

Estimating salmon harvest with coded-wire tags

David R. Bernard and John E. Clark

Abstract: We present a simple, comprehensive method to estimate harvest from one or more hatchery-produced or wild cohorts of Pacific salmon (*Oncorhynchus* spp.) caught simultaneously in one or more stratified commercial or recreational fisheries. The estimator is based on the return of coded-wire tags as modeled with multivariate compound probability distributions for catch sampling programs in which some samples are lost and some tags are not decoded. Knowledge of catches in strata and of tagging rates of cohorts need not be exact as in previously developed methods. Examples concerning historical and hypothetical commercial and recreational fisheries exploiting four cohorts of chinook (*Oncorhynchus tshawytscha*) and coho salmon (*Oncorhynchus kisutch*) are provided. Estimated covariances among cohorts proved to be negligible to the point of nonexistence for estimated harvests in commercial fisheries, but could be large and positive for estimated harvests in recreational fisheries. Normal approximations of confidence intervals were almost identical to confidence intervals developed with the parametric bootstrap for estimates in commercial fisheries and were less similar when significant harvest came from recreational fisheries.

Résumé : Nous présentons une méthode simple et complète pour estimer l'exploitation d'une ou de plusieurs cohortes de saumons du Pacifique (*Oncorhynchus* spp.), sauvages ou d'élevage, les prises se faisant simultanément dans une ou plusieurs pêcheries commerciales ou récréatives stratifiées. L'estimateur est fondé sur les retours de micromarques magnétisées codées et se conforme à la loi de probabilité composée à plusieurs variables pour les programmes d'échantillonnage des prises où certains échantillons sont perdus et où certaines marques ne sont pas décodées. On doit connaître avec exactitude les prises par strate et les taux de marquage des cohortes comme c'était le cas avec les méthodes développées antérieurement. Nous fournissons des exemples de pêches commerciales et récréatives, historiques et théoriques, exploitant quatre cohortes de saumon quinnat (*Oncorhynchus tshawytscha*) et coho (*Oncorhynchus kisutch*). Il s'avère que la covariance entre les cohortes est négligeable, au point de ne pas exister dans le cas des estimations des prises commerciales; elle pourrait cependant être positive et importante dans le cas des pêches récréatives. Les approximations normales des intervalles de confiance sont presque identiques aux intervalles de confiance déterminés par la méthode paramétrique de bootstrap appliquée aux estimations des prises commerciales; cela se vérifie moins lorsqu'une part importante de la récolte provient des pêches récréatives.

[Traduit par la Rédaction]

Introduction

Information on specific groups (cohorts) of Pacific salmon (*Oncorhynchus* spp.) exploited in fisheries along the northwest coast of North America is gathered through recovery of coded-wire tags (CWTs) (Johnson 1990; Lapi et al. 1990). A CWT is a 1 mm long stainless steel wire that is injected into the nasal cartilage of a smolt or fingerling. A coded, binary number unique to a cohort of salmon is microscopically etched on each CWT. A subset of a cohort is tagged, with each tagged salmon externally marked by excising its adipose fin. Each year, samples of adult salmon caught in fisheries are inspected to find fish without adipose fins. Whenever possible, heads from these recaptured salmon are retrieved and sent to laboratories. Heads that arrive at laboratories are passed through a magnetometer to detect a CWT and dissected if the presence of metal is

indicated. If a CWT is found and the tag is undamaged, its code is read under a microscope.

In analysis of information on recovered CWTs, harvest from a cohort is either fixed or a variable, depending on the intended use of the result. Harvest is a variable when the process of a fish surviving and being caught in a fishery is of interest, as is the case when describing spatial and temporal patterns of harvests in fisheries, when estimating survival rates of smolts, or when testing hypotheses concerning the efficacy of hatchery practices. Log-linear models of recovered CWTs as proposed in Green and Macdonald (1987) and Cormack and Skalski (1992) encompass variation from processes such as survival and exploitation plus variation from sampling. In these models, the observed pattern of recovered CWTs is considered to be a single realization of an infinite number of such patterns. In contrast, harvest from a cohort is fixed whenever it is treated as a finite quantity, as is done in stock-recruit analyses or when allocating harvests among users. For example, catches of salmon limited by treaty in provincial and state fisheries can be increased by the harvest of salmon produced in local hatcheries (Pacific Salmon Treaty 1985, annex IV, chapter 3, paragraph 2). In this exercise, process variation is irrelevant to estimates of harvest; sampling is considered as the sole source of variation in the estimation. Because there is no process variation in a finite quantity, log-linear models are inappropriate for estimating a fixed harvest. Clark and Bernard (1987) investigated bias in the intuitive estimator for a fixed

Received July 14, 1995. Accepted March 29, 1996.
J13000

D.R. Bernard.¹ Sport Fish Division, Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK 99518-1599, U.S.A.

J.E. Clark. Commercial Fisheries Management and Development Division, Alaska Department of Fish and Game, 802 3rd Street, Douglas, AK 99824-0200, U.S.A.

¹ Author to whom all correspondence should be addressed.
e-mail: davidbe@fishgame.state.ak.us

harvest with compound binomial-hypergeometric probability distributions and developed an unbiased estimate of variance. Geiger (1990), Newman (1990), and Schnute (1992) further developed variations on this estimator and its estimated variance from similar probability distributions. In their developments, catch (N) and the fraction of the cohort with CWTs (θ) are presumed known, a situation common to hatchery-produced salmon harvested in commercial fisheries.

In this paper, we present a simple, comprehensive method for estimating fixed harvests with multivariate compound probability distributions describing the recovery of CWTs from cohorts of salmon. Our estimator is appropriate if catch and the probability that a member of the cohort carries a CWT are both known or are estimated with significant error, as is often so for recreational fisheries and for cohorts of wild salmon. The estimator encompasses temporal and spatial stratification in several simultaneous fisheries exploiting one or more cohorts. Examples concern real and hypothetical fisheries exploiting four cohorts of chinook (*Oncorhynchus tshawytscha*) and coho salmon (*Oncorhynchus kisutch*). Normal approximations of confidence intervals are compared with approximate intervals generated with a parametric bootstrap of data for each example.

Method

Stratification

Total harvest (T) from one or more cohorts in one or more fisheries is the sum of harvests from individual cohorts and strata (r_{ij}):

$$(1) \quad T = \sum_i \sum_j r_{ij}$$

where individual strata (i) may be defined as different fisheries, fishing periods, or fishing districts, and cohorts (j) are combined according to need. The intuitive estimator for r_{ij} , the harvest of cohort j in stratum i , is the number of CWTs from cohort j recovered in stratum i (m_{ij}) expanded by rates at which the catch is sampled (ϕ_i), heads from marked fish in the sample reach the laboratory (a'_i/a_i), and detected CWTs are successfully decoded (t'_i/t_i), and by the fraction of the cohort with CWTs (θ_j):

$$(2) \quad \hat{r}_{ij} = \frac{a_i t'_i m_{ij}}{a'_i t'_i \phi_i \theta_j} = \frac{m_{ij}}{\lambda_i \phi_i \theta_j}$$

where a_i and a'_i are the number of sampled fish missing adipose fins and the number of their heads that reach a laboratory, and where t_i and t'_i are the number of CWTs detected at a laboratory and the number of these tags decoded (this and all subsequent notation about the estimator are defined in Table 1). Because samples are independently drawn among strata, variance of \hat{T} conditional on all r_{ij} is

$$(3) \quad \text{Var}(\hat{T}|\{r\}) = \sum_i \sum_j \text{Var}(\hat{r}_{ij}|r_{ij}) + 2 \sum_i \sum_j \sum_{k>j} \text{Cov}(\hat{r}_{ij}, \hat{r}_{ik}|r_{ij}, r_{ik})$$

where j and k denote different cohorts. Covariances between estimated harvests for two cohorts are included here because

Table 1. Notation used to describe parameters involved in estimators of harvest from cohorts of salmon based on the return of coded-wire tags (CWTs).

Symbol	Definition
a_i	Number of marked fish in the sample from stratum i
a'_i	Subset of a_i for which heads reach a laboratory
λ_i	Decoding rate of CWTs for marked fish in the sample from stratum i
m_{ij}	Number of decoded CWTs from cohort j in the sample from stratum i
M_j	Number of juvenile salmon in cohort j at their release
n_i	Size of sample taken from the catch in stratum i
N_i	Size of the catch in stratum i
p_{ij}	Probability that a fish caught in stratum i has a CWT from cohort j
ϕ_i	Fraction of the catch inspected (sampled) for stratum i
r_{ij}	Number of fish harvested from cohort j in stratum i
R_j	Subset of M_j juvenile salmon released with CWTs
t_i	Number of heads sampled from stratum i with CWTs detected magnetically at a laboratory
t'_i	Subset of t_i for which CWTs are decoded
$T_{(i)}$	Harvest from several cohorts during a fishing season (or in stratum i)
θ_j	Probability of a fish in cohort j having a CWT
x_{ij}	Number of fish from cohort j with CWTs caught in stratum i
y_{ij}	Subset of x_{ij} in the sample from stratum i
z_{ij}	Subset of y_{ij} for which heads arrive at a laboratory

both estimates are calculated from the same sample within a stratum, making them dependent.

Strata are usually defined to follow fishing periods and fishing districts, or to promote random sampling. Catch sampling that begins at a dock or tender and ends under a dissecting microscope is not a random process. Samplers cannot access all of the catch at a particular time, but sample an unsteady stream of fish. Estimates of T and its variance obtained through crude or no stratification will be unbiased only if sample size is proportional to catch or if cohorts are dispersed throughout the catch. Proportional sampling is difficult to achieve because catch sampling can be density dependent, and recoveries of CWTs from a cohort are clustered in space and time as the cohort migrates through fisheries. However, cohorts are more likely to be dispersed within a stratum of limited duration and geography. Cormack and Skalski (1992) present the model $E(m_{ij}) = R_j \phi_i \psi_{ij}$ in which R_j is the number of tagged salmon in the cohort when released and ψ_{ij} is a log-linear function of factors and their interactions, factors that include stratification. The m_{ij} follow a scaled Poisson distribution such that $\text{Var}(m_{ij}) = \rho E(m_{ij})$. In a fit of the most complex model to data (m_{ij}) from cohorts with replicated codes, $\rho \leq 1$ indicates that cohorts are completely dispersed within strata whereas $\rho > 1$ indicates over dispersion and clustering within strata.

For some strata and for some cohorts, all uncertainty in \hat{r}_{ij} comes from catch sampling, but in other instances, additional uncertainty arises from estimation of catch and the fraction of a cohort tagged. For hatchery-produced cohorts exploited in commercial fisheries, θ_j and N_i are often estimated with so little uncertainty that variances for their estimates can be ignored without meaningfully biasing statistics. In recreational

fisheries, sampling rate of the catch ϕ_i is often estimated with considerable uncertainty because the catch itself is estimated with relatively large variance. Catch is not tallied as in commercial fisheries, but estimated by interviewing a sample of anglers during on-site creel surveys, with postal surveys, or through logbooks (see Guthrie et al. 1991; Pollock et al. 1994). When cohorts represent wild stocks of salmon, the fraction marked is often low and estimated with considerable error because capturing fingerlings, smolts, or adults is difficult and expensive. Whenever θ_j or N_i (and subsequently ϕ_i) is estimated with relatively large variance, $\text{Var}(\hat{r}_{ij}|r_{ij})$ is a function of $\text{Var}(m_{ij}|r_{ij})$ plus $\text{Var}(\hat{N}_i)$, $\text{Var}(\hat{\theta}_j)$, or both. In the subsequent development of our estimator, we label N_i and θ_j as being "known" when estimated with negligible variance and as "estimated" when estimated with large variance.

Catch and fraction tagged are known

The first step in developing an unbiased estimator for harvest of two or more cohorts in a stratum concerns tagged salmon in the catch. The probability of x_{ij} tagged fish being caught in a stratum (fishery) follows the hypergeometric distribution $\text{Hyper}(x_{ij}|r_{ij}, M_j, R_j)$ where M_j is the number of juveniles in the cohort when released and R_j is the number of these juveniles that carried CWTs. Because M_j is usually known with considerable uncertainty for both hatchery-produced and wild cohorts, $\text{Hyper}(x_{ij}|r_{ij}, M_j, R_j)$ is not a useful beginning for developing an estimator of harvest. Clark and Bernard (1987), Geiger (1990), Schnute (1992), and we start with the binomial approximation $\text{Binom}(x_{ij}|r_{ij}, \theta_j)$ with its fewer and easier to estimate parameters. Because of low survival rates for juvenile salmon and less than complete exploitation of the survivors in multiple fisheries, $r_{ij} \ll M_j$, which makes $\text{Binom}(x_{ij}|r_{ij}, \theta_j)$ a virtually unbiased approximation to $\text{Hyper}(x_{ij}|r_{ij}, M_j, R_j)$.

The next step in the development concerns the number of tagged fish in a sample from a catch. If catch sampling acts like a random process, the joint probability of y_{ij} and y_{ik} tagged fish from two cohorts being sampled follows the joint hypergeometric distribution $\text{Hyper}(y_{ij}, y_{ik}|n_i, N_i, x_{ij}, x_{ik})$, where n_i is the sample size. Expression of this joint probability conditioned on harvests from both cohorts is

$$\text{Prob}(y_{ij}, y_{ik}|r_{ij}, r_{ik}) = \sum_{x_{ij}} \sum_{x_{ik}} \text{Binom}(x_{ij}|r_{ij}, \theta_j) \text{Binom}(x_{ik}|r_{ik}, \theta_k) \\ \times \text{Hyper}(y_{ij}, y_{ik}|n_i, N_i, x_{ij}, x_{ik})$$

This is the multivariate form of the compound probability distributions in Geiger (1990) and Newman (1990). If all fish heads sent to the decoding laboratory arrive, and if all CWTs detected at the laboratory are decoded, this joint distribution is a sufficient expression of probability for recovered tags. However, some heads may not reach the laboratory, and some tags may not be decoded. The joint probability distributions for the z_{ij} and z_{ik} heads in a sample that reach the laboratory and the m_{ij} and m_{ik} tags successfully decoded from these heads are $\text{Hyper}(z_{ij}, z_{ik}|a'_i, a_i, y_{ij}, y_{ik})$ and $\text{Hyper}(m_{ij}, m_{ik}|t'_i, t_i, z_{ij}, z_{ik})$. If some heads are lost or some detected CWTs are not decoded, the joint probability distribution of observing m_{ij} and m_{ik} tags in a sample is

$$\text{Prob}(m_{ij}, m_{ik}|r_{ij}, r_{ik}) = \sum_{y_{ij}} \sum_{y_{ik}} \sum_{z_{ij}} \sum_{z_{ik}} \text{Prob}(y_{ij}, y_{ik}|r_{ij}, r_{ik}) \\ \times \text{Hyper}(z_{ij}, z_{ik}|a'_i, a_i, y_{ij}, y_{ik}) \text{Hyper}(m_{ij}, m_{ik}|t'_i, t_i, z_{ij}, z_{ik}).$$

From this multivariate compound probability distribution, Clark and Bernard (1987) showed the intuitive estimate of harvest $\hat{r}_{ij} (= m_{ij}(\lambda_i \phi_i \theta_j)^{-1})$ to be unbiased (for reasons explained in Geiger (1990)) and unbiased estimates of variance and covariance with a second cohort to be

$$(4) \quad \hat{\text{Var}}(\hat{r}_{ij}|r_{ij}) = \left(\frac{1}{\lambda_i \phi_i \theta_j} - C_i \right) \frac{\hat{r}_{ij}}{C_i} - (1 - C_i) \frac{\hat{r}_{ij}^2}{C_i}$$

$$(5) \quad \hat{\text{Cov}}(\hat{r}_{ij}, \hat{r}_{ik}|r_{ij}, r_{ik}) = -(1 - C_i) \frac{\hat{r}_{ij} \hat{r}_{ik}}{C_i}$$

$$\text{where } C_i = \frac{t_i(t'_i - 1)a_i(a'_i - 1)N_i(n_i - 1)}{t'_i(t_i - 1)a'_i(a_i - 1)n_i(N_i - 1)}, \lambda_i = \frac{t'_i a'_i}{t_i a_i}, \text{ and } \phi_i = \frac{n_i}{N_i}.$$

Under even relatively small sample sizes, $C_i \approx 1$, and if treated as equal to one, covariances become zero, and unbiased estimates of variance can be simplified to a large-sample approximation:

$$(6) \quad \hat{\text{Var}}_{\text{LS}}(\hat{r}_{ij}|r_{ij}) = \frac{\hat{r}_{ij}}{\lambda_i \phi_i \theta_j} (1 - \lambda_i \phi_i \theta_j)$$

Note that if $m_{ij} = 1$, the unbiased estimate of variance and its large-sample approximation are equivalent. If $\lambda_i = 1$, our large-sample approximation is the same formulation developed by Geiger (1990, eq. 6) and by Schnute (1992, eq. 4.18). The term $1 - \lambda_i \phi_i \theta_j$ in eq. 6 acts as a correction for sampling from a finite population (an fpc).

Catch and fraction tagged are estimated

Our development of a comprehensive estimator for harvest from wild cohorts or in recreational fisheries begins with the estimated fraction \hat{p}_{ij} of tagged salmon in a catch from a cohort. The actual fraction $p_{ij} (= x_{ij}/N_i)$ has an intuitive estimate $\hat{p}_{ij} (= m_{ij}(\lambda_i n_i)^{-1})$ such that

$$(7) \quad \hat{r}_{ij} = N_i \left(\frac{m_{ij}}{\lambda_i n_i} \right) \theta_j^{-1} = N_i \hat{p}_{ij} \theta_j^{-1}$$

In a catch sampling program, the joint probability of y_{ij} and y_{ik} tagged fish from two cohorts being sampled follows a multinomial distribution $\text{Multinom}(y_{ij}, y_{ik}|n_i, p_{ij}, p_{ik})$ if p_{ij} and p_{ik} can be considered fixed. Note that knowledge of N_i , θ_j , and θ_k is not needed in this formulation. Remembering that $\text{Hyper}(z_{ij}, z_{ik}|a'_i, a_i, y_{ij}, y_{ik})$ and $\text{Hyper}(m_{ij}, m_{ik}|t'_i, t_i, z_{ij}, z_{ik})$ are the joint probability distributions for fish heads that reach the laboratory and for detected tags successfully decoded from these heads, the joint probability distribution of observing m_{ij} and m_{ik} tags in a catch sample becomes

$$\text{Prob}(m_{ij}, m_{ik}|n_i, p_{ij}, p_{ik}) = \sum_{y_{ij}} \sum_{y_{ik}} \sum_{z_{ij}} \sum_{z_{ik}} \text{Multinom}(y_{ij}, y_{ik}|n_i, p_{ij}, p_{ik}) \\ \times \text{Hyper}(z_{ij}, z_{ik}|a'_i, a_i, y_{ij}, y_{ik}) \text{Hyper}(m_{ij}, m_{ik}|t'_i, t_i, z_{ij}, z_{ik})$$

Joint factorial moments of $\text{Prob}(m_{ij}, m_{ik}|n_i, p_{ij}, p_{ik})$ were used to show that the intuitive estimate \hat{p}_{ij} is unbiased, and to derive unbiased estimates of variance for \hat{p}_{ij} and covariance with \hat{p}_{ik} (assuming p_{ij} and p_{ik} are fixed). Detailed descriptions of these

derivations can be obtained from the senior author upon request. Estimated variances and covariances are

$$(8) \quad \hat{\text{Var}}(\hat{p}_{ij}|n_i, p_{ij}) = \left(\frac{1}{D_i} \frac{n_i}{(n_i - 1)} \right) \times \left[\left(\frac{1}{\lambda_i n_i} \right) \hat{p}_{ij} - \left(1 - D_i \frac{(n_i - 1)}{n_i} \right) \hat{p}_{ij}^2 \right]$$

$$(9) \quad \hat{\text{Cov}}(\hat{p}_{ij}, \hat{p}_{ik}|n_i, p_{ij}, p_{ik}) = - \left(\frac{1}{D_i} \frac{n_i}{(n_i - 1)} \right) \times \left(1 - D_i \frac{(n_i - 1)}{n_i} \right) \hat{p}_{ij} \hat{p}_{ik}$$

where $D_i = \frac{t_i(t'_i - 1)a_i(a'_i - 1)}{t'_i(t_i - 1)a'_i(a_i - 1)}$. However, x_{ij} and hence p_{ij} are

not fixed, but are variables as noted earlier, with the result that $\hat{\text{Var}}(\hat{p}_{ij}|n_i, p_{ij})$ and $\hat{\text{Cov}}(\hat{p}_{ij}, \hat{p}_{ik}|n_i, p_{ij}, p_{ik})$ are uncorrected for sampling from a finite catch. An exact correction can be found by solving the relationship $\hat{\text{Var}}(\hat{r}_{ij}|r_{ij}) = N_i^2 \theta_j^{-2} \hat{\text{Var}}(\hat{p}_{ij}|n_i, p_{ij})$ (fpc) for fpc. The unbiased estimate of variance for \hat{p}_{ij} conditioned on harvest and accurately corrected for sampling from a finite population is

$$(10) \quad \hat{\text{Var}}(\hat{p}_{ij}|r_{ij}) = \hat{\text{Var}}(\hat{p}_{ij}|n_i, p_{ij}) \times \left[\frac{1 - m_{ij}(1 - C_i)}{E_i - m_{ij}(E_i - C_i)} - \lambda_i \phi_i \theta_j \frac{C_i}{E_i - m_{ij}(E_i - C_i)} \right]$$

where $E_i = \frac{N_i}{N_i - 1}$ and the relationship in square brackets is the

fpc. Because of the way in which it was developed, this fpc is the exact correction in variance for using $\text{Binom}(y_{ij}, y_{ik}|n_i, p_{ij}, p_{ik})$ instead of $\text{Hyper}(y_{ij}, y_{ik}|n_i, N_i, x_{ij}, x_{ik})$ in development of the estimator. In the same fashion, the corrected unbiased estimate of covariance is

$$(11) \quad \hat{\text{Cov}}(\hat{p}_{ij}, \hat{p}_{ik}|r_{ij}, r_{ik}) = \hat{\text{Cov}}(\hat{p}_{ij}, \hat{p}_{ik}|n_i, p_{ij}, p_{ik}) \left(\frac{1 - C_i}{E_i - C_i} \right)$$

Because C_i , D_i , and $E_i \equiv 1$ in most instances, the large-sample approximation to $\hat{\text{Var}}(\hat{p}_{ij}|r_{ij})$ is

$$(12) \quad \hat{\text{Var}}_{\text{LS}}(\hat{p}_{ij}|r_{ij}) = \frac{\hat{p}_{ij}}{\lambda_i n_i} (1 - \lambda_i \phi_i \theta_j)$$

Note again that when $m_{ij} = 1$, the unbiased estimate of variance and its large-sample approximation are equivalent. Also, note that the large-sample approximation for estimated covariance between \hat{p}_{ij} and \hat{p}_{ik} is 0.

For recreational fisheries in which N_i is estimated with significant error and for wild cohorts for which θ_j is known imprecisely, \hat{N}_i can be substituted for N_i and $\hat{\theta}_j^{-1}$ for θ_j^{-1} in the intuitive estimator to calculate $\hat{r}_{ij} (= \hat{N}_i \hat{p}_{ij} \hat{\theta}_j^{-1})$. Because estimates \hat{N}_i , \hat{p}_{ij} , and $\hat{\theta}_j^{-1}$ come from independent sampling programs, they are independent variates whose product has the unbiased estimate for variance (from Goodman (1960)):

$$(13) \quad \hat{\text{Var}}(\hat{r}_{ij}|r_{ij}) = \hat{r}_{ij}^2 [G(\hat{p}_{ij}) + G(\hat{N}_i) + G(\hat{\theta}_j^{-1}) - G(\hat{\theta}_j^{-1}) G(\hat{N}_i) - G(\hat{\theta}_j^{-1}) G(\hat{p}_{ij}) - G(\hat{N}_i) G(\hat{p}_{ij}) + G(\hat{\theta}_j^{-1}) G(\hat{N}_i) G(\hat{p}_{ij})]$$

where $G(\cdot)$ is the estimated squared coefficient of variation for the specified estimate. The expression $\hat{r}_{ij} \hat{r}_{ik} [G(\hat{N}_i) + G(\hat{p}_{ij}, \hat{p}_{ik}) - G(\hat{N}_i) G(\hat{p}_{ij}, \hat{p}_{ik})]$ is an unbiased estimate of covariance between \hat{r}_{ij} and \hat{r}_{ik} and was derived by plugging the equation above into the expression of $\text{Var}(\hat{T})$ (eq. 2), taking expectations, and rearranging the result. A detailed description of this derivation can be obtained from the senior author upon request. Noting that $G(\hat{p}_{ij}, \hat{p}_{ik}) = \text{Cov}(\hat{p}_{ij}, \hat{p}_{ik}|r_{ij}, r_{ik}) \hat{p}_{ij}^{-1} \hat{p}_{ik}^{-1}$ and remembering that the large-sample approximation for $G(\hat{p}_{ij}, \hat{p}_{ik}) = 0$, the large-sample approximation to covariance between \hat{r}_{ij} and \hat{r}_{ik} becomes

$$(14) \quad \hat{\text{Cov}}_{\text{LS}}(\hat{r}_{ij}, \hat{r}_{ik}|r_{ij}, r_{ik}) = \hat{r}_{ij} \hat{r}_{ik} G(\hat{N}_i)$$

This completes our development of a comprehensive, unbiased estimator for a fixed harvest of salmon based on CWTs. Listed in Table 2 is a series of formulations of the estimator and large-sample approximations of variances to fit common situations. In these formulations, θ_j was presumed known for hatchery-produced fish and estimated with error for wild cohorts. Similarly, catch was presumed known for commercial and estimated in recreational fisheries. When these presumptions do not hold, the formulations are still correct, only the labelling needs to be changed to fit the situation. Measurement error in estimated variances from plugging-in estimates for ϕ_i and θ_j in fpcs will be negligible whenever sampling, tagging, or decoding rates are low and will become relatively unimportant as variances are summed across strata. Large-sample approximations were used in the formulations in Table 2 because their simplicity and almost nonexistent bias, as is evident in our examples, make them the preferred method for estimating variance.

Examples

Harvests in 1989 from four cohorts of chinook and coho salmon in commercial troll and recreational fisheries were chosen as examples. Two cohorts of chinook salmon were released in 1986, one from Whitman Lake and one from Neets Bay hatcheries near Ketchikan, Alaska, with 44.86 and 5.04% carrying CWTs, respectively. Tags for each cohort of chinook salmon carried one of five codes to distinguish replicated subsets within each cohort. Two cohorts of coho salmon were released in 1988 from the same two hatcheries with 17.68 and 1.81% carrying CWTs, with only one code representing the cohort released from Neets Bay Hatchery. Two hypothetical fisheries for coho salmon were developed to demonstrate effects of $\lambda_i \ll 1$ and sample size on unbiased estimates of variance and their large-sample approximations. Information on both cohorts of coho salmon described above was used to create both hypothetical examples. Statistics from catch sampling programs are listed by fishing period (stratum) in Table 3 along with the number of recovered CWTs by cohort. Each stratum is considered a separate example.

Results from fitting log-linear models to recoveries are consistent with cohorts being dispersed within strata as defined in our historical examples of commercial fisheries. In the fit of the most complex log-linear model ($\log(\psi_{ij}) = \alpha_i + \tau_j + (\alpha\tau)_{ij}$ with α and τ representing strata and cohorts, respectively) to recoveries of CWTs from chinook salmon in the SE quadrant, $\rho = 0.75 (= \chi^2 / \text{df})$. Normalized, transformed residuals (Anscombe's transformation) form a relatively straight line (Fig. 1), which is further evidence of cohorts having been

Table 2. Formulations to estimate harvest of salmon from cohorts in a stratum with some members having CWTs for four common situations.

Situation	Formulations
Hatchery-produced salmon in a commercial fishery ($G(\hat{N}_i) \equiv 0, G(\hat{\theta}_j^{-1}) \equiv 0$)	$\hat{r}_{ij} = N_i \left(\frac{m_{ij}}{\lambda_i n_i} \right) \theta_j^{-1}$ $G(\hat{p}_{ij}) = \frac{(1 - \lambda_i \phi_i \theta_j)}{m_{ij}}$ $\hat{\text{Var}}_{\text{LS}}(\hat{T}_i) = \sum_j \hat{r}_{ij}^2 G(\hat{p}_{ij})$
Hatchery-produced salmon in a recreational fishery ($G(\hat{N}_i) > 0, G(\hat{\theta}_j^{-1}) \equiv 0$)	$\hat{r}_{ij} = \hat{N}_i \left(\frac{m_{ij}}{\lambda_i n_i} \right) \theta_j^{-1}$ $G(\hat{p}_{ij}) = \frac{(1 - \lambda_i \phi_i \theta_j)}{m_{ij}}$ $\hat{\phi}_i = \frac{n_i}{\hat{N}_i}$ $\hat{\text{Var}}_{\text{LS}}(\hat{T}_i) = \sum_j \hat{r}_{ij}^2 [G(\hat{p}_{ij}) + G(\hat{N}_i) - G(\hat{p}_{ij}) G(\hat{N}_i)] + 2 \sum_j \sum_{k>j} \hat{r}_{ij} \hat{r}_{ik} G(\hat{N}_i)$
Wild cohort in a commercial fishery ($G(\hat{N}_i) \equiv 0, G(\hat{\theta}_j^{-1}) > 0$)	$\hat{r}_{ij} = N_i \left(\frac{m_{ij}}{\lambda_i n_i} \right) \hat{\theta}_j^{-1}$ $G(\hat{p}_{ij}) = \frac{(1 - \lambda_i \phi_i \hat{\theta}_j)}{m_{ij}}$ $\hat{\text{Var}}_{\text{LS}}(\hat{T}_i) = \sum_j \hat{r}_{ij}^2 [G(\hat{p}_{ij}) + G(\hat{\theta}_j^{-1}) - G(\hat{p}_{ij}) G(\hat{\theta}_j^{-1})]$
Wild cohort in a recreational fishery ($G(\hat{N}_i) > 0, G(\hat{\theta}_j^{-1}) > 0$)	$\hat{r}_{ij} = \hat{N}_i \left(\frac{m_{ij}}{\lambda_i n_i} \right) \hat{\theta}_j^{-1}$ $G(\hat{p}_{ij}) = \frac{(1 - \lambda_i \phi_i \hat{\theta}_j)}{m_{ij}}$ $\hat{\phi}_i = \frac{n_i}{\hat{N}_i}$ $\hat{\text{Var}}_{\text{LS}}(\hat{T}_i) = \sum_j \hat{r}_{ij}^2 [G(\hat{p}_{ij}) + G(\hat{N}_i) + G(\hat{\theta}_j^{-1}) - G(\hat{p}_{ij}) G(\hat{N}_i) - G(\hat{p}_{ij}) G(\hat{\theta}_j^{-1}) - G(\hat{N}_i) G(\hat{\theta}_j^{-1}) + G(\hat{p}_{ij}) G(\hat{N}_i) G(\hat{\theta}_j^{-1})] + 2 \sum_j \sum_{k>j} \hat{r}_{ij} \hat{r}_{ik} G(\hat{N}_i)$

Note: Estimated variances are large-sample approximations. Note that $\hat{\text{Var}}_{\text{LS}}(\hat{T}) = \sum_i \hat{\text{Var}}_{\text{LS}}(\hat{T}_i)$.

dispersed within strata (Cormack and Skalski 1992). Although similar analysis is not possible for cohorts of coho salmon released in 1988 (no replicate codes were used at Neets Bay Hatchery), Cormack and Skalski (1992) found no evidence of clustering in recoveries from three cohorts of coho salmon released in 1984 from Whitman Lake Hatchery and harvested in troll fisheries a year later.

Bootstrap confidence intervals

Sampling distributions (F) and 95% confidence intervals for \hat{r}_{ij} and \hat{T} were approximated for our examples through parametric and nonparametric bootstrapping (see Efron and Tibshirani (1993) as source of basic methods). For commercial fisheries, each bootstrap sample of size n_i was drawn without replacement from the N_i records of caught salmon tallied according to their membership in one of seven groups (Table 4). Marked and tagged fish "inspected" in the bootstrap sample (drawn from groups 1–3 in Table 4) were each subjected to two Bernoulli trials with estimated probability \hat{q}_{ai} ($= \hat{a}_i' / a_i$) of their simulated heads reaching the laboratory and \hat{q}_{ti} ($= \hat{t}_i' / t_i$) of

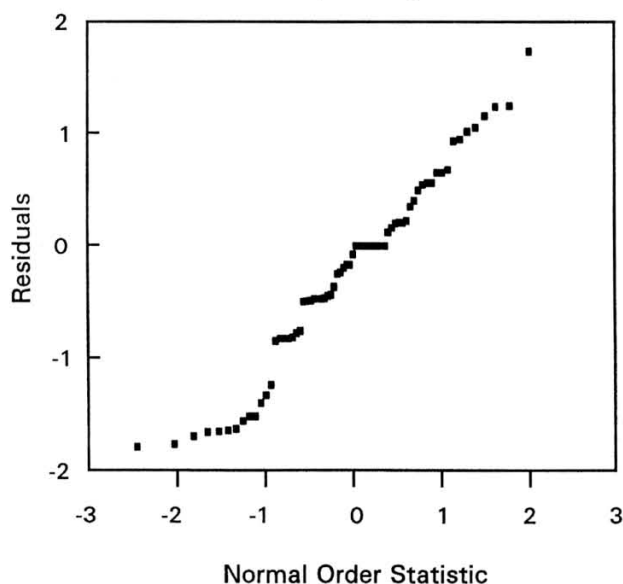
having an imbedded CWT decoded. Marked fish in the bootstrap sample that had no CWTs (drawn from group 4) were subject only to the first trial. For each bootstrap sample, tallies of simulated salmon comprised a vector of bootstrap statistics $\{a_i^*, a_i'^*, t_i^*, t_i'^*, m_{ij}^*, m_{ik}^*\}_b$ from which harvests from cohorts j and k were reestimated. Ten thousand bootstrap samples were drawn for each stratum and were used to form an estimate $F^*(\hat{r}_{ij}, \hat{r}_{ik})$ of the joint frequency distribution $F(\hat{r}_{ij}, \hat{r}_{ik})$. Simulations for recreational fisheries proceeded in the same way, except N_i was also simulated by sampling without replacement from data collected through a conveniently simplified creel survey. Relative frequencies of days spent fishing by anglers ending with a creel of 0, 1, or 2 chinook salmon (2 was the daily limit in 1989) were arbitrarily set at 0.80, 0.15, and 0.05, respectively, making the expected creel per day of fishing 0.25 fish and the variance 0.2875. The number of days fishing (U_i) was calculated as $\hat{N}_i / 0.25$ (\hat{N}_i from Table 3), making the expected frequencies of creels with 0, 1, and 2 fish $(0.80)U_i$, $(0.15)U_i$, and $(0.05)U_i$, respectively. The number of anglers interviewed in the simplified creel survey (u_i) was calculated

Table 3. Statistics from sampling chinook and coho salmon caught in selected fisheries in southeast Alaska in 1989.

Stratum	\hat{N}_i	$CV(\hat{N}_i)$	n_i	a_i	a'_i	t_i	t'_i	m_{ij}	m_{ik}
Chinook salmon: commercial troll fishery (SE quadrant)									
1 Jan. – 14 Apr.	2 487	—	888	61	59	52	52	4	1
4–10 June	2 669	—	1 310	111	111	106	106	5	8
11–17 June	826	—	113	23	23	23	23	3	0
18–24 June	3 291	—	1 832	144	141	132	131	5	1
25 June – 2 July	772	—	192	20	20	16	16	2	2
1 Oct. – 31 Dec.	8 865	—	3 806	251	249	235	235	6	5
Chinook salmon: sport fishery (Ketchikan)									
14 Apr. – 4 June	1 011	0.12	329	34	34	32	32	5	2
5–18 June	1 329	0.20	193	18	18	17	17	3	2
19 June – 2 July	935	0.20	88	15	13	13	13	3	2
3 July – 30 Sept.	2 220	0.15	238	18	18	17	17	2	0
Coho salmon: commercial troll fishery (NW quadrant)									
2–15 July	46 229	—	8 933	80	79	63	63	2	1
16 July – 19 Aug.	500 638	—	150 896	1 861	1 793	1 423	1 423	35	6
20 Aug. – 30 Sept.	354 920	—	113 679	2 176	2 155	1 827	1 827	25	13
Coho salmon: hypothetical fisheries									
	500 000	—	100 000	1 000	500	400	200	10	10
	1 000	—	500	10	5	4	2	1	1

Note: The first cohort of each species (j) was released from Whitman Lake Hatchery whereas the second (k) was released from Neets Bay Hatchery. Notation is defined in Table 1.

Fig. 1. Normalized, transformed residuals (Anscombe's transformation) from the model $\log(\psi_{ij}) = \alpha_i + \tau_j + (\alpha\tau)_{ij}$ fit to recoveries of CWTs from chinook salmon caught in troll fisheries in the SE quadrant in 1989. Effects of stratification in fisheries and of cohorts of chinook salmon are represented by α and τ . Residuals were assumed to follow a Poisson probability distribution.



by solving the equation of variance from a simple random sample from a finite population, in this case $(\hat{N}_i CV(\hat{N}_i))^2 = U_i (U_i - u_i) u_i^{-1}$ (0.2875). For instance, $U_i = 4044$ and $u_i = 296$ for the first listed example of a recreational fishery. Each simulated estimate of catch ($\hat{N}_{i(b)}^*$) is the product of U_i and the average of creels drawn for that bootstrap sample. For recreational fisheries, $N_i \leftarrow \hat{N}_{i(b)}^*$ and $\phi_i \leftarrow \phi_{i(b)}^*$ ($= n_i / \hat{N}_{i(b)}^*$) in Table 4 for each bootstrap sample. Regardless of the type of fishery, the

Table 4. Numbers of salmon in each of seven groups that comprise a catch of N_i in a stratum.

Group	Numbers
Marked and tagged fish in cohort j	$x_{ij(b)}^*$
Marked and tagged fish in cohort k	$x_{ik(b)}^*$
Marked and tagged fish in cohorts other than j or k	$\frac{t'_i}{\lambda_i \phi_i} - x_{ij(b)}^* - x_{ik(b)}^*$
Marked fish that have no tags	$\frac{a_i(a'_i - t_i)}{a'_i \phi_i}$
Unmarked fish in cohort j	$\hat{r}_{ij}^* - x_{ij(b)}^*$
Unmarked fish in cohort k	$\hat{r}_{ik}^* - x_{ik(b)}^*$
Unmarked fish in cohorts other than j or k	Remainder

Note: Asterisks denote estimates drawn from probability distributions for each bootstrap sample ($x_{ij(b)}^*$ and $x_{ik(b)}^*$ from $\text{Binom}(\hat{r}_{ij}, \theta_j)$ and $\text{Binom}(\hat{r}_{ik}, \theta_k)$, respectively) whereas subscripts b refer to specific bootstrap samples. Remaining notation is defined in Table 1.

percentile method was applied to each estimated frequency distribution to produce 95% confidence intervals for \hat{r}_{ij}^* , \hat{r}_{ik}^* and \hat{T}_i^* ($\hat{T}_{i(b)}^* = \hat{r}_{ij(b)}^* + \hat{r}_{ik(b)}^*$). Confidence intervals for \hat{T} across historical examples by species were obtained with the percentile method applied to $\hat{F}^*(\hat{T})$ in which $\hat{T}_{(b)}^* = \sum_i \hat{T}_{i(b)}^*$.

Results

Unbiased estimates of variance and large-sample and bootstrap approximations are similar for all examples (Table 5). Large-sample approximations are slightly conservative against unbiased estimates of variance, with the largest difference being 3.2% (the first hypothetical fishery). In all but two examples from both commercial and recreational fisheries, $0.99 < C_i \leq 1$, $0.99 < D_i \leq 1$, and $1 < E_i < 1.01$. The largest departure ($C_i \approx D_i = 0.59$ for the second hypothetical example) had no effect on the accuracy of the large-sample approximation because only one tag was recovered for each cohort in that stratum. In

Table 5. Estimates of harvest for two cohorts of chinook and coho salmon with unbiased estimates of variances (Unbias) and their large-sample (LS) and bootstrap (Boot) approximations along with unbiased estimates of covariances (\hat{Cov}).

Whitman Lake				Neets Bay				\hat{Cov}
\hat{r}_{ij}	Unbias	LS	Boot	\hat{r}_{ik}	Unbias	LS	Boot	
Chinook salmon: commercial troll fishery (SE quadrant)								
26	140	141	138	57	3 243	3 243	3 146	-2
23	80	80	80	323	12 714	12 750	12 508	-3
49	735	748	736	0	0	0	0	0
21	64	64	65	37	1 309	1 309	1 377	0
18	142	143	140	160	12 250	12 570	12 292	-11
31	133	133	131	233	10 610	10 618	10 635	-1
Chinook salmon: sport fishery (Ketchikan)								
34		214	219	122		7 422	7 703	60
46		719	752	273		38 565	40 563	502
82		2 341	2 395	486		122 604	119 921	1 594
42		844	867	0		0	0	0
Coho salmon: commercial troll fishery (NW quadrant)								
59	1 697	1 698	1 728	290	83 541	83 541	86 434	-4
682	12 585	12 596	12 272	1 142	216 011	216 038	226 960	-19
446	7 501	7 503	7 086	2 264	392 062	392 111	397 313	-11
Coho salmon: hypothetical fisheries								
1 131	122 775	126 835	113 250	11 050	11.811 ^a	12.199 ^a	11.430 ^a	-0.044 ^a
45	2 002	2 002	2 031	442	194 912	194 912	171 364	-0.014 ^a

Note: Examples (strata) are listed in the order established in Table 3.

^a×10⁶.

Table 6. Hypothetical contribution of covariance to summed variances in the first three examples of chinook salmon caught in recreational fisheries if θ_j and θ_k had both been 0.4486 and $CV(\hat{N}_i)$ had been 0.5 for all examples.

$\hat{Var}_{LS}(\hat{r}_{ij})$	$\hat{Var}_{LS}(\hat{r}_{ik})$	$2\hat{Cov}_{LS}(\hat{r}_{ij}, \hat{r}_{ik})$	$\hat{Var}_{LS}(\hat{T}_i)$	Covariance contribution (%)
444	4 342	2 112	6 897	31
1 026	22 059	6 362	29 447	22
3 299	70 209	20 165	93 673	22

the second departure ($D_i = 0.86$ for the third stratum in the recreational fishery), large-sample approximations are similar in comparison to bootstrap variances. Bootstrap variances are smaller than large-sample approximations in 13 instances, larger in 14, and the same in 1. Unbiased estimates of covariances are virtually zero for our examples of commercial fisheries and are relatively small for our examples of recreational fisheries. However, if θ_k for the cohort from Neets Bay Hatchery had been the same as that for the cohort from Whitman Lake Hatchery, and if $CV(\hat{N}_i)$ had been 50% for all strata, estimated covariance would have represented a considerable portion of estimated variances in the recreational fishery (Table 6). The estimated harvest of chinook salmon across all historical strata and cohorts (\hat{T}) is 2063, with a large-sample variance of 218 786 ($= \hat{Var}_{LS}(\hat{T})$). Just over half of this harvest (1085) and 81% of its estimated variance (177 021) came from the recreational fishery. The estimated harvest of coho salmon and its large-sample variance for the three historical examples are 4882 and 713 419.

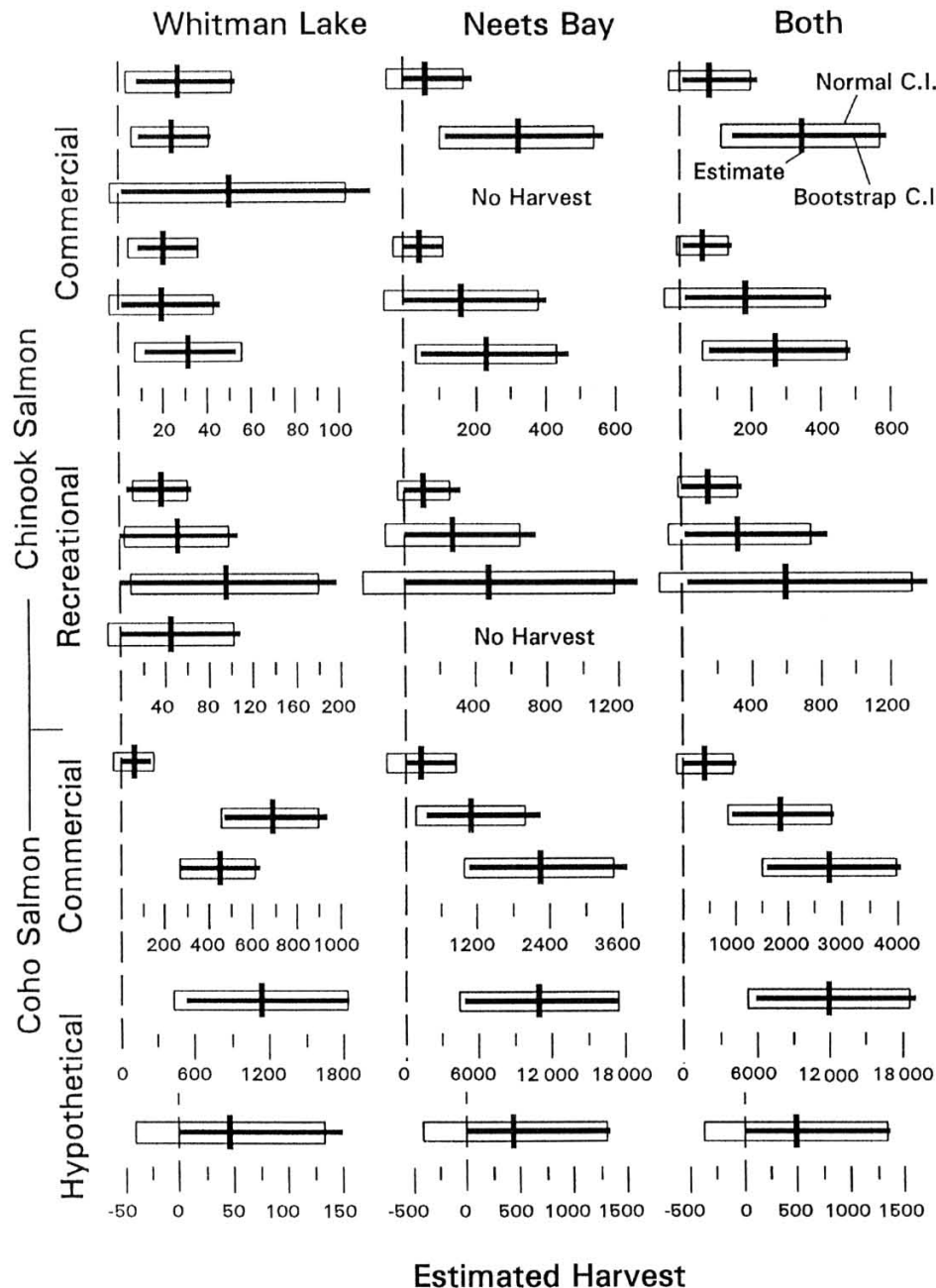
Bootstrap confidence intervals of estimated harvests are skewed for individual and combined cohorts in individual strata (Fig. 2), and for summed estimates of harvest across all

strata and cohorts in examples for chinook salmon (Fig. 3). Because chinook salmon from Neets Bay Hatchery in the third stratum of the recreational fishery represents 28% of estimated harvest of all chinook salmon and 59% of the variance, imprecision from this estimate is largely responsible for the poor correspondence of the confidence intervals for this species (Fig. 3). Bootstrap confidence intervals are severely skewed for the second hypothetical example, another instance in which sample sizes are small and $\lambda_i \ll 1$. In contrast, bootstrap confidence intervals are almost symmetrical for the first hypothetical example in which sample sizes are large, even though $\lambda_i \ll 1$. Normal approximations of 95% confidence intervals are not good approximations for \hat{r}_{ij} or \hat{T}_i in that lower limits of many extend below zero and fall short of the upper limit (Fig. 2). However, the normal approximation is almost identical to the bootstrap interval surrounding \hat{T} for coho salmon (Fig. 3).

Bias in $\hat{\theta}_j^{-1}$

Use of $\hat{\theta}_j$ in the general estimator comes with the warning that $\hat{\theta}_j^{-1}$ is biased because $E(\hat{\theta}_j^{-1}) > \theta_j^{-1}$ by Jensen's inequality (Ferguson 1967). As shown in Geiger (1990) and is evident from Table 7, bias in $\hat{\theta}_j^{-1}$ wanes with increasing sample size s_j and larger θ_j . Because bias is partially a function of θ_j , no safe correction for bias is possible without a priori knowledge of θ_j . Since sampling salmon in hatcheries is relatively cheap, large numbers of salmon are often handled, and bias in $\hat{\theta}_j^{-1}$ along with $\hat{Var}(\hat{\theta}_j^{-1})$ is often reduced to inconsequence. In contrast, capturing wild juvenile and adult salmon in rivers and lakes is expensive, and fewer young salmon can be tagged and fewer adult salmon can be inspected than in a hatchery. For this reason, problems with bias in $\hat{\theta}_j^{-1}$ are more likely for a wild

Fig. 2. Approximate 95% confidence intervals for estimated harvest of coho and chinook salmon in examples. For each example, boxes correspond to approximations based on the normal distribution; solid horizontal lines are comparable intervals from bootstrapping; and solid vertical lines correspond to estimates from original data. Examples are listed in the order established in Table 3. Species and type of fishery are specified along the vertical axis.

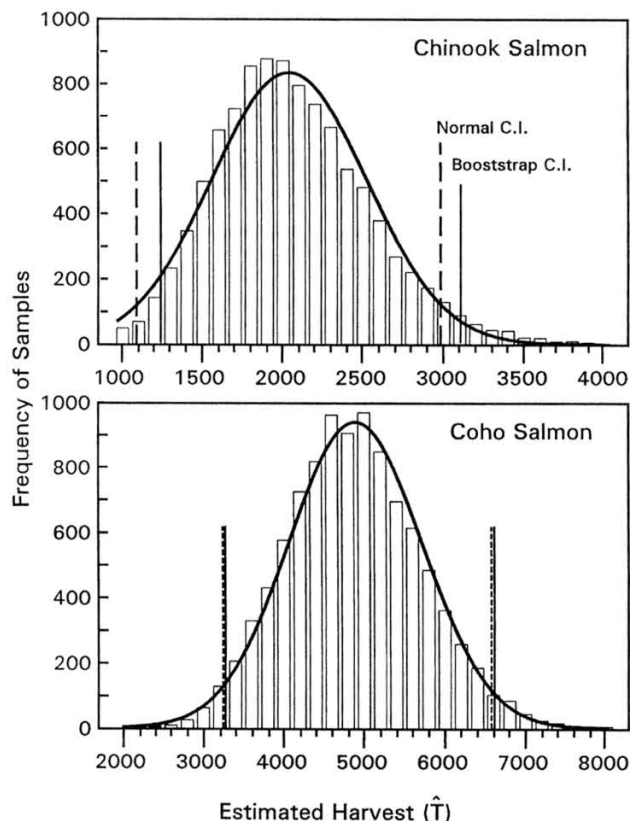


cohort. The only insurance is to increase θ_j and s_j enough so that bias in $\hat{\theta}_j^{-1}$ becomes negligible. Negligible bias, however, does not necessarily mean negligible variance. For example, the simulated statistic $\text{Var}(\hat{\theta}_{j(i)}^{*-1})$ is an estimated mean square error that is the sum of $\text{Var}(\hat{\theta}_j^{-1})$ and bias squared (plus some simulation error). When $s_j = 2000$ and $\theta_j = 0.02$, $\text{Var}(\hat{\theta}_{j(i)}^{*-1}) = 71$ and the estimated bias squared is 2, or about 3% of the MSE.

Bias in $\hat{\theta}_j^{-1}$ can also arise from salmon shedding CWTs. The

background rate of marked salmon without CWTs implied in Table 3 whenever $a_i' > t_i$ does not bias \hat{r}_{ij} provided that none of the tagged salmon in cohort j have lost their CWTs. If they have, \hat{r}_{ij} will be biased downward in direct proportion to the rate of tag loss whenever θ_j is estimated from juveniles. If θ_j is estimated by passing adults (or their heads) through a magnetometer, the estimate is automatically discounted for any tag loss, and \hat{r}_{ij} is unbiased. Occurrence in samples of salmon missing adipose fins for reasons other than tagging will not

Fig. 3. Bootstrap frequency distributions for \hat{T} using historical examples of fisheries for chinook and coho salmon with superimposed normal probability density functions Norm(\hat{T} , $\hat{SE}(\hat{T})$). Solid vertical lines correspond to limits of 95% confidence intervals for \hat{T} based on bootstrapped approximations (percentile method). Broken vertical lines correspond to limits of 95% confidence intervals based on normal approximations.



bias $\hat{\theta}_j^{-1}$ provided the statistic is not based on this secondary mark.

Conclusions

Estimates of harvest from our method are unbiased with unbiased estimates for their variance when catch and probability of being marked are known; when either are estimated, the $\{\hat{r}_{ij}\}$ are still unbiased, only their estimated variance is approximate because of plug-in substitution of $\hat{\phi}_i$ for ϕ_i and $\hat{\theta}_j$ for θ_j in the fpc. The approximation improves with fewer samples and fewer tagged juveniles, as the product $\lambda_i \phi_i \theta_j$ quickly nears zero. These circumstances are typical of harder to sample catches in recreational fisheries and more difficult to capture juveniles from wild stocks of salmon.

The large-sample approximation $\hat{\text{Var}}_{LS}(\hat{p}_{ij}) = \hat{p}_{ij}(\lambda_i n_i)^{-1}(1 - \lambda_i \phi_i \theta_j)$ is the preferred formulation of estimated variance from catch sampling. This simple approximation expanded by estimates of catch proved to be within a few salmon of unbiased and simulated variances in our examples. The approximation proved negligibly conservative across a wide spectrum of catches ($772 \leq N_i \leq 500\,638$), sample sizes ($88 \leq n_i \leq 150\,896$), sampling rates ($0.09 \leq \phi_i \leq 0.56$), and decoding rates ($0.59 \leq \lambda_i \leq 1$) in both commercial and recreational fisheries.

Table 7. Estimated relative bias (%) in simulated estimates of θ_j^{-1} .

s_j	θ_j		
	0.02	0.05	0.10
200	>31 ^a	12	5
300	24	7	3
400	17	5	2
500	13	4	2
2 000	3	1	<1

Note: Each estimate is based on 100 000 bootstrap samples, with each sample b drawn from $\text{Binom}(s_j, \theta_j)$. Estimates were calculated as $(\hat{\theta}_{j(c)}^{-1} - \theta_j^{-1})\theta_j \times 100$, with $\hat{\theta}_{j(c)}^{-1}$ being the average of $\hat{\theta}_{j(b)}^{-1}$.

^aIn some bootstrap samples, $\hat{\theta}_{j(b)}^{-1}$ was undefined, making this estimate of relative bias a minimum when calculated from the remaining samples.

The compound rate λ_i of heads not lost and of detected tags decoded has an effect on the precision of estimated harvest equal to that of the sampling rate of the catch and the tagging rate of juveniles. Fractions λ_i , ϕ_i , and θ_j each appear twice in large-sample approximations of variance, but always together in the product $\lambda_i \phi_i \theta_j$. Recent experience in Alaska has been that almost all tags are decoded once detected with the magnetometer, and few heads in samples from commercial fisheries are lost. However, the potential for sampled heads not to reach the dissecting laboratory is considerably greater when sampling recreational fisheries, especially when larger salmon are considered trophies. Also, including λ_i in the estimator allows for conscious subsampling at sampling sites and at laboratories.

For commercial fisheries, at least those for which catch is known, covariances between harvests can be ignored with little penalty. Dropping covariances from calculations will only make summed variances negligibly more conservative, even for very small strata with few recaptured fish. Whether fractions of cohorts are known or estimated is irrelevant to the significance of covariances when summing harvest across fishing districts and fishing periods in commercial fisheries.

Covariances should not be ignored in recreational fisheries, at least not in those in which cohorts are well tagged and catch is estimated imprecisely. As shown in our examples, the significance of covariance between two cohorts is directly related to θ_j and θ_k (their magnitudes, not the precision of their estimates), and $\text{CV}(\hat{N}_i)$. Because tagging juveniles in a hatchery is usually easier than along a stream, covariance will be more important between cohorts from hatcheries than between cohorts representing wild production. Under these circumstances, ignoring covariance when summing harvests could seriously underestimate variance when recreational fisheries are involved in calculations.

Normal approximations of confidence intervals improve as the scope of estimates broadens. Normal approximations of confidence intervals are demonstrably poor when harvest comes from sparsely tagged cohorts in recreational fisheries, as in our examples involving harvests of chinook salmon from Neets Bay Hatchery in the recreational fishery near Ketchikan. Under these circumstances, methods based on the parametric bootstrap or likelihood profiles (Venzon and Moolgavkar 1988), as used in Schnute (1992), produce better confidence intervals. However, both of these methods require considerable computation, which can be a real burden when there are many cohorts being caught in many fisheries. When estimates

are summed across strata and cohorts to produce seasonal statistics, such as harvest of all hatchery-produced salmon in a year, the normal approximation will be nearly identical to confidence intervals determined from profile likelihoods or from bootstrapping (as in our example for coho salmon). In this situation, ease of computation would usually make the normal approximation the preferred method for determining confidence intervals about estimated harvest.

Acknowledgements

We thank David Evans, Hal Geiger, Allen Bingham, and Bob Marshall for their reviews of earlier drafts. We are grateful to Gail Heineman and Bob Marshall for their assistance with programming and to Allen Bingham and Paul Suchanek for their assistance with data in our examples. We also thank our anonymous reviewers, especially the one who reminded us of the easier method of joint factorial moments to investigate estimators based on discrete probability distributions. This work was partially supported with funding authorized under the Federal Aid to Sport Fish Restoration Act.

References

- Clark, J.E., and Bernard, D.R. 1987. A compound multivariate binomial-hypergeometric distribution describing coded microwire tag recovery from commercial salmon catches in southeastern Alaska. Informational Leaflet No. 261. Alaska Department of Fish and Game, Juneau, Alaska.
- Cormack, R.M., and Skalski, J.R. 1992. Analysis of coded wire tag returns from commercial catches. *Can. J. Fish. Aquat. Sci.* **49**: 1816–1825.
- Efron, B., and Tibshirani, R.J. 1993. An introduction to the bootstrap. Chapman Hall, New York.
- Ferguson, T.S. 1967. Mathematical statistics, a decision theoretic approach. Academic Press, New York.
- Geiger, H.J. 1990. Parametric bootstrap confidence intervals for estimating contributions to fisheries from marked salmon populations. *Am. Fish. Soc. Symp.* **7**: 667–676.
- Goodman, L.A. 1960. On the exact variance of products. *J. Am. Stat. Assoc.* **55**: 708–713.
- Green, P.E.J., and Macdonald, P.D.M. 1987. Analysis of mark–recapture data from hatchery-raised salmon using log-linear models. *Can. J. Fish. Aquat. Sci.* **44**: 316–326.
- Guthrie, D., Hoenig, J.M., Holliday, M., Jones, C.M., Mills, M.J., Moberly, S.A., Pollock, K.H., and Talhelm, D.R. (Editors). 1991. Creel and angler surveys in fisheries management. *Am. Fish. Soc. Symp.* **12**.
- Johnson, J.K. 1990. Regional overview of coded wire tagging of anadromous salmon and steelhead in northwest America. *Am. Fish. Soc. Symp.* **7**: 782–816.
- Lapi, L., Hamer, M., and Johnson, B. 1990. Data organization and coding for a coastwide mark–recovery data system. *Am. Fish. Soc. Symp.* **7**: 720–724.
- Newman, K.B. 1990. Variance estimation for stock-contribution estimates based on sample recoveries of coded-wire-tagged fish. *Am. Fish. Soc. Symp.* **7**: 677–685.
- Pollock, K.H., Jones, C.M., and Brown, T.L. 1994. Angler survey methods and their applications in fisheries management. *Am. Fish. Soc. Spec. Publ.* No. 25.
- Schnute, J.T. 1992. Statistical analysis of embedded replicates in mark–recovery experiments. *Can. J. Fish. Aquat. Sci.* **49**: 432–442.
- Venzon, D.J., and Moolgavkar, S.H. 1988. A method for computing profile-likelihood based confidence intervals. *Appl. Stat.* **37**: 87–94.