NIWA
Taihoro Nukurangi

Implications of the Precision Seafood Harvesting 'Modular Harvest System' on snapper stock yield relative to standard trawl

Prepared for Precision Seafood Harvesting

Revised December 2017

30 March 2018 9.11 AM

Prepared by:
Jeremy McKenzie
Russell Millar

For any information regarding this report please contact:

Jeremy McKenzie
Fisheries Stock Assessment Scientist
s 9(2)(a) [redacted]

National Institute of Water & Atmospheric Research Ltd
Private Bag 99940
Viaduct Harbour
Auckland 1010

Phone +64 9 375 2050

NIWA CLIENT REPORT No: 2017370AK
Report date: September 2017
NIWA Project: PSH16302

Quality Assurance Statement		
s 9(2)(a) [redacted]	Reviewed by:	s 9(2)(a) [redacted]
	Formatting checked by:	
s 9(2)(a) [redacted]	Approved for release by:	s 9(2)(a) [redacted]

Contents

Tables

Figures

Executive summary

A comparative analysis of Under-sized snapper Caught per Kilo (UCK) for MHS and standard trawl gears, factoring in uncertainty in the method selectivity estimates, found strong statistical evidence that the MHS UCK derived ratio was “worse than” that of standard trawl.

The point estimate discard survival MHS would need to attain achieve equivalence to standard trawl in B40% equilibrium yield was 63% as derived from a stock assessment model. This contrasts to the point survival estimate of 35% MHS would need to attain to achieve an equivalent UCK ratio to standard trawl assuming the “available” snapper population length structure to both gears was the same as that observed in the recent PSH Kaharoa selectivity survey. The reason the UCK and model B40% point survival estimates differ is due to the SA model predicting a greater proportion of available snapper being above the 25 cm MLS at B40% to that observed in the Kaharoa selectivity survey. This difference highlights a limitation with using UCK as a comparative measure for commercial fishing gears in that UCK ratios can change if the available snapper length composition to the gear changes.

Taking into account uncertainty in the selectivity estimates, the B40% equilibrium yield modelling predicted MHS would need to achieve discard survival rates in the order of 80% to achieve statistical equivalence at a 95% or better confidence level.

A similar analysis looking at shifting the MHS selectivity curve closer to that of standard trawl indicated a right-shift greater than 40 mm would be needed to achieve statistical equivalence at a 95% or better confidence level. However the degree of shift could be reduced to ~30 mm assuming mean discard survival is 30%.

Through-mesh snapper mortality levels of 5% or higher have the potential to significantly alter MHS’s predicted snapper impact relative to standard trawl. We believe a consideration of through-mesh mortality needs to be factored into the MHS accreditation process.

1 Introduction

The Precision Seafood Harvesting (PSH) programme is developing innovative fish capture technologies to catch seafood in a way that maximises fish quality, and reduces mortality of unintended catch (i.e. undersized fish of target species and unwanted species). For this new gear design to be accepted for operational use, the Ministry for Primary Industries (MPI) needs to be satisfied that the Modular Harvesting System (MHS) performs at least as well as the traditional commercial trawling gear currently authorised for use under the Fisheries Act 1996 and supporting regulations. In other words, it must perform “no worse than existing gear” (i.e. standard trawl). As part of the MPI’s Enabling Innovative Trawl Technologies (EITT) initiative, a number of criteria have been proposed against which new trawl gear should be assessed. These include species composition of the catch, size composition of the catch, impact on protected species, and benthic impacts (MPI: Fisheries Innovative Trawl technologies Notice 2017).

If satisfied with the performance of the new MHS, MPI will need to change the regulatory framework so that the new gear can be lawful.

It is reasonable to assume that ***if*** the MHS gear catches and kills a higher proportion of sub-MLS (Minimum Legal Size) fish than standard trawl (ST) for a given weight of retained legal-sized fish the method is likely to have a "worse" overall impact on stock productivity.

Two factors determining a fishing gear's net sub-MLS mortality impact are:

1. Selectivity i.e. the relative number of sub-MLS fish the gear catches relative to the number available;

2. Mean survival rate of discarded sub-MLS fish.

Although a number of fin-fish species are taken by the northern inshore trawl fisheries, the species of highest economic and social value is snapper. For this reason snapper has been the primary focus of the **inshore** MHS selectivity and survival studies conducted to-date (Jones and Millar 2017). The general view of MPI seems to be that if MHS gear can be shown to have "no worse" impact to snapper sustainability than trawl the technology is likely be deemed "acceptable" for inshore fisheries use.

A recent inshore snapper selectivity study for trawl and MHS gears found MHS gear was more selective of sub-MLS snapper than standard trawl at the 0.10 level of significance (Jones and Millar 2017). MPI have requested PSH commission additional analytical work to investigate the implication of selectivity differences between standard trawl and MHS on snapper sustainability and yield. The main purpose of these simulations is to determine if additional (potentially costly) survival studies are necessary/warranted.

At a recent meeting attended by MPI, PSH, NIWA, and Plant & Food, two analytical approaches for assessing the "no worse" criteria for MHS were recommended:

1. Empirical simulations applying the estimated selectivity curves to the length-frequency data collected during the inshore trials (e.g., cover-net + traditional/MHS net). Such simulations would estimate the potential retained catch of sub-MLS snapper with each net and calculate the number of sub-MLS (undersized) snapper per kilo of above-MLS fish [UCK]).
2. Theoretical stock level simulations comparing population trajectories for scenarios when trawl catches are taken with the two selectivity curves estimated from the trials (to bound the potential impacts).

The simulation analysis described in this report were directed at answering the following questions:

1. Is the MHS method likely to be "no worse or better" than standard trawl in proportional catch of sub-MLS snapper on the basis of:

    a. the available size composition of snapper as seen in the recent Kaharoa net-selectivity trials;

    b. the predicted Hauraki Gulf stock size composition as derived from the 2013 SNA 1 assessment model (Francis & McKenzie 2015, MPI 2016).

2. Is the net difference in predicted equilibrium stock yield at 40% virgin biomass (B40%) of a MHSonly commercial Hauraki Gulf snapper fishery "no worse or better" than single trawl-only

commercial fishery (pursuant to recreational harvest allocation remaining fixed at 40% of TAC [Total Allowable Catch]) assuming:

   a. sub-MLS survival of snapper caught and released at the surface (i.e. ‘discarded’) from both MHS and standard trawl is zero percent;

   b. sub-MLS survival of released snapper from MHS is 100% and standard trawl is zero percent?

3. Assuming MHS yield proves to be “no worse or better” under the 100% sub-MLS survival assumption; what is the minimum level of MHS discard survival needed to still satisfy the “no worse or better” criteria?

2 Methodology

2.1 Incorporating uncertainty in the Jones and Millar selectivity estimates

The selectivity curves for MHS and standard gears, as estimated by Jones and Millar (2017) differ, suggesting MHS gear retains more snapper less than the 25 cm (current commercial MLS) than standard trawl (Figure 2-1). The implication of the Figure 2-1 MHS selectivity curve **pursuant to a 100% discard mortality assumption** is MHS will have a higher detrimental impact on snapper stock Maximum Sustainable Yield (MSY) than standard trawl, i.e. that MHS is “worse than” trawl.

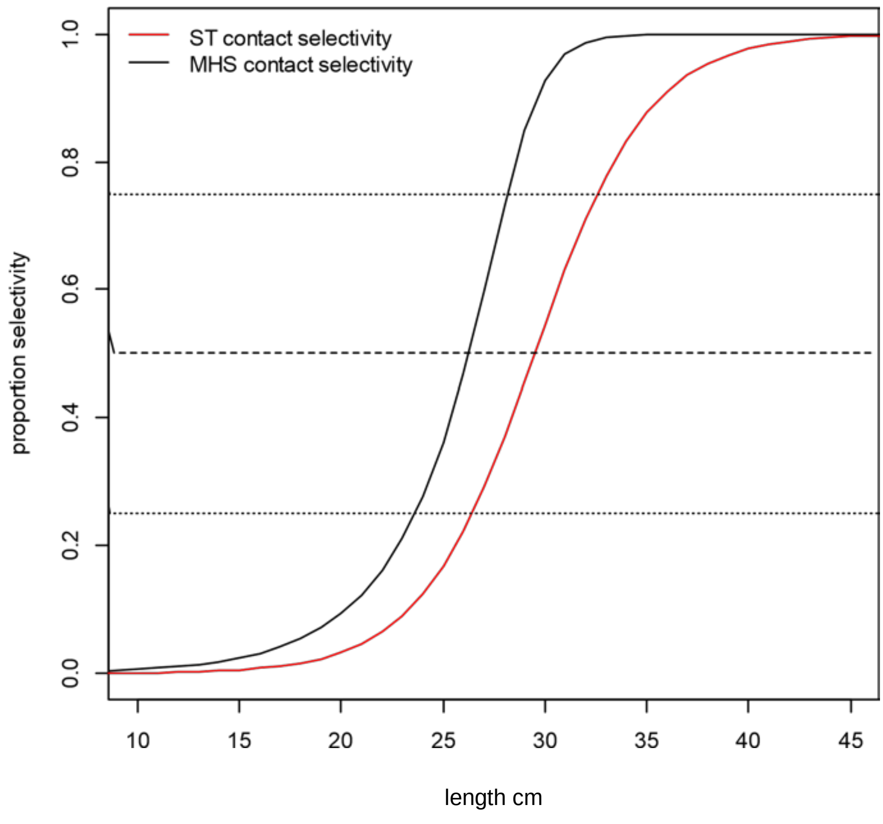
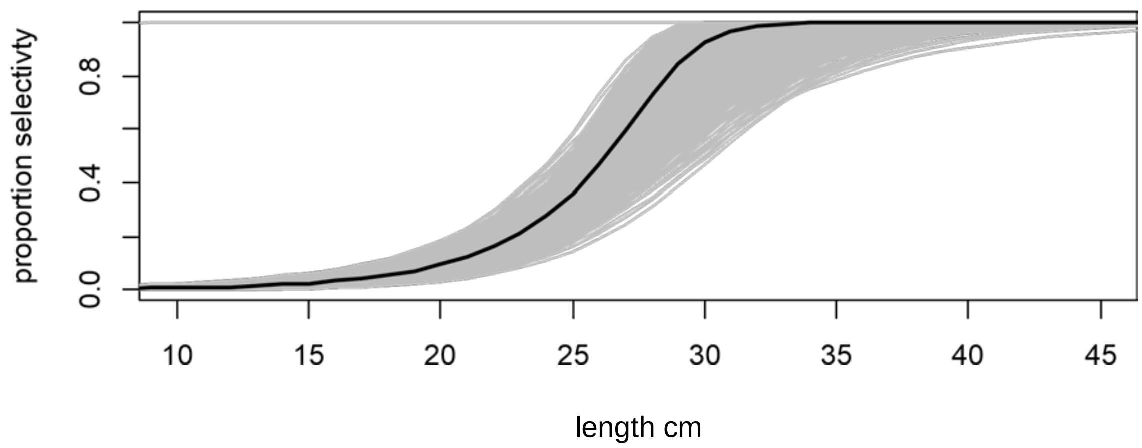


Figure 2-1: Differences seen in the MHS and standard trawl (ST) mean selectivity curves for snapper as estimated by Jones and Millar (dotted lines correspond to respectively 25, 50, & 75% relative selection).

However, in evaluating “no worse than standard trawl” criteria for MHS based on the differences in gear selectivity seen in the Jones and Millar analysis it is necessary to take account uncertainty in the selectivity estimates (Figure 2-2).

MHS selectivity



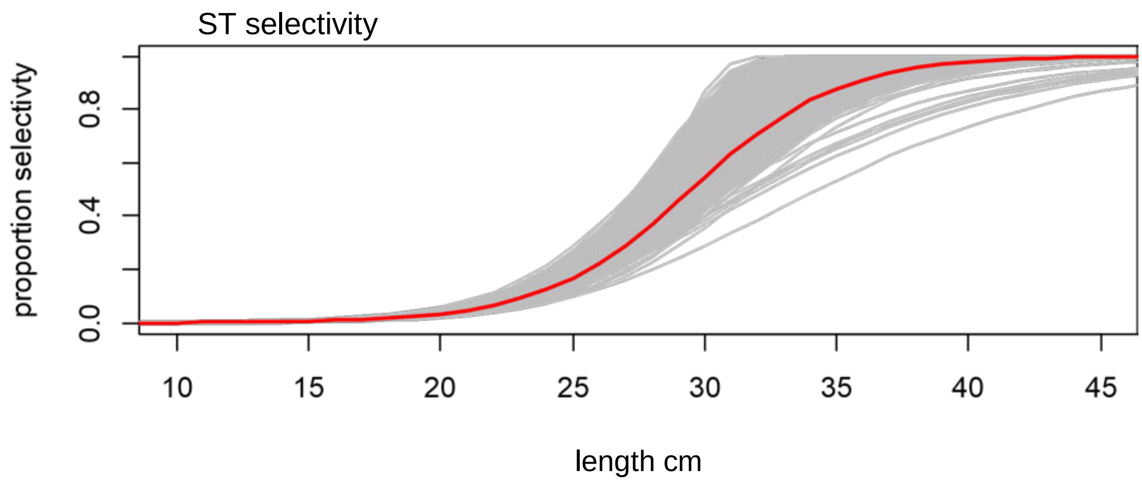


Figure 2-2: 1,000 (bootstrap) alternative snapper selectivity curves for MHS and standard trawl (ST) illustrating the range of uncertainty in the Jones and Millar selectivity estimates (thick lines).

The core approach for evaluating the "no worse" criteria involves applying the bootstrap selectivity curves given in Figure 2-2 to length frequency estimates of the proportion of the Hauraki Gulf snapper stock "available" to both gears. Uncertainty as to the length composition of the "available" Hauraki Gulf snapper stock (as derived from surveys or stock assessment models described below) was assumed to be same for each of the two gear types and therefore could be ignored in the simulations, i.e. the length compositions of snapper "available" to the two methods for the purpose of evaluating MHS "no worse than" criteria were assumed to be "know without error".

2.2 Testing the statistical significance of the "no worse than standard trawl" criteria taking in account uncertainty in the selectivity estimates

The “specified outcome” or hypothesis we wish to test is that MHS impact on snapper is “no worse than trawl”. The specific statistics we are going to use to test this hypothesis are: UCK based on observed survey length frequency data; and equilibrium yield at B40% derived from a stock assessment model.

Statistical hypothesis tests are framed in terms of quantifying the plausibility of a specified hypothesis**,** being the outcome you are **most interested or concerned about being true.** The burden of proof is therefore in establishing that this “desired outcome” is highly likely at some predefined statistical level of confidence (e.g. better than a 95% likelihood of being true). This is done by establishing that the contrary hypothesis (that the desired outcome in not true) is implausible. The contrary hypothesis is called the **null hypothesis**. If the likelihood of the **null hypothesis** being true is very low, say less than 5%, then we reject it and can say that the desired outcome (which is the opposite of the null hypothesis) is plausible. Failure to reject the null hypothesis does not necessarily mean the null hypothesis is true, but rather that there was insufficient data collected to disprove it.

In comparing MHS and standard trawl gear there are three possible conclusions:

1. MHS is worse than trawl;
2. MHS is equivalent to trawl;
3. MHS is better than trawl.

Given the MPI requirement of for MHS gear is to be “no worse than” standard trawl, statistical proof of either conclusions 2 & 3 would meet this requirement. Whereas a statistical proof of conclusion 1 would clearly establish MHS did not meet the requirement. Traditional statistical tests are usually aimed at establishing if two populations differ in some way (e.g. “worse than” or ”better than”). However, not being able to establish a difference between two populations does not constitute statistical proof of equivalence. Proof for equivalence requires the use of a special class of hypothesis testing called an “Equivalence testing” (Appendix C). If MHS gear cannot be shown to be better or worse than standard trawl it will still need to pass a statistical test of equivalence in order to satisfy the MPI requirement. An inability to statistically prove either MHS equivalence or difference to standard trawl would simply mean there are insufficient selectivity data to derive a definitive conclusion the solution being to collect more data.

2.3 Empirical UCK simulations applying the estimated selectivity curves to the length-frequency data collected during the selectivity inshore trials

A useful measure of the impact of the two fishing gears is the ratio of the number of **U**nder-sized snapper **C**aught per **K**ilo of legal sized snapper landed (UCK; Bentley 2015).

UCK was calculated for each experimental gear using the length-weight relationship given by $w = 0.04467 \times l^{2.793}$ used in SNA assessments (weight in grams, length in cm). Bootstrapping was performed to emulate the sampling variability in the calculated UCK.

The following three approaches were used to test for UCK differences between the two methods

2.3.1 UCK based on actual tow data

For each experimental gear, a bootstrap sample was generated by resampling the experimental tows of that gear (with replacement from the 18 conventional tows or 16 MHS tows), and then resampling from the snapper frequencies for each sampled haul. UCK was then recalculated using the bootstrap sample. This was done 1000 times, resulting in 1000 bootstrapped UCK values.

2.3.2 UCK calculated from selection curves using experimental availability

During the recent Hauraki Gulf at-sea MHS gear selectivity trails half of the trawl tows were made with an enclosing fine-mesh cover. The cover's purpose was to retain all small fish passing either through the MHS selection holes or the cod-end meshes of the standard trawl net. It is reasonable to assume that the combined length frequency from the mesh cover and the retained MHS and trawl catch were representative of the total available length frequency of snapper and other fin-fish in the area the selectivity trials were taking place (Figure 2-3).

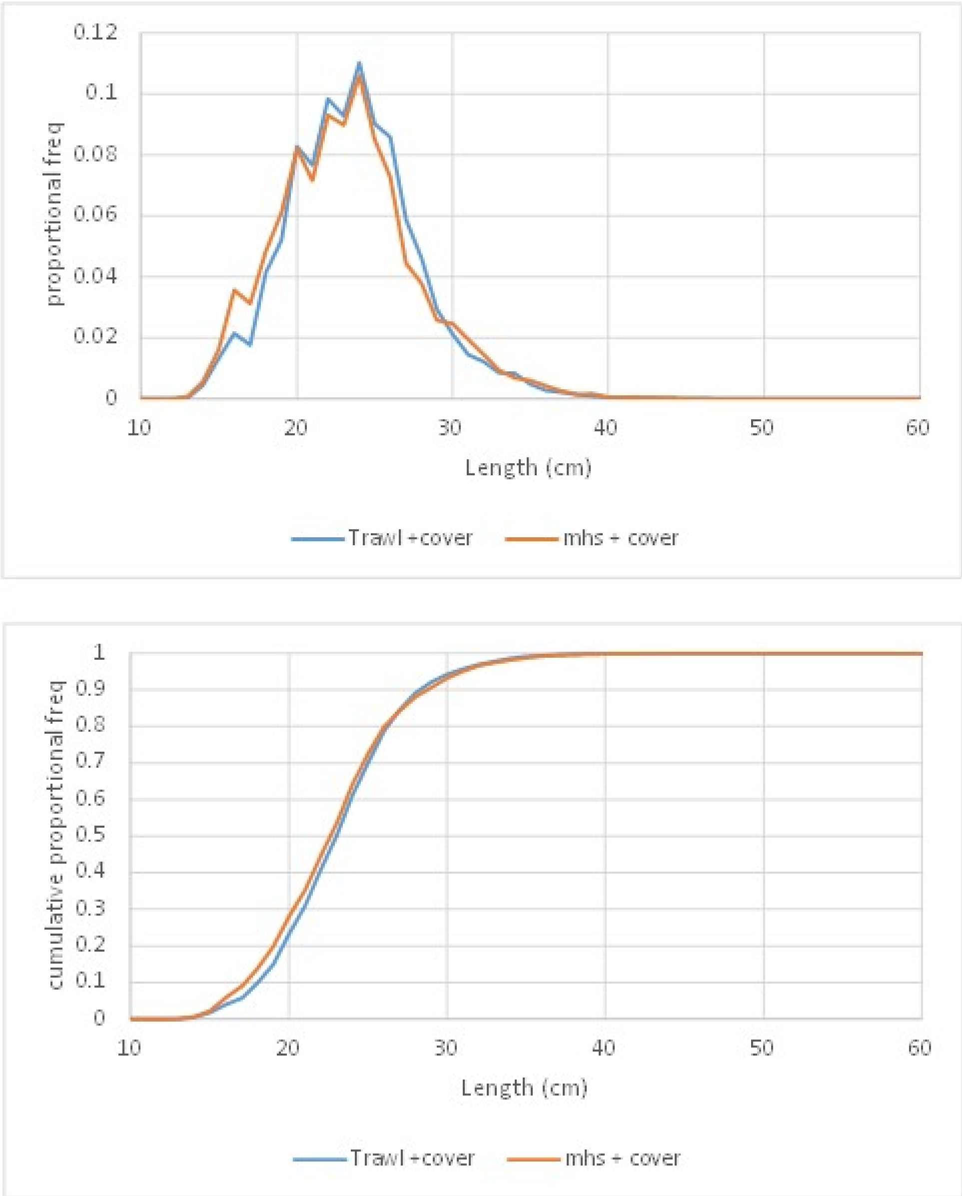


Figure 2-3: Combined proportional length frequency and cumulative frequency of snapper from MHS and combined cover Kaharoa tows and standard trawl and combined cover Kaharoa tows.

The length frequency of all snapper available to the gears was calculated by summing over all 34 covered-gear tows. For both conventional and MHS gears, a set of 1000 bootstrapped selection curves was applied to these length frequencies. This gives the length frequency that would be caught in the gear under each of the 1000 bootstrap selection curves. For each gear, UCK was calculated from each of these 1000 sets of length frequencies, to obtain 1000 bootstrap UCK values.

2.3.3 UCK calculated from selection curves using assessment availability

An alternative estimate of the size composition of the Hauraki Gulf snapper stock available to trawl gear in the 2012-13 fishing year (Figure 2-4) was derived from a modified version of the 2013 SNA 1 stock assessment model (Francis & McKenzie 2015; Appendix A). The bootstrap methodology of the previous section was then applied to obtain 1000 bootstrap UCK values calculated for each gear.

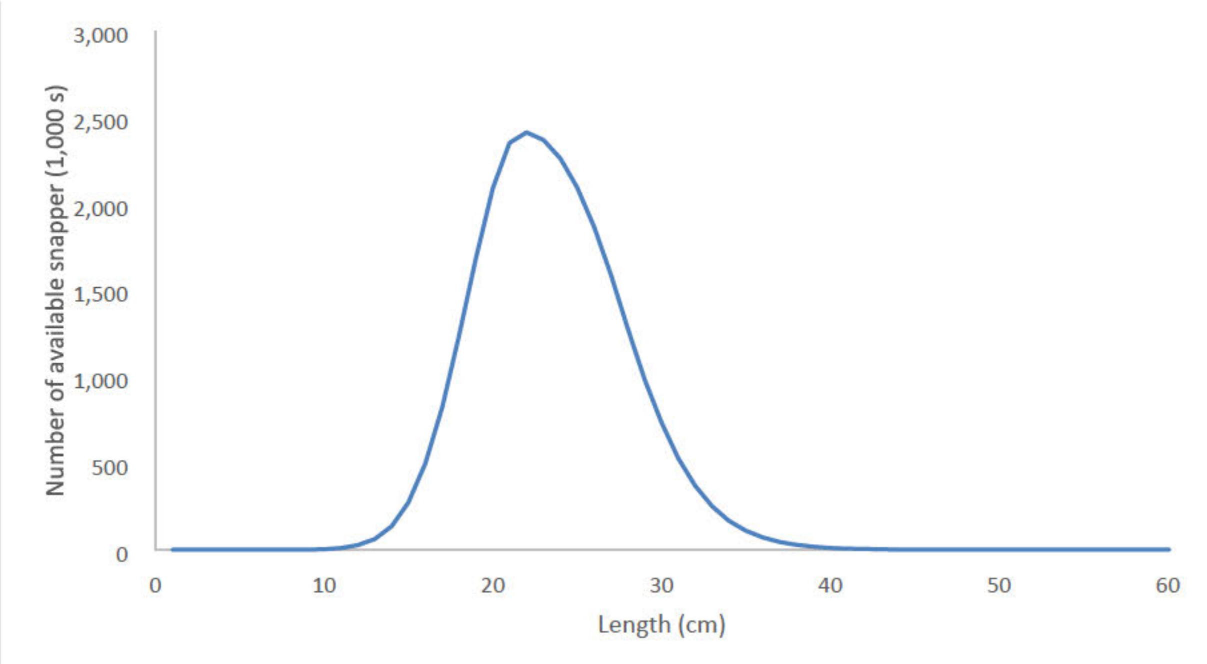


Figure 2-4: Estimated length frequency of Hauraki Gulf snapper available to trawl gears in 2013 from the 2013 Hauraki Gulf assessment model.

2.4 Theoretical stock level simulations comparing population trajectories for scenarios when trawl catches taken with the two selectivity curves estimated from the trials.

The main limitation of available population length frequency analyses described in 2.3 is that these analyses provide little insight into the degree of impact gear selectivity has on snapper-stock yield and long-term sustainability. To better understand the respective gear selectivity yield implications of MHS and trawl it was necessary to look at method selectivity differences within the confines of a lengthbased stock assessment model. The stock assessment modelling approach enabled us to evaluate trade-offs between leaving a fish in the water to grow and breed (gain in yield-per-recruit) against the risk of it dying through natural causes (loss of yield-per-recruit). Using this approach it was possible to compare yield gains and losses associated with harvesting fish at different sizes, including putting them back in to water pursuant to various levels of assumed survival.

NIWA developed a suite of yield simulation modelling tools for investigating alternative selectivity and survival harvest strategies for snapper as part of a 2015 snapper management review (MPI 2016). These methods and approaches have been accepted by MPI and minimal modifications for undertaking MHS and trawl implied yield comparisons were required.

The estimates of equilibrium yield at B40% pursuant to either MHS or standard trawl estimated selectivities were derived using the length-based Hauraki Gulf assessment model described in Appendix A.

Using this model it was possible to look at the yield implications of MHS selectivity under different levels of assumed sub-MLS discard survival (Appendix A). To do this required deriving estimates of total selectivity for MHS and standard trawl for application in the model as the Jones and Millar selectivity estimates are gear contact selectivity estimates only and therefore do not reflect snapper availability to the gear (See Appendix B for description).

Simulation scenarios were undertaken for the following assumed levels of MHS discard survival:

1. B40% yield and catch length comparisons assuming zero sub-MLS survival for both methods
2. B40% yield and catch length comparisons assuming 100% sub-MLS survival for MHS and zero survival for trawl
3. B40% yield and catch length comparisons assuming 75% sub-MLS survival for MHS and zero survival for trawl
4. B40% yield and catch length comparisons assuming 50% sub-MLS survival for MHS and zero survival for trawl
5. B40% yield and catch length comparisons assuming 25% sub-MLS survival for MHS and zero survival for trawl

Uncertainty on the MHS and standard trawl derived selectivity estimates were incorporated into the model yield estimations using the same bootstrap approach as described in Section 2.3.

2.5 UDK a more useful comparative statistic than UCK

UDK being the ratio of number of undersized snapper **D**eaths per landed kilo of catch is arguably a better impact measure for comparing MHS and standard trawl gears in that it takes into account fish survival. By definition the UDK ratio can never be greater than the UCK ratio and it is the number of undersized fish dying after encountering a gear not the number initially caught that should be of concern or interest.

3 Results

3.1 UCK estimation results

3.1.1 Method 1: UCK of actual tows

The observed UCK for conventional gear was 0.51 with 95% bootstrap confidence interval of 0.41 to 0.62 (Figure 3-1). Observed UCK for MHS was 0.65 with 95% confidence interval 0.51 to 0.84 (Figure 3-1). The UCK of the MHS gear was higher than that of the conventional gear on 93.3% of the 1000 bootstraps (Figure 3-2). These results constitute strong statistical evidence that the MHS UCK ratio is “worse than” standard trawl.

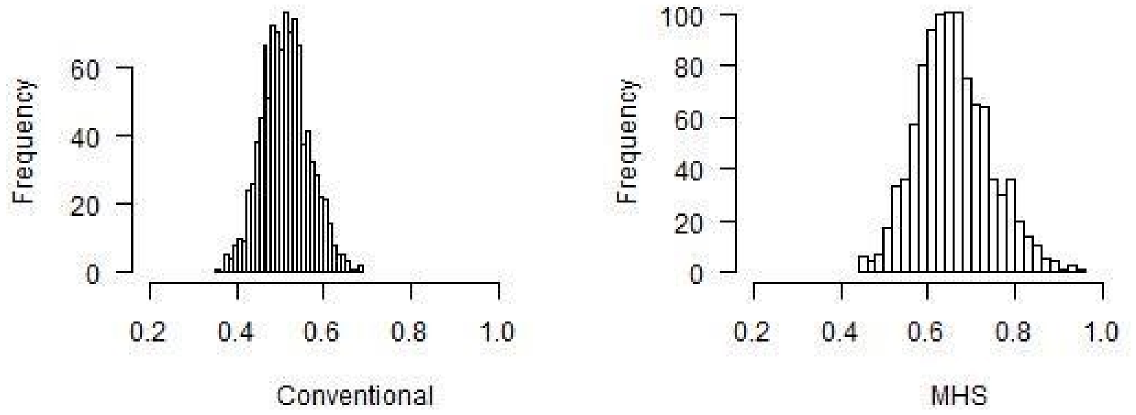


Figure 3-1: Bootstrap distribution of UCK.

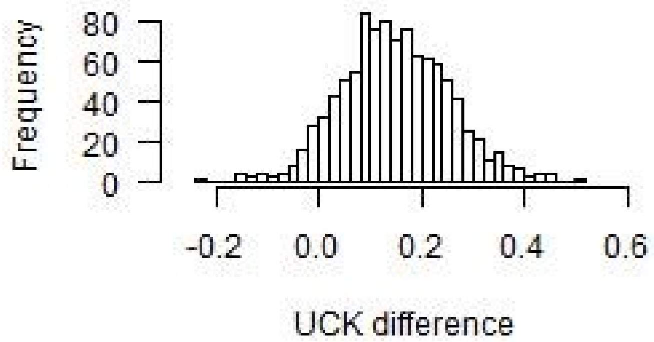


Figure 3-2: Bootstrap distribution of the MHS UCK minus the conventional UCK.

3.1.2 Method 2: UCK calculated from selection curves using experimental availability

UCK for conventional gear had 95% bootstrap confidence interval of 0.36 to 0.54 (Figure 3-3). MHS had 95% confidence interval 0.49 to 0.96 (Figure 3-3). The UCK of the MHS gear was higher than that of the conventional gear on 98.2% of the 1000 bootstraps (Figure 3-4). Again, the statistical conclusion is that the MHS UCK ratio is "worse than" standard trawl.

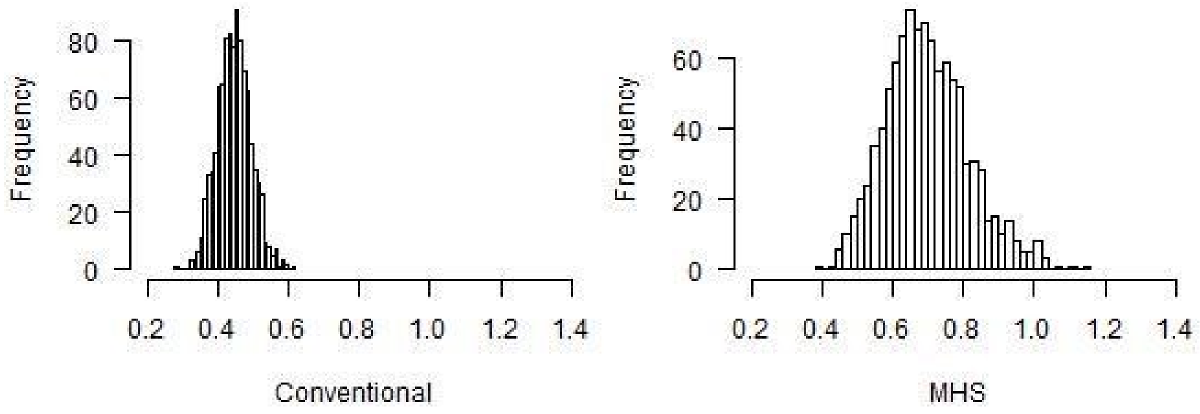


Figure 3-3: Bootstrap distribution of UCK.

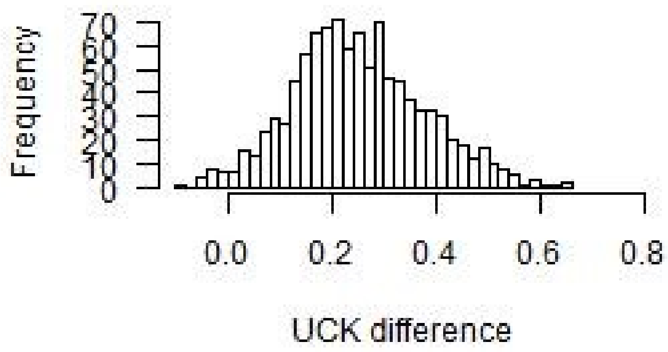


Figure 3-4: Bootstrap distribution of the MHS UCK minus the conventional UCK.

3.1.3 Method 3: UCK calculated from selection curves using 2013 Hauraki Gulf stock assessment model predicted availability to trawl

UCK for conventional gear had 95% bootstrap confidence interval of 0.34 to 0.50 (Figure 3-5). MHS had 95% confidence interval 0.45 to 0.89 (Figure 3-5). The UCK of the MHS gear was higher than that of the conventional gear on 97.9% of the 1000 bootstraps (Figure 3-6). Again, the statistical conclusion is that the MHS UCK ratio is "worse than" standard trawl.

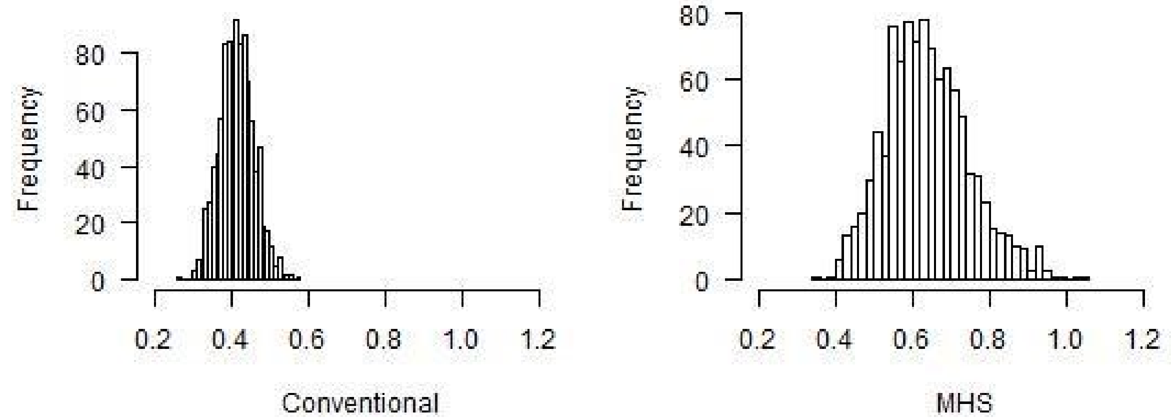


Figure 3-5: **Bootstrap distribution of UCK.**

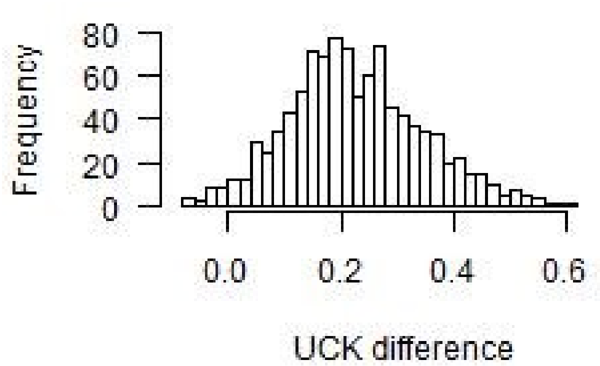


Figure 3-6: **Bootstrap distribution of the MHS UCK minus the conventional UCK.**

For the MHS gear there is little difference in the bootstrap distribution of UCK under the three methods used. For conventional gear, there is little difference between Methods 2 and 3, but the bootstrap distribution of UCK shifts to the left (i.e. UCKs tend to decrease) under Methods 2 and 3 compared to Method 1. This can occur due to the high variability between hauls and size distribution of fish available to the hauls. For example, selectivity is unaffected by the relative numbers of snapper that enter the covered on uncovered gear in any haul pair (because this is treated as an incidental parameter that is included). However overall UCK is affected since catch length frequencies are summed over all uncovered hauls.

3.2 Point estimates of MHS sub-MLS proportional survival required to achieve an equivalent UDK to standard trawl

We can legitimately convert UCK ratios to UDK ratios by discounting sub-MLS fish that can be assumed to have survived the capture and release process, thus it is possible to determine what level of MHS discard survival would be needed to achieve UDK equivalence with standard trawl. To do this we need to know the length composition of snapper "available" to both gears and apply the respective selectivity curves to this distribution to obtain an estimate of catch for both methods. If we assume an "available" population with a length composition as per the Kaharoa selectivity survey, the pointestimate level of survival required to achieve UDK equivalence to standard trawl is ~35% for an MLS of 25 cm (Figure 3-7). The required proportional survival changes as MLS changes, the lowest required level corresponding to a 26 cm MLS (Figure 3-7).

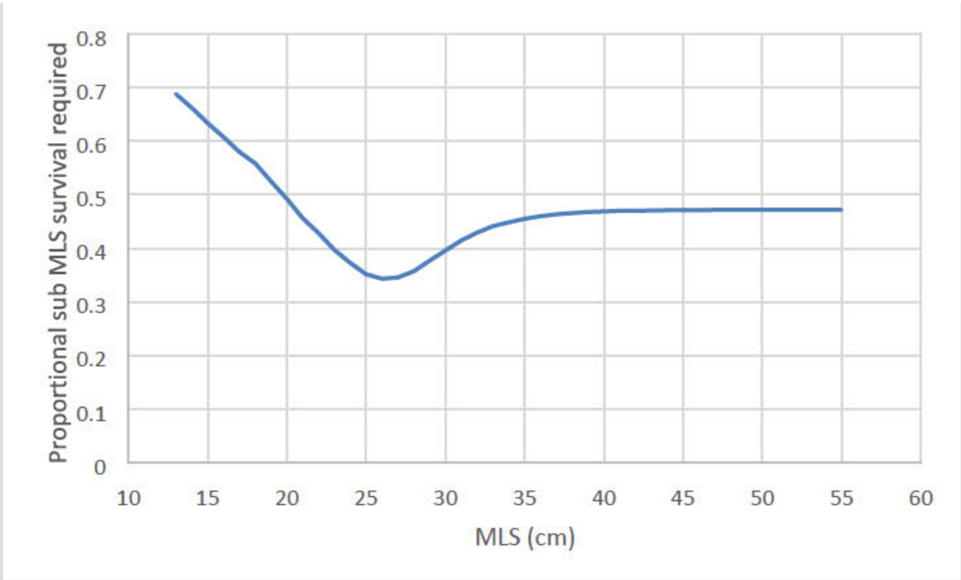


Figure 3-7: Proportional sub-MLS survival MHS needed to attain to achieve UDK equivalence with standard trawl (assumed ST survival being zero) by MLS. (Note: these values are point estimates and do not take into account uncertainty in the Jones and Millar (2017) selectivity curves).

3.3 B40% Equilibrium yield comparisons based on alternative MHS discard survival levels

3.3.1 Point Estimate Yields as based on Jones and Millar estimated selectivity curves (Figure 2-1).

MHS point estimate equilibrium yields at B40% as derived from the Hauraki Gulf assessment model were lower than standard trawl at fixed assumed discard survival below 63% (Table 3-1). At zero survival MHS yield is 90% that of trawl (Table 3-1). This implies MHS is 10% less efficient than ST in stock yield terms.

Table 3-1: Stock assessment model point estimate equilibrium yields at B40% for MHS over a range of discard survival levels. % change in yield is relative to the ST equilibrium yield value (2,406 t).

% survival	annual yield (tonnes)	% change in yield
0	2,172	-10%
25	2,258	-6%
50	2,350	-2%
63	2,406	0%
75	2,446	2%
100	2,549	6%

3.3.2 What is the difference in B40% yield between MHS and ST above the MLS (25 cm)?

We can compare the B40% yield efficiencies of MHS and ST by running the projection model assuming all snapper discarded from both gears below the MLS of 25 cm survive. The result from this simulation found that the B40% equilibrium yield for MHS of is still lower than ST (5%). MHS has a lower "legal-range" yield than ST because MHS removes more snapper in the 25-30 cm range, meaning MHS is less optimal than ST in terms of yield per recruit.

To summarise: half the difference in B40% equilibrium yield between MHS and ST is due to yield per-recruit selection inefficiencies above the MLS.

3.3.3 Predicted levels of MHS required discard survival accounting for uncertainty in the Jones and Millar selectivity estimates

Incorporating selection curve estimation uncertainty (Figure 2-2) into the Hauraki Gulf snapper model projections resulted in wide variation in MHS and standard trawl predicted yields (Figure 3-8). A high degree of overlap in yield distributions associated with different discard survival levels is also observed (Figure 3-8).

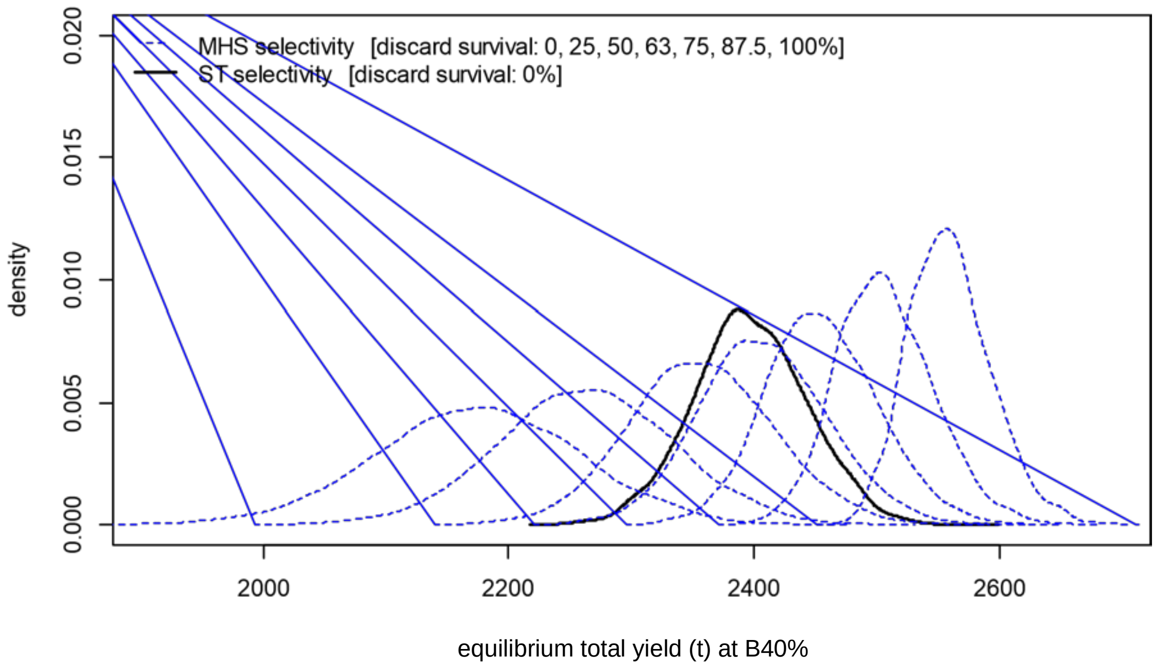
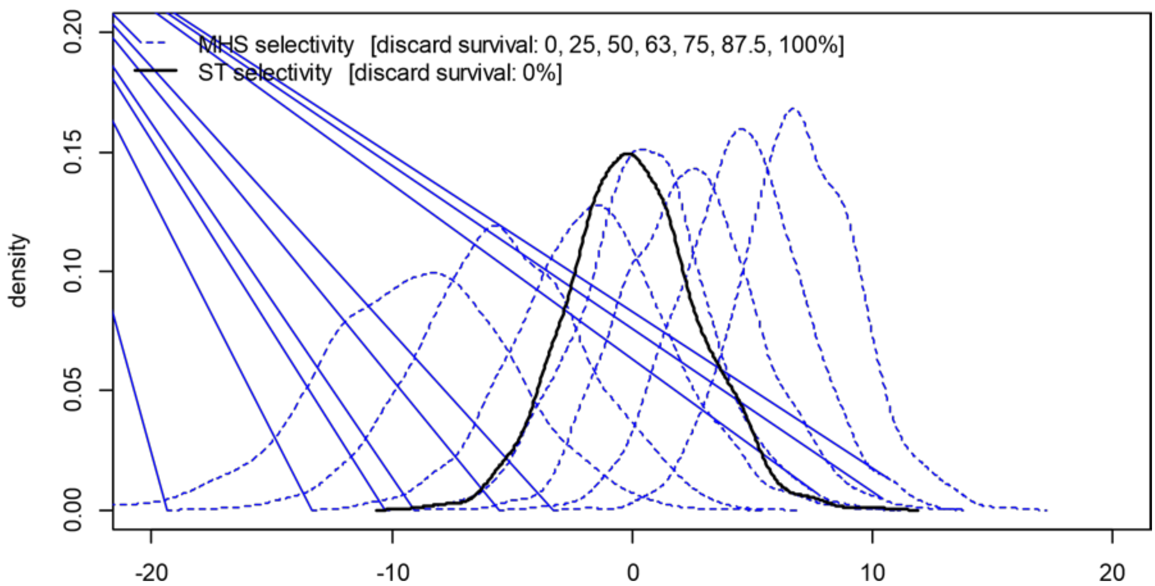


Figure 3-8: Distribution of predicted Hauraki Gulf model B40% yield (tonnes) relative to gear type and varying levels of assumed discard survival allowing for uncertainty in the selectivity curves.

A more informative way to examine the "no worse than standard trawl" criteria for MHS is to express MHS yield as a percentage difference relative to standard trawl yield (Figure 3-9). Although the point estimate reduction in yield of MHS over trawl under pursuant to a zero survival assumption was 10% (Table 3-1) factoring in uncertainty in MHS shows the relative reduction in yield could be as high as 20% (Figure 3-9).



MHS % difference in B40% equilibrium yield relative to ST

Figure 3-9: MHS % difference in equilibrium B40% yield relative to standard trawl at different assumed discard survival levels (note standard trawl discard survival is zero percent in all comparisons.

MHS equilibrium yield equivalence to standard trawl hypothesis test results

The test of equivalence in B40% between MHS and standard trawl is made pursuant to the assumption of "no worse" yield, the null hypothesis for the test, which we are hoping to be able to reject, being "worse yield". This means our test is effectively "one-tailed", the test equivalence range being from "no-worse" to "infinitely better than" (Figure 3-10).

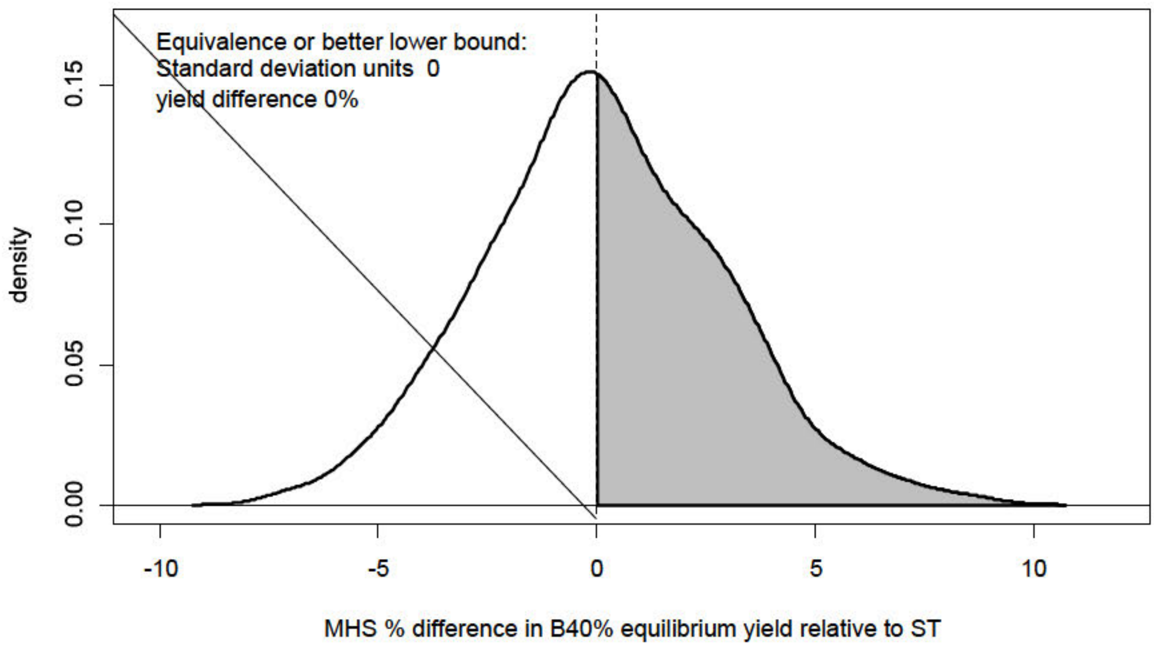
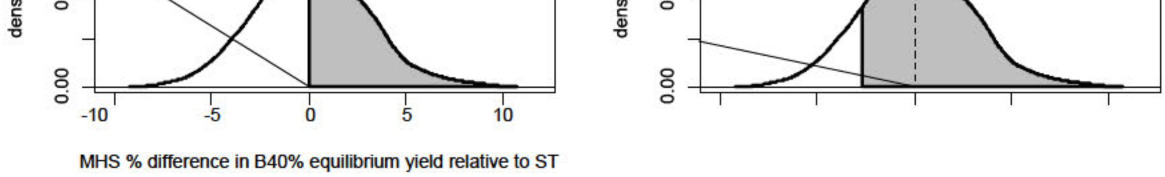
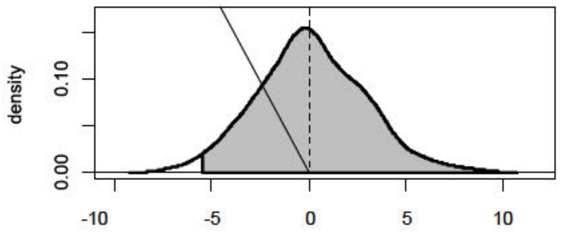


Figure 3-10: Range of deemed equivalence from "no-worse" (0% difference) to "infinitely better than". Graph shows the expected B40% equivalence distribution derived from comparing ST with ST (itself).

We ran tests of equivalence for lower bound tests of equivalence at 0, 1, & 2 standard deviations which equated to 0%, -2.76%, and -5.52% reduction in B40% yield (Figure 3-11) as derived from the standard trawl expected equivalence distribution (Figure 3-10).





MHS % difference in B40% equilibrium yield relative to ST -
10 -5 0 5 10

MHS % difference in B40% equilibrium yield relative to ST

Figure 3-11: Graphical representation equivalence ranges tested these corresponding to "no-worse" reductions in B40% yield of 0%, -2.76%, and -5.52% (0,1,2 standard deviations as derived from the expected ST equivalence distribution).

MHS needs to achieve to an average discard survival rate in excess 75% in order to attain statistical equivalence in B40% equilibrium yield with standard trawl (Table 3-2). The -2.76 and -5.52% equivalence test lower-bound cut-off criteria had the expected effect of reducing the level of discard survival required to achieve MHS yield equivalence with standard trawl (Table 3-2). The required MHS survival matched the point estimate prediction (63%: Table 3-1) when the lower bound of the equivalence range was -5.52% (Table 3-2).

Table 3-2: Level of discard survival MHS would need to achieve to satisfy the "no worse" hypothesis. Shaded cells correspond to survival rates where the "no worse than" criteria is satisfied at the < 5% level of significance.

	MHS % difference in yield to ST equivalence cut off %		
survival	0.00%	-2.76%	-5.52%
0	0.994	0.957	0.851
25	0.955	0.806	0.525
50	0.7	0.371	0.096
63	0.451	0.122	0.017
75	0.178	0.024	0.001
87.5	0.021	0.002	0
100	0.001	0	0

3.4 B40% Equilibrium yield comparisons based on right-shifting MHS selectivity curve in combination with alternative minimum discard survival levels.

An overall average MHS discard survival level in the order of 60-80% is likely to be very difficult to achieve in practice and also very costly to demonstrate. Therefore it is recommended that PSH should also consider the alternative mitigation option of shifting the MHS selectivity to be closer to that of standard trawl. To aid this consideration a series of B40% yield simulations were undertaken to determine how many millimetres the MHS selectivity curve would need to shift to achieve yield equivalence with ST. Shifts were undertaken to make the centre point of the curve (the L50 parameter) closer to that of ST. Only shifting the L50 selectivity curve parameter meant the shape of the MHS selectivity curve remained unchanged.

Under the premise that average MHS discard survival levels equal to or less than 30% might be more easily attainable and less costly to demonstrate, B40% yield simulations were also undertaken to determine the amount of MHS selectivity curve shift required to achieve yield equivalence with standard trawl assuming 10, 20 and 30% MHS discard survival.

3.4.1 Point estimate MHS selectivity curve shifts

On the basis of the MHS and ST point-estimate selective curves (Figure 2-1) the MHS selectivity curve would require a 25.5 mm shift to attain equivalence in B40% equilibrium yield with standard trawl (Table 3-3, Figure 3-12). At 30%MHS discard survival the amount of selectivity shift required reduces to 15.5 mm (Table 3-3, Figure 3-12).

Table 3-3: Amount the L50 MHS selection parameter would need to shift to achieve B40% Equilibrium yield statistical equivalence with standard trawl relative different assumed levels of MHS discard survival.

	assumed MHS discard survival			
	0%	10%	20%	30%
L50 shift (mm)	25.5	22.8	19.6	15.5

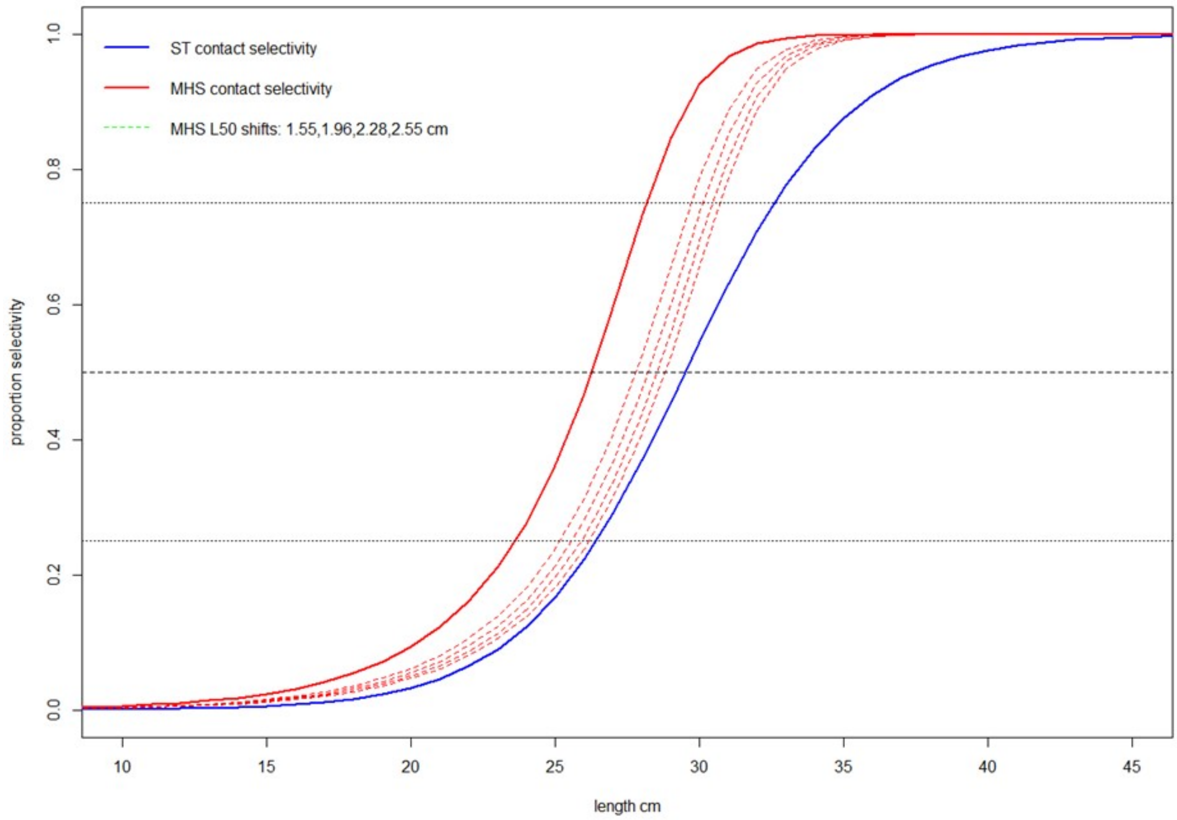


Figure 3-12: Comparative MHS L50 curve shifts (cm).

3.4.2 Predicted amount of MHS selectivity shift required accounting for uncertainty in the Jones and Millar selectivity estimates

The MHS selectivity shift in excess of 40 mm is required to attain statistical equivalence in B40% equilibrium yield with standard trawl (Table 3-4), when considering uncertainty in the selectivity estimates. The -2.76 and -5.52% equivalence test lower-bound cut-off criteria had the expected effect of reducing the level of selectivity curve shift required to achieve MHS yield equivalence with standard trawl (Table 3-4). The amount of required MHS curve shift approaches the point estimate prediction (25.5 mm; Table 3-3) when the lower bound of the equivalence range was -5.52% (Table 3-4).

Table 3-4: Amount MHS selectivity curve shift needed (mm) to satisfy the "no worse than ST" hypothesis test at 0% MHS discard survival. Shaded cells correspond to survival rates where the "no worse than" criteria is satisfied at the < 5% level of significance.

L50 shift (mm)	MHS % difference in yield to ST equivalence cut-off 0.00%	-2.76%	-5.52%
0	0.992	0.933	0.828
10	0.918	0.744	0.467
20	0.688	0.408	0.165
25.5	0.505	0.199	0.083
30	0.324	0.137	0.047
40	0.136	0.034	0.009
50	0.044	0.005	0.001

A MHS selectivity shift between 30 - 40 mm is required to attain statistical equivalence in B40% equilibrium yield with standard trawl at 30% MHS discard survival (Table 3-5). At 30% discard survival, the required MHS selectivity shift matched the point estimate prediction (15.5 mm; Table 3-3) when the lower bound of the equivalence range was -5.52% (Table 3-5).

Table 3-5: Amount MHS selectivity curve shift needed (mm) to satisfy the "no worse than ST" hypothesis test pursuant to 10% and 30% average MHS discard survival levels. Shaded cells correspond to survival rates where the "no worse than" criteria is satisfied at the < 5% level of significance.

10% MHS discard survival

L50 shift (mm)	MHS % difference in yield to ST equivalence cut-off 0.00%	-2.76%	-5.52%
0.917	0.732		0
10	0.856	0.662	0.358

0.984
0.910

30% MHS discard survival

L50 shift (mm)	MHS % difference in yield to ST equivalence cut-off 0.00%	-2.76%	-5.52% 0
0.715	0.374		
10		0.307	0.6850.090

22.8	0.447	0.195	0.073
30	0.238	0.064	0.022
40	0.066	0.017	0.002
50	0.015	0.001	0.000

20	**15.5**	0.491	0.173	0.049
	20	0.310	0.072	0.018
	30	0.098	0.025	0.001
	40	0.017	0.001	0.000
	0.292	0.564	0.114	

4 A consideration of mortality associated with snapper passing through MHS and ST meshes relative to a UDK ratio

The above simulations do not take account of mortality associated with snapper passing through the meshes of MHS and standard trawl nets. There is virtually no information on through-mess mortality for New Zealand snapper, however based on overseas research for similar sparid-type species it is probably "unreasonable" to assume mesh contact survival in snapper is 100% (Harley et al 2000).

Again using the Kaharoa observed mesh and cover length frequency data as a proxy for Hauraki Gulf snapper available to the trawl, it is possible to derive simple UDK-type catch ratios for MHS and standard trawl taking into account different assumed levels of both discard mortality and throughmesh mortality. The MHS and ST point estimate selectivity curves (Figure 2-1) suggest a through-mesh mortality of only 5% in both gears would be sufficient to negate the UDK disparity between MHS and standard trawl (Table 4-1). At a 5% through-mesh mortality level standard trawl is predicted to kill more snapper from through-mesh encounters than the gear kills by surface discarding (Table 4-1). For the same weight of landed catch twice as many snapper are predicted to die passing through standard trawl meshes than MHS meshes (Table 4-1).

Table 4-1: Predicted UDK and mortality ratios for MHS and ST at 100% and 95% through-mesh mortality rates.

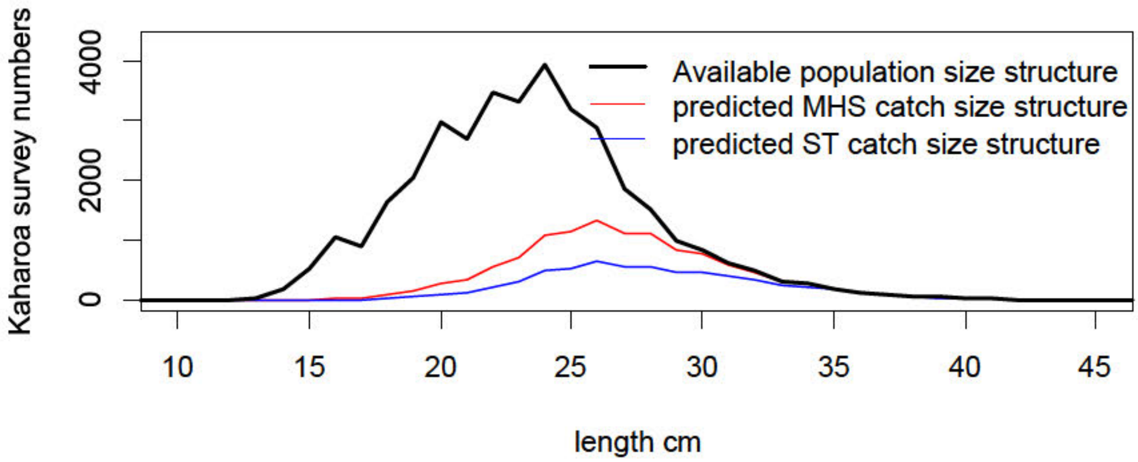
	MHS	ST	MHS	ST
discard survival	0%	0%	0%	0%
mesh survival	100%	100%	95%	95%
UDK	0.686	0.445	0.942	0.950

Expected catch breakdown

% snapper landed	72%	79%	66%	64%
% snapper killed through mesh	0%	0%	9%	19%
% snapper not landed but killed by discard	28%	21%	25%	17%

5 Conclusions

There was strong statistical evidence that the MHS UCK ratio as based on the available size structure from the Kaharoa survey, is "worse than" standard trawl UCK ratio, the UCK ratio between MHS and standard trawl in all three simulation tests, with UCK ratios for MHS being "worse-than" the standard trawl in 93-98% of the bootstraps. This result is not surprising given the Jones and Millar point estimate selectivity for MHS implies higher selectivity for sub-MLS snapper than standard trawl (Figure 2-1). Differences in these selectivity curves, as applied to the Kaharoa survey cover + cod-end length frequency data, imply that the catch rate of 25 cm snapper by MHS is double that of standard trawl, increasing to over 5 times that of standard trawl for the smallest commonly caught length-class (Figure 5-1).



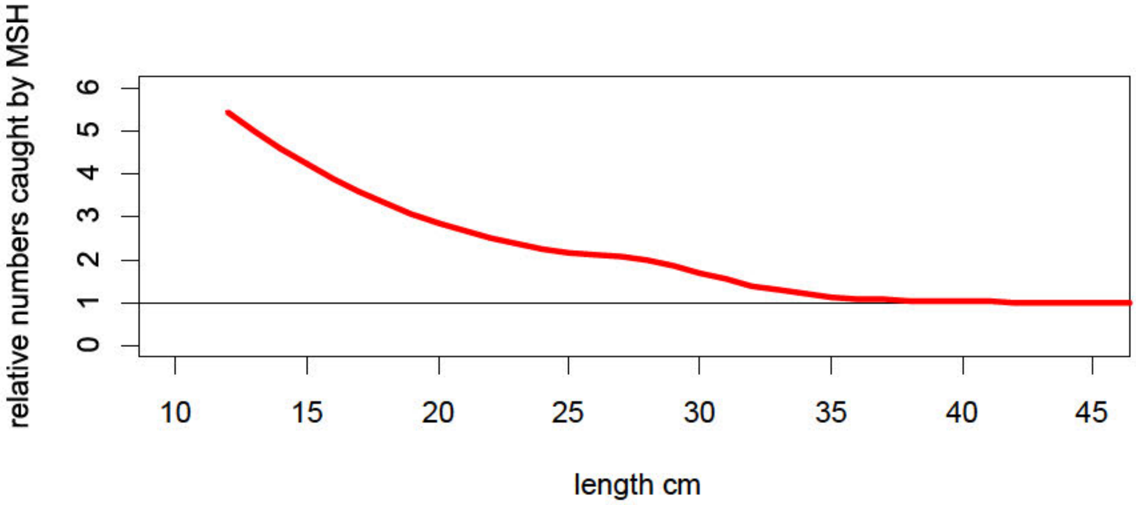


Figure 5-1: Top graph shows the expected length compositions of MHS and ST after application of the Jones and Millar point estimate selectivity curves to the combined Kaharoa cod-end cover data (black line). Bottom graph shows the relative numbers of snapper caught by MHS compared to ST by length.

Assuming a snapper population "available" length structure similar to that observed in the Kaharoa survey the point estimate survival MHS would need to achieve UDK equivalence to standard trawl was ~35%.

Applying the Jones and Millar point estimate selectivity curves (Figure 2-1) in a Hauraki Gulf snapper stock assessment model, the model predicts MHS would need a discard survival level of 63% to achieve yield equivalence with standard trawl. Taking into account uncertainty in the selectivity estimates MHS would need to achieve discard survival rates in the order of 80% to achieve statistical equivalence at a 95% confidence level or better.

The reason the UDK and B40% point survival estimates differ (35 c.f. 63%) is in part due is due to MHS yield pre-recruit inefficiencies above the MLS. Another limitation of UDK as a comparative measure for commercial gear is these ratios can change if the available length composition to the gear changes.

Yield-based simulations indicate MHS will need to achieve an average discard survival rate in the order of 60-80% to mitigate the yield disparity with standard trawl. This level of discard survival is likely to be very difficult to achieve in practice and also very costly to demonstrate. Therefore it is recommended that PSH should consider dual mitigation options involving shifting the MHS selectivity to be closer to that of standard trawl in conjunction with undertaken less costly survival studies aimed at establishing better-than 10-30% discard survival.

On the basis of the Kaharoa selectivity trial data, a through-mesh mortality level of only 5% has the potential to mitigate simple UDK-type ratio disparities between MHS and standard trawl. The predicted level of incidental mortality at the 5% through-mesh mortality level for standard trawl is similar to that derived by Harley et al (2000).

The level of through-mesh mortality required to mitigate disparity in B40% equilibrium yield has yet to be determined but likely to be at least a similar order of magnitude to that indicated by UDK ratio comparisons (e.g. 10-20%). Through-mesh mortality should be a consideration in the establishment of a “no worse than” evaluation criteria for MHS. It is recommended PSH invest in research aimed at establishing credible snapper through-mesh mortality lower limits for MHS and standard trawl.

6 Acknowledgements

The authors would like to acknowledge s 9(2)(a) [redacted] for critical review of the final draft.

7 Glossary of abbreviations and terms

EITT	MPI's Enabling Innovative Trawl Technologies Initiative
L_{50}	Length at which 50% of fish are retained in the codend
MHS	Modular Harvesting System
MPI	Ministry for Primary Industries
PSH	Precision Seafood Harvesting
SR	Selection range.
UCK	Undersized snapper Caught per kilogram of legal snapper
UDK	Undersized snapper Deaths per kilogram of legal snapper

8 References

Bentley, N. 2015. An analysis of catches and fish survival associated with the Modular Harvest System (MHS) and conventional commercial trawl gear. Unpublished Trophia contract report for Precision Seafood Harvesting Ltd.

Francis R.I.C.C.; McKenzie, J.R. (2015). Assessment of the SNA 1 stocks in 2013. New Zealand Fisheries Assessment Report 2015/76. 82 p.

Harley, S.J., T.B. Millar, and B.H. McArdle. 2000. Estimating Unaccounted Fishing Mortality Using Selectivity Data: an Application in the Hauraki Gulf Snapper (*Pagrus auratus*) Fishery in New Zealand. Fisheries Research 45(2):167–78

Jones, E. Millar, R. 2017 SNA 1 Selectivity Trials Report, report prepared for Precision Seafood Harvesting NZ Ltd

Ministry for Primary Industries 2016 Snapper (SNA1) Management Plan Report prepared for SNA1 Strategy Group held by the Ministry for Primary Industries Wellington (https://www.mpi.govt.nz/protection-and-response/sustainable-fisheries/snapper-1management-plan)

Appendix A YPR model description

The model used for the YPR simulations was a length-based proxy of the age-based model used in the 2013 SNA 1 assessment, having similar growth, mortality and gear selectivity characteristics as the Hauraki Gulf sub-stock. The key differences between the YPR and SNA 1 assessment model are given in Table 1.

Table 1: Key differences between the YPR and 2013 SNA 1 assessment models.

YPR Hauraki Gulf proxy model	2013 SNA 1 assessment model
single area (Hauraki Gulf characteristics) length –based	three area with movement age-based
deterministic model runs	stochastic model runs
dead discarded fish included in catch	dead discarded fish not included

In the YPR model harvest was specified relative to four fisheries in the model:

1. Commercial longline (LL)
2. Commercial single trawl (ST)
3. Commercial Danish seine (DS)
4. Recreational line (REC)

Model estimates of yield represented the total weight of fish LANDED by each of the four method fisheries. The weight of discarded dead fish (discard mortality) was accounted for in the model but did not contribute to the measure of yield.

Since snapper are known to have different vulnerabilities to fishing gears depending on their length, it was critical that the YPR model accommodated the different selectivity characteristics of the various fishing methods. The selectivity characteristics of each fishery in the model can be represented as a curve of relatively selectivity by length (Figure 1).

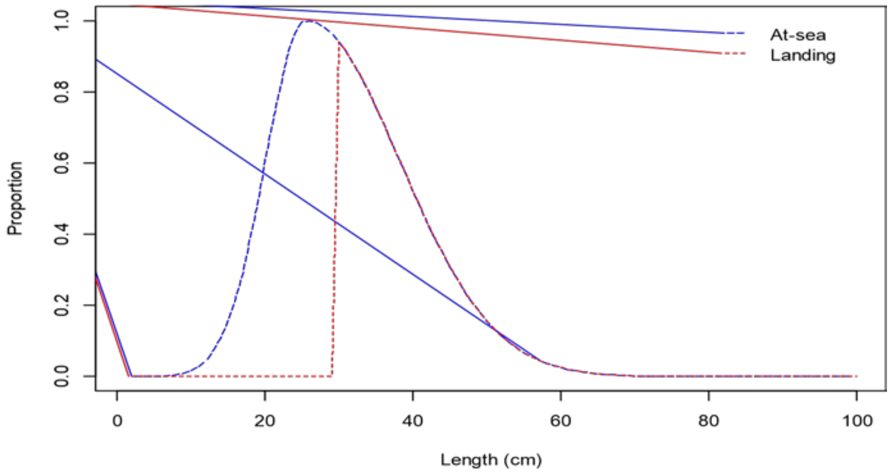


Figure 1: Selectivity curve for the recreational line fishery. The blue line shows the total selectivity of the method by length. From the graph it can be seen that 25 cm snapper are 100% (1.0) selected whereas 60 cm fish are poorly selected. The dotted green line shows the selectivity relative to a 30 cm MLS. A high proportion of snapper less than the MLS are selected by this method implying that a large proportion of the snapper taken by recreational line are likely to be returned to the water.

The model selectivity curves for each method are given in Figure 2, with the blue lines denoting the total at-sea selectivity and the black lines showing the proportion of fish discarded under the MLS that are likely to have died.

The YPR results were tabulated relative to the "status quo" scenario – i.e. the current commercial and recreational MLSs (25; 30cm) and current assumed levels of discard mortality of 100% for commercial Danish seine and trawl and 10% for commercial longline and recreational line (Figure 2).

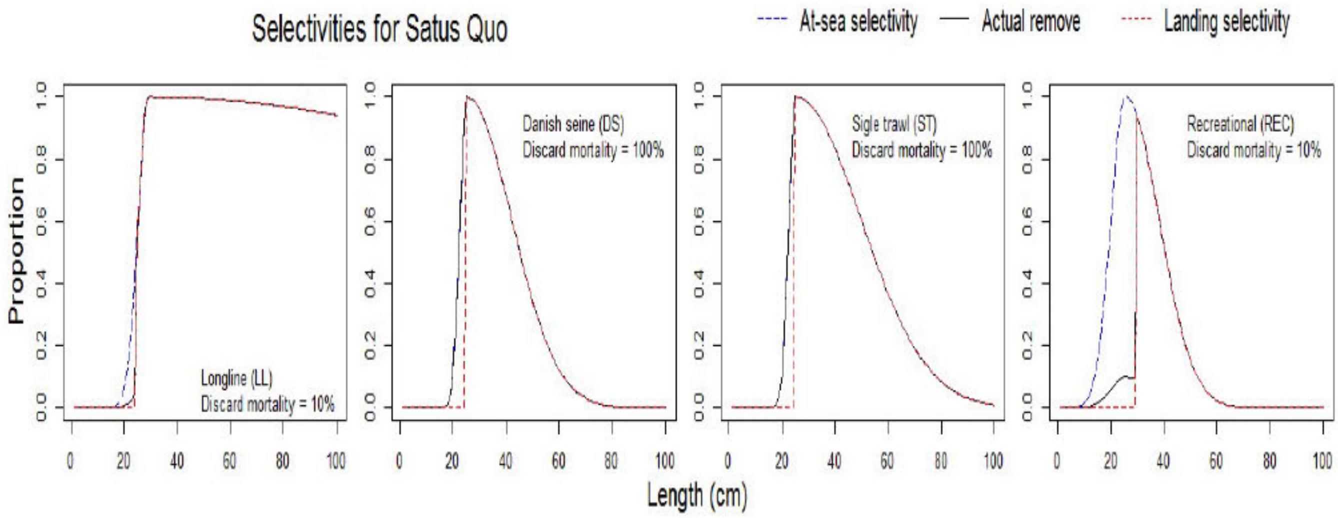


Figure 2: Selectivity curves for the four main fishing gears. The blue lines denote the total at-sea selectivity and the black lines show the proportion of fish discarded under the MLS that are likely to have died.

The yield statistics derived for each harvest strategy pertain to the stock at the target of 40% B0 with the relative landed catch component by fishing method being:

- Standard trawl or MHS: 60 %
- Recreational line : 40 %

Appendix B Derivation of “Total” MHS and Standard Trawl Derivation of “Total” MHS and Standard Trawl selectivity estimates for use in the model yield simulations

In order to compare the relative yield implications of MHS and standard-trawl in a stock assessment 'YPR-type' simulation we need to have estimates to **Total-selectivity** for the two methods.

At the stock removal level fishing-gear total-selectivity defines the difference between the "true" size or age structure of the stock and the size and age structure of the proportion of the stock the gear removes. Total-selectivity can be thought of as being a combination of three component selectivities:

1. **Available selectivity:** being the difference between the “true" stock size or age and the size and age composition of the population in the areas "available" to the gear
2. **Vulnerable selectivity:** being the difference between the size and age composition of the population in the areas "available" to the gear and fish that are Vulnerable to the gear i.e. fall into it. An example of available fish that are not vulnerable to it are fish that can out-swim the gear and hence never interact with it.
3. **Contact selectivity:** being the difference between the size and age composition of fish 'vulnerable' to the gear and the size and age structure of fish retained by it.

If we have measures of each of the three component selectivities we can derive "total selectivity" as follows:

*Available selectivity*Vulnerable selectivity*Contact selectivity = Total-selectivity*

Similarly we can derive estimates of any one of the above four selectivities given estimates of the other three (or combinations thereof).

Note: Selectivity estimated by stock assessment models are 'typically' total-selectivity estimates usually moderated by the effect of a MLS and are therefore 'retained' total-selectivity estimates.

PROBLEM:

To compare the relative yield implications of MHS and standard trawl in a stock assessment 'YPRtype' simulation we need to have estimates of Total-selectivity for the two methods but the bootstrap selectivity estimates we have from the Kaharoa selectivity trials (Figure 2-2) are estimates of 'contact selectivity' only.

PROPOSED SOLUTION:

The combined length frequency of snapper from the cod-end and cover from the Kaharoa selectivity trials represent the size composition of available and vulnerable snapper in the area the trails were conducted.

If it can be assumed that these compositional data are also representative of available and vulnerable stock to single trawl in 2017 it is then 'reasonable' to fit these data in the stock assessment (SA) model and derive a selectivity curve representing snapper both 'available and vulnerable' to trawl (henceforth termed AV selectivity).

We can therefore derive estimates of trawl 'Total selectivity' on the basis of the Kaharoa trail observational estimates as follows:

Total selectivity (MHS/ST) = AV selectivity (as derived from SA model) * MHS/ST Contact selectivity (as estimated from the trials).

To do this we need to make the following assumptions about the recent Kaharoa snapper data:

1. The survey data is representative of the trawl and MHS fishery as a whole in 2017.
2. That the available and vulnerable selectivity of trawl and MHS gear (as represented by the covered tow data) are the same, i.e. the difference in total selectivity between the two gear types is solely due to contact selectivity.

Test of assumption 1

Although we do not have data from the wider trawl and MHS fishery in 2017 to compare to the Kaharoa length data, we do have at-sea MPI observer data from the Hauraki Gulf trawl fishery in 2015.

The size range of the two data sets is broadly similar (Figure B-1). The strong peak of 24 cm fish seen in the 2015 fishery data might be due to a strong year class in the fishery, and the higher proportion of 25 -30 cm fish in the 2017 data is not inconsistent with this assumption. Overall it is "reasonable" to conclude that the 2017 Kaharoa survey length frequencies are representative of the wider trawl fishery.

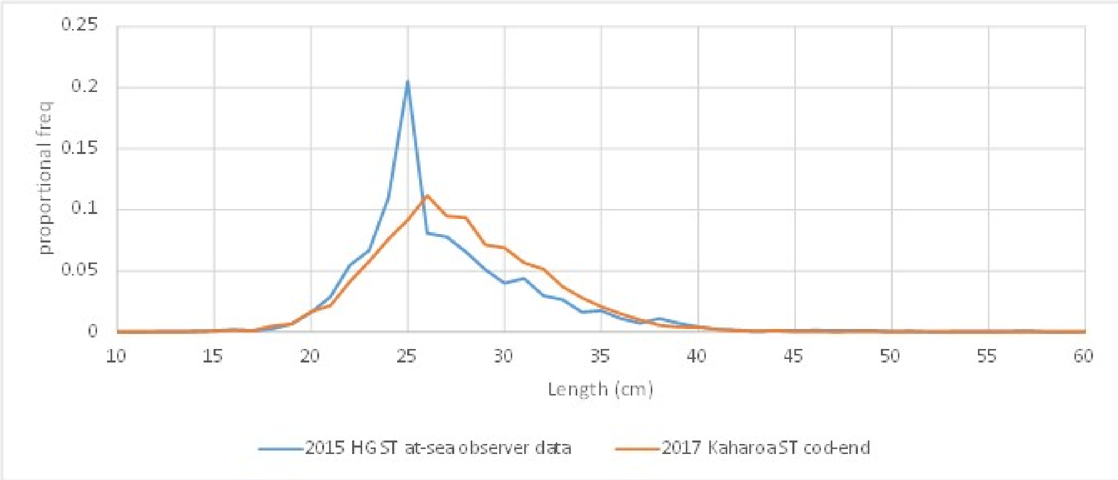


Figure B-1: Proportional length frequency of Hauraki Gulf trawl snapper as derived from the 2015 MPI at-sea observer programme and the 2017 Kaharoa selectivity trials.

Test of assumption 2

Again assumption 2 appears to hold as the combined MHS and standard trawl plus cover length frequencies seen in the Kaharoa survey were similar (Figure B-2).

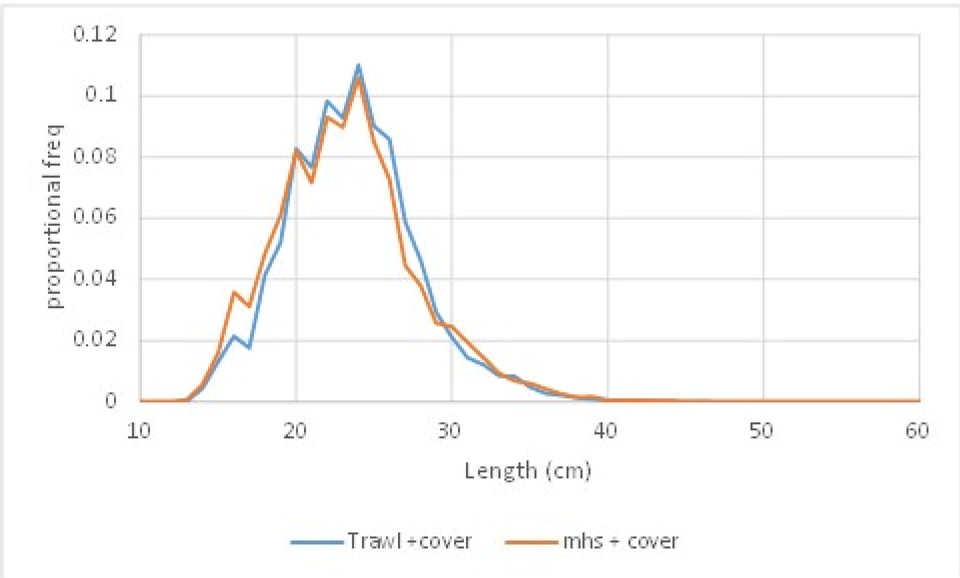


Figure B-2: Proportional length frequencies of the Kaharoa survey MHS cod-end + cover and standard trawl cod-end + cover

Conclusion:

The 2017 Kaharoa covered codend length data is likely to be broadly representative of the available and vulnerable Hauraki Gulf snapper population to the inshore trawl and MHS fishery in 2017.

Snapper trawl AV selectivity estimates from 2013 Hauraki Gulf model fits to the combined Kaharoa cover length frequency data.

AV snapper trawl selectivity was estimated as a 3 parameter double normal curve in the 2013 Hauraki Gulf model to the combined Kaharoa cover and cod-end data (Figure B-2). The model achieved a good fit to the Kaharoa data suggesting the AV selectivity parameters (22.1916, 3.79841, 7.30383) were reasonably well estimated (Figure B-3).

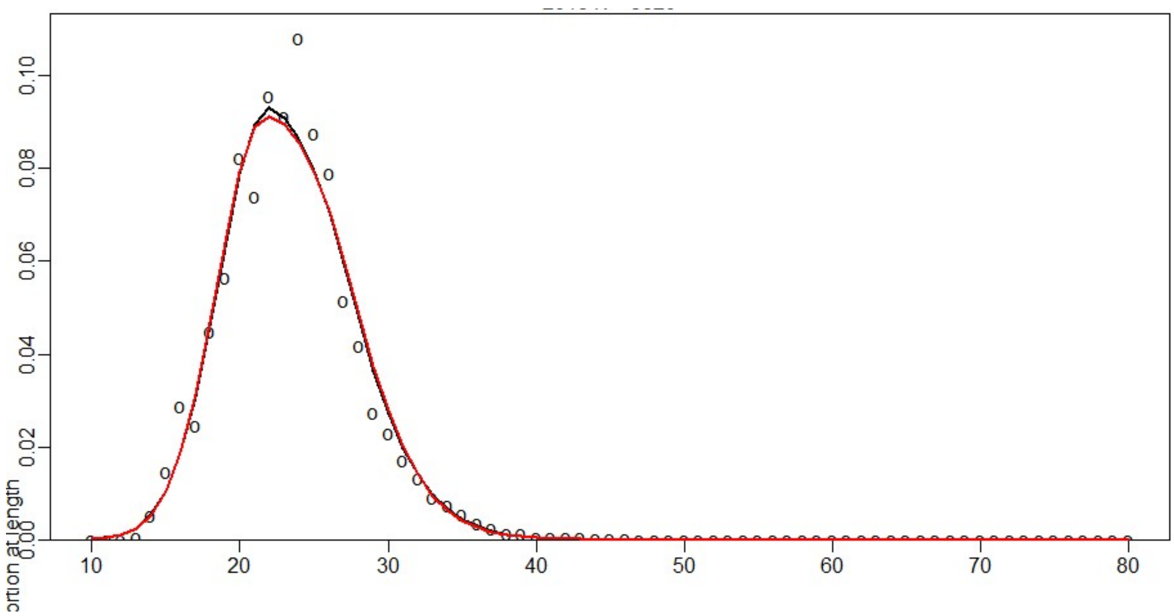


Figure B-3: 2013 Hauraki Gulf snapper assessment model fit to the combined Kaharoa cod-end and cover length-frequency data.

The estimated AV selectivity curve is strongly domed the length at maximum selectivity being 22.2 cm (Figure B-4), meaning both very small and very large snapper are not available and vulnerable to trawl in the areas these gears are likely to operate in the Hauraki Gulf. The combined total selectivity curve estimates for MHS and standard trawl are not markedly dissimilar (Figure B-5), with uncertainty in the respective contact selectivity curves resulting in a high degree of overlap (Figure B-6).

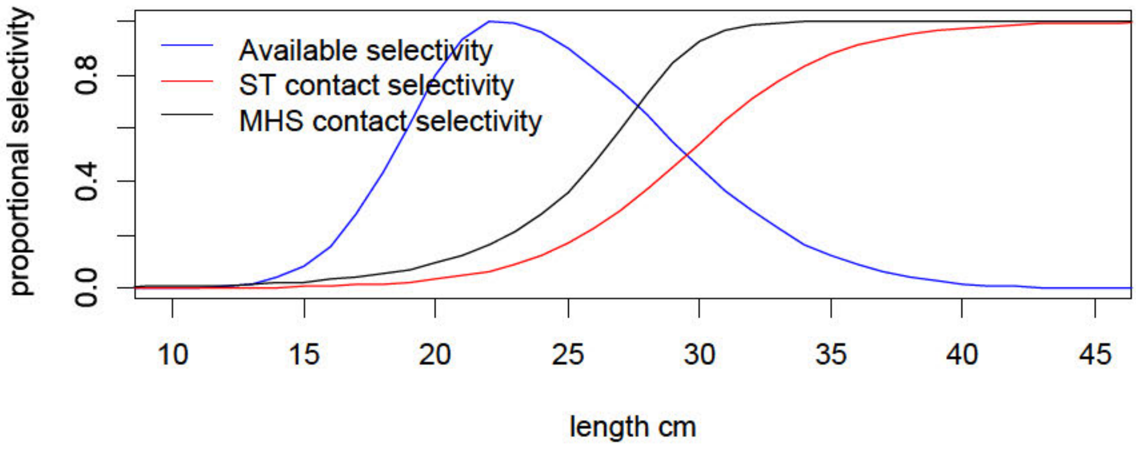


Figure B-4: Comparison of the AV, MHS and standard trawl selectivity curves.

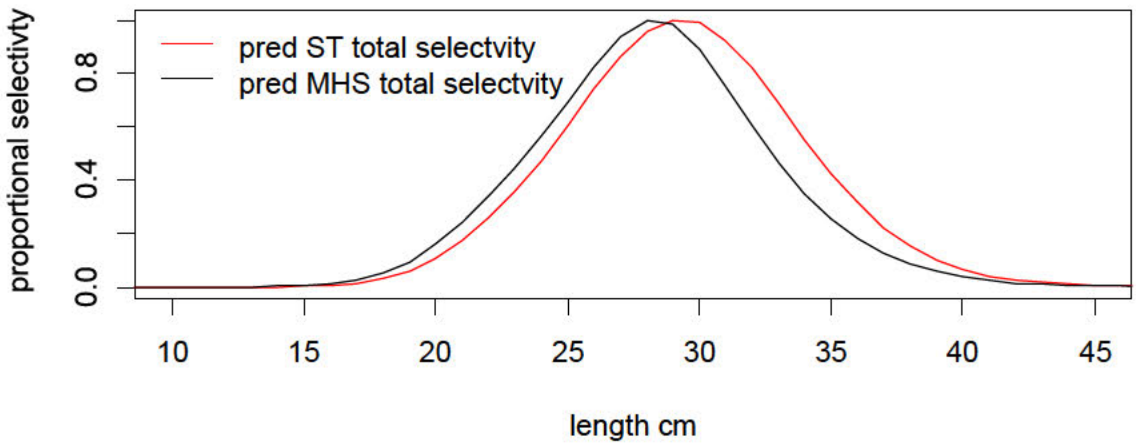


Figure B-5: Point estimates of MHS and standard trawl total selectivity derived from combining the individual curves in Figure B-4.

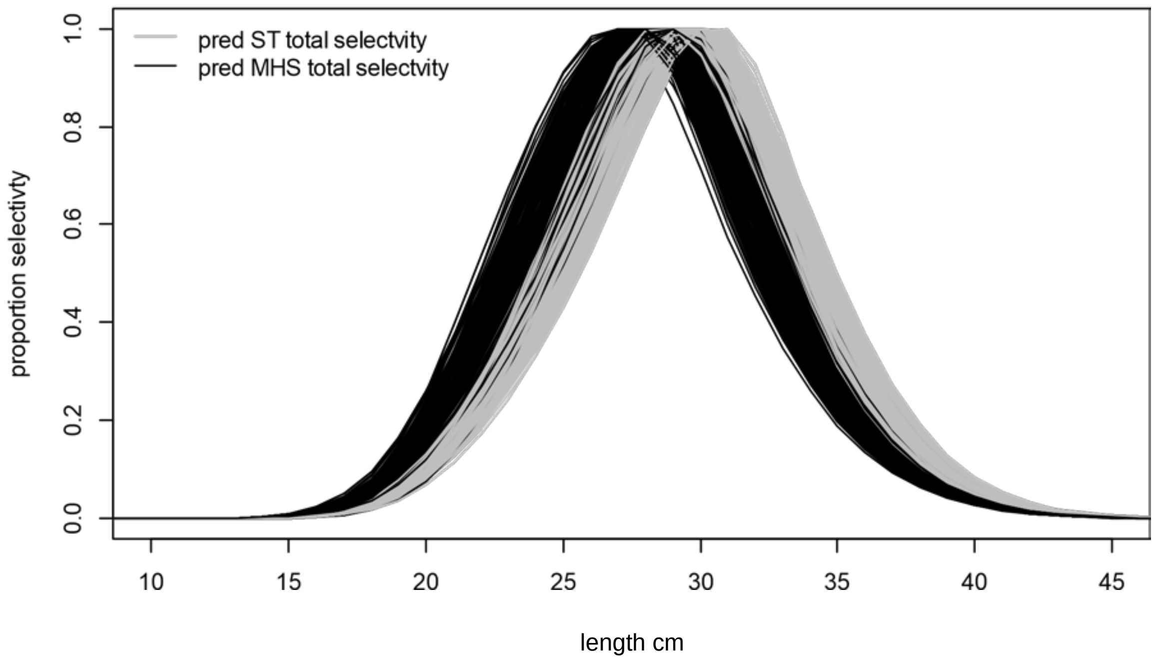


Figure B-6: 1,000 (bootstrap) alternative snapper Total selectivity curves for MHS and standard trawl (ST)

Appendix C Equivalence Hypothesis Testing

In **traditional hypothesis testing** we are usually interested in knowing with high statistical certainty if two sets of values (u_1 and u_2) come from **different parent distributions**. The mistake we are most concerned about making is concluding u_1 and u_2 come from different distributions when in fact they don't (termed in statistical parlance as making a Type I mistake). We formulate the problem as a test of the null hypothesis **(H_0) that u_1 and u_2 come from the same parent distributions** and set about trying to find the probability that this null hypothesis is true. If we find the probability of our equivalence null hypothesis is less than 5% ($p < 0.05$) then we can be 95% sure of the two distributions being different. It is important to realise that acceptance probabilities of the null hypothesis greater than 5% do not necessarily mean equivalence in u_1 and u_2 is true it simply means we can't reject this as a hypothesis. By way of a simple example; if were interested in knowing if aphids on tree A had on average larger wings than aphids on tree B a sample of one aphid from each tree would be unlikely be sufficient to establish the difference. Not being able to establish a difference in wing-size would not prove the equivalence null hypothesis is true it simply would means we were unable to reject it.

Equivalence hypothesis testing applies if we are more interested in knowing with high statistical certainty if two sets of values (u_1 and u_2) come from same parent distribution i.e. the mistake we are most concerned about making is concluding u_1 and u_2 come from the same distribution when in fact they don't (termed in statistical parlance as making a Type II mistake). Under equivalence testing we formulate our null hypothesis **(H_0) as being that u_1 and u_2 come from different parent distributions** and set about trying to find the probability associated with the null hypothesis being true. Again, we would need to establish that the probability of our difference null hypothesis being true was less than 5% ($p < 0.05$) in order to be >95% sure of the two distributions were the same.

In Traditional hypothesis testing (i.e. testing for difference) we typically are testing two distributions for congruence against a single reference statistic. For example if we were take a random sample of 100 Hispanic adult males and 100 white adult males and subtract the difference in their height you would fail to reject the equivalence null hypothesis if zero difference occurred within the 95% range of the data (Figure C-1). Equivalence testing requires specifying a range over which the equivalence criteria must be met. The observed set of values must fall within this range for the equivalence criteria to be met (Figure C-1).

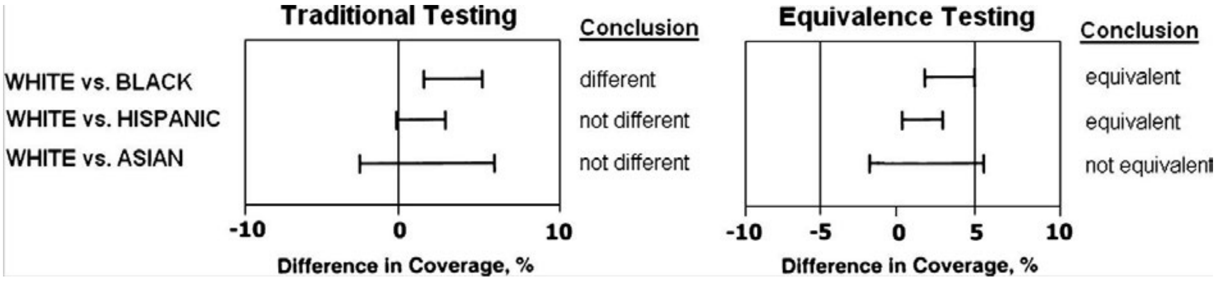


Figure C-1: Comparison of Traditional hypothesis testing which involves testing if a comparative reference point falls within and observed distribution and Equivalence testing which tests whether the observed distribution lies within an equivalence range.

Setting the equivalence testing interval:

Equivalence testing requires defining:

a. a statistic to compare between the two populations;
b. an "acceptable" equivalence interval for that statistic[1].

One way to do this is to set the interval based relative to the standard deviation of the "expected" equivalence distribution. The expected equivalence distribution can be derived by randomly comparing one of the distributions with itself (Figure C-2). Statistical equivalence (i.e. rejection of the non-equivalence hypothesis) is achieved if a statistically significant percentage of the observed comparative values fall within the equivalence range (e.g. > 95%).

[1] Typically defined as an upper and lower bound

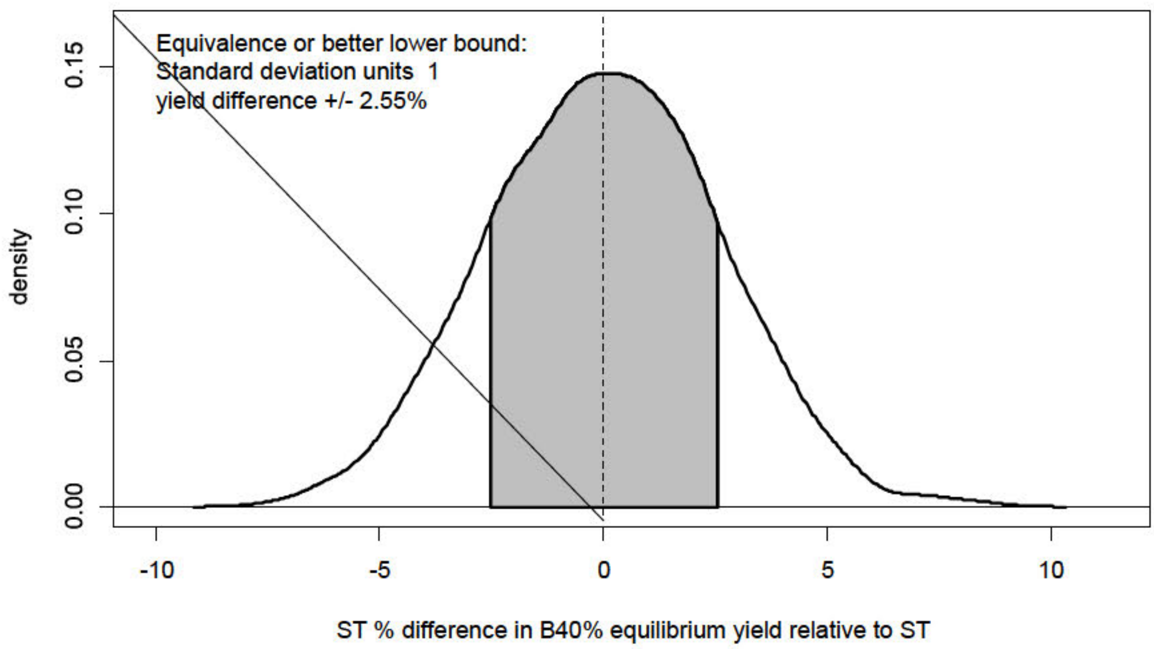


Figure C-2: Shaded area shows equivalence acceptance being +/- 1 standard deviation as generated by comparing the ST B40% difference to ST (i.e. itself).