Exxon Valdez Oil Spill
Restoration Project Annual Report

Otolith Marking of Pink Salmon in Prince William Sound Salmon Hatcheries, 1996

Restoration Project 96188
Annual Report

This annual report has been prepared for peer review as part of the *Exxon Valdez* Oil Spill Trustee Council restoration program for the purpose of assessing project progress. Peer review comments have not been addressed in this annual report.

Timothy L. Joyce
David G. Evans
Kristen M. Munk

Alaska Department of Fish and Game
Division of Commercial Fisheries Management and Development
401 Railroad Avenue
Cordova, Alaska 99574

March 1997

Otolith Marking of Pink Salmon in
Prince William Sound Salmon Hatcheries, 1996

Restoration Project 96188
Annual Report

Study History: Hatchery and wild stocks in Prince William Sound salmon fisheries have been assessed through use of an extensive coded wire tagging program. Due to the expense of applying coded wire tags and the various assumptions that need to be met since tags could only be applied to a relatively small portion of the total population being studied, scientists have been trying to develop a mass marking technique that would avoid these problems. Thermal marking of otoliths is a relatively new technology in which specific patterns can be laid down on the otoliths of incubating fish. The technique promises to improve the precision and accuracy of hatchery contribution estimates. In 1995 and 1996, it was used at Prince William Sound hatcheries to mark otoliths of all incubating pink salmon (*Oncorhynchus gorbuscha*). It is believed that the method will replace coded wire tagging as a means of identifying pink salmon produced in hatcheries.

Abstract: In the fall of 1995 and 1996, thermal marks were applied to the otoliths of all hatchery pink salmon in Prince William Sound. Otolith marks were highly visible on voucher samples taken in the spring 1996 from hatchery swim-up fry. Mixtures of hatchery and wild swim-up fry were sent to the Otolith Processing Laboratory in Juneau, Alaska. A double blind experiment indicated that about 98% of the thermally applied otolith marks could be accurately identified. In preparation for return of these marked fry as adults, a catch sampling protocol was assessed using results of a sampling experiment with finclipped adult pink salmon and a computer simulation. Known numbers of finclipped pink salmon were added to holds of tenders as they received catches from seiners participating in a Solomon Gulch cost recovery fishery. Sample estimates of the proportion of finclipped pink salmon aboard tenders were compared to actual population proportions. A computer simulation, using catch data from the 1996 fishery, is being completed to determine the likely precision of estimated contribution rates when populations of fish within a tender are structured. Results obtained suggest that the precision of estimates of hatchery contributions will be greater than specified in sample size calculations.

Project Data:
Data pertaining to the double blind test are stored in Microsoft Excel worksheets, ASCII files, and a Microsoft Access database. Software code used to analyze the data (SAS™, GAUSS™) is available in ASCII format. Data pertaining to the finclip study are stored in Excel worksheets and ASCII files. Simulation results are available in ASCII format, as is the GAUSS™ code used to generate them.

Data Custodian: David Evans
e-mail address: davide@fishgame.state.ak.us
telephone: (907) 267-2176

fax: (907) 267-2442
address: Alaska Dept. of Fish and Game
333 Raspberry Road
Anchorage, AK 99518-1599

Key Words: Commercial harvest, hatchery, *Oncorhynchus gorbuscha*, otolith, pink salmon, Prince William Sound, thermal mark, voucher sample, wild stock.

Citation:

Joyce, T.L., D.G. Evans, and K. M. Munk. 1997. Otolith marking of pink salmon in Prince William Sound hatcheries, 1996, *Exxon Valdez* Oil Spill Restoration Project Annual Report (Restoration Project 96188), Alaska Department of Fish and Game, Division of Commercial Fisheries Management and Development, Cordova, Alaska.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

EXECUTIVE SUMMARY

This report documents Restoration Study 96188, one of the projects designed to restore the pink salmon *Oncorhynchus gorbuscha* resource of Prince William Sound to its pre-spill status. Three objectives were outlined for this study. The first objective, to apply unique and distinct thermal marks to the otoliths of developing pink salmon embryos at all four pink salmon hatcheries in Prince William Sound, was met in both the 1995 and 1996 brood years using equipment purchased and installed in 1995. The second objective, to measure the quality and uniqueness of otolith marks applied in 1995, and to identify any problems pertaining to specific mark assignments, was met upon completion of the double-blind test in which Otolith Laboratory personnel successfully identified the origin of otoliths from hatchery and wild fry 98% of the time. It is reasonable to assume that recovered otoliths can be identified with little error. The third objective, to test a method for collecting representative samples from tender boats unloading salmon at the processing plants, was met by use of a finclipping experiment and simulation study. No evidence was found to suggest that the sampling methodology was biased, and simulation studies revealed that precision of estimates under ordered populations may be substantially greater than that expected under populations with no structure.

INTRODUCTION

Between 1961 and 1976, when hatcheries were absent from Prince William Sound, the commercial seine harvest of wild pink salmon *Oncorhynchus gorbuscha* averaged about 3.4 million fish. In the early 1970's, run failures led to an aggressive enhancement program which included construction of hatcheries. By 1986, five hatcheries were operating: Solomon Gulch hatchery, producing pink salmon, and later, also chum *O. keta*, coho *O. kisutch* and chinook salmon *O. tschawytscha*; A. F. Koernig hatchery, producing pink salmon; W. H. Noerenberg hatchery, producing pink salmon, and later, also chum, coho and chinook salmon; Cannery Creek hatchery, producing pink salmon; and Main Bay hatchery which produced chum and presently raises sockeye salmon *O. nerka*. From the late 1980's to the present, returns to these facilities have contributed approximately 20 million pink salmon to the annual run.

Prince William Sound hatchery parent stocks were selected from local native populations, and the migratory timing of adult hatchery and wild runs are similar. Furthermore, all of these stocks migrate to their natal streams or hatcheries through corridors in the southwestern and western areas of Prince William Sound. Since both timing and migratory corridors of the large hatchery runs and the much smaller wild runs are similar, and since the hatchery runs can be harvested at a greater rate than wild runs, there is a great danger of overexploiting wild runs. Indeed, shortfalls in wild escapements occurred in more than half of the 15 years prior to hatchery production, when the average exploitation rate was 42%, a figure considerably lower than the 60% considered appropriate today for returning hatchery salmon.

To protect wild stocks in a fishery dominated by hatchery salmon, managers needed information pertaining to the temporal and spatial distributions of hatchery and wild salmon. In 1986, a coded wire tagging program was initiated for hatchery releases of pink salmon to meet this need with the first recovery of tagged, returning adults in the commercial and cost recovery fisheries beginning in 1987. This tag recovery data enabled managers to obtain estimates hatchery and wild contributions to catches from selected temporal and spatial strata within the fishery.

The March 24, 1989, *Exxon Valdez* oil spill exacerbated the problems faced by fishery managers. The spill contaminated intertidal portions of streams where most wild pink salmon stocks in western Prince William Sound spawn as well as the marine waters traversed by juvenile salmon on their migration seaward. Detrimental effects have been found from oil contamination upon pink salmon embryos, pre-emergent fry, and juvenile salmon in wild populations (Sharr *et al.* 1994, Willette *et al.* 1994). The decisions made by fishery managers suddenly became more complex since they now affected wild populations injured by the oil spill.

The coded wire tag program was continued after the spill and was funded by Natural Resource Damage Assessment study F/S 3 through 1991 (Sharr *et al.* 1995a). During this period, the program continued to provide information pertaining to the stock composition of the commercial salmon catch. The pink salmon tagging program was supported from 1992 through 1996, by

Restoration Studies R60A (Sharr *et al.* 1995b), R93067 (Sharr *et al.* 1995c), R94320B (Sharr *et al.* 1995e), R95320B (Riffe *et al.* 1996) and R96186 (in preparation), along with contributions from the Prince William Sound Aquaculture Corporation, Valdez Fisheries Development Association and the State of Alaska.

Coded wire tag hatchery contribution estimates are based on several assumptions. The most contentious of these pertain to an adjustment factor used to account for differential mortality and tag-shedding. Adjustment factors are calculated based on the assumptions that 1) brood ponds contain only salmon reared at the hatchery in question, and 2) for a given cohort, the tagging rate calculated for the brood stock is equal to that experienced in the commercial fishery. Immigration of wild fish into brood stocks may occur (Sharr *et al.* 1995c), which would tend to inflate catch estimates of hatchery salmon, and tags may induce straying (Habicht 1996) , which might tend to inflate estimates of wild stock escapement. In light of these studies, it became clear that hatchery contribution estimates based on coded wire tags may be flawed, and an alternative marking technology was sought

Munk *et al.* (1993), Mosegard (1987) and Volk (1990) have demonstrated that chinook, coho, sockeye, chum, pink, and Atlantic *Salmo salar* salmon otoliths in embryos can be marked by carefully controlled changes in water temperature, while Hagan *et al.* (1995) have successfully incorporated the technology into a mixed stock fisheries assessment program. In 1995 and 1996, thermal marks were applied to the otoliths of all pink salmon incubating in Prince William Sound hatcheries, with support from R95320C and R96188, respectively. Using otolith marks will eliminate problems associated with tag loss and differential mortality, and, therefore, the need for applying adjustment factors. The cost of applying otolith marks is also substantially less than that of applying coded wire tags to an equal number of fish. In 1997, simultaneous recovery of coded wire tags and thermally marked otoliths will allow us to examine some of the assumptions made in the coded wire tag program.

In the otolith program, every salmon receives a thermal mark, and the proportion of the catch that must be examined for a given level of precision will be much smaller than that needed in a coded wire tag program, where only about 2% of released salmon are marked. However, when sample sizes are small, issues concerning representative random sampling and correct identification of otolith origins become more important A proposed sampling methodology for recovering marked otoliths from commercial harvests was tested by comparing sample estimates of the proportions of externally marked salmon to known population proportions aboard a tender. Our ability to accurately identify thermally marked otoliths was measured through a double-blind test conducted at the Otolith Laboratory. Finally, computer simulations were used to assess the effect of structured populations on the precision of estimates.

This report documents application of thermal marks for the 1996 brood year salmon and presents an early assessment of mark quality for the 1995 brood year salmon. Field and computer experiments designed to assess the proposed sampling methodology are also reported.

OBJECTIVES

1. To apply unique and distinct thermal marks to the otoliths of developing pink salmon embryos at all four pink salmon hatcheries in Prince William Sound.

2. To measure the quality and uniqueness of otolith marks applied in 1995, and to identify any problems pertaining to specific mark assignments.

3. To test a method for collecting representative samples from tender boats unloading salmon at the processing plants.

METHODS

Application of Thermal Marks-Fall 1996

Thermal marks were laid down after the primordial stage of otolith development (approximately 275 TU) or, equivalently, at the 'eyed' stage. Methods followed those of Munk (1993) and Munk *et al.* (1993). Marking was completed prior to hatching to eliminate confusion with the 'hatch mark', and to prevent problems associated with gas supersaturation. Each ring within a mark was created by a temperature induced modification of the rate of deposition of otolith material. This modification was accomplished by raising the ambient temperature of the incubation water for 24 hours by 4 °C, and then rapidly returning it to its original value. Each repetition of this process induced one ring.. Later in the season as the ambient temperature dropped, 36-hour, alternating cycles were used at the Cannery Creek and W.H. Noerenberg hatcheries to insure proper spacing between rings. Marking schedules were staggered for pairs of incubators so that the oil fired boilers ran continuously. This schedule marked the maximum number of embryos in the shortest time.

The thermal marks were classified using a "Region, Band, and ring" (RBr) code, written numerically as 'R:B.r' (Table 1). The region (R) of the mark denotes the general location of the mark within the otolith and has three designations. A '1' implies that the mark occurs in the area after the primordial stage and before the hatch mark, a '2' that the mark occurs in the area after the hatch mark, and a '3' that the mark may occur in one area or the other. A mark may consist of more than one band of rings and the 'B' designation of the RBr code indicates the number of these bands. The number of rings associated with each band is denoted by 'r', which immediately follows the band designator in the code. Generally speaking, bands will have a minimum of three rings to avoid confusion with normal growth sequences.

Table 1 Thermal mark codes and associated thermal schedules

Hatchery	Schedule	R:B.r	Ring pattern
A.F.KOERNIG	(4X)24H:24C Base	1:1.4	IIII
	(3X)24H:24C Accessory	1:1.4+1.3	IIII III
CANNERY CREEK	(3X)24H:24C,(1X)72H: 36C,2(X)24H:24C	1:1.3,2.3	III III
W.H. NOERENBERG	(8X)24H:24C Base	1:1.8	IIIIIII
	(3X)24H:24C Accessory	1:1.8+1.3	IIIIIII III
SOLOMON GULCH	(6X)24H:24C	1:1.6	IIIII

Prince William Sound hatchery basemarks appear on the otolith in the Region 1 section of the otolith. They were chosen to distinguish hatchery of origin. The W.H. Noerenberg and A.F. Koernig facilities also applied accessory marks which differentiate size at release. The accessory mark is identified in the RBr coding by a '+' prefix to the band number, and indicates an interval of greater than five rings.

Voucher samples were taken at the time of emergence from each lot at each hatchery so that thermal mark codes could be verified, and any confounding marks laid down during the remaining incubation period could be documented.

Determination of the readability of otoliths marked in 1995

Our ability to successfully determine the origin of otoliths was measured, along with information on within and between reader variability, and reader agreement, through double blind tests conducted at the Otolith Laboratory. An identification matrix was used to highlight specific tendencies readers may have had when making erroneous assignments.

Sample Collection

In the spring of 1996, pink salmon fry were collected from incubators at the W.H. Noerenberg, A.F. Koernig, Cannery Creek and Solomon Gulch hatcheries, and from twenty streams located throughout Prince William Sound. Approximately 600 hatchery fry and 600 wild fry were sent to the Otolith Laboratory where otoliths were extracted, mounted on glass microscope slides with thermoplastic cement, and placed in labeled slide boxes. These slide boxes were sent to personnel at the Anchorage office of the Alaska Department of Fish and Game where slides were coded and mixed. Twelve boxes of one hundred coded slides were then shipped back to the Otolith Laboratory for identification. Coded information was not made available to Otolith Laboratory personnel.

Experimental design

During the study four readers were assessed in four different mark interpretation events (Table 2). In the first event, the twelve boxes of 100 slides each were assigned to readers in a random fashion with the restriction that each reader grind an equal number of slides. Once a slide had been ground, no further grinding by other readers was permitted. It was assumed that each reader was capable of grinding an otolith to the degree that an interpretation was possible. In the subsequent three events, time and funding constraints limited analyses to four boxes randomly chosen from the twelve original boxes.

Table 2 Experimental layout for test of otolith-reading laboratory. Numbers differentiate slide-boxes.

First Event				Second Event				Third Event				Fourth Event			
Reader				Reader				Reader				Reader			
A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
3	4	7	11	10	10	6	7	7	10	6	10	7	10	6	6
1	2	5	10	6	6	3	3	6	6	10	3	6	7	3	3
9	1	6	8	7	7	10	10	10	7	7	7	3	3	10	7
6	10	2	12	3	3	7	6	3	3	3	6	10	6	7	10
11	7	3	2												
10	5	1	7												
4	3	10	5												
5	11	9	4												
8	6	12	6												
2	1	8	3												
12	8	11	1												
7	9	4	9												

Along with determination of the origin of an otolith, each reader also recorded a measure of the confidence with which the determination was made. Upon completion of the readings, all determinations were sent back to the Anchorage office for analysis.

Data Analysis

The overall ability to correctly identify otoliths was determined, along with an examination of the level of agreement between readers and of trends in misclassifications, by comparing readers interpretations of marks to the known origin of marks.

Success Rates in Otolith Identification

The success rate for identification of a population of otoliths is defined as the probability that readers will determine the origin of a randomly selected otolith without error. Success rates were

estimated for six different populations. These consisted of the overall population of otoliths of hatchery and wild origin, the populations associated with each of the four hatcheries, and that associated with the wild population alone.

The estimated success rate for a population , $\hat{p}$, is calculated as

$$\hat{p} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{t} p_{ij}}{rt}, \qquad (1)$$

where p_{ij} is the success rate measured for the i^{th} reader and j^{th} event.

To calculate an appropriate variance estimate for $\hat{p}$, account must be taken of sampling variability in the initial selection of otoliths from the population, and of variability encountered in the laboratory estimation process which is derived from the random nature of reader and event effects. The entire process was simulated and approximate confidence intervals for success rates obtained using the percentile method.

Each iteration of the simulation for a given population was conducted in two stages (simulated quantities are identified with '*'):

1) Simulation of initial sample selection : A binomial (n, $\hat{p}$) random variable, X^*, was generated with n = number of otoliths initially selected and $\hat{p}$ = estimated success rate for the population in question (Equation 1). A simulated success rate $p^* = X^*/n$ was then calculated.

2) Simulation of laboratory estimation of p^*: The quantity $\hat{p}^*$ was generated from $N(p^*, \hat{\sigma}^2_{\hat{p}_s})$, where $\sigma^2_{\hat{p}_s}$ is the variance of the laboratory estimate and $\hat{p}_s$ is the success rate given the selected sample. The quantity $\hat{p}_s$ is the mean of rt=16 observations (r=4 readers and t=4 events), and operation of the central limit theorem ensures an approximation to normality. While $\hat{p}_s$ is bounded by unity, the estimate of its variance, $\hat{\sigma}^2_{\hat{p}_s}$, is such that the normality assumption is reasonable. Under the assumption that reader effects are random, the quantity $\sigma^2_{\hat{p}_s}$ was estimated from the original data using a two-stage random effects model:

$$p_{ij} = \mu + R_i + \tau_{ij}, \qquad (2)$$

where R_i and τ_{ij} are random variables representing effects of the i^{th} reader and j^{th} event, respectively: i=1,..,r; j=1,..,t. Estimates of the variance of R_i, $\hat{\sigma}^2{}_R$, and τ_{ij}, $\hat{\sigma}^2{}_\tau$, were obtained from a two-way ANOVA (SAS™, 1996). An estimate of $\sigma^2_{\hat{p}_s}$ was calculated as:

$$\hat{\sigma}^2_{\hat{p}_s} = \frac{\hat{\sigma}^2_R}{r} + \frac{\hat{\sigma}^2_\tau}{rt} \quad . \tag{3}$$

With reader effects considered fixed, the estimate of $\sigma^2_{\hat{p}_s}$ is given by

$$\hat{\sigma}_{\hat{p}_s} = \frac{\hat{\sigma}^2_\tau}{rt} \quad . \tag{4}$$

For each simulated population, an empirical distribution function (EDF) was constructed for both the random and fixed-reader cases, and 95% confidence intervals for the success rates were obtained as $EDF^{-1}(0.025)$ and $EDF^{-1}(0.975)$. Estimated variances were calculated using the common sample variance formula on the vector of simulated success rates. Confidence intervals and variance estimates calculated for the fixed-effect case will be more appropriate for the 1997 return of pink salmon, providing the current group of readers remain at the Otolith Laboratory until that time. Those calculated for the random-effect case may be more appropriate for comparisons of success rates over years. For both analyses, it is assumed that the Otolith Laboratory is 'stable' in that readers did not improve through the testing period. This is a reasonable assumption given that extensive training was given prior to the tests. A multiple regression of success rate versus event, accounting for readers, was performed to examine this question.

Reader Agreement

While perfect agreement can occur simultaneously with complete failure in identification, the degree of consistency among readers is nevertheless an important parameter.. Cohen's kappa was used to assess agreement between readers. This statistic compares the observed agreement to that expected if the ratings were independent, and thus accounts for agreement occurring by chance alone. For $\Gamma_o = \sum \hat{\pi}_{ii}$ and $\Gamma_e = \sum \hat{\pi}_{i+}\hat{\pi}_{+i}$, where π_{ii} is the probability of a classification in category i by both readers, and π_{+i} is the marginal probability for category i for one of the readers and π_{i+} for the other reader, Cohen's kappa is calculated as

$$\kappa = \frac{\Gamma_o - \Gamma_e}{1 - \Gamma_e} \quad . \tag{5}$$

The ratio is a measure of the agreement in excess of that expected by chance to the excess under perfect agreement. The distribution of κ is asymptotically normal for multinomial sampling , and 95% confidence intervals were calculated as κ +/1.96*standard error (see Agresti, 1990 for variance formula).

Identification matrix

An identification matrix was produced in order to identify any trends in errors. A 5x5 matrix was constructed with true and observed origin describing rows and columns, respectively. In addition

to examining the matrix visually, a quasi-independence model was fitted to the data, whereby the main diagonal of the matrix was fitted perfectly, and a test of independence conducted in the off-diagonal portions of the matrix. The model assumes that, conditional on disagreement, odds ratios among all rectangularly formed 2x2 tables equal 1.0.

The model is described as

$$\log \mu_{ij} = \lambda + \lambda_i^T + \lambda_j^O + \delta_i I(i,j) \ , \qquad (6)$$

where $I(i,j)=1$ when $i=j$ and 0 otherwise. T refers to the variable associated with the true identification of the otolith, and O refers to that associated with the observed identification. If significant lack-of-fit existed when this model was fitted, Fisher's exact test of independence was used in 2x2 tables formed in the off-diagonal areas of the matrix.

Assessment of Proposed Catch-Sampling Technique

Finclip study

A comparison was made between sample estimates of the proportions of salmon bearing an external mark and known population proportions. During the cost recovery harvest in District 221 in 1996, known proportions of salmon loaded onto tenders were finclipped, and estimates of these proportions were made using the proposed sampling method. Bias in the sampling scheme was assessed by examining the proportion of the (1-α) confidence intervals containing the known population proportions. The exercise also served as a field test of the sampling technique.

Pectoral fins were removed from known numbers of salmon *en route* to the holds of six tenders delivering to the Peter Pan processor in the port of Valdez, Alaska. Four independent pseudo-systematic samples were taken from each tender at the conveyor belt during delivery to the processor. An electronic wrist watch with a 'count-down' feature served as the signal for technicians to sample a salmon. Salmon were sampled in this manner until the tender was unloaded. Each technician sampled approximately 500 to 600 salmon.

An estimate of the population proportion associated with the i^{th} tender derived from the j^{th} sample ($\hat{p}_{ij}$) was made as

$$\hat{p}_{ij} = \frac{m_{ij}}{n_{ij}} \ , \qquad (7)$$

where, n_{ij} = number of salmon examined for a missing pectoral fin from the i^{th} tender by sampler j and, m_{ij} = number of marked fish found in n_{ij}.

An estimate of the population proportion associated with the i^{th} tender over the four systematic samples was calculated as

$$\bar{p}_i = \frac{1}{4}\sum_{j=1}^{4}\hat{p}_{ij} \quad . \tag{8}$$

The variance of $\bar{p}_i$ was estimated as $s^2{}_i/4$, where $s^2{}_i$ (multi-start systematic sampling variance estimate) was calculated as

$$s_i^{\,2} = \frac{\sum_{j=1}^{4}(\hat{p}_{ij} - \bar{p}_i)^2}{3} \quad . \tag{9}$$

Assuming that sample estimates were normally distributed, a 50% confidence interval was calculated for each $\bar{p}_i$ as

$$\bar{p}_i \pm t_{0.25,3df} * \sqrt{\frac{s^2{}_i}{4}} \quad . \tag{10}$$

For each of the six independent confidence intervals, an assessment was made to determine whether the interval included the corresponding population mean. Under the null hypothesis that samples were unbiased, the number of times coverage was achieved, x, is a realization of binomial random variable, X, with parameters n=6 and p=0.5. A p-value for the null hypothesis was calculated as $2P(X < x)$ for $x < np$ and $2P(X < x)$ for $x > np$.

Another *ad hoc* test was performed in which two pairs of systematic sample means were chosen from each group of four means associated with each tender. For each pair of means, one estimate was made of the population proportion associated with the tender, along with an estimate of its variance. The calculations are similar to those described by Equations 8 and 9. A total of twelve proportions and corresponding variances were independently estimated from the six tenders. Confidence intervals were calculated in a manner identical to that above, except that the t parameter was associated with 1 degree of freedom (each s^2 was based on two means). A determination was made of the number of times the corresponding population mean was included within the twelve intervals, and the hypothesis test described above was conducted under the assumption that the random variable was binomial ($n=12$, $p=0.5$). Averaging the number of coverage instances over all 3^6 (=729) possible constructions solved the question regarding the construction of the pairs of means, *i.e.*, from which means pairs are formed.

Simulation study

Computer simulation studies were undertaken to determine the influence of structured populations on the precision of the estimated proportion of hatchery salmon caught in a fishery

opening. It was hypothesized that in structured populations, the systematic sampling methodology would lead to estimates that were more precise than those obtained under random sampling. An extreme example follows. In a catch of 10,000 salmon, a systematic sample of ten salmon taken from the conveyor belt would consist of one salmon from each contiguous set of 1,000 salmon. If 5,000 hatchery salmon were unloaded first, followed by 5,000 wild salmon, the variance of the estimated proportion of hatchery salmon in the load is zero. The variance of an estimate derived from a random sample would be of hypergeometric form, and would be greater than zero (about 0.025). The simulation examined the degree to which the precision of estimates based upon random sampling theory might be underestimated in the presence of structured populations.

The 1996 pink salmon fishery statistics were used as the framework for the simulation study, which was conducted in the following manner.

1) Catches and coded wire tag estimates of hatchery proportions were obtained for each of 29 major harvest-district-week openings occurring in 1996.

2) For each harvest-district-week openings, a group of tender loads was selected which was representative of those delivered in that stratum. Loads were randomly selected from this group until the total accumulated load approached the catch associated with the stratum. Any difference between the catch and the accumulated load was split equally among the selected tenders, so the total selected load and catch were equal. The tenders were selected only once during the simulation and are a representative realization of the 1996 fishery.

3) For each selected tender load, a specific structure was imposed on the population of fish moving along the processor conveyor belt. The nature of the population was determined by a 'mixing factor' which controlled the degree to which hatchery and wild fish were mixed. Only one mixing factor was used within each simulation. The mixing algorithm operated as follows:

   A) Each salmon population within a tender was first ordered (e.g. all hatchery salmon, then all wild salmon).
   B) For a mixing factor of 0.7, a random sample of 70% of the ordered population was selected and randomized within itself . A mixing factor of 0.0 indicated that the final population was completely ordered while a mixing factor of 1.0 indicated it was arranged randomly.
   C) The randomized salmon were returned to the positions in the population from which the ordered sample had been taken.

4) For each iteration of the simulation, a systematic sample of s=100 salmon was taken from each tender load, N. The group of starting points required for the set of systematic

samples taken from each tender was randomly selected without replacement from all possible *N/s* starting points.

5) For the *t* tenders associated with each harvest-district-week stratum within an iteration, a composite sample of 150 salmon were randomly selected from the group of *t* samples of size *s* in a manner proportional to the loads aboard the tenders. The proportion of hatchery salmon in this final sample was calculated and stored until completion of the simulation.

6) After all iterations had been completed, the standard deviation was calculated for the simulated proportions of hatchery salmon for each harvest-district-week stratum. It was then compared to theoretical values obtained with random sampling assumptions. The simulation was conducted using GAUSS™ (1996) code is presented in Appendix E.

7) The exercise was repeated for five mixing factors, with fifty iterations conducted for each factor.

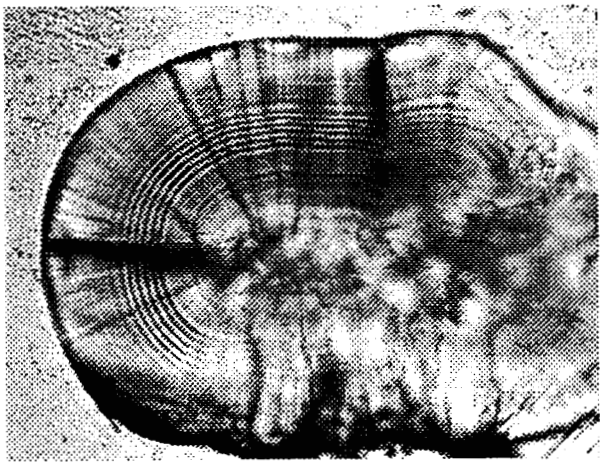
RESULTS

Application of Thermal Marks-Fall 1996

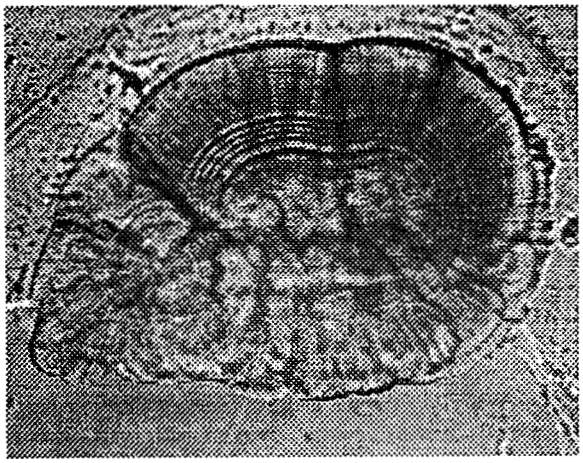
Incubation water temperatures were maintained at 3.8° to 4.0°C above ambient at all of the Prince William Sound hatcheries when required by the marking schedule When marking system problems occurred, they were fully documented by hatchery staff and the Otolith Laboratory was notified. Modifications to mark schedules were made when appropriate, resulting in only minor variations to basemarks. None of these modifications compromised mark integrity of any hatchery basemark.

Samples taken three weeks after completion of the marking process revealed that high quality thermal marks had been laid down at each of the four hatcheries (Figure 1) .

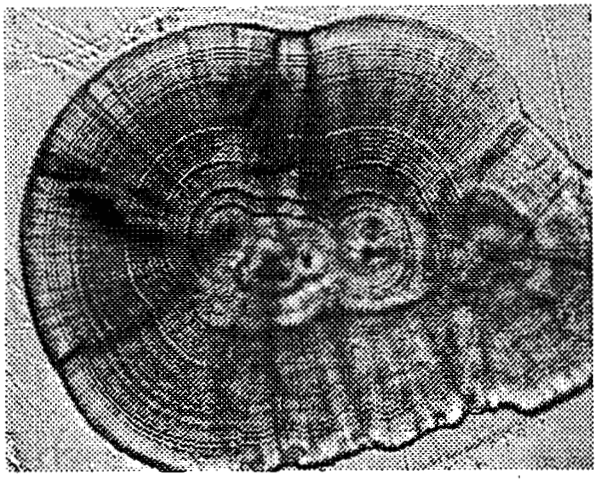
Figure 1 Thermally marked pink salmon otoliths sampled from Prince William Sound hatcheries.



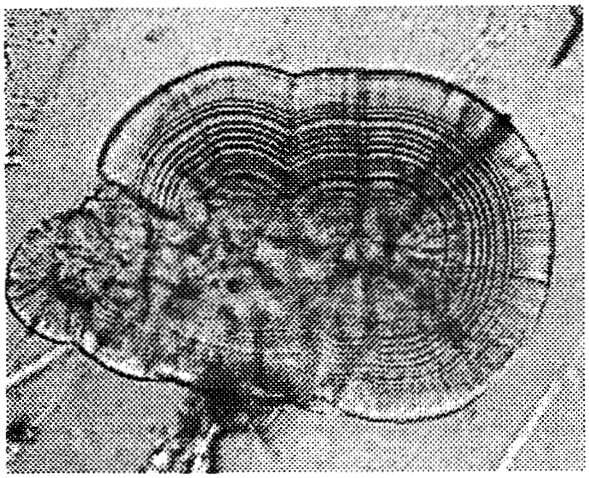
Solomon Gulch - 96



A.F. Koernig - 96



W. H. Noerenberg - 96



Cannery Creek - 96

Success Rates

Estimated mean success rate was 0.988 over all readers and all events (Table 3). Treating readers as random or fixed factors had little bearing on empirical confidence intervals. The estimate is associated with a 95% confidence interval of {0.975, 0.997}. The expected success rate was 0.20 (five choices for each slide) for a reader having no ability to determine the origin of the otoliths on a slide.

Table 3 Overall success rates of readers by event[a].

	Event				
Reader	1	2	3	4	Mean
A	0.968	0.983	0.988	0.993	0.983
B	0.989	0.983	0.985	0.990	0.987
C	0.992	0.995	0.988	0.990	0.991
D	0.994	0.993	0.993	0.985	0.991
Mean	0.986	0.988	0.988	0.989	0.988

a Means may not be reproducible from the table due to rounding.

Overall success rate improved when calculated on a hatchery-wild distinction. The average success rate was 0.996 over all readers and events. Success rates fell below 0.992 only once (see Appendix A). No confidence intervals were calculated for this measure.

Estimated success rates and 95% confidence intervals by origin ranged from 0.967 {0.918, 1.000} to 0.993 {0.977, 1.000} (Table 4). Again, treatment of reader effects as random or fixed had little bearing on the results.

Table 4 Success rates by origin of otolith.

Origin[a]	Lower 95% Bound	Point Estimate	Upper 95% Bound
AFK	0.927	0.984	1.000
CC	0.959	0.990	1.000
SG	0.918	0.967	1.000
WN	0.961	0.992	1.000
WILD	0.977	0.993	1.000
ALL	0.975	0.988	0.998

a AFK=A.F. Koernig; CC=Cannery Creek; SG=Solomon Gulch; WN=W.H. Noerenberg.

No evidence was found to suggest that success rates improved with event for three of the four readers (p=0.76, 0.45, 0.12). An improvement was noted in the success rate of the fourth reader from the first event to the second. In spite of this, the assumption of random variation in reader success rates from one event to another was retained.

Reader Agreement

Measures of agreement between readers ranged from 0.962 to 0.996 (Table 5). The measurements pertain to the first event (1,199 unique slides).

Table 5 Agreement between readers. Single entries are values of Kappa. Entries in parentheses are estimated 95% confidence intervals.

Reader	B	C	D
A	0.964 {0.951,0.977}	0.963 {0.950,0.976}	0.962 {0.948,0.975}
B		0.992 {0.985,0.998}	0.993 {0.987,0.999}
C			0.996 {0.992,1.000}

Identification matrix

The identification matrix for all readers for the first event indicated that, among off-diagonal cells, the SG/CC, the WILD/AFK and the CC/WN cells are where error rates occurred worthy of consideration (Table 6; Appendix B).

Table 6 Identification matrix for all readers for the first period[a].

		OBSERVED				
	ORIGIN	AFK	CC	SG	WN	WILD
TRUE	AFK	574	0	2	0	0
	CC	2	529	0	17	0
	SG	0	20	624	0	0
	WN	0	3	3	642	4
	WILD	14	4	0	0	2358

a AFK=A.F. Koernig; CC=Cannery Creek; SG=Solomon Gulch; WN=W.H. Noerenberg.

There was significant lack-of-fit of the quasi-independence model (χ^2=107.5 with 11 df; p=0), suggesting the existence of dependencies among the rectangularly-formed 2x2 tables in the off-diagonal areas of the matrix. The interpretation of the dependencies is as follows: Given that a mistake has been made in identification of an otolith, the assignment of the otolith origin is dependent on its true identification. Inspection of Table 6 suggests that the contributions to the lack of fit derive from 2x2 tables such as that with diagonals formed from the cells corresponding to SG (True):CC (Observed) and Wild (True):AFK (Observed). A Fisher-exact test yields a p-value of 0 for the hypothesis of independence for this table. If a mistake is made identifying a Solomon Gulch otolith, it is most likely to be classified as a Cannery Creek otolith while a misidentified otolith from a wild salmon is most likely to be classified as an A.F. Koernig otolith. Another similar example exists in the misidentification of Cannery Creek otoliths as W.H. Noerenberg otoliths.

Finclip Study

It was assumed that the number of coverage events is a binomial random variable, X (n=6, p=0.5) in order to test the null hypothesis that unbiased samples are obtained. A p-value was calculated as $2*P(X \geq 3)=1.0$. No evidence was found to suggest that samples were biased (Table 7; Appendices C and D).

When a similar exercise was performed for two-at-time selections (twelve trials), the average number of coverage events was seven out of twelve over all 729 contingencies (p=0.81). Again no evidence was found to suggest that samples were biased.

Table 7 Mark and sample data for finclip experiment

Tender	Load	# Marks	Mark	Sample Estimate	S.E.(mean)	50% C.I.	Coverage[a]
1	18,070	1,363	0.075	0.089	0.0061	0.084, 0.094	0
2	15,174	2,047	0.135	0.140	0.0118	0.131, 0.149	1
3	33,568	2,769	0.082	0.077	0.0065	0.072, 0.081	0
4	25,935	2,159	0.083	0.071	0.0059	0.066, 0.076	0
5	20,777	1,903	0.092	0.095	0.0059	0.090, 0.099	1
6	25,628	2,226	0.087	0.089	0.0062	0.084, 0.094	1
Total							3

a = 1 if confidence interval contains population proportion.
= 0 if confidence interval does not contain population proportion

Simulation Study

When the simulation was executed with random populations, the simulated measures of precision (d values) agreed fairly well with their theoretical counterparts (Table 8). When order was imposed on the populations, the simulated d values decreased . For a mixing factor of 0.7, the average d value was approximately half the value derived from theoretical calculations which assumed random populations and a 50:50 mix of hatchery and wild salmon.

Table 8 Effect of ordered populations upon precision of estimates.

Population Structure		Min/Max d^a value over 29 openings		Average d value over 29 openings
		Min(d)	Max(d)	
Random				
Theoretical		0.026	0.080	0.063
Simulated		0.025	0.091	0.066
Ordered				
Level[b]	0.1	0.028	0.095	0.062
	0.3	0.023	0.082	0.060
	0.5	0.026	0.077	0.055
	0.7	0.019	0.062	0.045

a d is defined : $P(p-d<Estimated(p)<p+d)=0.95$, where p=Population proportion

b A level of 0.3 means that 30% of the fish in the tenders were ordered, *i.e.* 70% of the fish were randomized among the ordered fish.

DISCUSSION

Thermal Mark Application and Detection

Our success in determining the origin of an otolith (*i.e.* readability) is of interest when hatchery contributions to a catch are estimated from recovered otoliths. The double-blind study assessed the readability of otoliths extracted from brood year 1995 fry sampled from hatcheries and streams in the spring of 1996.

It was reassuring to find that readers were able to correctly identify otoliths of unknown origin in a blind test 98% to 99% of the time. Success rates were even higher when simply measured in terms of whether the otolith was identified as coming from either a hatchery or wild pink salmon. Success rates should be as good or better during actual use of this method because readers will grind every otolith they read to their 'personal specifications' and 'second-readings' during postseason analysis will correct some inseason errors. Also, identification of thermal patterns that caused inaccurate determinations in the blind test will help readers working on otoliths extracted from adults returning in 1997. Since measures of agreement between readers were very high, it appears that we will be able to very accurately identify otoliths recovered from the commercial pink salmon catch in 1997.

It is instructional to consider several factors when extrapolating test results to returning adults. The first relates to readability of otoliths extracted from fry versus those extracted from adults. The second concerns an observation by one reader that otoliths extracted from hatchery fry were larger than those taken from wild fry. The third involves evidence indicating that some misclassifications were not made randomly, but reflected a pattern.

With respect to the first factor, two years of three-dimensional growth will bury the thermal mark and require additional grinding. However, an experienced reader has little problem coping with removing this additional deposition, and is often able to reveal the core microstructure within 10 seconds. In addition, no degradation of a thermal mark has been noted over time.

With respect to the second factor, a recognized difference in size between otoliths of emergent hatchery and wild fry could bias estimates of the ability of a readers to identify a salmon from its thermal mark. In practice, however, otolith readers had overwhelming confidence in identifying strong thermal marks, and size differences were only considered secondarily, if at all. When the blind test analysis was restricted to otoliths only of hatchery origin, which were of similar size, the overall success rate was still over 98%. We don't know whether traits such as otolith size will allow us to differentiate hatchery and wild salmon in the adult return, but recognition of their existence may increase the accuracy of classification.

With respect to the third factor, some evidence suggested that certain classification errors were systematic. Some confusion of AFK with wild otoliths was expected from a preliminary examination of hatchery voucher otoliths and a small sample of wild fry otoliths. In fact, 20% of all misclassifications in the first period were due to this error (Table 6). Some assignment of WN otoliths to SG was also foreseeable, although it was a very small component of the error (4%). This problem was caused by a small fraction of the WN release being marked with an aberrant pattern, which was thought to most probably result in misclassifying this WN mark to SG. The remaining errors involved combinations of SG-CC, CC-WN, AFK-SG, CC-AFK,WN-wild. These misclassification errors were not expected, and the readers will be instructed to bear the results in mind when identifying otoliths of returning adults in 1997.

The above three factors are relatively inconsequential with respect to our overall ability correctly identify thermal marks. The high quality of marks applied in 1995 and the ability of Otolith Laboratory staff to identify these marks with little error provides us with a high degree of confidence in being able to use this tool successfully for stock identification during the 1997 commercial fishery.

Assessment of Proposed Catch-Sampling Technique

The catch-sampling experiment showed that presence or absence of finclips did not influence the samplers selection of salmon being unloaded from tenders. The experiment did not address the possibility that the samplers might have made selections based on other traits, such as size. While conditions aboard tenders precluded doing experiments in which only large or small salmon were

finclipped, the experiment did show that the samplers were able to objectively sample a population consisting of two visually distinct classes of salmon (clipped and unclipped) without any indication that these samples were biased. We feel that since the samplers were unbiased regarding visually distinct classes, they are probably not going to be affected by more subtle differences, such as size, when sampling pink salmon.

The relatively small sample of six tenders, was still sufficiently large to allow rejection of the hypothesis of unbiased samples at α=0.03 (*p*-value associated with zero or six coverage events) The twelve trial analysis yielded a wider selection of possible *p*-values , but still failed to provide evidence suggesting that the null hypothesis should be rejected.

The utility using a digital watch mounted in a headset as a sampling device was another important finding from the finclip experiment. The digital watch worked extremely well, and the 'Count-Down' feature signal was audible over the loudest of the processor activities. The sampling device caused no physical discomfort to samplers, which is important if it is to be worn for prolonged periods each day.

The simulation study showed that the actual precision of estimates of hatchery proportions will likely be underestimated. The estimated proportion of hatchery salmon in a stratum catch will likely be within 0.08 (*d* value) of the population proportion 95% of the time with current sample sizes, random sampling, and a 50:50 partition of hatchery and wild salmon. For the 1996 fishery, departure from the 50:50 partition would have yielded a mean theoretical *d* value of about 0.06 with random sampling. The simulation showed that if, in addition to a departure from the worst-case 50:50 arrangement, marked and unmarked populations had exhibited some structure, a systematic sampling scheme would yield smaller *d* values. The average *d* value was almost half of the worst case theoretical value when the mixing factor was 0.3.

The calculated precision was always known to be for a worst-case 50:50 hatchery-wild event. The simulation study has revealed that the precision may be further underestimated when systematic sampling is used with ordered populations. Systematic sampling is easy to implement, guarantees a sample that is spatially uniform, and provides a low variance estimate (although estimation of the degree to which this low variance estimate occurs is debated in the literature). The technique can lead to erroneous estimates with respect to the variable being measured, if the sampling interval coincides with any periodicity within the sampled population. However, chances are remote for any delivery process to lead to pulses of hatchery and wild salmon at the same frequency at which they will be sampled.

CONCLUSIONS

The major objective of this project was to apply unique and distinct thermal marks to all pink salmon embryos produced in Prince William Sound hatcheries, and, by doing so, to allow the identification of hatchery salmon in mixed harvests of hatchery and wild salmon. Samples taken three weeks after marking indicated that unique and distinct marks had been applied in 1996. Results of a double blind test indicated that Otolith Laboratory readers had few problems successfully differentiating hatchery otoliths from those obtained from wild populations.. Field tests of the proposed sampling methodology indicated that accurate and precise estimates of hatchery contributions to mixed stock harvests could be obtained. Simulation studies further revealed that under certain conditions, the precision of estimates may be greater than expected.

ACKNOWLEDGMENTS

We would like to thank the following ADF&G Otolith Laboratory staff members for their valuable assistance in this project. Rick Berning supervised specimen processing and data entry phases of the project, including database design, barcode specimen labels, data entry screens, and process control that provided assurance of blindness and data integrity. Peter Hagan provided useful advice and suggestions concerning mark application and otolith processing. Becky Achten and George Koenig dissected otoliths and prepared slide specimens. Joseph Cashen, Dione Cuadra, and William Foster, Jr. processed otoliths, making mark determinations and providing insight regarding otolith patterns. We would also like to thank the following ADF&G Prince William Sound Area office staff member for their help. Renate Riffe and Jay Johnson provided help and advice for many portions of the field work for this project. Al Cox, Jim O'Rourke, Larry Gondek, Dave Waters, Josh Noah, Tim Torell and Ellen Simpson spent long hours sampling and marking slimy wriggling fish. We also thank Stephen Fried, Regional Research Biologist, who provided assistance over the course of this study and reviewed the manuscript.

LITERATURE CITED

Gauss. 1996. Aptech Systems Inc., Maple Valley Washington.

Habicht, C., Sharr, S. and J. Seeb. 1996. Coded Wire Tag Placement Affects Homing Precision of Pink Salmon. *In press.*

Hagan, P., Munk, K., Van Alen, B. and B. White. 1995. Thermal Mark Technology for Inseason Fisheries Management: A Case Study. Alaska Fishery Research Bulletin Vol. 2, No. 2.

Joyce, T., Evans, D. & R. Riffe. 1996. Otolith Marking of Pink Salmon in Prince William Sound Salmon Hatcheries, 1995. Exxon Valdez Oil Spill Restoration Project Annual Report

(Restoration Project R95320C), Alaska Department of Fish and Game, Commercial Fisheries Management and Development Division, Cordova, Alaska.

Mosegaard, H., N.G. Steffner, and B. Ragnarsson. 1987. Manipulation of otolith microstructure as a means of mass-marking salmonid yolk sac fry. Pages 213-220 in S.O.Kullander and B. Fernholm, editors. Proceedings: fifth congress of European ichthyologists. Swedish Museum of Natural History, Stockholm.

Munk, K., (*in* prep), Thermal Marking Manual: A Guideline to the Induction of Thermal Marks in Otoliths for the Purpose of Mass-Marking Hatchery Stocks. Alaska Department of Fish and Game, Special Publication, Juneau.

Munk, K., W. Smoker, Beard, D. and R. Mattson. 1993. A Hatchery Water-Heating System and its Application to 100% Thermal Marking of Incubating Salmon. The Progressive Fish-Culturist. 55:284-288.

Riffe, R., Gehlbach, S., Evans, D. and B. Bue. 1996. Coded wire tag recoveries from Prince William Sound salmon fisheries, 1995. Exxon Valdez Oil Spill Restoration Project Annual Report (Restoration Project R95320B), Alaska Department of Fish and Game, Commercial Fisheries Management and Development Division, Cordova, Alaska.

SAS, 1996. SAS Institute Inc., SAS Campus Drive, Cary, North Carolina.

Sharr, S., Bue, B., Moffit, S., Craig, A. and D. Evans. 1994. Injury to salmon eggs and pre-emergent fry in Prince William Sound. Exxon Valdez Oil Spill State/Federal Natural Resource Damage Assessment Project Final Report (Fish/Shellfish Study #2), Alaska Department of Fish and Game, Commercial Fisheries Management and Development Division, Cordova, Alaska.

Sharr, S., Peckham, C., Sharp, D., Peltz, L., Smith, J., Willette, T., Evans, D. and B. Bue. 1995a. Coded wire tag studies on Prince William Sound salmon, 1989-1991. Exxon Valdez Oil Spill State/Federal Natural Resource Damage Assessment Project Final Report (Fish/Shellfish Study #3), Alaska Department of Fish and Game, Commercial Fisheries Management and Development Division, Cordova, Alaska.

Sharr, S., Peckham, C., Sharp, D., Smith, J., Evans, D., and B. Bue. 1995b. Coded wire tag studies on Prince William Sound salmon, 1992. Exxon Valdez Oil Spill Restoration Project Final Report (Restoration Project R60A), Alaska Department of Fish and Game, Commercial Fisheries Management and Development Division, Cordova, Alaska.

Sharr, S., Peckham, C., Sharp, D., Evans, D. and B. Bue. 1995c. Coded wire tag recoveries from Prince William Sound salmon, 1993. Exxon Valdez Oil Spill Restoration Project Final Report (Restoration Project R93067), Alaska Department of Fish and Game, Commercial Fisheries Management and Development Division, Cordova, Alaska.

Sharr, S., Peckham, C., Sharp, D., Smith, J., Willette, M., Evans, D., Bue, B., Gehlbach, S. and R. Riffe. 1995d. Stock identification of chum, sockeye, chinook, and coho salmon in Prince William Sound salmon. Exxon Valdez Oil Spill Restoration Project Final Report (Restoration Projects R93068/R94137/R95137), Alaska Department of Fish and Game, Commercial Fisheries Management and Development Division, Cordova, Alaska.

Sharr, S., Riffe, R., Gehlbach, S., Evans, D. and B. Bue. 1995e. Coded wire tag recoveries from Prince William Sound salmon, 1994. Exxon Valdez Oil Spill Restoration Project Annual Report (Restoration Project R94320B), Alaska Department of Fish and Game, Commercial Fisheries Management and Development Division, Cordova, Alaska.

Volk, E.C., S.L. Schroder, and K.L. Fresh. 1990. Inducement of unique otolith banding patterns as a practical means to mass-mark juvenile Pacific salmon. American Fisheries Society Symposium 7: 203-215

Willette, T.M. and G. Carpenter. 1994. Early marine salmon injury assessment in Prince William Sound. Exxon Valdez Oil Spill State/Federal Natural Resource Damage Assessment Project Final Report (Fish/Shellfish Study Number 4), Alaska Department of Fish and Game, Commercial Fisheries Management and Development Division, Cordova, Alaska.

APPENDICES

Appendix A Success rates of readers by event when determinations are on a hatchery-wild basis only[a].

Reader	Event 1	2	3	4	Mean
A	0.988	0.993	0.995	0.998	0.993
B	0.997	0.993	0.995	0.9988	0.995
C	0.998	1.000	0.995	0.995	0.997
D	0.999	1.000	1.000	0.995	0.999
Mean	0.995	0.996	0.996	0.996	0.996

a Means may not be reproducible from the table due to rounding.

Appendix B Identification matrices by reader for the first event[a].

		OBSERVED				
Reader A		AFK	CC	SG	WN	WILD
	AFK	143	0	1	0	0
	CC	2	123	0	12	0
TRUE	SG	0	5	156	0	0
	WN	0	2	3	157	1
	WILD	9	4	0	0	581
Reader B		AFK	CC	SG	WN	WILD
	AFK	143	0	1	0	0
	CC	0	135	0	2	0
TRUE	SG	0	5	156	0	0
	WN	0	1	0	161	1
	WILD	3	0	0	0	591
Reader C		AFK	CC	SG	WN	WILD
	AFK	144	0	0	0	0
	CC	0	135	0	2	0
TRUE	SG	0	5	156	0	0
	WN	0	0	0	162	1
	WILD	2	0	0	0	592
Reader D		AFK	CC	SG	WN	WILD
	AFK	144	0	0	0	0
	CC	0	136	0	1	0
TRUE	SG	0	5	156	0	0
	WN	0	0	0	162	1
	WILD	0	0	0	0	594

a AFK=A.F. Koernig; CC=Cannery Creek; SG=Solomon Gulch; WN=W.H. Noerenberg.

Appendix C Individual sample estimates for finclip experiment.

Tender	Load (#fish)	# Marks	Mark Rate	Sample Estimates
1	18,070	1,363	0.075	0.091 0.099 0.072 0.094
2	15,174	2,047	0.135	0.161 0.127 0.113 0.158
3	33,568	2,769	0.082	0.067 0.077 0.068 0.095
4	25,935	2,159	0.083	0.084 0.073 0.073 0.056
5	20,777	1,903	0.092	0.106 0.104 0.081 0.089
6	25,628	2,226	0.087	0.072 0.088 0.094 0.101

Appendix D Population and sample means for the finclip experiment.