Uncertainty in Length Measurements of Live Coral Trout: Implications for Compliance to and Enforcement of Minimum Legal Size Limits

B. D. Mapstone[1], G. Carlos[1], C. Lunow[2] & A.E. Reid[2].

Project No. B4.8

[1] CRC Reef & School of Tropical Environment Studies and Geography, Sir George Fisher Research Building, James Cook University, QLD. 4811. Australia.
[2] Queensland Department of Primary Industries, Oonoonba, QLD. 4810. AUSTRALIA

National Library of Australia Cataloguing-Publication entry

Uncertainty in Length Measurements of Live Coral Trout: Implications for Compliance to and Enforcement of Minimum Legal Size Limits.

Bibliography.

Includes index.

ISBN x xxxxxx xx x.

1. Fisheries – Queensland – Great Barrier Reef. 2. Fishery management – Live Reef Fish. I. Mapstone, Bruce D. (Bruce David). II. James Cook University of North Queensland.

338.372709943

This publication should be sited as Mapstone *et al.* (2003).
Uncertainty in Length Measurements of Live Coral Trout: Implications for Compliance to and Enforcement of Minimum Legal Size Limits.

CRC Reef Research Centre
Townsville, xxpp.

Published by the Cooperative Research Centre for the Great Barrier Reef World Heritage Area @ 2002

Non-Technical Summary

B4.8: Uncertainty in Length Measurements of Live Coral Trout: Implications for Compliance to and Enforcement of Minimum Legal Size Limits

Authors: **B. D. Mapstone, G. Carlos, C. Lunow & A.E. Reid**

Objectives

1. Assess whether post-capture changes in length occur in live coral trout held on board commercial line-fishing vessels.
2. Estimate the magnitude of within and between individual measurement error or bias.

Summary

Harvest of reef finfish from the Great Barrier Reef (GBR) is an important commercial and recreational activity. Minimum sizes for retention are an important instrument in managing the harvest of many species. Fishers' compliance with minimum legal sizes, and enforcement of these by officials, however, assumes the reliable and consistent measurement of the lengths of landed fish.

Recently, breaches by commercial fishers retaining allegedly undersized live coral trout have been challenged based on assertions that fish kept on board vessels effectively 'shrink' after capture and that measures of live fish are inherently variable. Prosecutions have failed due to arguments that either of these factors may occur, but there is currently little understanding of their magnitudes. Failed legal proceedings are cause for concern for fishers, who object to unproved accusations of illegal behaviour, regulators, who seek certainty in the prosecution of regulations, and society, which ultimately bears the cost of enforcement.

In this project we sought to assess the veracity of claims that live common coral trout (*Plectropomus leopardus*) change size following capture and to estimate the magnitude of variations associated with repeated measurements of coral trout, such as would occur when fishers measured fish at the time of capture and enforcement officers did so subsequently during compliance inspections.

We found that fish kept on board a commercial fishing vessel during normal fishing operations varied in size both positively and negatively over a 10 day trip. There was an overall average "shrinkage" of 1.76 $\pm$ 0.76 mm (95% CI). Fish showed a decrease in average size from soon after they were brought aboard the vessel, with the greatest reduction in size occurring in the first 1-2 days. Fish subsequently continued to "shrink" by smaller amounts, reaching their smallest size on the third last day of the trip when they were on average 3.2 $\pm$ 1.02 mm (95% CI) below their initially measured size. The greatest daily change in mean size was a positive change, or elongation, of 1.43 $\pm$ 0.31 mm ($\pm$SE) during the final day of the trip. It was notable that if the large changes in size over the first 1-2 days were ignored, the significant relationship between length of the fish and time at sea disappeared.

Large variations in length measurements were also observed between repeated measurements by different observers (three commercial fishers and two research officers) over short periods (<22.5 hours). These measurement errors, or uncertainties, were generally greater than the effect of fish shrinkage over the 10 days of the trip. Average 95% and 99% confidence intervals on measurements for common coral trout by different observers were 2.91mm and 4.83mm respectively. Average confidence intervals attributable to variation amongst repeated measurements by the same observer were 1.72mm (95% CI) and 2.35mm (99% CI).

Whilst there was a real overall uncertainty due to measurement error, much of this could be eliminated by careful attention to the techniques used to measure the fish. For example, the two researchers, who were using a pre-determined and consistent technique for length

estimation, had similar measurement biases and considerably greater precision of measurements than the fishers. The 95% confidence intervals on measurements attributable to between- and within-observer variations for observers using similar measurement techniques were reduced to 2.08mm and 2.16mm respectively.

The fundamental inability to determine an exact "true length" of a fish means that enforcement of minimum legal sizes cannot be 'knife-edged' and definite. With knowledge of the magnitude of this uncertainty, however, it is possible to determine the numbers of a given sample of fish that would be reasonably expected to fall below the legal size on repeated measurement simply because of measurement uncertainty. It is also possible then to stipulate how many fish would have to appear to be undersize before a reasonable conclusion would be made that the observations were not attributable to measurement uncertainty and that at least some of those fish were indeed being kept illegally. We present exemplar tables of such 'unlikely numbers' to illustrate how knowledge of measurement uncertainty might be used to construct rules for prosecution that include allowances for 'reasonable' margins of error. These tables allow fishers and enforcement officers to determine how many fish, from a given sample, they would expect to appear undersized by given amounts due only to measurement variation and see the points at which records of undersized fish become sufficiently frequent to be considered improbable and so warrant prosecution. Very many such tables can be generated, however, each associated with a different level of risk that a prosecution might proceed in error. The choice of what is an acceptable margin for such error ultimately resides with legislators and / or the judiciary. Once such a criterion is set, however, application of a standardised measurement methodology by all fishers and enforcement officers is likely to make enforcement of minimum legal sizes considerably more precise and robust than is currently feasible.

Conclusions and Recommendations:

- Uncertainty in measurement of live reef fish has a clear empirical basis and we have quantified both between-observer uncertainty and within-observer uncertainty and their effects of compliance monitoring.
- Evidence was found for short-term (days-weeks) changes in sizes of live coral trout held aboard a commercial fishing vessel, but changes were generally small (<5mm).
- We provide tables of the expected frequencies of fish measured to be under the legal size because of measurement uncertainty in a format designed to provide decision-making criteria for when prosecution is justified on the balance of probabilities.
- Enforcement, administrative, legislative & industry personnel should be advised of the potential magnitudes both of the uncertainty in measurements and the frequency with which given discrepancies in measurement might be expected during compliance and enforcement monitoring.
- Further work should be done to verify whether rudimentary instruction in measurement procedures is effective in reducing uncertainties in the measurement of fish and, more importantly, reducing the amount of variation between measurements by different people (e.g., at capture and at compliance inspection).
- A standard measuring technique should be defined and sent to all fishers and enforcement officers. In the interest of reducing the potential variation between observers' measures the authors suggest the following: the fish's mouth should be closed or gently and carefully forced closed, there should be no lateral compression of the body or caudal peduncle, and the length should be measured from the tip of the lower jaw to the posterior margin of the dorsal lobe of the tail. If the dorsal lobe is damaged, a measurement to the end of the ventral lobe should be used.

Contents

Acknowledgments

Funding for this work was provided by the Queensland Department of Primary Industries and the Cooperative Research Centre for the Great Barrier Reef World Heritage Area (CRC Reef). The fishing vessel, crew and facilities for the fieldwork were provided by Mr Vince Nielsen. We gratefully acknowledge the assistance of Vince Nielsen, Dennis Mahr and Bryan Woodhen in catching and measuring fish under normal fishing conditions. We thank Bob Grimley, Peter Tanner, Ian Brown, Gavin Begg and Andrew Tobin for their helpful comments on drafts of this report.

Background and Need

Harvest of reef fin fish from the Great Barrier Reef is an important recreational, commercial and tourist activity. Minimum sizes for retention are an important instrument for managing the harvest of several species. Compliance with minimum legal sizes by fishers and enforcement of them by officials assumes reliable measurement of the lengths of landed fish.

Since 1993 significant amounts of the commercial harvest have been retained alive for export to lucrative markets in Hong Kong. Breaches of fishers for retaining undersized live fish have been challenged on a number of occasions based on the assertion that fish kept on board vessels effectively 'shrink' in length after capture. Whilst the amount of 'shrinkage' might be small, the minimum legal size regulations are assumed to be knife-edged regulatory instruments and, in theory, enforced without margin for errors in measurement. This means that any real or perceived change in length of fish near to the minimum legal size between capture and measurement by enforcement officers may lead to erroneous prosecution of fishers. The potential for such errors has resulted in the failure of a number of prosecutions of fishers for possession of various numbers of apparently undersized fish. Such failed legal proceedings are cause for concern for both fishers, who object to unproved accusations of illegal behaviour, regulators, who seek certainty in the prosecution of regulations, and society, which ultimately bears the cost of enforcement. Without any information by which to judge if differences between measurements by fishers and enforcement officers are real or artefacts of measurement errors, effective enforcement of the minimum size regulations will continue to be difficult to effect. Calls for clarification of this issue have arisen from fishers, managers, enforcement officers and the judiciary.

In this research, therefore, we sought to assess the veracity of claims that live common coral trout (*Plectropomus leopardus*) 'shrink' following capture and to estimate the magnitude of random variations associated with repeated measurements of coral trout. We chose to work with common coral trout because it is the mainstay of the live reef food fish trade from the GBR and has consistently been the subject of disputed breaches.

Benefits

Benefits of the research will accrue to participants in the fishery, enforcement agents for the fishery and the judiciary. Fishers will benefit by being aware of the risks they take in 'hedging' and keeping fish that are questionably close to legal size and understanding the margin of caution they need to allow to avoid erroneous prosecution. Enforcement agents will benefit by being informed about the consequences of random variations in measurements of fish and the likelihood that a few fish that are apparently undersize could have been classed as such because of chance events. This knowledge will facilitate more informed judgements about when prosecutions are likely to succeed or fail. The judiciary will benefit from advice about when matters brought before the courts are likely to be beyond reasonable doubt and when they are within expectations of measurement uncertainty, rather than having to dismiss cases routinely because of lack of information.

Objectives

1. To assess whether post-capture changes in length occur in live coral trout held on board commercial line-fishing vessels.
2. To estimate the magnitude of within- and between-person measurement error or bias.

Introduction

Line fishing is a major commercial, tourism (charter fishing) and recreational activity on the Great Barrier Reef (GBR) (Gwynne 1991; Higgs 1996; Mapstone *et al*. 1996, 1997). A regime of minimum legal sizes for harvest is one of the primary instruments for managing the GBR line fishery (Anon 1999). Broadly, the rationale for setting minimum legal size limits revolves around protection of individuals until they have had the opportunity to spawn at least once, thereby ensuring the conservation of sufficient biomass of spawning fish to provide replenishment of populations. Enforcement of the minimum legal size provisions is an important consideration, therefore, since the harvest of under-size fish impacts on this planned reserve of spawning stock biomass.

The enforcement of minimum (or maximum) size limits in fisheries frequently assumes a knife-edged assessment of whether a fish is above or below a size limit. That is, there is little or no formal allowance for uncertainty in the measurement of fish. Uncertainty here equates with unavoidable variation between successive measurements of a fish arising from a variety of sources. Successive measurements of each fish are to be expected because the fisher must measure the fish at the time of capture to decide whether to keep or release it, and then the enforcement officer must measure the fish to decide whether the fisher is complying with the law. In the commercial fishery, other people (such as buyers) might also measure the fish to verify whether they are being sold legal product.

Variation in measurements of each fish can arise from three main sources:

1. Differences between observers in the way the fish is held and measured or differences in the interpretation of length;
2. Slight inconsistencies in the way an individual records length on successive occasions;
3. Changes in the actual length of the fish between measurements, either through growth or 'shrinkage', possibly because of physiological reactions to the stress of being caught and handled that result in small changes in tissue torpor;

In the absence of real changes in fish length, 1 (between observer variation) would mean that different people measuring a fish at the same or different times would sometimes, perhaps often, record different lengths. These differences might reflect either random variation or observer-specific measurement bias or a combination of both. Bias would mean that one measurer's observations are always greater than (or less than) another's, whereas random variation means that the differences between observers are not consistent and cannot be predicted.

The second source of variation (2, within-observer variation) would mean that the same fisher measuring the same fish repeatedly over short intervals would sometimes record different lengths. Between- and within-observer variation are referred to collectively as measurement process error or measurement uncertainty. Both will be represented in the comparison of any measurements of the same fish taken by different people and they are expected to be additive.

The third source of variation (3, real change in length) is largely beyond the control of the measurement process. If other sources of variation were known to be absent, this real change might be specified relatively easily and would be expected to be small over short periods. In reality, however, apparent real changes in length are likely to be confounded with measurement process errors. This confounding may mean that measurement uncertainties may either obscure or amplify real changes in length and render small changes in length difficult to estimate unequivocally.

The uncertainties introduced by the above sources of variation have significant implications for decisions made by fishers and enforcement officers. Operationally, they mean that each fisher may need to allow for a margin of error in length measurement when deciding whether to keep each fish or enforcement officers may need to allow for a margin of error when deciding whether to prosecute a fisher in possession of an apparently under-size fish, or both. Contention about measurement procedures, poor understanding of the uncertainty in length measurements, and preliminary evidence that fish apparently shrink slightly over the several days between capture and landing (V. Nielsen, unpub. video footage) have resulted in failed prosecutions for retention of undersized fish, largely because the judiciary is prone to give the benefit of doubt to the defendant. Such failed legal proceedings are cause for concern for fishers, who object to unproved accusations of illegal behaviour, regulators, who seek certainty in the prosecution of regulations, and society, which ultimately bears the cost of enforcement.

In this study we sought to estimate the magnitudes of the above three sources of variation in measured lengths of the common coral trout, *Plectropomus leopardus*, and document their relative and combined implications for the enforcement of legal limits on the size of common coral trout harvested from the Great Barrier Reef. We then used estimates of measurement error and post-capture changes in fish size to estimate the expected frequencies of minimum legal size violations arising from chance alone during the measurement of specific numbers of fish sampled from catches.

Materials and Methods

Field Methods

All data presented here were collected over 10 days at sea on board a commercial fishing vessel during normal fishing operations on the Great Barrier Reef (GBR). Common coral trout to be used in the study were caught by standard hook and hand-line gear from the commercial fishing vessel. A total of 42 common coral trout of near to minimum legal size (370.3mm-465.3mm mean Total Length (TL)) were selected from daily catches from the main vessel over 5 days. Only fish caught from the main vessel, rather than from fishing dories, were used in order to standardise handling over all fish and to minimise the delay between capture and measurement for this study.

Each fish was measured at the time of capture, bathed in freshwater for one minute to remove ectoparasites, and placed with two other fish either in one of six 63L blue plastic Nally bins or one of eight hard plastic meshed 'lug-basket'. The Nally bins were kept under shade on the deck of the vessel and supplied with fresh flowing seawater. The lug-baskets were immersed in the vessels large holding tank below deck. The three fish in each bin or basket differed in length by approximately 20mm. These differences in length were conspicuous and, together with identifiable marks such as colour morph, hook damage, etc., enabled unambiguous identification of individual fish without tagging or otherwise marking them. All measurements of length were made using an approved standard measuring board[3] of 52cm length, with a solid vertical head plate and horizontal measuring plate graduated at 0.5mm intervals. Data were recorded on pre-prepared waterproof data sheets.

1. Within- and between-observer variation

Sampling Design

Within-observer and between-observer variation in measurements of length were investigated by a structured repeated measures sampling design in which each of five observers measured five times each of the 18 fish held in the Nally bins on deck. Fish were removed one by one from holding bins by a researcher using a net and taken to a measuring station that was out of sight of the holding area. Each fish was measured once by one of the observers and then placed in a temporary holding tank at the measurement station but out of sight of the observer. When all three fish from a bin had been measured, they were transferred from the temporary holding bin back into their original holding bin. A complete set of measurements (18 fish) was completed as quickly as possible for one observer before repeating the process for the next observer, with a total of five observers (two research staff and three commercial fishers) measuring all fish. Observers left the measuring station after they had completed a 'run' of measurements so that they could not observe measurements being made by other observers.

The entire process was then repeated five times, with the order of measuring by the five observers held constant throughout the 5 replicate 'runs' of measurements. The order in which each bin was processed was changed for each run but was the same for each observer within a run. The order of removal of fish from each bin was haphazard. Logistics of set up and availability of the fishers whilst they were doing other duties meant that the experiment had to be run over two days. The first three runs of measurements were completed on the day of set up and the next two on the following morning, meaning that the period between the first and last measurements was 22.5 hours. Fish were maintained in their holding tanks overnight with continuous flow of fresh seawater. This sampling design minimised the risk that observers would remember the lengths of individual fish between

[3] Queensland Boating and Fisheries Patrol, measuring board number QA 858, certificate of variation number G5207 for measurement board QA 858 (measuring accuracy of +/- 0.2 millimetres), Chief Inspector of Trade Measurement empowered under the *National Measurement Act 1960.*

successive measurements, allowed comparison of measurements from all observers at each run of measurements, and allowed for the separation of any trends in measurement (*e.g.*, because of handling stress) from random variations in measurements by each observer. The sampling design is summarised in Table 1.

Table 1: The structured repeated measures sampling design for estimation of within-observer and between-observer variations in measurements of common coral trout. Fish were sampled haphazardly from each bin for each measurement.

Run	Observer	Bin order
1	1	1, 2, 3, 4, 5, 6
	2	1, 2, 3, 4, 5, 6
	3	1, 2, 3, 4, 5, 6
	4	1, 2, 3, 4, 5, 6
	5	1, 2, 3, 4, 5, 6
2	1	6, 3, 4, 2, 1, 5
	2	6, 3, 4, 2, 1, 5
	3	6, 3, 4, 2, 1, 5
	4	6, 3, 4, 2, 1, 5
	5	6, 3, 4, 2, 1, 5
3	1	4, 3, 1, 2, 6, 5
	2	4, 3, 1, 2, 6, 5
	3	4, 3, 1, 2, 6, 5
	4	4, 3, 1, 2, 6, 5
	5	4, 3, 1, 2, 6, 5
4	1	5, 1, 3, 4, 6, 2
	2	5, 1, 3, 4, 6, 2
	3	5, 1, 3, 4, 6, 2
	4	5, 1, 3, 4, 6, 2
	5	5, 1, 3, 4, 6, 2
5	1	6, 5, 4, 3, 2, 1
	2	6, 5, 4, 3, 2, 1
	3	6, 5, 4, 3, 2, 1
	4	6, 5, 4, 3, 2, 1
	5	6, 5, 4, 3, 2, 1

The commercial fishers were told to measure fish as they would normally to assess whether a fish met legal size requirements. This generally meant placing the fish on the board with mouth (open or closed, depending on the natural inclination of the fish and the usual practice of the fisher) touching the head-plate and some manoeuvring of the caudal lobes to obtain the maximum reading of length. The two fisheries researchers used a slightly different technique in an effort to reduce variability, with the fish's mouth closed, no lateral compression of the body and the length measured from the tip of the lower jaw to the posterior margin of the dorsal lobe of the tail. The commercial fishers did not witness the researchers' techniques and were not given any guidance as to their measuring technique.

Analytical Methods

The primary focus of this part of the work was the variation in measurement of length, rather than variations in lengths of fish *per se*. Accordingly, we were most interested in the degree to which the measurements of the same fish by different observers varied and the degree to which repeated measurements of the same fish by each observer differed from each other, or from their mean. An underlying assumption of these estimates was that the fish did not systematically change length during the 22.5 hour period over which measurements were taken. To test this assumption, the five measurements of each fish by each observer were tested for linear trends in measured length over the five runs. Any fish that showed a statistically significant ($p<0.05$) trend in length with run for more than one observer was excluded from subsequent analyses.

Two facets of inter-observer variation were examined: relative bias and imprecision.

Relative Bias Among Observers

Differences in relative bias would result in systematic variation among observers and would be expected to manifest as consistent differences among observers in their measurements of each fish, representing over- or under-estimation of length by one or more observers relative to the others. This would result in significantly different mean deviations of an observer's measurements of each fish from the mean length for that fish calculated across all observers. These deviations (*d*) were calculated as:

$$d_{ijk} = l_{ijk} - \bar{l}_{i..}$$

where l_{ijk} is the length measured for fish i (i =1-18) by observer j (j = 1-5) on run k (k = 1-5), and $\bar{l}_{i..}$ is the mean of all measurements by all observers for fish i.

Measured lengths and their deviations from mean length for each fish were analysed by repeated measures analyses of variance to test for consistent differences (biases) in measurements among observers on different runs. The analyses comprised the Between Subjects (Subjects = fish) effects of Bin and fish nested within Bin (fish(Bin)) and Within Subjects effects of Observer and Run. All factors were considered random variables. This meant that tests for the main effects of Observer and Run were possible only if some interaction terms were demonstrably trivial and could be pooled with other effects, therefore effectively being dropped from the analyses. Effects were pooled in the models when $p>0.3$.

Precision of Measurements

Variation among repeated measurements of the same fish by each observer would represent within-observer uncertainty, or imprecision, in length estimation and was measured by the Coefficient of Variation (CV = Standard Error/mean, Andrew & Mapstone 1987) of each observer's five measurements of each fish. Differences in the CV among observers would signal that different observers measured fish with differing consistency. Accordingly, CVs were compared among observers by a repeated measures analysis of variance comprising the Between Subjects (fish) effects of Bin and fish(Bin) and Within Subjects effect of Observer. As before, all factors were considered random variables. We explored the potential that uncertainty in the measurement of fish might have varied with the size of the fish by doing an analysis of covariance comparing among observers regressions of CV on mean fish length.

In the above analyses, F-ratio denominators were determined following the Cornfield-Tukey method (Winer *et al.* 1992) and effects considered statistically significant if $p<0.1$. The liberal significance criterion of $p<0.1$ was chosen to improve the power of the tests because strong inferences would be made from either significant or non-significant results (Andrew & Mapstone 1987; Peterman 1990, Mapstone 1995; Keough & Mapstone 1995).

Between and Within Observers Variation in Measurements

Between and within-observer variances were estimated separately for each fish by Residual Maximum Likelihood (REML) methods for models involving only the effects of Observer and run(Observer). Estimates of Standard Errors (SE) and 95% and 99% Confidence Intervals (CI) were then derived from each variance estimate to indicate the contributions of between- and within-observer variation to uncertainty in measurements of live coral trout. Because separate estimates of measurement CIs were estimated for each fish, we were able to estimate the uncertainty in our estimates of CIs and a measure of a reasonable upper bound (‘worst case’) for CIs on length measurements. We used the 90th percentile of the CIs estimates as this reasonable upper bound.

Empirical Estimates of Differences Between Pairs of Measurements

Finally, empirical distributions of the differences between pairs of measurements of the same fish were calculated and plotted. Differences were calculated as the chronologically first measurement subtracted from the second in each pair so that the sign in the result would indicate whether the fish would be deemed to have ‘shrunk’ (negative values) or ‘grown’ (positive values) between the first and second measurement. Differences were calculated between pairs of observations by the same observer on different runs to estimate ‘within observer’ effects, between pairs of observations by different observers within the same run (‘between observer’ effects), and between all pairs of measurements of the same fish by either the same or different observers over all runs (between + within observer effects). The relative frequency of differences was taken as an estimate of the probability that such a difference would be observed had a fish not actually changed in length. The cumulative relative frequency was taken as an estimate of the probability that a difference of that size or more negative (i.e., fish had apparently ‘shrunk’ since first measured) would have arisen because of measurement error alone when a fish had not actually changed in length. We used the latter to calculate from a binomial distribution the probability of recording *n* or more differences of *d* or greater between two measurements of length when different numbers of fish were measured. We considered values of *d* of -2mm, -5mm and -10mm when samples of 20, 50 and 100 fish were measured, as might be expected during inspections of catch by law enforcement officers. We considered only negative differences because we were interested only in the prospect that a fish might have been deemed of legal size when first measured but less than legal size subsequently, since ‘growth’ of a ‘legal sized’ fish between first and second measurements would not result in a breach.

2. Short-Term Trends in Fish Length

Sampling Design

Systematic changes in lengths of fish were assessed from repeated measurements of the 24 fish held in the mesh baskets suspended in the vessel’s main holding tank. The fish were caught over three consecutive days (10, 9 and 5 fish caught on days 1, 2 and 3 respectively). Measurements commenced on the day of capture and all fish present on each day thereafter were measured at the same time.

Fish were measured only by researchers, as described above. Measurements were taken at seven times: immediately after capture, just prior to transfer to the main holding tank (within 30 minutes after the first measurement), on days 3, 5, 7, and 9 of the trip, and on day 10 of the trip, prior to unloading in port (Table 2). This timetable meant that the fish caught on the first, second and third days had their first two measurements on different days and had two, one and zero days respectively between their second and third measurements (Table 2). Both researchers measured most fish on the first and second occasions, and three fish on the third occasion. Subsequent measurements were alternated between the researchers to minimise potential handling stress associated with measurement. The order in which baskets were processed was randomised on each measurement occasion and the fish from each basket were taken haphazardly by one researcher for measurement by the other researcher. Ambient sea water temperature (at 2m depth) was recorded at time of capture and the water temperature in the vessel’s holding tank was recorded on all other measuring occasions.

Table 2: Sampling design for serial measurements of fish over 10 days to estimate systematic changes in length. '✓' indicates that the nominated *Fish* were measured by the indicated *Observer*(s) on the measurement occasion indicated by *Measurement* N^o. Note that fish 24 was not measured on the 3rd and 7th measurement occasions.

	Day of Trial													
	1		2		3		3	4	5	6	7	8	9	10
Measurement N^o	1	2	1	2	1	2	3		4		5		6	7
Observer	Both		Both		Both		A		B		A		B	A
Fish: 1 – 10	✓	✓					✓		✓		✓		✓	✓
11 – 19			✓	✓			✓		✓		✓		✓	✓
20 – 23					✓	✓	✓		✓		✓		✓	✓
24					✓	✓	-		✓		✓		✓	-

Analytical Methods

Preliminary Analyses

Effects of holding basket on measurements of fish were tested for using Residual Maximum Likelihood methods (REML, Patterson and Thompson 1971, Welham & Cullis 1999) with the random model basket + fish(basket).

Because water temperature varied considerably during the trip, a grouped linear regression was done to test for a relationship between fish length and water temperature. The response variable was total length and the explanatory variables were temperature (random variate) and fish (factor with 24 levels) and their interaction.

Comparison of Initial Measurements by Both Observers

Different researchers measured the fish on successive occasions after day 3 of the trial. Thus, any inference of systematic changes in lengths of fish over all measurements rested on the assumption that observer biases in measurements were negligible and did not influence estimates of change in length. We tested for differences in measurement bias between the two researchers in two ways. First, we analysed the differences between measurements of the same fish by both observers on those occasions (days 1 - 3, Table 2) when both observers measured most of the fish. Second, mean lengths of each fish were compared between observers on the relevant days by REML analyses of a mixed linear model comprised of the fixed effects Day (3 levels), Observer (2 levels), and Measurement Number (Occasions, 2 levels) and the random effect of fish. Tests for significance were by likelihood ratio tests following the procedures of Welham and Thompson (1992).

Comparison of Length Measurements over Time

The sequences of measurements over the 10 days of the trial by the two observers were compared in three ways, based on three different classifications of measurement number. Measurement number was considered as: a) a simple ordered sequence of measurements; b) the calendar 'day of trial'; and c) the number of days since capture.

In the first case, measurement numbers were treated as discrete levels in a repeated measures factor 'time', effectively ignoring the fact that different fish were caught and initially measured on different calendar days. When both observers measured each fish (times 1 and 2), the multiple records of lengths for each fish were averaged within each time prior to this analysis. Data were analysed by repeated measures analysis of variance to test for differences in mean fish length among 'times' 1-7.

Second, we fitted a model comprising the random variable fish and the fixed effects of observer (A, B) and day of trial (1-10) to the data by REML procedures to derive mean lengths of fish classified by an observer on each day of the trial. The estimated means produced by the REML analysis were then used in a grouped linear regression to test for difference between the trends in the length measurements by the two observers over the 10 days of the trial. In the absence of observer-related differences in slopes of the regressions, a common slope was fitted to the data. This treatment of time had the limitation that not all fish were present on days one and two of the trial (Table 2).

Because the day of capture varied among fish, meaning that on any calender day of the trial different fish would have been captive for different periods, we repeated the preceding analysis replacing the factor 'day of trial' with 'days since capture' (designated 0, for day of capture, to 9). This meant that measurements being compared on each day (since capture) were from fish that had been captive for equal periods, even though those measurements would have been made on different dates. This might be considered the most appropriate interpretation of time since fish would be expected to vary in length with the period they had been in captivity rather than with the date of capture or measurement. We also included day of capture (effectively defining groups of fish caught on May 24th, 25th and 26th) as a covariate in the model from which mean lengths for each observer on each day were estimated. Thus, estimates were adjusted for any effects related to the calender day on which the fish were caught. As above, the estimated means produced by the REML analysis were then used in a grouped linear regression to compare trends in the length measurements between observers over time since capture.

Finally, we plotted the series of mean differences between successive measurements and the distribution of all pairs of differences between the first measurement, at time of capture, and each subsequent measurement of the same fish on another day. In the later case, the second measurement was excluded because it was taken on the same day as the first. These plots would indicate the likely distribution of results of measuring fish at the time of capture and once more on some subsequent day, perhaps when the fish were landed.

3. Estimates of Uncertainty from Changes in Length and Measurement Variation.

In field situations, comparisons of two measurements of a fish are most likely to occur between one measurement, by a fisher, at the time of capture and a second, by an enforcement officer or buyer, at the time of landing the fish in port. These measurements may be separated by one to several days, depending on the duration of a fishing trip and when during the trip the fish was caught. Thus, perceived differences between the two measurements will include effects of any real changes in length of the fish and variation arising from within- and between-observer differences in measurement procedure.

To estimate the properties of these combined effects, we combined the time-series of measurements taken over the duration of the trip with the series of short-term repeated measurements taken within a 24 hour period and calculated the distribution of potential differences in length. These calculations involved the following steps:

1. Calculate the difference between each pair of measurements for a fish by the same or different observers over the short interval (<24 hours, as described previously);
2. For each fish observed over the 10 days of the trip calculate the difference between the n^{th} measurement (n = 2, 3, 4, etc.) and the first measurement of that fish;
3. Add each of the differences calculated in (1) to each difference calculated in (2);
4. Use the resultant series of combined differences, incorporating both differences in measurements over <= 24 hours + differences over 1-10 days, to estimate the probability of measurements taken at capture and landing being different by specified amounts (see below).

Note that day-to-day differences between measurements taken over the course of the trip were not included in step 2 above (except for the differences between days 2 and 1). That is, the difference between day 5 and day 4, day 7 and day 6 etc were not included - only the differences between day 5 and day 1, day 4 and day 1, day 7 and day 1 etc were used. This approach was taken to simulate the effects of varying periods between capture and landing and to avoid biasing the results toward differences that would apply only to fish caught (and measured) one day before landing (and re-measurement). All differences were calculated as the chronologically first measurement subtracted from the second in each pair so that the sign in the result would indicate whether the fish would be deemed to have 'shrunk' (negative values) or 'grown' (positive values) between the first and second measurement.

The relative frequency of the combined difference estimates was taken as an estimate of the probability that such a difference would be observed because of the combination of measurement 'error' and potential real changes in length. The cumulative relative frequency was taken as an estimate of the probability that a difference of that size or more negative (*i.e*., fish had apparently 'shrunk' since capture) would have arisen, given our observations on the trip from which these data arose. We used the latter to calculate from a binomial distribution the probability of recording *n* or more differences of *d* or greater between two measurements of length when different numbers of fish were measured. We considered values of *d* of -5mm, -10mm, -15mm and -20mm for samples of from 5 to 200 fish being measured, as might be expected during inspections of catch by law enforcement officers. We considered only negative differences because we were interested only in the prospect that a fish might have been deemed of legal size when first measured but less than legal size subsequently, since 'growth' of a 'legal sized' fish between first and second measurements would not result in a breach.

Results

1. Within- and between-observer variation

Three of the 18 fish showed significant linear decreases in measured lengths over the five runs of measurements for 2 or more observers ($p<0.05$). Accordingly, data from these fish were dropped from subsequent analyses because they did not satisfy the underlying assumption that was central to this part of the work: i.e., that fish did not change length significantly or consistently during the 24 hours over which repeated measurements were taken. Averaged over all observers measurements, these fish apparently decreased in length by 2.4mm (0.52%), 5.5mm (1.25%) and 3.0mm (0.73%) over the 22.5 hours within which the five measurement runs were done. Since two of the fish were from one bin, all data from that bin were excluded from analyses involving tests for effects of Bin, meaning that such analyses were based on data for 14 fish from 5 bins. Analyses not including bin effects were based on 15 fish.

Relative Bias Among Observers

Both total length data and deviations data satisfied the assumption of sphericity of variance-covariance matrices (Mauchly's criterion, $p>0.4$ in all cases), and so univariate tests of effects are reported hereafter (Winer *et al*. 1992).

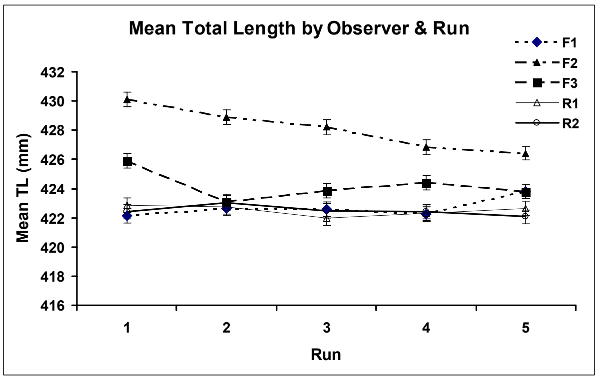
Significant interactions occurred between Observers and Run and Observers and fish for the length data, indicating the relative biases of fishers changed with measurement run (Observer*Run interaction: $F_{16,64}=3.30$, $p=0.0003$) and the fish being measured (Observer*fish: $F_{36,144}=1.95$, $p=0.003$). The same results were found for analyses of the deviation data. There were no significant effects of the bin in which fish were held, either overall or in interaction with Run or Observer ($P>0.14$ for all terms).

Separate analyses of runs for each observer indicated that the relative biases of two observers changed significantly among runs during the trial ($F_{4,16}=8.56$, $p=0.0007$, $F_{4,16}=2.64$, $p=0.073$; Fig. 1), whilst those of the other observers remained relatively stable over the five measurement runs ($F_{4,16}=1.64$, $p=0.213$; $F_{4,16}=0.80$, $p=0.545$; $F_{4,16}=0.82$, $p=0.532$) (Fig. 1). Comparisons among observers at each run showed that the two researchers measured fish with consistently similar biases ($p>0.3$) at all times and were similar to one of the fishers in 4 of the 5 runs ($p>0.3$). The other two fishers demonstrated relative biases different from their colleague and both researchers during some or all runs ($p<0.1$, Fig. 1). One of these fishers showed consistent reduction in his relative bias with time, suggesting some cumulative change in measurement methodology.

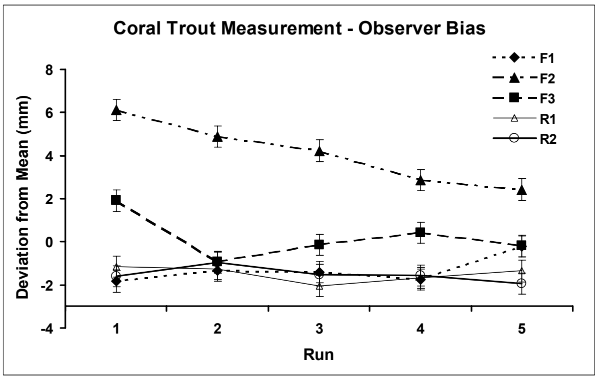
Despite the significant Observer*Fish interaction in analyses, deviations in measurements were relatively consistent among fish for each observer ($F_{13,56}=1.00$, 1.58, 0.99, 1.10, 1.10; $p=0.464$, 0.120, 0.473, 0.378, 0.377 respectively for the five observers). As expected from the above results and Figure 1, however, strong effects of observer on measurement deviations were present for most fish ($p<<0.01$ for 12 of 14 fish) and for only one of the 14 fish were differences among observers relatively slight ($F_{4,24}=2.35$, $p=0.09$).

Figure 1: Plots of average total length (a) and average deviations of measurements from the overall mean measurement for each fish (b) for each observer at each run of measurements. Error bars are Standard Errors. Overlapping lines indicate that the corresponding observers shared similar measurement biases.

a)



b)



Precision of Measurements

Coefficients of Variation (CVs) calculated from the five measurements of each fish by each observer (ie., reflecting only within-observer variation) varied significantly among observers ($F_{4,16}$=2.34, p=0.095). Researchers showed similar precision in measurement (average CV=0.00141) but fishers showed variable precision of measurement, with two fishers similar to each other (CV=0.00175, 0.00196) but significantly different (P<0.05) from the third (CV=0.00235). Precision of measurements by all fishers was significantly poorer (P<0.05) than that for researchers.

Slopes of regressions of CV on fish length were homogeneous among observers ($F_{4,60}$=0.79, p=0.537) and the overall regression was significant ($F_{1,68}$=5.33, p=0.024) with a slope of $-1.09*10^{-5}$, indicating that larger fish were measured with greater precision The absolute magnitudes of Standard Errors (SE) were relative constant across the range of fish lengths in our trials ($F_{1,68}$=2.07, p=0.154), meaning that the declining CVs were the result of changes in the numerator (mean length) rather than the denominator (SE) of the CVs.

Mean CVs adjusted for the common slope ranged between 0.00176 and 0.00236 for fishers, whilst those for the researchers were 0.00146 and 0.00131. The measurement of a 38cm coral trout by a fisher would have expected average 95% confidence intervals (CIs) of ±2.39mm (calculated from the average SE), although uncertainty could be as high as ±2.56mm (95%CI derived from largest SE). The same fish measured by a researcher would be expected to have 95% confidence intervals of ±1.29mm.

Between and Within Observers Variation in Measurements

REML estimates of between all observers variance ranged from 1.70 to 9.15 over the 15 fish included in analyses with a mean of 5.63, whilst within observer variances ranged between 1.30 and 6.87 with mean 3.76. On average, 95% and 99% CIs on measurements of the same fish by different observers due solely to observer-related effects were 2.91mm and 4.83mm respectively. Measurements of the same fish by an individual observer would be expected to vary with observer because of the errors in measurement specific to each observer. Over all observers, however, average 95% and 99% CI of measurements by a single observer were 1.72mm and 2.35mm respectively. The 90^{th} percentiles (reasonable upper limits) of estimated 95% CIs were 3.75mm and 2.15mm respectively for between and within observer errors, whilst the 90^{th} percentiles of estimated 99% CIs were 6.22mm and 2.93mm respectively.

When only similarly biased observers (the two researchers and one fisher) were considered, between observer variances ranged from 0 to 1.70 with mean 0.27 and within observer variances ranged from 0.80 to 4.90 with a mean of 2.43. For these observers, our likely upper limit (the 90th percentile of our estimates) to 95% confidence intervals on measurements attributable to between- and within-observer variation were 2.08mm and 2.16mm respectively, whilst the corresponding figures for 99% CI were 4.80mm and 3.02mm respectively.

Empirical Estimates of Differences between Pairs of Measurements

Empirical distributions of differences in pairs of measurements are shown in Fig. 2. From these data it is clear that most measurements were within about 5mm of each other, but there were relatively long tails of large differences between observers, within observers and overall. Negative differences were slightly more frequent than positive differences, with 90% of overall differences between –7.5mm and +5.5mm. This skew was likely to be mainly the result of differences between fisher 2 and other observers since fisher 2 measured fish before 3 of the other observers and routinely estimated fish to be longer than the other observers (Fig. 1). On 1% or less of occasions, two such measurements would differ by more than -11mm.

The probabilities that a second measurement would be at least 2mm, 5mm or 10mm less than an earlier measurement were 0.459, 0.197 and 0.019 respectively when all observers were included in calculations. When only data from the observers with relatively homogeneous biases (2 researchers and 1 fisher) were used, the probabilities were reduced to 0.096, 0.012 and <0.0001.

The likelihood of observing various numbers of such differences in samples of 50 and 100 measured fish are shown in Figure 3. For all data combined, there was a very high probability of observing differences in measurement of only 2mm because of measurement uncertainty alone in approximately 50% of fish measured twice. There was approximately a 50% chance of returning differences in measurement of 5mm or more in around 25% of fish measured. Only for differences in measurement of 10mm or over was there a relatively low probability of more than a few individuals (~5% of the sample) varying in estimated length because of measurement uncertainty. This would mean, for example, that if 50 fish that had been measured at 38cm on capture were re-measured when landed in port, on 6.8% of occasions at least 3 fish would be measured as 37cm or smaller, on 10% of occasions at least 14 fish would be measured as 37.5cm or smaller, and on 9.8% of occasions 28 or more fish would be measured as 37.8cm or smaller simply because of uncertainties in measurement practices. When the data from more consistent observers were considered, there would be a near zero probability that any fish measured as 38cm on capture would be re-measured as 37cm or less, approximately 2% chance of measuring 3 or more out of 50 such fish as 37.5cm or less, and only a 10.5% chance of measuring 8 or more fish out of 50 as 37.8cm or smaller.

Figure 2: Relative frequency distributions of differences between pairs of measurements of the same live coral trout. Differences were calculated as the chronologically first measurement in each pair taken from the second, meaning that negative values represent 'shrinkage' and positive values represent 'growth'. '**Overall**' refers to differences between all pairs of observations, including observations made by the same and different observers during any run. '**Between**' indicates differences between pairs of observations made at the same time (same run) by any two different observers. '**Within**' refers to differences between pairs of observations made by the same observer during different runs. ***Left*** - data for all observers; ***Right*** - data for three observers (2 researchers and 1 fisher) with similar biases.

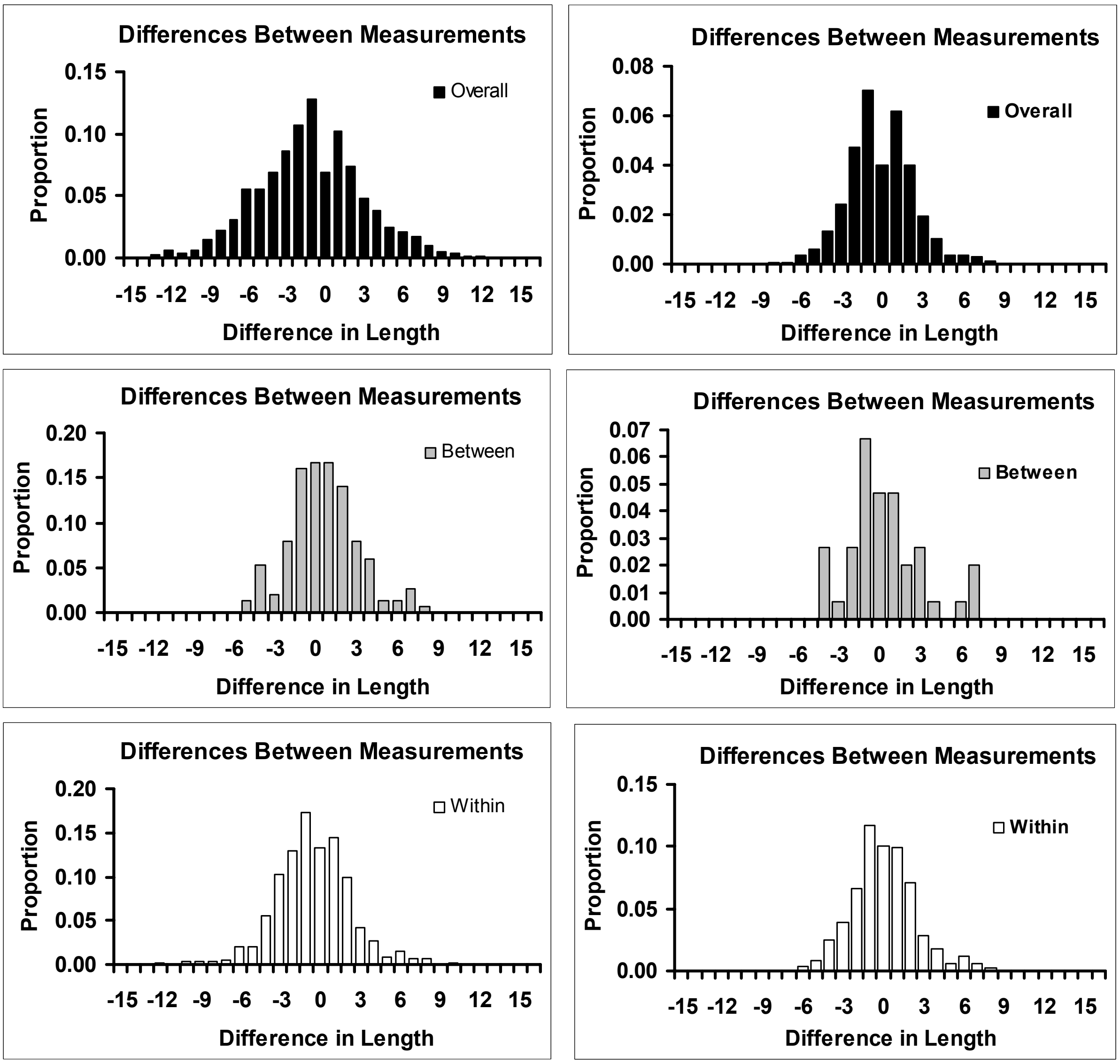
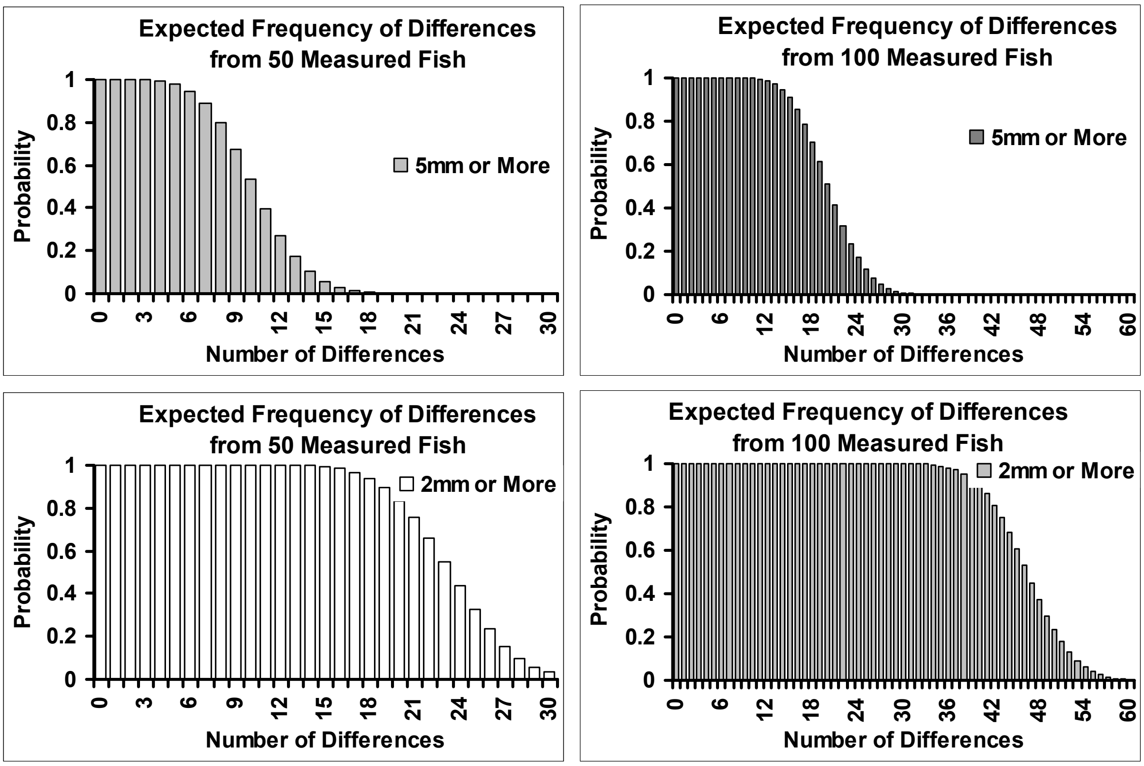
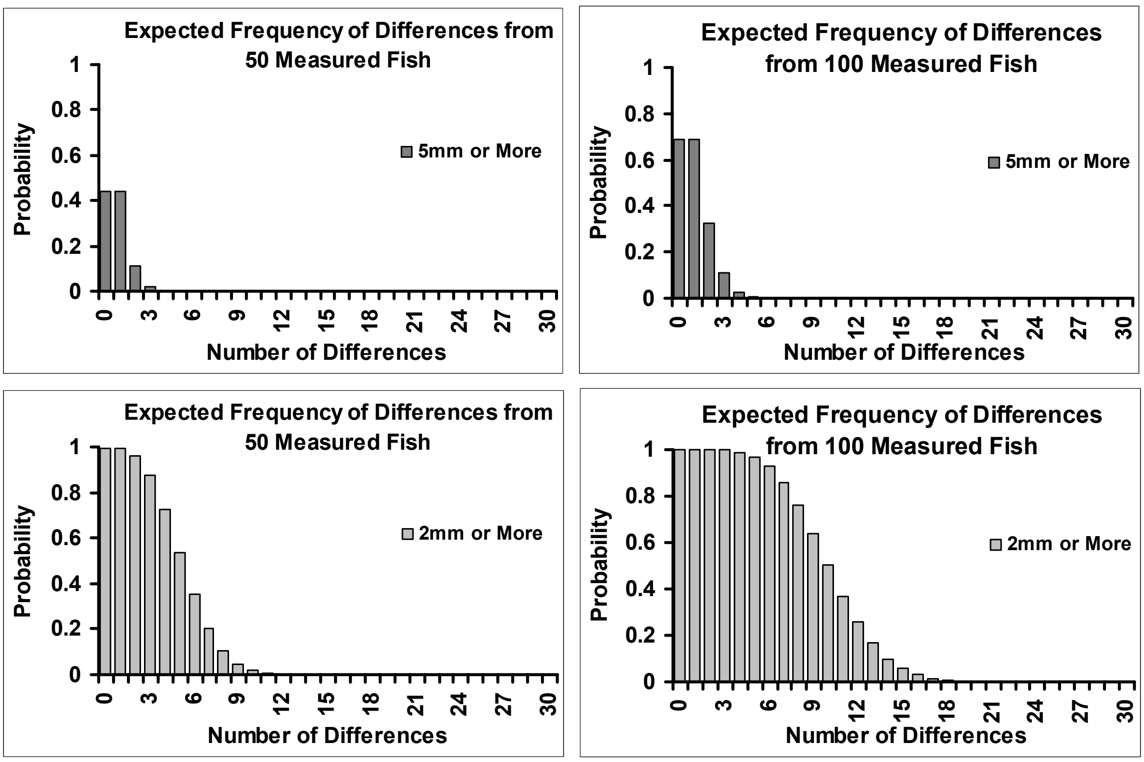


Figure 3: Probability of observing at least the given '**Number of Differences**' between 2 measurements of the same fish where the second measurement was less than the first by '**5mm or More**' or '**2mm or More**' when either 50 (left) or 100 (right) fish were measured. ;

A **- data from all observers**.



B **- data from three observers (2 researchers and 1 fisher) with similar relative biases.**



2. Short-Term Trends in Fish Length

Preliminary Analyses

Being held in different baskets accounted for no variation in fish length (p=0.92). Accordingly, effects of baskets were not included in subsequent analyses.

Temperature during the trial varied from 18.9°C to 25.5°C. Slopes of the regression of fish length on water temperature did not differ significantly among the 24 fish (p=0.999) and the combined slope was estimated to be 0.0301 (SE=0.0648) which did not differ significantly from zero (p=0.643). Thus, there was no evidence that changes in water temperature within the above range significantly affected fish length during the trial.

Comparison of Initial Measurements Between Observers

There were no significant interactions among Measurement Number, Day or Observer on those occasions when fish were measured by both researchers (Day*Observer*Occasion, p = 0.1; Day*Observer, p=0.6285; Day*Occasion, p=0.6597; Observer*Occasion, p=0.7101). Accordingly, these terms were dropped and a main effects model fitted to the data, resulting in a significant effect of Measurement Number (p=0.0326) but non-significant effects of Observer and Day. Averaged over all fish and both observers, the mean length of fish approximately 30 minutes after capture (measurement number 2) was 0.6 mm less than at the time of capture. Based on the absence of any significant, or even non-trivial, differences in measurements between observers, we considered observers to be consistent in their measurements during the first 3 days of the trial.

Comparison of Fish Length Measurements over Time

Time Considered as a Discrete Ordered Sequence

Repeated measures analysis of variance comparing mean length amongst seven measurement numbers (‘times’, irrespective of whether each occasion occurred on the same day for all fish) indicated significant differences in average lengths among times (p<0.001). Degrees of freedom were adjusted by the Greenhouse-Geisser epsilon (0.5802) to adjust for non-sphericity in the error variance-covariance matrix before testing for the effect of time (Payne 1996). The least square means estimated from this analysis are shown in Table 3. The difference between the averages of the 1st and 7th measurements was 1.8mm, with a maximum difference between any two measurements of 3.2mm.

Table 3: Mean fish lengths estimated following repeated measures analysis of variance comparing mean length over seven measurement numbers. Common super-scripts indicate means that did not differ significantly (p<0.1) in pair-wise comparisons.

Measurement N°	Mean length (mm)
1	405.4^{a}
2	404.5^{b}
3	403.4^{ce}
4	403.2^{ce}
5	402.9^{c}
6	402.2^{d}
7	403.6^{e}

Time Considered as “Day of the trial”, 1-10

The slopes of the linear regressions fitted to measurements of fish over the 10 days of the trial did not differ between observers, irrespective of whether time was coded as calender day of trial (p=0.16) or days since capture (p=0.15). Accordingly, we inferred that trends in measurement of the fish were not influenced by the person taking the measurements and common slopes were fitted to the data from both observers. The regression of fish length on calender day of trial fitted to all data was significant (p=0.004, adjusted R^2 = 0.627, Figure 4) with the relationship

Length of the fish = 405 mm – 0.27 * *Day of trial,*

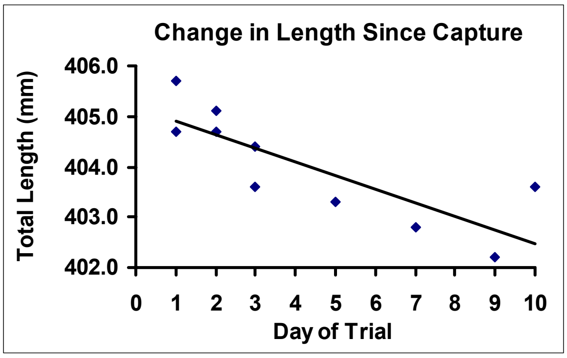
indicating that fish decreased in length on average by 0.27 mm per day, or 2.43 mm between days 1 and 10.

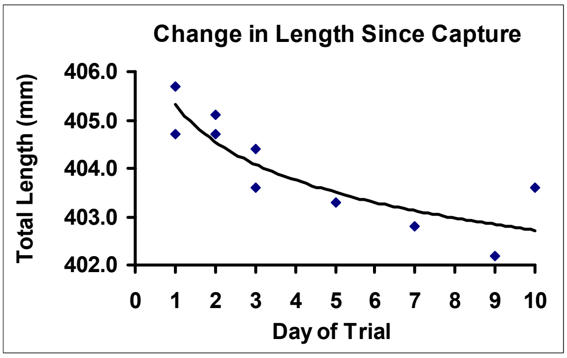
We also tested the fit of an exponential relationship for these data, which also showed no significant observer effect (p=0.148) and that the single exponential curve explained a significant amount of variation in fish length (p=0.004, adjusted R^2 = 0.729, Figure 4). The fitted equation for both observers was:

Length of the fish = 403 mm + 3.74 * $0.724^{\textit{Day of trial}}$

The average change in length over the 10 day period predicted by this equation was 2.6 mm.

Figure 4: Linear (left) and exponential (right) regressions of fish length (measured by both observers) on day of trial.





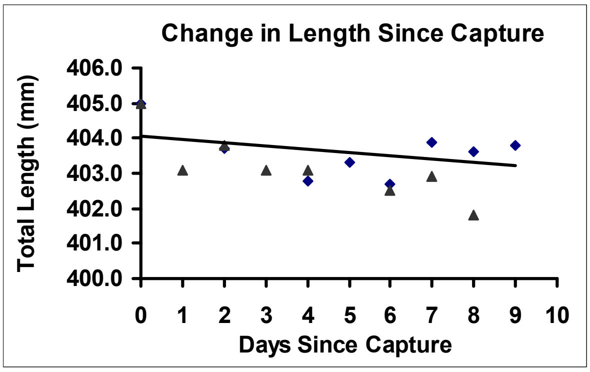
Time Considered as "Days Since Capture"

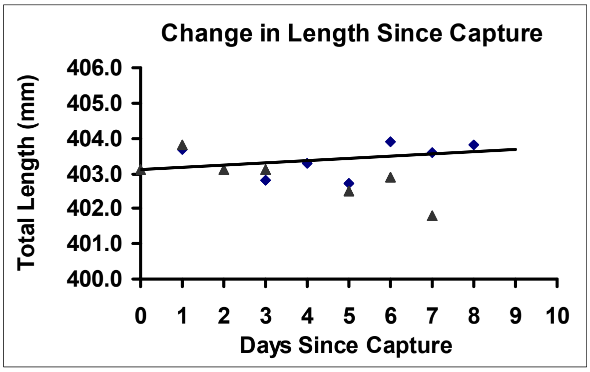
The slope of the regression of mean fish length on days since capture also was significant (p=0.034, adjusted R^2 = 0.233, Figure 5). The fitted equation was:

Length of the fish = 404 mm– 0.15 * *days since capture* (Figure 5)

The points from the days of capture (zero days since capture) had relatively high leverage in this fit. If these data were omitted, the linear regression was not significant (p=0.589, *Length of the fish* = 403.3 + 0.04 * *days since capture,* Figure 5).

Figure 5: Linear regressions of fish length (measured by both observers) against days since capture, after adjusting for date of capture, including (left) and omitting (right) measurements taken on the day of capture.

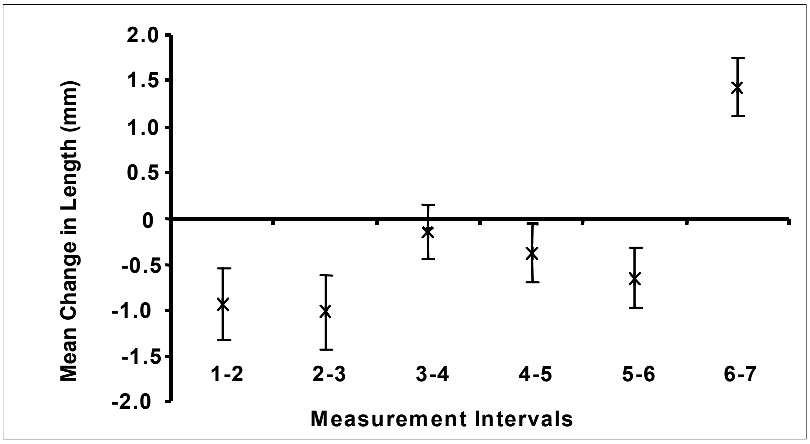




Differences between Successive Measurements

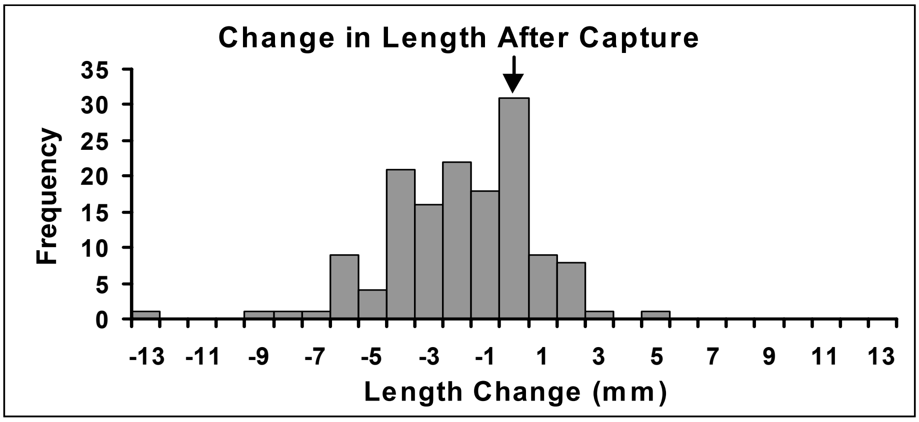
Coral trout on average apparently changed in length between successive measurements, with the mean change generally being below zero, though often not significantly so (Figure 6). Fish appeared to 'shrink' most between the first and second and second and third measurements. The conspicuous exception was between the final two measurements, when fish on average significantly increased in length by an average of 1.4 mm (± 0.6 mm 95% CI). Nevertheless, fish on average were measured to be significantly shorter by 1.76 (± 0.76 mm 95% CI) at the end of the trial than at the beginning (t-test, df = 22, p = <0.001). Thus, the late increase in length had not offset the accumulation of successive small decreases in length that on average totalled 3.20 (± 1.02 mm 95% CI) between capture and the penultimate measurement.

Figure 6: Mean change in size between successive measurements (± SE). Mean changes in length were significantly different from zero for intervals 2-3 and 6-7 only.



The frequency distribution of differences in measured length between the day and time of capture and each subsequent day indicated that whilst the modal difference was within 0.5mm of zero, negative (shrinking) changes accounted for 65.3% of values and positive values accounted for only 13.2% of values. The maximum reduction in length was -13 mm, with four values exceeding -6 mm, whilst the maximum 'growth' in length was +5 mm and only two values exceeded +2 mm (Figure 7).

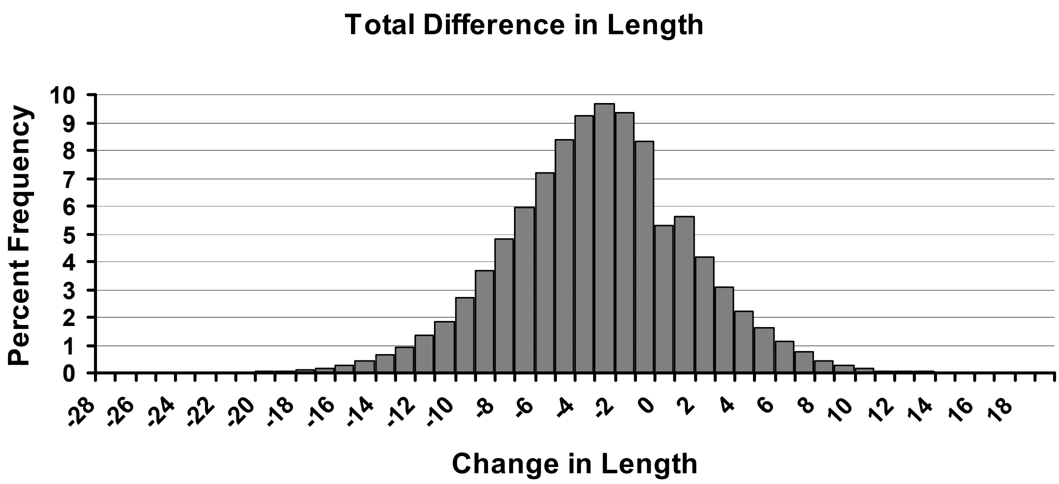
Figure 7: Frequency distribution of changes in length for each fish from the day and time of capture to subsequent days of measurement. Negative values indicate the fish apparently 'shrank', whilst positive values indicated 'growth' in total length. The arrow indicates no change in measured length between days.

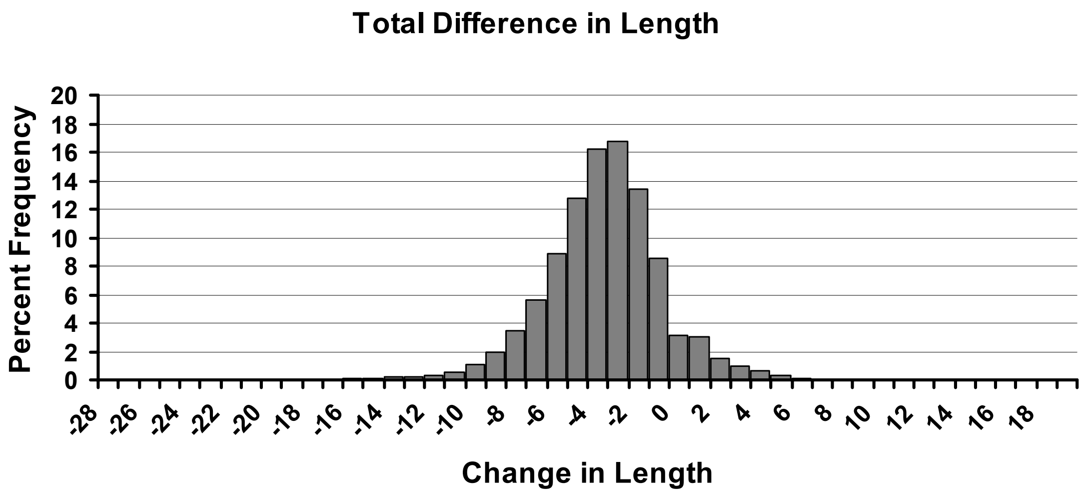


3. Estimates of Uncertainty from Changes in Length and Measurement Variation.

The distribution of estimated differences arising from the combined effects of measurement error and change in length during a fishing trip is shown in Figure 8. Results from including all observations and those from only relatively consistent observers are shown. It is clear that most measurements were within about 10mm of each other, but negative differences were more frequent than positive differences. When consistency of measurement method was not considered, 90% of overall differences lay between –28mm and +2.5mm, with a median value of approximately -3mm and less than 0.1% of differences of -20mm or more negative. When only measurements from the three consistent observers (2 researchers and 1 fisher) were considered, however, 90% of overall differences lay between –21 mm and -1mm, with a median value of approximately -4mm and less than 0.1% of differences of -17 mm or more negative.

Figure 8: Relative frequency distributions of differences between pairs of measurements of the same live coral trout combined with observed differences in length during a fishing trip. Differences were calculated as the chronologically first measurement in each pair taken from the second, meaning that negative values represent ‘shrinkage’ and positive values represent ‘growth’. The upper distribution is for data from all observers, irrespective of differences in measurement method or consistency. The lower distribution is for data from only the three observers who used similar methods and produced relatively consistent measurements.





Figures 9 and 10 show the cumulative frequency distribution of these estimated differences, with emphasis on the most negative values. The plots indicate that there would be less than a 5% chance of a repeated measurement of a fish being 12mm or more less than an earlier measurement because of measurement error and 'shrinkage' alone, even if different people used different measurement practices (Fig. 9). There would be approximately a 1% chance of such measurements being 15mm or more shorter than a previous measurement.

Figure 10 suggests that if a consistent measurement practice was used, there would be less than a 5% chance of a repeated measurement of a fish being 9mm or more less than an earlier measurement because of measurement error and 'shrinkage' alone and a 1% chance of a measurement being 12mm or more shorter than a previous measurement of the same fish. Collectively, Figures 8, 9 and 10 illustrate the potential for improvement in length estimation, and reduction in measurement uncertainty, if standardisation of measurement practices could be adopted by all fishers and enforcement officers.

Finally, in Tables 3 and 4 we tabulate the numbers of fish measured as undersize by 5mm, 10mm, 15mm or 20mm that would trigger a breach given four rates of desired confidence that a breach was warranted. The numbers in Table 3 are based on the measurement uncertainty we observed across all observers (2 researchers & three fishers) and so might be seen as a 'worst case' scenario – i.e., most generous to the fishers. Those in Table 4 reflect what might be expected if a consistent method of measuring fish was adopted across the fishery and applied with relatively uniform precision by all people.

Figure 9: Cumulative relative frequency distributions of differences between pairs of measurements of the same live coral trout combined with observed differences in length during a fishing trip. *Differences calculated for all observers (2 researchers and 3 fisher) irrespective of similarity in measurement variations.* Differences were calculated as the chronologically first measurement in each pair taken from the second, meaning that negative values represent 'shrinkage' and positive values represent 'growth'. The top panel (a) shows the entire distribution, whilst the lower two panels show 'magnified' views of the left parts of the distribution representing approximately the lowest (and most negative) 50% (b) and 10% (c) of observations respectively.

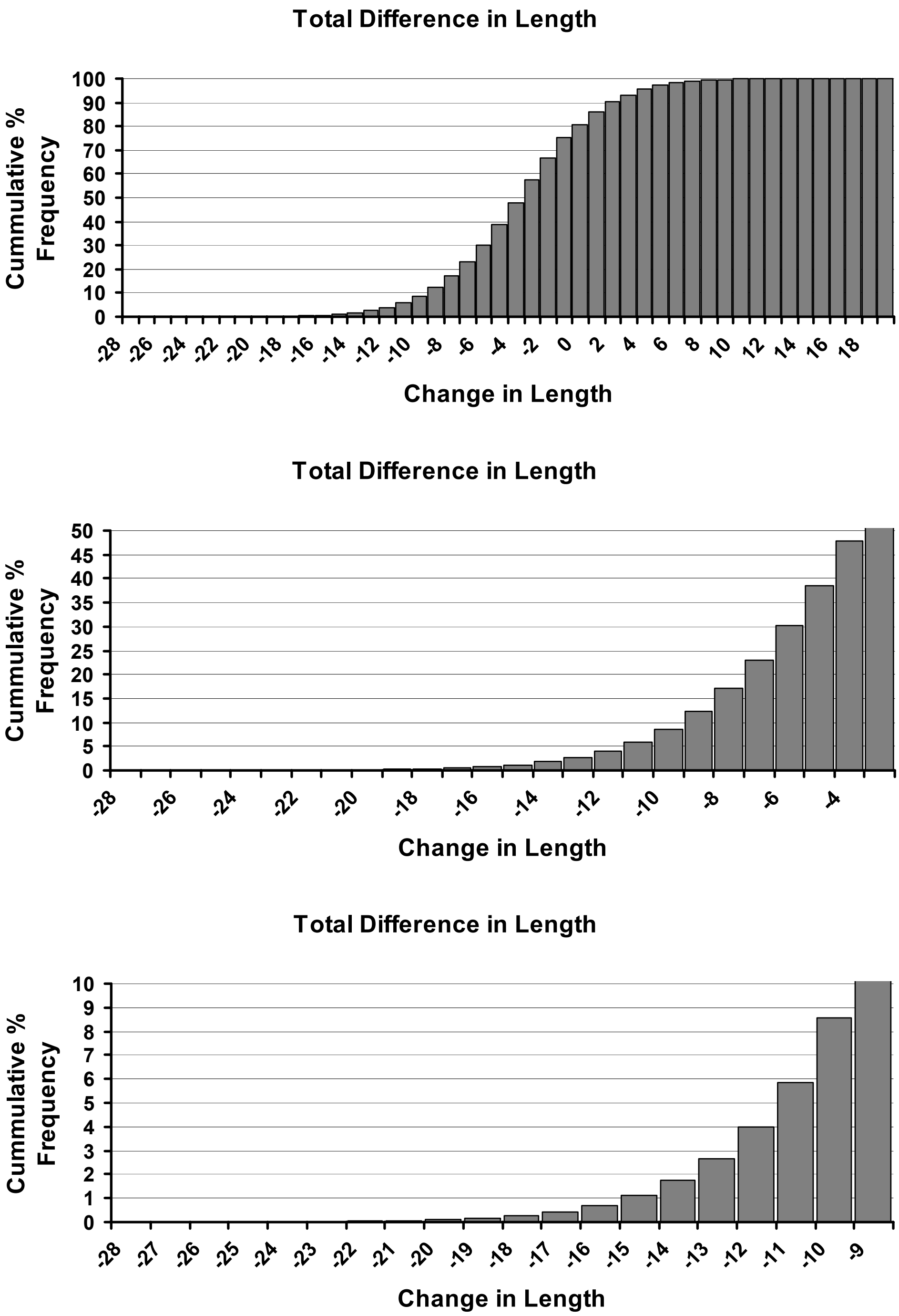


Figure 10: Cumulative relative frequency distributions of differences between pairs of measurements of the same live coral trout combined with observed differences in length during a fishing trip. *Differences calculated for observers (2 researchers and 1 fisher) with similar measurement variations only*. Differences were calculated as the chronologically first measurement in each pair taken from the second, meaning that negative values represent 'shrinkage' and positive values represent 'growth'. The top panel (a) shows the entire distribution, whilst the lower two panels show 'magnified' views of the left parts of the distribution representing approximately the lowest (and most negative) 50% (b) and 10% (c) of observations respectively.

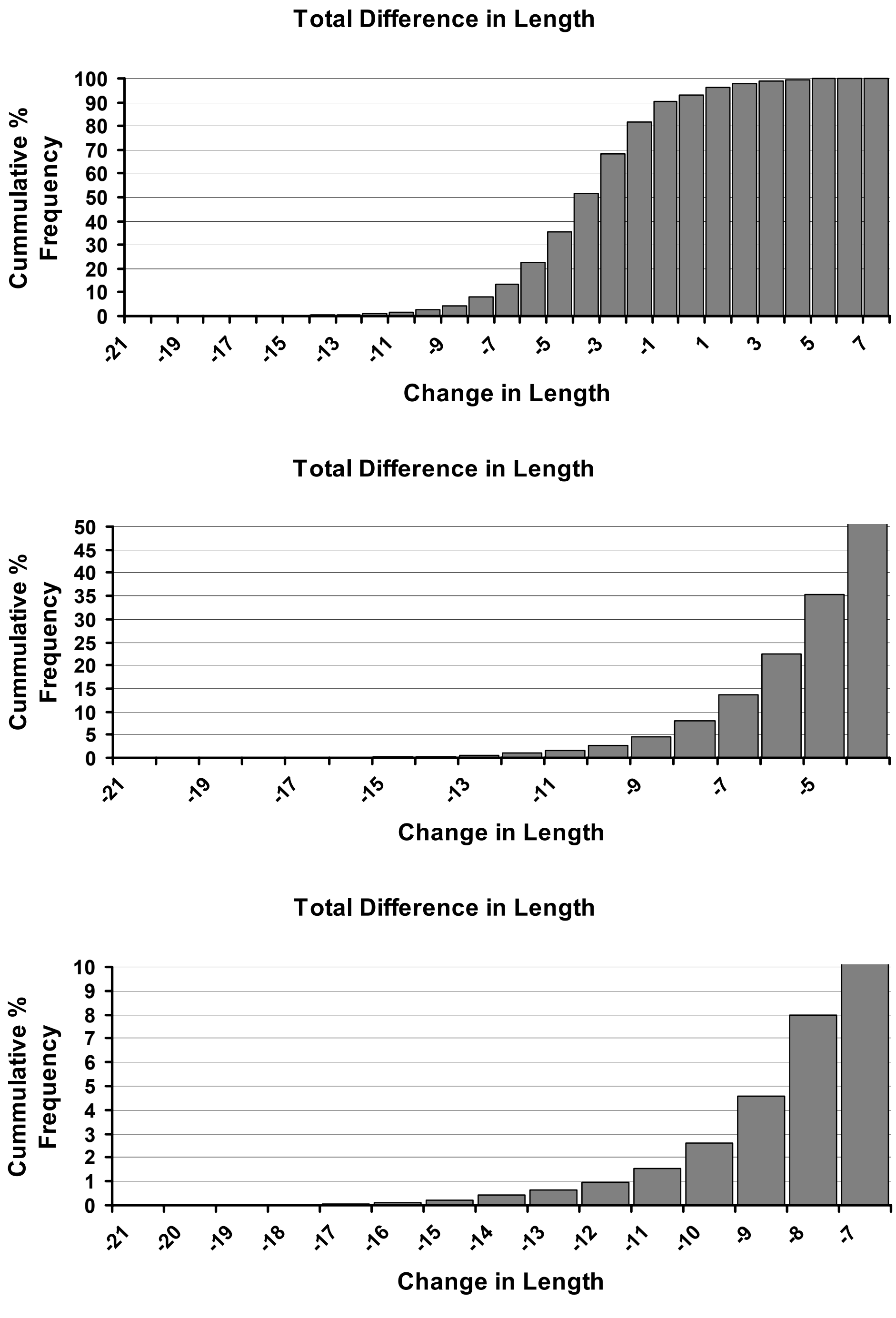


Table 3: if the **Measured** number of fish were suspected of being close to 38cm long, data in each other column show how many of those fish would have to be measured as less than or equal to the length indicated at the top of that column (**≤ 37.5, ≤ 37, ≤ 36.5, ≤ 36.0** in cm) in order to trigger a breach if the desired confidence in the breach was no less than indicated at the top of each set of columns (**Confidence in Breach = xx%**) had all fish been exactly 38cm at the time of capture. For example (shaded row), if 25 suspect fish were measured and it was important to be 99% confident that *at least one of the fish* was indeed below the minimum legal size, a breach would be warranted if *either* 16 (or more) fish were ≤ 37.5cm long *or* 7 (or more) fish were ≤ 37cm long *or* 3 (or more) fish were ≤ 36.5cm long *or* 2 (or more) fish were ≤ 36.0cm long. Tabulated data allow for inconsistencies in measurement method as well as variation among and within observers and changes in length.

Confidence in Breach = 99.9%				
Measured	**≤37.5**	**≤37.0**	**≤36.5**	**≤36.0**
5	5	4	3	2
10	10	5	3	2
15	13	6	3	2
20	16	7	4	2
25	18	8	4	2
30	21	9	4	2
40	26	11	5	2
50	31	12	5	2
60	36	14	5	3
70	41	15	6	3
80	45	17	6	3
90	50	18	6	3
100	55	19	6	3
120	64	22	7	3
140	73	24	8	3
160	82	27	8	3
180	91	29	9	3
200	100	31	9	3

Confidence in Breach = 99%				
Measured	**≤37.5**	**≤37.0**	**≤36.5**	**≤36.0**
5	5	3	2	0
10	8	4	2	0
15	11	5	3	2
20	14	6	3	2
25	16	7	3	2
30	19	8	3	2
40	24	9	4	2
50	28	10	4	2
60	33	12	4	2
70	38	13	4	2
80	42	14	5	2
90	47	15	5	2
100	51	17	5	2
120	60	19	6	2
140	68	21	6	2
160	77	23	7	2
180	86	26	7	3
200	94	28	7	3

Confidence in Breach = 95%				
Measured	**≤37.5**	**≤37.0**	**≤36.5**	**≤36.0**
5	5	3	2	0
10	7	3	2	0
15	10	4	2	0
20	12	5	2	0
25	15	6	2	0
30	17	6	2	0
40	22	8	3	0
50	26	9	3	0
60	30	10	3	0
70	35	11	3	2
80	39	12	4	2
90	43	13	4	2
100	48	14	4	2
120	56	17	4	2
140	64	19	5	2
160	73	21	5	2
180	81	23	6	2
200	89	25	6	2

Confidence in Breach = 90%				
Measured	**≤37.5**	**≤37.0**	**≤36.5**	**≤36.0**
5	4	2	0	0
10	7	3	2	0
15	9	4	2	0
20	12	4	2	0
25	14	5	2	0
30	16	6	2	0
40	20	7	2	0
50	25	8	3	0
60	29	9	3	0
70	33	10	3	0
80	37	11	3	0
90	42	12	3	0
100	46	13	4	0
120	54	15	4	0
140	62	17	4	2
160	71	19	5	2
180	79	21	5	2
200	87	23	5	2

Table 4: Data as for Table 3 (above), but calculated for observers with a relatively consistent measurement method and consistency – such as might be expected if a consistent measurement method was used for all measurements of fish, both at the time of capture and during inspection by enforcement officers or buyers.

Confidence in Breach = 99.9%				
Measured	**≤37.5**	**≤37.0**	**≤36.5**	**≤36.0**
5	5	3	2	0
10	9	4	2	0
15	12	4	2	0
20	15	5	2	0
25	17	5	3	0
30	20	5	3	0
40	25	6	3	2
50	29	7	3	2
60	34	7	3	2
70	38	8	3	2
80	43	9	3	2
90	47	9	4	2
100	51	10	4	2
120	60	11	4	2
140	68	12	4	2
160	76	13	4	2
180	85	13	4	2
200	93	14	5	2

Confidence in Breach = 99%				
Measured	**≤37.5**	**≤37.0**	**≤36.5**	**≤36.0**
5	5	2	2	0
10	8	3	2	0
15	11	3	2	0
20	13	4	2	0
25	16	4	2	0
30	18	4	2	0
40	22	5	2	0
50	27	5	2	0
60	31	6	2	0
70	35	6	3	0
80	39	7	3	0
90	43	7	3	0
100	48	8	3	0
120	56	9	3	0
140	64	10	3	0
160	72	10	3	0
180	80	11	3	0
200	87	12	3	0

Confidence in Breach = 95%				
Measured	**≤37.5**	**≤37.0**	**≤36.5**	**≤36.0**
5	5	2	0	0
10	7	2	0	0
15	9	3	0	0
20	12	3	0	0
25	14	3	2	0
30	16	3	2	0
40	20	4	2	0
50	24	4	2	0
60	28	5	2	0
70	32	5	2	0
80	36	6	2	0
90	40	6	2	0
100	44	6	2	0
120	52	7	2	0
140	60	8	2	0
160	67	9	2	0
180	75	9	3	0
200	83	10	3	0

Confidence in Breach = 90%				
Measured	**≤37.5**	**≤37.0**	**≤36.5**	**≤36.0**
5	4	2	0	0
10	6	2	0	0
15	9	2	0	0
20	11	2	0	0
25	13	3	0	0
30	15	3	0	0
40	19	3	0	0
50	23	4	2	0
60	27	4	2	0
70	31	5	2	0
80	35	5	2	0
90	39	5	2	0
100	42	6	2	0
120	50	6	2	0
140	58	7	2	0
160	65	8	2	0
180	73	9	2	0
200	80	9	2	0

Discussion

We have demonstrated significant variations in measurement of live coral trout both among observers and within individual observers' measurement procedures. We have also shown that coral trout kept alive at sea can change significantly in apparent length over several days, though such changes were smaller than the apparently random errors in repeated measurements by the same or different observers over short periods. Of the total variation in measurements, variance among all observers was 1.50 times larger than the average variance among repeated measurement by the same observer. Our analyses provided initial insights into the amount of reduction in between-observer variation that might be expected with appropriate training of observers, however, with between-observer variance for a group of observers with similar measurement biases being only 11% that of the within-observer variance. This result suggests that rudimentary training in a standardised measurement method may reduce significantly variations among repeated measurements of live coral trout by different observers.

Observer Bias and Measurement Error

There were significant and consistent differences in relative bias among observers. The most likely explanation of such biases is that different observers use slightly different measurement techniques. Such bias also may be made more consistent among observers by the establishment of standard recommended procedures for measuring fish and rudimentary training in those methods. The observation that the two researchers, who had agreed on a standard measurement procedure, had consistently very similar biases may indicate that standardised measurement methods could indeed remove observer-dependent bias as a source of measurement variation. It is noteworthy also that the relative biases of the researchers were frequently similar to at least one of the fishers, suggesting that relatively little training may be needed to produce consistent measurement results among people. There was also evidence of change in relative bias in at least one observer, meaning that they tended to become more similar to other observers over time. This trend might have been the result of casual discussions among the fishers involved in the study that resulted in one fisher changing their measuring practice toward those of other fishers, possibly further indicating that relatively little training would be required to produce relatively uniform measurements among different observers.

Short-term Changes in Fish Length

We have demonstrated also that measured lengths of fish changed significantly with time following capture, in general tending to decrease for several days following capture. Despite this 'shrinking' effect, average length increased significantly on the last day of the trial by 1.43 ± 0.31 mm ($\pm$SE).

The estimated linear rate of reduction in length ranged between 0.15 and 0.27 mm per day, depending on how time was considered (day of trial or days since capture). It is notable that considering changes in length relative to days since capture clearly indicated a disproportionate effect in the first 24 hours after capture, and even within 30 minutes of capture. Removing the measurements taken on the day of capture (which were always the greatest lengths recorded) eliminated the evidence of a downward trend in length with time since capture. This result is suggestive of an acute effect of capture and / or handling that is not repeated during holding and subsequent handling. It is unclear whether stress of capture might have caused the first measure to be inflated or the stress of being held on the vessel and handled caused all subsequent measurements to be under-estimates of the fishes' true lengths. In either case, however, the net effect would be that decisions about the size of fish made at the time of capture would be slightly liberal with regard to the legal size limit compared to decisions made subsequently.

Implications of Measurement Uncertainty and Change in Length for Enforcement of Minimum Legal Sizes

The presence of different biases among observers is likely to exacerbate the difficulties in compliance and enforcement of minimum legal size regulations, especially given that there is no practical way to allow for individual biases when fish are being measured by many different people. The impact of differential bias was demonstrated convincingly by removing the measurements of the two most 'aberrant' fishers from the calculations of paired differences. Doing so meant that the range of differences in pairs of measurements was reduced from ±15mm to ±8mm, and the likelihood of 3 or more fish being recorded as having 'shrunk' by 5mm or more when 50 fish were measured was reduced from 99.85% to 2.0%.

Clearly, any standardisation of measurement procedures and appropriate training that increases bias similarity among fishers and enforcement officers will improve compliance and enforcement certainty considerably. Reduction in measurement variation similar to what we saw when eliminating the 'aberrant' observers would mean that a 2% allowance for error in compliance monitoring or enforcement would involve a limit on the number of fish under-size by 5mm or more being reduced from 31 in 50 to 3 in 50. If rudimentary training also reduced the within-observer variation, as might be expected, then the margin of error attached to compliance with minimum legal size limits would be even further reduced.

The apparently stochastic sources of error in measurements of length of coral trout were considerably greater than almost all of the variations in length we recorded during the 10 day period for which fish were held on the fishing vessel. With few exceptions (e.g., the 4 fish that apparently 'shrank' by 7-13mm), we would not expect to be able to distinguish real changes in length from measurement error in routine measurements by untrained or differently biased, observers. Although our data indicate that some fish apparently do change in length by relatively large amounts (>6mm), in general, changes in length are likely to be small beyond 1-2 days after capture. Discrepancies in measurements of lengths between different observers, however, are likely to be comparatively large and frequent.

Given that those changes in length that did occur tended to happen shortly after capture, it may be necessary for fishers to measure fish close to the minimum legal size 1-2 days after they have been caught to verify that initial impressions of their size were not over-estimates. As discussed above, standardisation of measurement procedures and basic, readily accessible instructional material about those procedures are likely to be of greatest benefit in reducing uncertainty in compliance to and enforcement of minimum legal sizes.

Even with relevant training and standardisation of measurement procedures, however, uncertainties in measured length are likely to persist. Further, existing fish measuring techniques which eliminate handling have not been successful in appreciably reducing measurement errors. For example, video imaging indicates that repeatability of computer measuring was similar to manual measuring (Tipping, 1994), and although labour saving for some situations, is costly and not portable or practical for enforcement officers. More sophisticated computer approaches using stereoscopic imaging also suffer errors/variability of around 3%. It seems unlikely, therefore, that currently available technological 'fixes' would obviate the need for developing standardised manual measurement procedures or remove the need to formally allow for measurement uncertainties when assessing compliance with minimum size limits.

Sample Criteria for Specified Rates of Confidence in Breaches

The above results indicate that the measurement of fish on two or more occasions is unlikely to result in identical length estimates. The question remains, then, of how allowances for such uncertainties should be included in decisions by enforcement officers about whether to issue a breach. The fish selected for measurement by enforcement officers will be those fish that appear to be close to the minimum legal size limit. If we assume that all such fish were measured as exactly 38cm when caught, and so legally harvestable, guidelines for the tolerance of apparently undersized fish at a later time of measurement can be specified.

The numbers of those fish that appear to be undersize subsequently and should be considered outside of the reasonable allowances for measurement uncertainty will depend on four things (see Table 3):

a. The estimated probabilities of a fish being smaller by specified amounts on a second measurement compared with a previous measurement. Two sets of such estimates are provided above, one for consistent observers and one for diverse observers;
b. The degree of certainty that breaches are warranted or legitimate that will be considered acceptable. That rate will be arbitrary and is a measure of the confidence that society, the judiciary or legislators demand in effecting a prosecution – or, conversely, the risk that they are willing to accept that sometimes a breach might occur simply because of measurement errors;
c. The number of suspect fish measured; and
d. The amount by which fish are estimated to be under the legal size limit.

The numbers of fish required to trigger a breach presented in Tables 3 and 4 were derived from this approach. Numbers in Table 3 are based on the measurement uncertainty we observed across all observers (2 researchers & three fishers) and so might be seen as a 'worst case' scenario – i.e., most generous to the fishers. Numbers in Table 4 were derived from data from only those observers who showed reasonable consistency with each other in their measurements and were believed to use a similar measurement method. These results indicate that the numbers of undersized fish required to trigger a breach would be reduced considerably if standardised measurement procedures were shown to reduce inter-observer measurement errors generally and were made compulsory across the fishery.

It is noteworthy that we have no evidence that a fish measured as 38cm at the time of capture would be subsequently measured as small as 35cm in length and it would be extremely unlikely that more than 2 or 3 fish would be erroneously classed as less than 36cm long, even allowing for non-uniformity in measurement practices. Without standardisation of measurement method, therefore, this observation might be taken as justification for issuing a breach if, in addition to the figures in Table 3, any fish was measured as 35cm or smaller during compliance monitoring. With standardisation of measurement method across the fishery, it would seem exceptionally unlikely that any fish that was actually 38cm long would be estimated to be less than 36cm in length. Thus, if a standard method was adopted, a breach might be considered justified if any single fish was measured as less than 36cm in length.

Finally, it is important to emphasise that the figures in Table 3 must not be taken to indicate the numbers of undersized fish it is safe to have in a catch. The tabulated numbers are based on the assumption that all fish selected for measurement a second time were actually 38cm when first measured. If some of those fish were measured as smaller than 38cm initially, the probability that the indicated numbers of fish will be subsequently measured as undersize will increase significantly. That is, taking the tabulated figures to indicate that a few undersized fish can be kept with safety will be highly likely to result in prosecution if the catch is later inspected.

Recommendations

We make the following recommendations in view of the results of this research.

1. Enforcement, administrative, legislative and industry personnel be informed that the issue of uncertainty in measurement of live fish has a clear empirical basis.
2. Enforcement, administrative, legislative & industry personnel be advised of the potential magnitudes both of the uncertainty in measurements (or the differences to be expected between two independent measurements) and the frequency with which given discrepancies in measurement might be expected during compliance and enforcement monitoring.
3. Further work be done to verify whether rudimentary instruction in measurement procedures is effective in reducing uncertainties in the measurement of length of fish and, more importantly, reducing the amount of variation to be expected in the comparison of measurements by different people (e.g., at capture and at compliance inspection).
4. Additional work be done to establish the similarity or difference in measurement uncertainty among different species of fish before seeking to apply these results to measurements of species other than coral trout. Additional work also should include replicating the work reported here to verify its generality.
5. Consideration be given to drafting regulatory and / or judicial guidelines for criteria by which observed frequencies (number of fish per sample) and magnitudes (numbers of millimetres) of apparent infringement of minimum legal sizes should be considered within the bounds of reasonable expectation or considered highly unlikely to occur by chance and therefore lead to prosecution.
6. An education and extension program be developed to adequately inform fishers and enforcement personnel of the need to a) check the measurement of ‘near legal’ sized fish after holding for 1-2 days, b) measure fish by the recommended standard method (when such method had been agreed), and c) allow for uncertainty in measurement of fish to minimise of the risk of erroneous prosecutions.
7. A standard measuring technique be defined and sent to all fishers and enforcement officers. In the interest of reducing the potential variation between observers’ measures the authors suggest the following; that the fish’s mouth be closed or gently and carefully forced closed, there should be no lateral compression of the body or caudal peduncle, and the length should be measured from the tip of the lower jaw to the posterior margin of the dorsal lobe of the tail, preferably gently compressed to the centre-line of the body. If the dorsal lobe is damaged (*e.g.*, due to injury or ‘fin rot’), a measurement to the end of the ventral lobe should be used if the latter is undamaged. If both lobes are damaged no reliable measurement of total length will be available.

References

Andrew N.L. and B.D. Mapstone. 1987. *Sampling and the Description of Spatial Pattern in Marine Ecology*. **Oceanog. Mar. Biol. Ann. Rev.**, Vol. 25: 39-90. 1987.

Anon, 1999. *Draft Management Plan and regulatory impact statement. Queensland Coral Reef Fin Fish Fishery.* Prepared by the **QFMA** with the Reef Fish MAC. July 1999, pp 80.

Genstat 5 Committee of the Statistics Department, Rothamsted Experimental Station (1998). "**Genstat 5, Release 4.1** Reference Manual Supplement." Oxford: Numerical Algorithms Group.

Gwynne, L. 1990. *A review of the reef line fishery and proposed management measures. A discussion paper*. Queensland Fish Management Authority, Brisbane.

Higgs, J.B. 1996. *A review of published fisheries dependent and independent surveys of the recreational Great Barrier Reef line fisheries and demersal reef fish stocks*. Report to the Queensland Fisheries Management Authority.

Keough M.J., and B.D. Mapstone. 1995. *Protocols for designing marine ecological monitoring programs associated with BEK Mills*. **National Pulp Mills Research Program Technical Report N^{o} 11**. CSIRO. Canberra. 185pp.

Mapstone B.D., J.P. McKinlay and C.R. Davies 1996b. *A description of the commercial reef line fishery logbook data held by the Queensland Fisheries Management Authority*. **Report to the Queensland Fisheries Management Authority.** 480pp

Mapstone B.D., C.R. Davies and J.W. Robertson. 1996. *The Effects of Line Fishing on the Great Barrier Reef: Available Evidence and Future Directions*. pp 178-192 in **Proceedings of The Great Barrier Reef: Science Use and Management - a National Conference**. Townsville.

Mapstone B.D. 1995. *Scalable decision rules in environmental impact assessment: effect size, Type I, and Type II errors.* **Ecological Applications**. 5(2): 401-410.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. **Biometrika** 58, 545-554.

Payne, R. W. 1996. *Analysis of Repeated measurements in Genstat*, Notes from workshop presented in conjunction with the Australasian Genstat Conference, Genstat 96.

Peterman, R. 1990. *Statistical power analysis can improve fisheries research and management*. **Can. J. Fish. Aquat. Sci.** 47: 2-15.

Tipping, J M 1994, *Measuring fish by video image processing.* **The Progressive Fish-Culturalist** 56:299-300

Welham, S. and Cullis, B. 1999. *Analysis of correlated data using REML*. Notes from workshop presented in conjunction with the Australasian Genstat Conference, Genstat 99.

Welham, S. J. and Thompson, R. 1992. *REML likelihood ratio tests for fixed model terms*. In: Royal Statistical Society Conference, University of Sheffield, 9-11 September 1992, Abstracts of Papers.

Winer B.J., Brown D.R., Michels K.M. (1992) *Statistical Principles in Experimental Design*. 3rd Ed. McGraw-Hill Kogakusha, Tokyo. 1057pp.