0.36) for the smallest fish (lengths 15–22), lower (0.19) for middle-sized fish (lengths 23–33), and high (1.00) for the largest fish (lengths greater than 33) (Table 9B).]

Two final points to note about the use of the catch-at-age software with data from these potting surveys. First, this software was not designed for such surveys, and so does not accommodate the two-stage sampling of pots within sets. Thus the user must choose whether to enter the length data by set, as was done by Carbines et al. (2008), or by pot. If we are just concerned with obtaining an AF and a point estimate of $Z$, it hardly matters whether the data are set-based or pot-based, because the results will be virtually identical. However, if we want to estimate the precision of the AF (as a mean-weighted c.v.) and of $Z$ (as a confidence interval) then it is important that the data be presented to this software as set-based (for reasons discussed above in Section 3.2). Second, this software requires length-weight parameters because for each set it scales the LF to the catch weight. These length-weight parameters have relatively little effect on the estimated AF or $Z$, but nonetheless it is important, for reasons of replicability, that the parameters used be documented in the survey reports (I don't think this was done in any of the reports; moreover Carbines et al. (2008) appear to have used the same parameter values for all areas, without any explanation). NIWA is currently developing a new program, CALA, for the analysis of length and age software, which will be more flexible, allowing a wide range of sampling structures, including that for BCO potting surveys. CALA will also allow set LFs to be scaled to the number, rather than the weight, of fish caught, which seems more sensible for these surveys.

**Table 9: Number of fish by sex and length in the aged samples from trips A, mys0401, and B, nav0501.**

A Trip mys0401

| Sex | Length (cm) | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 19 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 |
| M | 1 | 2 | 3 | 6 | 6 | 5 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| F | 1 | 6 | 6 | 6 | 6 | 6 | 6 | 5 | 6 | 6 | 6 | 6 | 6 | 6 | 5 | 6 | 5 | 6 | 6 | 6 | 6 | 6 |

| | Length (cm) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 |
| M | 5 | 6 | 4 | 3 | 3 | 2 | 3 | 1 | 0 | 3 | 2 | 1 |
| F | 3 | 4 | 2 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

B Trip nav0501

| | Length (cm) | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 15 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 37 | 38 |
| M | 0 | 1 | 1 | 1 | 1 | 2 | 2 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 3 | 2 | 1 | 3 | 4 |
| F | 1 | 2 | 2 | 3 | 2 | 4 | 0 | 2 | 4 | 5 | 3 | 5 | 5 | 5 | 4 | 4 | 1 | 5 | 0 | 0 | 0 | 0 |

| | Length (cm) | | | | | | |
|---|---|---|---|---|---|---|---|
| | 39 | 40 | 41 | 42 | 43 | 45 | 46 |
| M | 2 | 3 | 2 | 2 | 3 | 1 | 1 |
| F | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

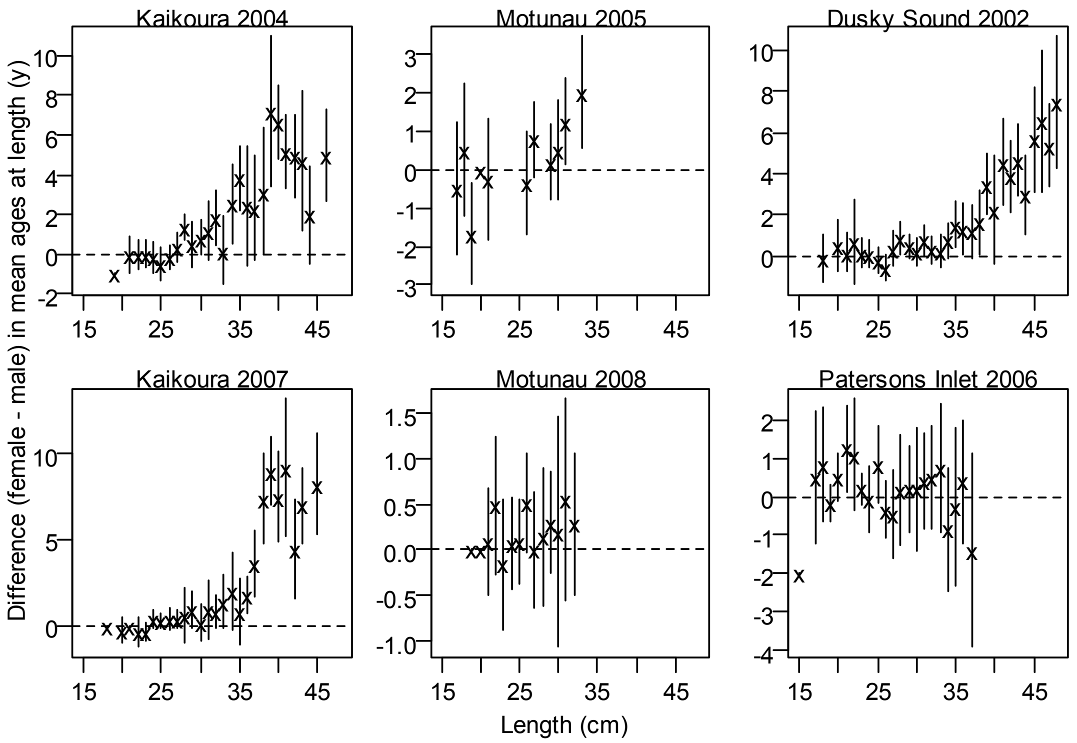


**Figure 6: Between-sex differences in mean age at length ('x', with vertical lines indicating 95% confidence intervals) for data from six surveys.**

### 3.3.2 The estimation of *Z* from the overall age frequency

Several methods exist to estimate $Z$ from an overall AF, but all involve inferring an age of (full) recruitment, $a_{\text{rec}}$, at or near the peak of this age frequency, and then calculating $Z$ from the shape of the recruited portion of the AF. The traditional regression method, used by Carbines (2007), equates $Z$ to minus the slope of a regression line fitted to the log proportions (or numbers) at age plotted against age (for ages greater than or equal to $a_{\text{rec}}$). Carbines et al. (2008) used a different method to estimate $Z$, that of Chapman & Robson (1960), in which

$$Z = \log_e \left( \frac{1 + \bar{a} - a_{\text{rec}} - 1/n_{\text{rec}}}{\bar{a} - a_{\text{rec}}} \right) \tag{12}$$

where $n_{\text{rec}}$ is the number of recruited fish in the AF and $\bar{a}$ is their mean age. Dunn et al. (2002) showed this latter method to be superior to both the regression method, and to two modifications of that method which involve ignoring some age bins with few data.

The problem with using Equation (12) in the current context is in determining what value to use for $n_{\text{rec}}$. This equation was developed for a situation in which the age frequency was derived from a simple random sample from the recruited fish, and $n_{\text{rec}}$ was the size of that sample. In contrast, the blue cod AFs are derived from a complex sample structure, in which $n_{\text{rec}}$ is not well defined. [Note that AFs calculated using Equations (9)–(11) are intended to represent the numbers of fish of each age in the whole population, rather than a sample from that population. However, for BCO surveys we can't even interpret the AFs as numbers in the population because we don't know what area is fished by each pot.] I can see only two solutions to this problem – a simple one, and a complicated one. The simple solution, which I recommend, is to use the maximum-likelihood estimator, which doesn't use $n_{\text{rec}}$. With this estimator,

$$Z = \log_e \left( \frac{1 + \bar{a} - a_{\text{rec}}}{\bar{a} - a_{\text{rec}}} \right) \tag{13}$$

The more complicated solution is to calculate an effective sample size for the AF using equation (7) of Dunn et al. (2002), scale the whole AF to that sample size, and then set $n_{\text{rec}}$ to the number of recruited fish in the scaled AF. This is more complicated because the calculation of an effective sample size requires bootstrapping the survey data to calculate a mean-weighted c.v. That is not too onerous if we simply want to estimate $Z$, but it becomes very time consuming if we want to estimate the precision of our $Z$ estimate (where we would have to do bootstraps within bootstraps). Another reason for preferring the simple solution (i.e., using Equation (13), rather than (12)) is that I think the values of $n_{\text{rec}}$ obtained by the complicated solution will usually be high (in the hundreds) so the difference between $Z$ estimates from Equations (12) and (13) will be small.

### 3.3.3 Choice of recruitment age

As already noted, the estimation of $Z$ requires the recruitment age, $a_{\text{rec}}$, to be specified. A common approach is to set $a_{\text{rec}}$ as the age at the peak of the AF. This was done by Carbines (2007), who calculated a single $Z$ for $a_{\text{rec}} = 8$ y. A problem with this approach is that the peak of the AF might occur at different ages in successive surveys, so the mortality estimates from these surveys will not be comparable. This approach also fails to acknowledge uncertainty about $a_{\text{rec}}$.

A better approach was adopted by Carbines et al. (2008) who chose a range of plausible values for $a_{rec}$ (5 to 8 y), and calculated a $Z$ estimate for each of these.

### 3.3.4 Estimating the precision of *Z*

Carbines et al. (2008) estimated three 95% confidence intervals for each $Z$ estimate using a simulation procedure adapted from that of Dunn et al. (2002). This procedure incorporated four sources of uncertainty: sampling uncertainty associated with the survey length and age samples (as summarised by the AF mean-weighted c.v. calculated by the catch-at-age software); ageing error (expressed as a c.v., $cv_a$); year-to-year variation in $Z$ (expressed as a c.v., $cv_z$); and recruitment variability (expressed as an s.d. in log space, $\sigma_R$). Three confidence intervals were produced for each $Z$ because Carbines et al. (2008) considered three levels of uncertainty – labelled Low, Medium, and High – with parameter values as given in Table 10.

**Table 10: Parameter values for the three levels of uncertainty (Low, Medium, and High) used by Carbines et al. (2008) (see their table 3) in calculating confidence intervals for total mortality, *Z*.**

| | | Parameter values | | |
|---|---|---|---|---|
| Source of uncertainty | Parameter | Low | Medium | High |
| Ageing error | $cv_a$ | 0.10 | 0.15 | 0.20 |
| Variation in $Z$ | $cv_z$ | 0.00 | 0.10 | 0.20 |
| Variation in recruitment | $\sigma_R$ | 0.50 | 0.70 | 1.00 |

A plot of all $Z$ estimates in Carbines et al. (2008) and later reports shows that there is relatively little difference between the Low, Medium, and High confidence intervals, particularly when compared to the variation in $Z$ with age at recruitment, and between surveys (Figure 7). Therefore I recommend that only the Medium confidence intervals should be calculated in future.

I also recommend that Equation (13) be used, rather than Equation (12), in the calculation of these confidence intervals. Full details of the simulation procedure for estimating the precision of $Z$ are given by Beentjes & Francis (2011).

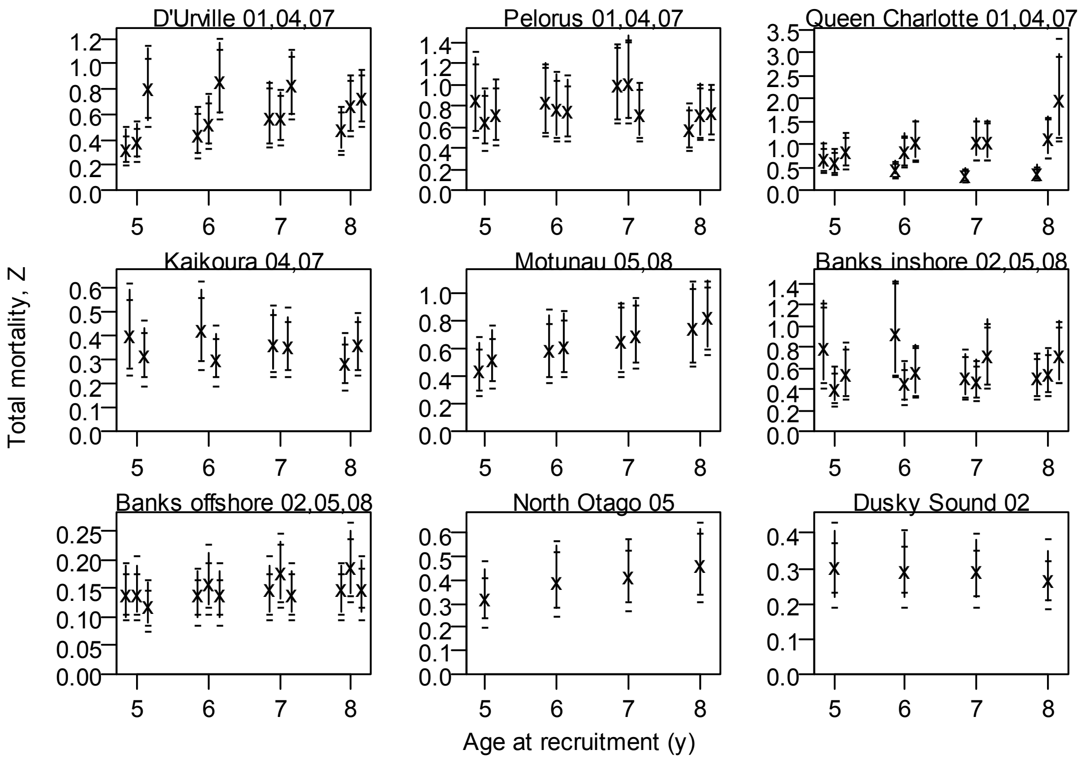


**Figure 7: Plot of all estimates of total mortality, *Z*, with confidence intervals (from Blackwell 2008, Carbines et al. 2008, Beentjes & Carbines 2009, Carbines & Beentjes 2009). Each panel shows *Z* estimates ('x') for four different recruitment ages in one ore more years, with the area and years given above the panel. The vertical lines indicate the 'Medium' 95% confidence intervals for *Z*, bounds for the 'Low' and 'High' confidence intervals as plotted as '-'.**

## 3.4 Investigation of age-length keys

The third outcome requested for this project was *To determine whether there are sufficient age data collected to develop a separate age-length key for each stratum*. This is probably not quite the right question to ask. An examination of the numbers of otoliths collected in surveys to date (Table 11) suggests that for most surveys there would be few strata in which there were sufficient otoliths to develop a separate ALK.

Perhaps a better question is, how many ALKs should be produced for each survey, and which subsets of the data require separate ALKs? The purpose of an ALK is to describe the distribution of age for each length class, and thus allow the conversion of LFs to AFs. [Note that variation in year-class strengths will cause these distributions to vary from year to year, even if growth rates don't change, so it is important not to use an ALK from one year with LFs from a different year.] Thus, how important it is to produce separate ALKs for two subsets of our data depends on how different the distributions of age at length are. Plots like Figure 6 are useful in investigating how big these differences are. However, these plots will be informative only if there are sufficient otolith data in both subsets.

I investigated the possibility of dividing into subareas each area that has historically had a separate ALK. In defining potential subareas I required strata within a subarea to be contiguous and was constrained by the requirement to have a reasonable number of otoliths in each subarea. With these constraints I defined 8 possible splits (Table 12), and made plots like Figure 6 for each (Figure 8).

**Table 11: Numbers of otoliths per stratum, and overall, in each survey. '–' = stratum not used in survey. Surveys are grouped by area.**

| | | Stratum | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Area | Year | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Overall |
| BANKS | 2002 | – | 110 | 4 | 0 | 44 | 3 | 54 | 118 | – | – | 333 |
| | 2005 | – | 47 | 18 | 11 | 54 | 20 | 39 | 68 | – | – | 257 |
| | 2008 | – | 57 | 1 | 24 | 54 | 53 | 85 | 41 | – | – | 315 |
| KAIK | 2004 | – | 101 | 36 | 82 | 68 | – | – | – | – | – | 287 |
| | 2007 | – | 30 | 141 | 62 | 39 | – | – | – | – | – | 272 |
| MOTN | 2005 | – | 6 | 11 | 80 | – | – | – | – | – | – | 97 |
| | 2008 | – | 236 | 6 | 14 | – | – | – | – | – | – | 256 |
| DUSK | 2002 | – | ? | ? | ? | ? | ? | – | – | – | – | 333[1] |
| OTAG | 2005 | – | 30 | 88 | 18 | 9 | 74 | – | – | – | – | 219 |
| PATE | 2006 | – | 66 | 67 | 2 | 82 | 12 | – | – | – | – | 229 |
| SNDS | 2001 | – | 32 | 77 | 82 | 95 | 57 | 115 | 27 | 48 | – | 533 |
| | 2004 | 0 | 20 | 80 | 82 | 99 | 69 | 84 | 26 | 46 | 101 | 607 |
| | 2007 | 19 | 0 | 80 | 154 | 203 | 156 | 252 | 33 | 90 | 295 | 1282 |

[1]Because of a temporary error in the age database it is not currently possible to assign the otoliths for this trip to strata.

**Table 12: Descriptions, for each of the eight areas that have historically had separate ALKs (excluding Dusky Sound), of a possible split into two subareas, and the number of otoliths per subarea in each year.**

| | Subarea 1 | | Subarea 2 | | | No. of otoliths | |
|---|---|---|---|---|---|---|---|
| Area | Name | Strata | Name | Strata | Year | Subarea 1 | Subarea 2 |
| BANKS inshore | western | 1–3 | eastern | 4–5 | 2002 | 114 | 47 |
| | | | | | 2005 | 76 | 74 |
| | | | | | 2008 | 82 | 107 |
| BANKS offshore | Le Bons | 6 | Pompeys | 7 | 2002 | 54 | 118 |
| | | | | | 2005 | 39 | 68 |
| | | | | | 2008 | 85 | 41 |
| Kaikoura | southern | 1–2 | northern | 3–4 | 2004 | 137 | 150 |
| | | | | | 2007 | 171 | 101 |
| Queen Charlotte | inner | 1–2 | outer | 3 | 2001 | 109 | 82 |
| | | | | | 2004 | 100 | 82 |
| | | | | | 2007 | 80 | 154 |
| Pelorus | inner | 4–5 | outer | 7–8 | 2001 | 152 | 75 |
| | | | | | 2004 | 168 | 72 |
| | | | | | 2007 | 359 | 123 |
| D'Urville Island | eastern | 6 | western | 9 | 2004 | 84 | 101 |
| | | | | | 2007 | 252 | 295 |
| North Otago | inshore | 1, 2, 4 | offshore | 3, 5 | 2005 | 127 | 92 |
| Paterson Inlet | outer | 4 | other | 1–3, 5 | 2006 | 147 | 82 |

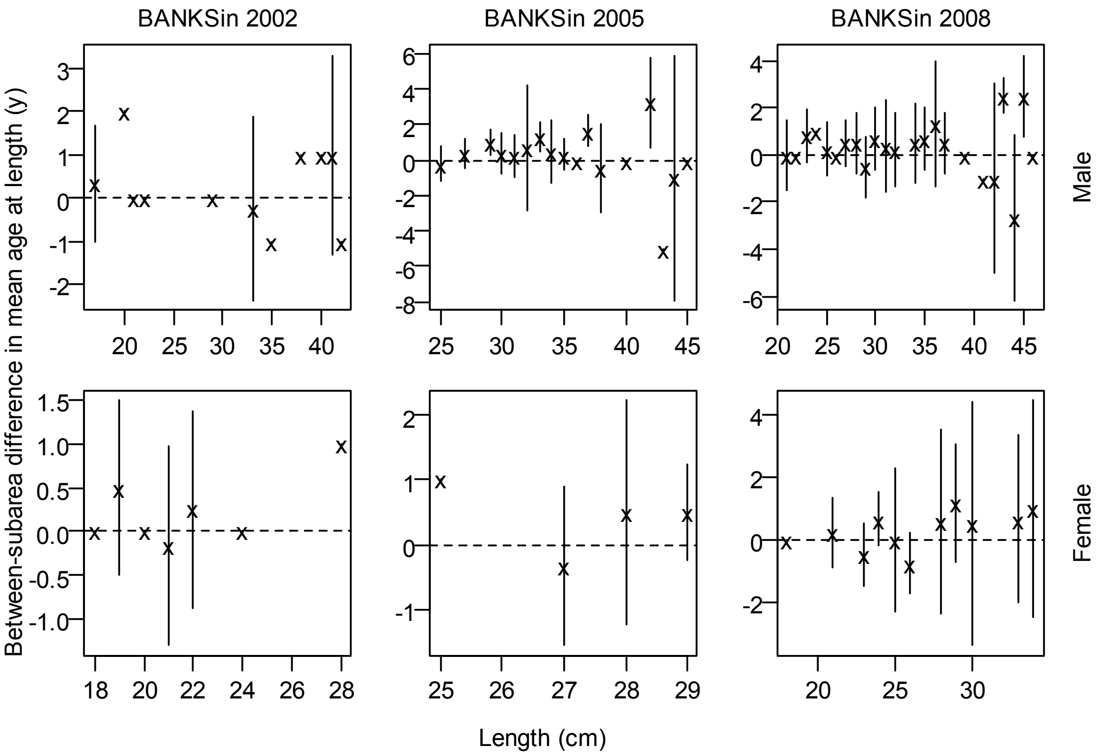


**Figure 8A: Differences in mean age at length ('x', with vertical lines indicating 95% confidence intervals), by sex and year, between western (strata 1–3) and eastern (strata 4–5) subareas within the Banks Peninsula inshore strata. Positive differences indicate that mean ages are higher in the second subarea. No confidence interval is plotted when there is no variation in age at length in the samples in both subareas.**

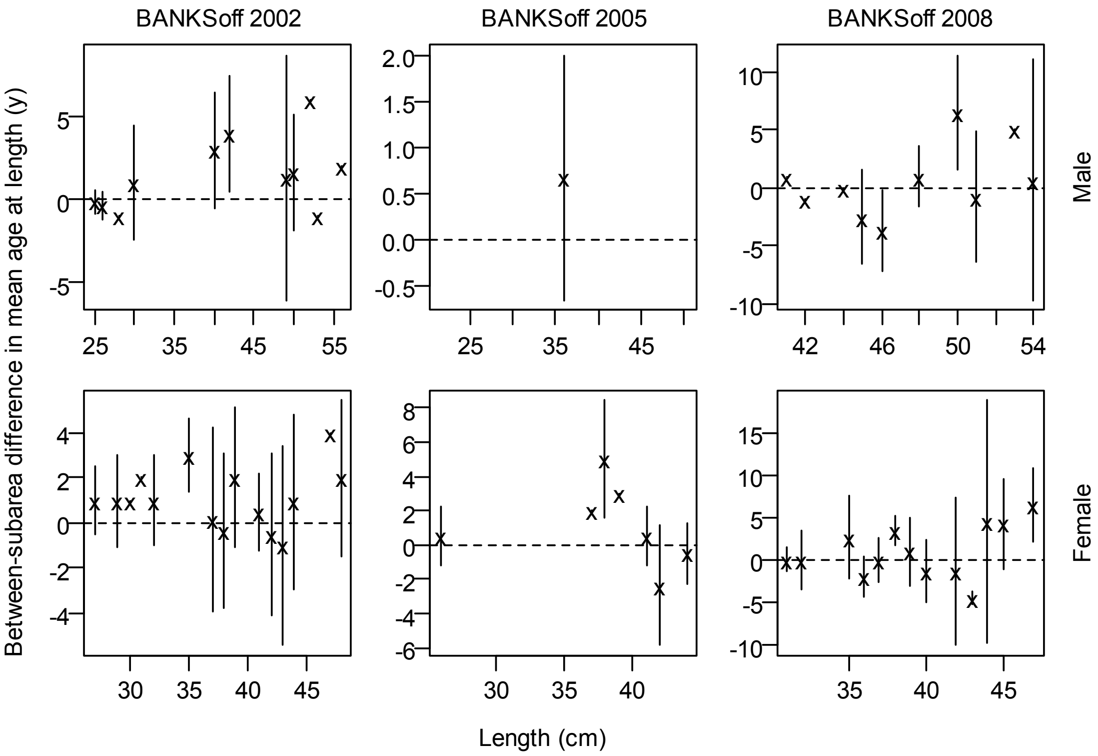


**Figure 8B: As for Figure 8A, but for Le Bons (stratum 6) and Pompeys (stratum 7) subareas within the Banks Peninsula offshore strata.**

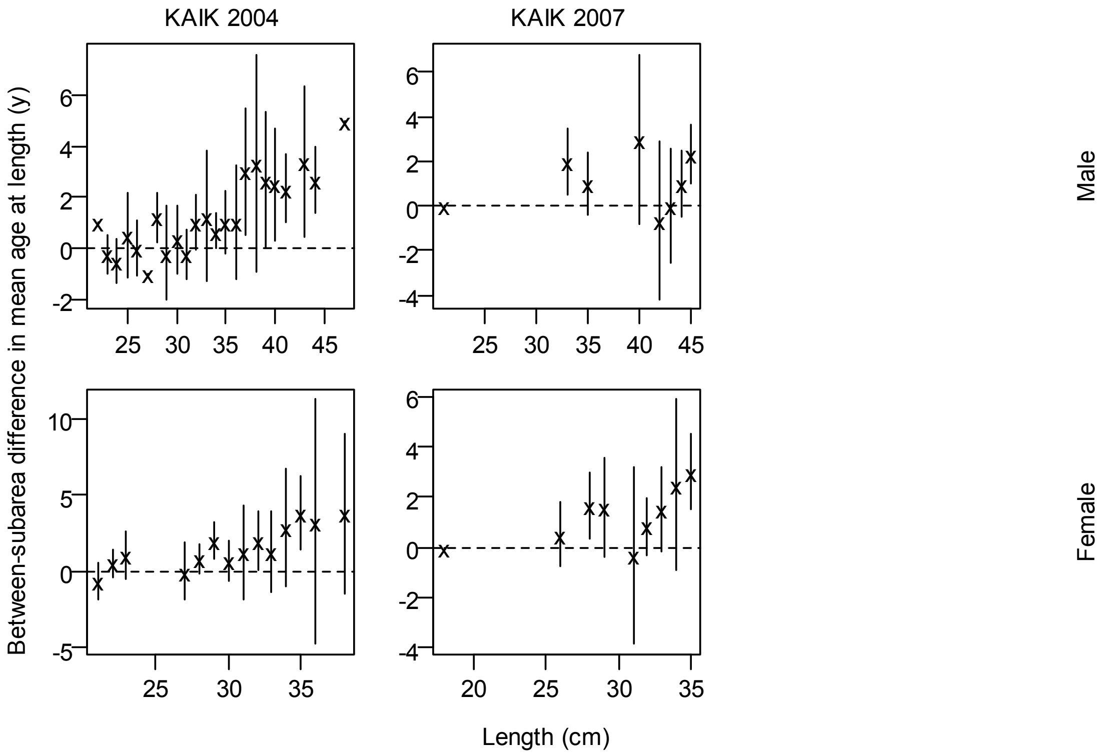


**Figure 8C: As for Figure 8A, but for southern (strata 1–2) and northern (strata 3–4) subareas within the Kaikoura strata.**

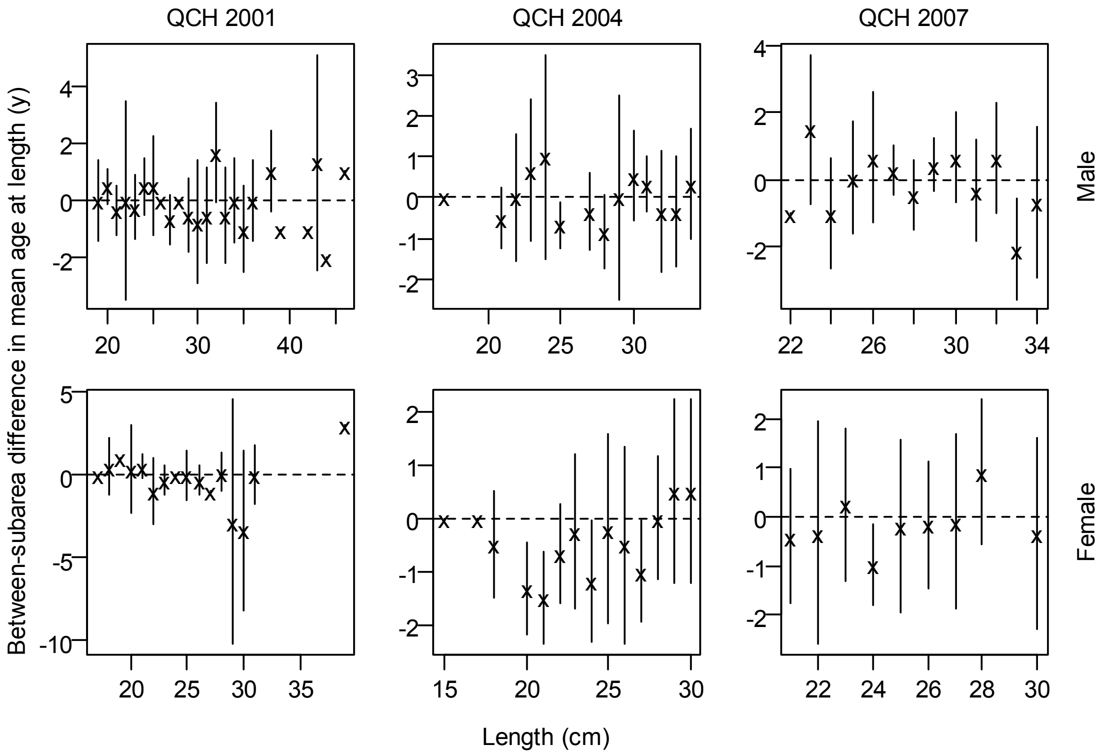


**Figure 8D: As for Figure 8A, but for inner (strata 1–2) and outer (stratum 3) subareas within the Queen Charlotte Sounds strata.**

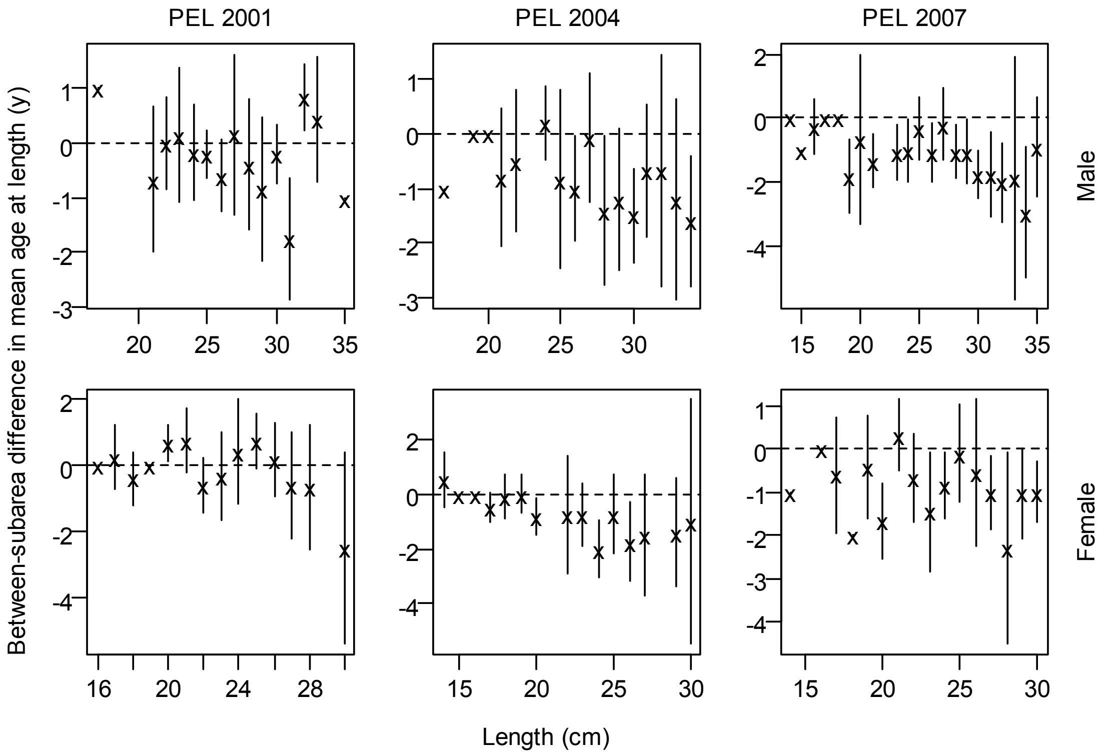


**Figure 8E: As for Figure 8A, but for inner (strata 4–5) and outer (strata 7–8) subareas within the Pelorus Sounds strata.**

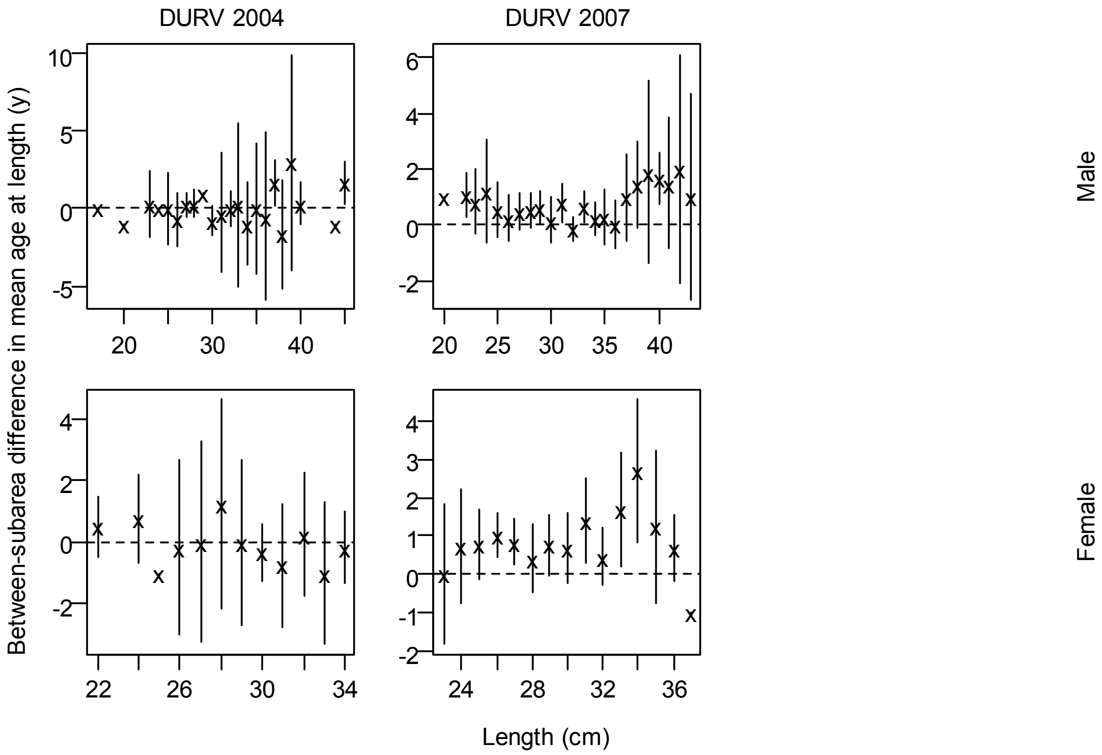


**Figure 8F: As for Figure 8A, but for eastern (stratum 6) and western (stratum 9) subareas within the D'Urville Island strata.**

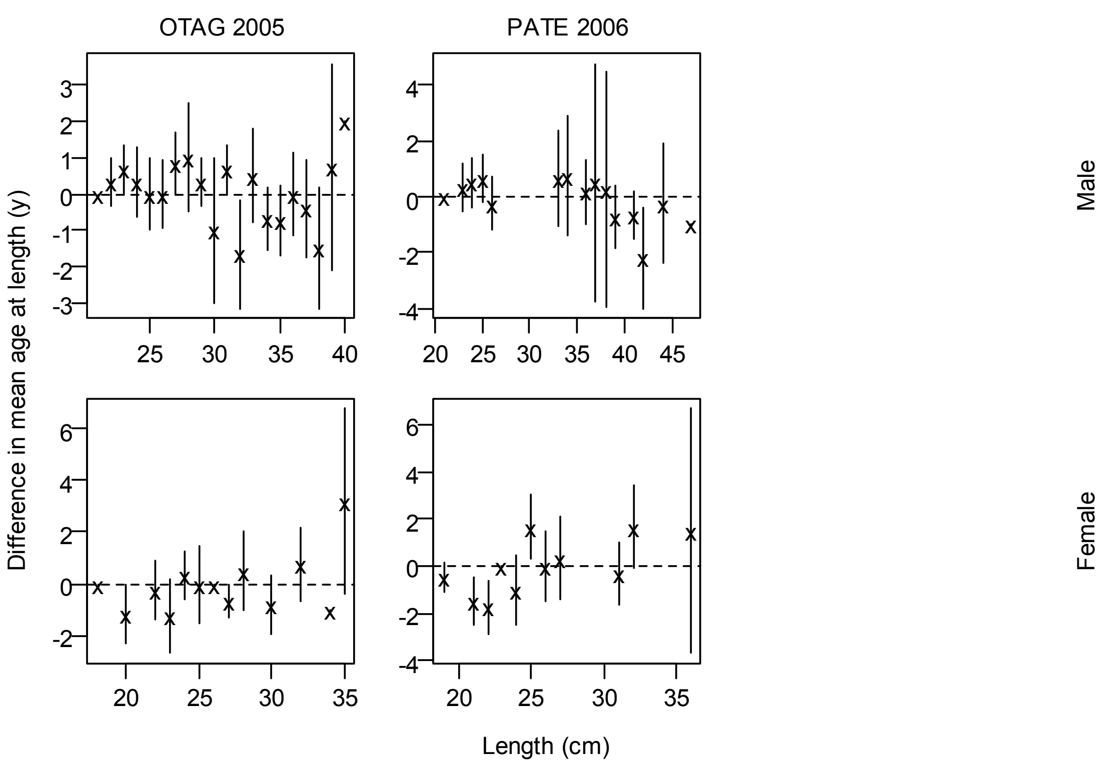


**Figure 8G: As for Figure 8A, but for one survey each in two different areas: inshore (strata 1, 2, 4) and offshore (strata 3, 5) in North Otago (left panels); and outer (stratum 4) and other (strata 1–3, 5) in Paterson Inlet (middle panels).**

Amongst these possible splits there were only two where there seemed clear evidence of a between-subarea difference in ALKs: mean age at length seemed consistently higher in the northern strata for Kaikoura (Figure 8C), and in the inner strata for Peloros (Figure 8E). For D'Urville Island, the results were equivocal, with the larger 2007 data set suggesting that western fish were older at a given length, while the smaller 2004 data set showed no clear difference (Figure 8F). For some areas and years the data were clearly insufficient to compare ALKs. Perhaps the most extreme example was BANKS offshore in 2005. Although 65 male otoliths were collected in this area, only 4 could be used in Figure 8B because there was only one length class (36 cm) in which there were otoliths from both subareas (Table 13).

**Table 13: Number of male fish by stratum and length in the aged samples for area BANKS offshore in 2005.**

| Stratum | Length (cm) | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 46 |
| 6 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 |
| 7 | 2 | 0 | 1 | 2 | 1 | 2 | 3 | 1 | 4 | 4 | 3 | 3 | 2 | 4 | 4 | 2 | 3 | 2 | 0 | 1 | 0 |

| Stratum | Length (cm) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 57 | All |
| 6 | 1 | 0 | 3 | 2 | 1 | 2 | 1 | 1 | 3 | 1 | 20 |
| 7 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 45 |

### 3.4.1 Should we create more age-length keys?

The above results suggest that there are three areas in which we may want to produce separate ALKs by subarea: southern and northern strata in Kaikoura; inner and outer strata in Pelorus; and (possibly) eastern and western strata around D'Urville Island. Our decision as to whether this would be worthwhile should depend on both gains and costs.

The main gain to be made from splitting ALKs by subarea is the removal of a possible source of bias in both the overall AF for the area, and the associated estimate of *Z*. To get an idea of the likely size of these gains I calculated AFs and *Z*s both with and without the split by subarea. In all three areas the changes in AFs and *Z*s was relatively small (Figure 9, Table 14). [Technical note: All AFs shown in Figure 9 were calculated using length-weight parameters given in Section 2.2 above; to calculate the AFs plotted as dashed lines I applied Equations (9)–(11) separately to the two subareas, and then summed the overall AFs (expressed as numbers, not proportions!) for the two subareas.] A secondary gain to be made from splitting ALKs is that it might allow more detailed modelling of growth and movement. The additional cost involved in splitting ALKs by subarea derives from the collection and reading of bigger otolith samples.

Whether the potential gains, as indicated by the changes in Figure 9 and Table 14, are sufficient to justify the additional costs is a decision best made by MFish, as the funder of these surveys. My feeling is that it is probably not worthwhile to increase the number of ALKs.

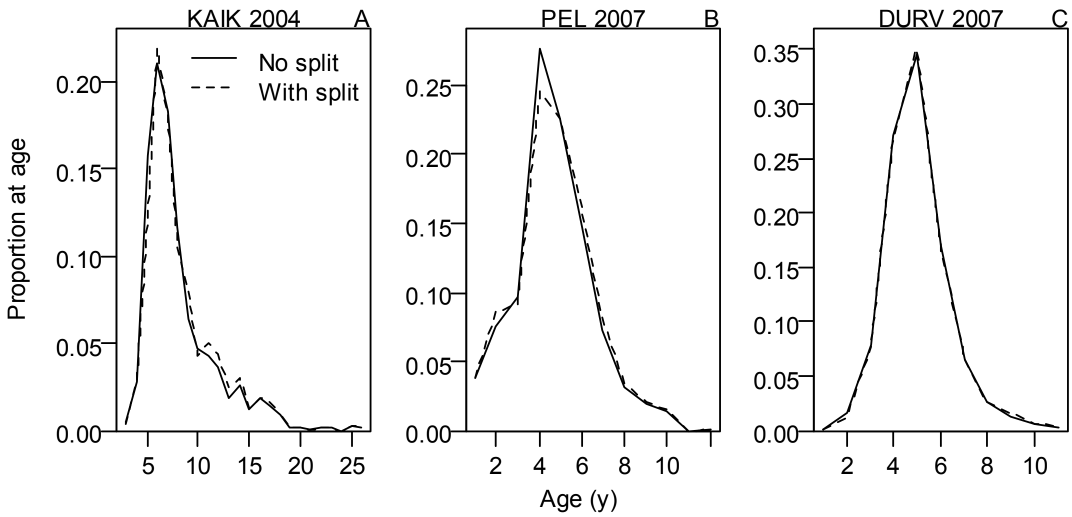


**Figure 9: Estimated age frequencies (AFs) for three areas in which there is some evidence of between-subarea differences in ALKs. The AFs were calculated (for the year in which there was most otolith data) both using a single ALK (solid lines), and also using a separate ALK for each subarea (broken lines).**

**Table 14: Estimates of total mortality, *Z*, calculated from the age frequencies of Figure 9.**

| | KAIK 2004 | | PEL 2007 | | DURV 2007 | |
|---|---|---|---|---|---|---|
| | No split | With split | No split | With split | No split | With split |
| 5 | 0.27 | 0.25 | 0.66 | 0.65 | 0.83 | 0.82 |
| 6 | 0.30 | 0.29 | 0.75 | 0.75 | 0.88 | 0.84 |
| 7 | 0.30 | 0.29 | 0.78 | 0.79 | 0.85 | 0.80 |
| 8 | 0.28 | 0.27 | 0.81 | 0.82 | 0.85 | 0.81 |

### 3.4.2 Structuring the otolith sample to reduce possible bias

In surveys like these there is always a potential for bias in the calculation of AFs and *Z*s if the distribution of age at size (i.e., the ALK) varies within and between strata. We have seen above that the available data allow only a very limited investigation of such variability. One way to reduce the possibility of bias from this source is to make sure that the otolith sample from any part of the survey area is reasonably representative of the vulnerable population therein. Ideally, the number of otoliths collected in each stratum should be proportional to the vulnerable population. Also, the structure of the sample (i.e., the length distribution and sex ratio of otolithed fish) should be similar to that of the vulnerable population in that stratum. There is no need to be overly rigorous in this matter, but any gross deviation from this ideal raises the risk of bias.

## 4. ACKNOWLEDGMENTS

This research was carried out by NIWA under contract to the Ministry of Fisheries (MFish Project IPA2009/18). I am grateful to Mike Beentjes, Trevor Willis, Glen Carbines, and Jeremy McKenzie for information about data collection and analysis from previous surveys, and to Vivian Haist for information, and useful discussions, about finite bootstrapping. Murray Smith reviewed a draft of this report.

## 5. REFERENCES

Beentjes, M.P.; Carbines, G.D. (2003). Abundance of blue cod off Banks Peninsula in 2002. *New Zealand Fisheries Assessment Report 2003/16.* 25 p.

Beentjes, M.P.; Carbines, G.D. (2006). Abundance of blue cod off Banks Peninsula in 2005. *New Zealand Fisheries Assessment Report 2006/1.* 24 p.

Beentjes, M.P.; Carbines, G.D. (2009). Abundance, size and age composition, and mortality of blue cod off Banks Peninsula in 2008. *New Zealand Fisheries Assessment Report 2009/25.* 46 p.

Beentjes, M.P.; Francis, R.I.C.C. (2011). Blue cod potting surveys: standards and specifications. Version 1. *New Zealand Fisheries Assessment Report 2011/29.*

Blackwell, R.G. (2002). Abundance, size and age composition of recruited blue cod in the Marlborough Sounds, September 2001. Final Research Report for Ministry of Fisheries Project BCO2001/01. (Unpublished report held by Ministry of Fisheries, Wellington.)

Blackwell, R.G. (2005). Abundance and size composition of recruited blue cod in the Marlborough Sounds, September 2005. Final Research report for the Ministry of fisheries Research Project BCO2003/01. (Unpublished report held by Ministry of Fisheries, Wellington).

Blackwell, R.G. (2008). Abundance and size composition of recruited blue cod in the Marlborough Sounds, September 2007. Final Research Report for Ministry of Fisheries Research Project BCO2006/01. 24 p. (Unpublished report held by Ministry of Fisheries, Wellington.)

Booth, J.G.; Butler, R.W.; Hall, P. (1994). Bootstrap methods for finite populations. *Journal of the American Statistical Association 89*: 1282–1289.

Bull, B.; Dunn, A. (2002). Catch-at-age user manual v1.06.2002/09/12. NIWA Internal report 114. 23 p. (Unpublished report held in NIWA library, Wellington.)

Carbines, G.; Dunn, A.; Walsh, C. (2008). Age composition and derived estimates of total mortality for blue cod in South Island potting surveys, 2002–2005. *New Zealand Fisheries Assessment Report 2008/68.* 74 p.

Carbines, G.D. (2007). Relative abundance, size, and age structure of blue cod in Paterson Inlet (BCO 5), November 2006. *New Zealand Fisheries Assessment Report 2007/37.* 31 p.

Carbines, G.D.; Beentjes, M.P. (2003). Relative abundance of blue cod in Dusky Sound in 2002. *New Zealand Fisheries Assessment Report 2003/37.* 25 p.

Carbines, G.D.; Beentjes, M.P. (2006a). Relative abundance of blue cod off north Canterbury in 2004–05. *New Zealand Fisheries Assessment Report 2006/30.* 26 p.

Carbines, G.D.; Beentjes, M.P. (2006b). Relative abundance of blue cod off North Otago in 2005. *New Zealand Fisheries Assessment Report 2006/29.* 20 p.

Carbines, G.D.; Beentjes, M.P. (2009). Relative abundance, size and age structure, and mortality of blue cod off north Canterbury (BCO 3) in 2007−08. *New Zealand Fisheries Assessment Report 2009/37.* 56 p.

Chao, M.-N.; Lo, S.-H. (1985). A bootstrap method for finite population. *Sankhya, Series A 47*: 399–405.

Chapman, D.G.; Robson, D.S. (1960). The analysis of a catch curve. *Biometrics 16*: 354–368.

Cole, R.G.; Alcock, N.K.; Tovey, A.; Handley, S.J. (2004). Measuring efficiency and predicting optimal set durations of pots for blue cod *Parapercis colias*. *Fisheries Research 67*: 163–170.

Dunn, A.; Francis, R.I.C.C.; Doonan, I.J. (2002). Comparison of the Chapman Robson and regression estimators of Z from catch-curve data when non-sampling stochastic error is present. *Fisheries Research 59(1–2)*: 149–159.

Francis, R.I.C.C. (2006). Measuring the strength of environment-recruitment relationships: the importance of including predictor screening in cross-validation. *ICES Journal of Marine Science 63*: 594–599.