Bias in common catch-curve methods applied to age frequency data from fish surveys

Gary A. Nelson

SEDAR58-RD47

Received: November 19, 2019

ICES Journal of
Marine Science

International Council for
the Exploration of the Sea
Conseil International pour
l'Exploration de la Mer

ICES Journal of Marine Science (2019), doi:10.1093/icesjms/fsz085

Bias in common catch-curve methods applied to age frequency data from fish surveys

Gary A. Nelson *

Massachusetts Division of Marine Fisheries, 30 Emerson Avenue, Gloucester, MA 01930, USA

**Corresponding author: tel: +1 9782820308; fax: +1 6177273337; e-mail: gary.nelson@mass.gov.*

Nelson, G. A. Bias in common catch-curve methods applied to age frequency data from fish surveys. – ICES Journal of Marine Science, doi:10.1093/icesjms/fsz085.

Received 29 January 2019; revised 19 April 2019; accepted 23 April 2019.

Catch curve analysis is often used in data-limited fisheries stock assessments to estimate total instantaneous mortality (Z). There are now six catch-curve methods available in the literature: the Chapman–Robson, linear regression, weighted linear regression, Heincke, generalized Poisson linear, and random-intercept Poisson linear mixed model. An assumption shared among the underyling probability models of these estimators is that fish collected for ageing are sampled from the population by simple random sampling. This type of sampling is nearly impossible in fisheries research because populations are sampled in surveys that use gears that capture individuals in clusters and often fish for ageing are selected from multi-stage sampling. In this study, I explored the effects of multi-stage cluster sampling on the bias of the estimates of Z and their associated standard errors. I found that the generalized Poisson linear model and the Chapman–Robson estimators were the least biased, whereas the random-intercept Poisson linear mixed model was the most biased under a wide range of simulation scenarios that included different levels of recruitment variation, intra-cluster correlation, sample sizes, and methods used to generate age frequencies. Standard errors of all estimators were under-estimated in almost all cases and should not be used in statistical comparisons.

Keywords: cluster sampling, fish surveys, total mortality.

Introduction

Catch curve analysis, the estimation of total mortality from a single sample of age composition data, is often used in fisheries stock assessments where limited data about the population are available. In such cases, researchers may use age data collected from fisheries-independent surveys because it is thought that age frequencies tend to be more representative of the population age structure. For example, catch curve analysis was used to estimate Z for the recreationally exploited *Lagodon rhomboides* (Nelson, 2002), the commercially exploited *Nemadactylus macropterus* (Wankowski *et al.*, 1988), and the formerly commercially exploited *Alosa aestivalis* and *Alosa pseudoharengus* collected during fisheries-independent seine and trawl surveys (ASMFC, 2017).

The common estimators used in catch curve analysis are the Heincke (1913), Chapman and Robson (1960), simple linear regression (Ricker 1975), and weighted linear regression (Maceina and Bettoli, 1998) and all have implicit model assumptions that have to be met to produce unbiased estimates of Z and associated standard errors. The main assumptions are that the population is in a steady state, implying that recruitment is constant over time, Z is constant over time and across ages, all fish are assumed equally vulnerable to sampling above a certain age (i.e. fully recruited age), and there are no errors in ageing.

Several studies have investigated the performance of the common Z estimators when these assumptions are violated. Through simulation, Dunn *et al.* (2002) investigated the sensitivity of the Chapman–Robson and simple linear regression methods to sources of variability in the true mortality rate, auto-correlated annual recruitment and ageing error. Similarly, Smith *et al.* (2012) explored the sensitivity of the Heincke, Chapman–Robson, simple linear regression, and weighted linear regression estimators and their associated variances to variability in uncorrelated annual recruitment and the choice of age at full recruitment under a range of Z values and sample sizes. In general, use of Chapman–Robson with the age at peak number plus one year rule for the age at full recruitment or the weighted linear regression estimator with the peak age rule provided the least biased estimates of Z, although Smith *et al.* (2012) recommended using the former estimator because the weighting procedure used in the latter method was purely *ad hoc*. More recently, Millar (2015) introduced the generalized Poisson model and the random-intercept Poisson log-linear mixed effect model, and showed that performance of the latter is superior to the Chapman–Robson, weighted linear regression, and generalized Poisson models under similar scenarios because it is a non-steady state model that accounts for annual variation in recruitment.

In fisheries research, age composition used in catch curve analysis has to be estimated from samples collected from a population. All aforementioned mortality estimators require that aged fish are sampled randomly from the population to obtain unbiased estimates of Z and associated standard errors from the underlying probability models (Chapman and Robson 1960; Seber 2002). The assumptions of simple random sampling require that each individual selected for the sample has the same nonzero probability of occurring in the sample, and that the selection of one individual is not influenced by other individuals already selected (Lohr 1999). To meet these assumptions, fish would have to be captured individually and at random from a population. This type of sampling is nearly impossible in fisheries research because fish populations are distributed heterogeneously over large areas and the types of gear generally used (e.g. seines, trawls, etc.) capture individuals in groups or *clusters* (Pennington and Volstad, 1994; Nelson 2014). Cluster sampling creates non-independence because the inclusion of an individual in a sample is related to the probability of selecting a cluster, not an individual, and the selection of one individual becomes dependent on the selection of another individual (Lohr, 1999). As a result, total mortality estimators that assume simple random sampling may produce biased estimates when applied to age composition estimated from cluster sampling because clustering is not taken into account by the underlying probability models. Jensen (1996) recognized this issue and modified the Heincke estimator to account for one-stage (all fish caught are aged) cluster sampling.

The estimation of age composition is complicated further because a multi-stage cluster sampling design is often used during fisheries-independent surveys to subsample fish for aging (Aanes and Volstad, 2015). At the first stage, the clusters (hauls or tows) are usually taken at random locations. If the number of fish in a tow is large, second-stage sampling may occur by taking a random subsample from each haul to measure characteristics of fishes such as length. Third-stage sampling may also occur in which a simple or a stratified (by length) random sample of the second stage individuals is taken to obtain individuals for estimation of age composition, and often the number of fish taken from each haul is fixed (ASMFC 1994; Aanes and Volstad, 2015). In these cases, the numbers-at-age are not those obtained by simply summing the number of each age in the sample; rather, they have to be derived by using proportion estimators that correctly weight samples for haul sizes (Aanes and Volstad, 2015).

A challenge that arises from multi-stage sampling is deciding which number (the total fish caught in all hauls or actual third-stage sample size) should be used to derive the age frequencies from the proportion estimates. A typical goal of a fisheries-independent survey is to estimate the age composition of the total catch and may be done by multiplying the total fish caught in all hauls by the estimated proportions. However, some investigators may decide to use the actual third-stage subsample size because age frequencies are usually derived from a sample and the actual sample size is required in the Chapman–Robson estimator of total mortality and standard error. In this case, if the total fish caught in all hauls was used, the resulting estimate of standard error may be greatly under-estimated. For estimators that use a linear model framework (e.g. simple linear regression, weighted linear regression, etc.), this may not be a significant issue because the number of ages, not the number of fish sampled, is used in the calculation of standard error of Z (slope of the regression) (Neter *et al.*, 1996).

Given that total mortality estimators are often used to assess stocks that are data-limited, an investigation of the performance of common estimators that researchers will likely apply to age frequency data collected in fish surveys is warranted. In this study, I used a simulation of a fish survey to explore the effects of the degree of clustering, multi-stage cluster sampling, subsample size and choice of the expansion number used to derive numbers-at-age from weighted proportions on the performance of catch curve estimators and standard errors.

Methods

Estimators

The total mortality estimators examined in this study were the Chapman and Robson (1960), simple linear regression (Ricker 1975), weighted linear regression (Maceina and Bettoli, 1998), the modified version of the Heincke method proposed by Jensen (1996) to correct for one-stage cluster sampling, the Poisson log-linear model (Millar 2015), and a random-intercept Poisson log-linear mixed model formulated by Millar (2015). It should be noted that the random effects mixed modeling framework has the potential to account for sources of variation in multistage sampling, but investigation of such models was beyond the scope of this paper. The details of each method explored herein are described below. The Chapman and Robson (1960) method determines total mortality from a natural log-transformed annual survival rate determined from age composition data modelled as a geometric distribution under the steady-state assumption. The formula for total mortality ($\hat{Z}_{CR}$) corrected for bias due to log-transform is

$$\hat{Z}_{CR} = \log_e\left(\frac{\sum_{a=0}^{r} a \cdot m_a}{m + (\sum_{a=0}^{r} a \cdot m_a) - 1}\right) - \frac{(m-1)(m-2)}{m\left(m + (\sum_{a=0}^{r} a \cdot m_a) - 1\right)\left((\sum_{a=0}^{r} a \cdot m_a) + 1\right)}$$

where a is the age of each fully recruited age-group, r is the oldest age, m_a is the number of fish of age a in the sample, and m is the total number of fish in the sample. The ages of the fully recruited fish are standardized to begin at $a = 0$. The estimated standard error of $\hat{Z}_{CR}$ is

$$SE(\hat{Z}_{CR}) = \sqrt{\frac{(1 - e^{-\hat{Z}_{CR}})^2}{me^{-\hat{Z}_{CR}}}}$$

Chapman and Robson (1960) noted under the steady state assumption that the age frequencies could be modelled as a Poisson

Table 1. The R package, function and code snippet used to fit each total mortality estimator to simulated age composition data.

Estimator	Package	Function	Code Snippet
CR	fishmethods	agesurv	agesurv(age=age, estimate="z", method="crcb")
LM	stats	lm	lm(log(number)∼age, data=datafile)
WLM	stats	lm	lm(log(number)∼age, weights=predict(lm(log(number)∼age, data=datafile)), data=datafile)
GLM[a,b]	stats	glm	glm(number∼age, family="poisson", link="log", data=datafile)
RE[a]	lme4	glmer	glmer(number∼age+(1|age), family="poisson", link="log", data=datafile)
JEN	–	–	written code

[a] The code snippet given in Millar (2015) was used to extend the age-frequency data.

[b] Additional code was written to calculate the over-dispersion parameter.

distribution as well, and Millar (2015) showed total mortality could be estimated by using a generalized Poisson log-linear model:

$$\mu_a = \exp^{\alpha - Z \cdot a}$$

where μ_a is Poisson mean at age *a*, α is the fixed intercept and *Z* is the fixed slope. The model is fitted to numbers-at-age with age as a covariate via maximum likelihood with a log-link function. The estimate of total mortality ($\hat{Z}_{GL}$) is the negative of the model slope.

The simple linear regression method is a common technique used to estimate total mortality (Ricker 1975; Seber 2002). A simple linear regression model is fitted to the natural-log of the number of fully recruited fish greater than zero with age as a covariate by using least-squares. The estimate of total mortality (Z_{LM}) is the negative of the slope coefficient. The weighted linear regression estimator ($\hat{Z}_{WL}$) introduced by Maceina and Bettoli (1998) determines total mortality in a two step procedure. A simple linear regression is first fitted to the natural-log of the number of fully recruited fish greater than zero with age as a covariate. A weighted linear regression model is then fitted to the same data but the predicted values of natural-log numbers-at-age from the simple linear regression are used as weights in the estimation. The estimate of total mortality ($\hat{Z}_{WL}$) is the negative of the slope coefficient.

Jensen (1996) adapted the Heincke (1913) total mortality estimator to first-stage cluster sampling (all fish in each haul are aged) by estimating annual mortality as a ratio of the sum of the number of the first fully recruited age from all hauls and the sum of the number of all fully recruited fish from all hauls.

$$\hat{Z}_{JN} = -\log_e\left(1 - \frac{\sum_{j=1}^{n} m_{1j}}{\sum_{j=1}^{n}\sum_{a=1}^{\max} m_{aj}}\right)$$

where m_{1j} is the number of the first fully recruited age in haul *j*, m_{aj} is the number of fully recruited age *a* fish in haul *j* and *n* is the total number of hauls. Improved standard errors are derived herein by using jackknife method where jackknifing is performed at the haul level (Pennington and Volstad, 1994).

Millar (2015) showed that a better model for estimating total mortality when the steady state assumption is violated is the random-intercept Poisson log-linear mixed effects model. This model takes into account annual variation in recruitment, making it a non-steady state model. The model form is

$$\mu_a = \exp^{\alpha + \beta_a - Z \cdot a}$$

where μ_a is Poisson mean number at age *a*, α is the fixed intercept, β_a are normally distributed ($N(0, \sigma^2 R)$) random effects and *Z* is the fixed slope coefficient. The model is fitted to numbers-at-age with age as a covariate via maximum likelihood with a log-link function. The estimate of total mortality ($\hat{Z}_{RE}$) is the negative of the model slope.

On the basis of Smith *et al.* (2012) and Millar (2015), the ages considered fully recruited were selected by using the "peak age plus one year" criterion for the $\hat{Z}_{CR}$, $\hat{Z}_{GL}$, $\hat{Z}_{RE}$, and $\hat{Z}_{JN}$ estimators and the "RG" criterion (peak age and all age groups with nonzero catch) for the $\hat{Z}_{LM}$ and $\hat{Z}_{WL}$ estimators. In addition, the standard errors of $\hat{Z}_{CR}$ and $\hat{Z}_{GL}$ were corrected for over-dispersion by multiplying each by the square-root of a variance inflation factor (Burnham and Anderson, 2002) calculated by using the chi-square goodness of fit statistic of each model. Only ages with an expected frequency of at least unity were used in the calculation of over-dispersion (Millar 2015).

All models were fitted to simulated age frequencies data (described below) by using statistical functions in R (R Development Core Team, 2018). The functions and model code used for each estimator are listed in Table 1.

Simulations

The performance of each estimator was explored over a range of true *Z* values, ρ, third-stage subsample sizes (fish aged per haul) and choice of expansion number (total fish caught in all hauls or third-stage sample size). Because variability in recruitment is an important factor affecting estimator performance (Dunn *et al.*, 2002; Smith *et al.* 2012; Millar, 2015), the impact of recruitment variation was also examined in conjunction with these factors. For comparison, bias of estimators using age frequencies from first-stage sampling (all fish captured were aged) was calculated under all scenarios. Under the $\rho = 0$ with first-stage sampling, the bias of each estimator was considered the baseline and is compared with bias under the remaining scenarios to determine the influence of a factor.

Data generation

A basic population dynamics model nearly identical to Millar (2015) was created to generate population numbers-at-age with true *Z* values ranging from 0.1 to 1.0 per year. The relative number of age-1 fish in the population in each year was generated assuming first-order autocorrelation following Dunn *et al.* (2002):

$$N_{t_0,1} = \exp^{\epsilon_1}$$
$$N_{t,1} = \exp^{\phi\log(N_{t-1,1})+\sqrt{1-\phi^2}\cdot\epsilon_t}$$

where t_0 is the first year of the simulation, $N_{t,1}$ is the relative number of fish of age *1* in year *t*, ϵ_t are independently distributed normal random deviates ($N(0,\sigma^2{}_R)$) and ϕ is the autocorrelation coefficient. Following Millar (2015), the effects of differing levels of recruitment variability were investigated by using a fixed ϕ of 0.37 and σ_R equal to 0.35, 0.67, and 1.17.

The relative numbers of fish age 2 and older in the population were determined by forward projection of the age 1 numbers:

$$N = N_{t-1,a-1} \cdot \exp^{-Z_{a-1}}$$

where Z_a is the total mortality for age *a*, which is the sum of the partially recruited instantaneous fishing (s_aF) and natural (*NM*) mortality rates ($Z_a = s_aF + NM$). Following Millar (2015), fully recruited *F* and *NM* were each set to half of the true *Z* value and the partial recruitment (s_a) vector was 0.25 for age-1, 0.75 for age-2, 0.964 for age-3, 0.996 for age-4, and 1.00 for ages 5 and older. The maximum number of ages in the population was set to 100. The population was projected for 200 years. The population age composition was expressed as probabilities obtained from the last year's abundances:

$$p_a = \frac{s_aN_a}{\sum_{a=1}^{100} s_aN_a}$$

To simulate haul-specific survey catches and age compositions, first-stage sampling was accomplished by first randomly generating catch for 150 hauls from a negative binomial distribution, parameterized with mean haul size and dispersion parameter set at 200 and 0.7, respectively, by using function *rnegbin* in the R package MASS (Venables and Ripley, 1999). The negative binomial parameters are intermediate values for catch distributions of the top five species in five trawl surveys examined in Nelson (2014). The first 20 hauls with positive catches were then used in the simulation. The age frequencies for each haul were then randomly generated from the Dirichlet-Multinomial distribution, a multivariate generalization of beta binomial distribution that incorporates correlation among observations, by using function *simPop* in R package *dirmult* (Tvedebrink, 2013). The Dirichlet-Multinomial distribution was parameterized with the total catch of each haul, age-probabilities (p_a), and an intra-cluster correlation value (ρ), which measures the degree of clustering (*within-cluster* similarity). Age compositions were generated with four levels of intra-cluster correlation ($\rho = 0.0$, 0.1, 0.2, and 0.3). Under $\rho = 0$, bias in estimates represents what would result under simple random sampling. The maximum value (0.3) is close to the maximum (0.26) estimated for Arctic cod by Aanes and Pennington (2003) from commercial catches. In reality, the intra-cluster correlation could be higher for fisheries-independent surveys, but there were no estimates available in the literature.

To investigate the impact of subsample size on the performance of the *Z* estimators when third-stage subsampling occurs, a random second-stage subsample equal to 30% (intermediate percentage of fish sampled for lengths in the surveys examined by Nelson 2014) of the total number in each haul was taken from each haul using a multinomial model parameterized with the haul catch age-probabilities. The third-stage sampling was accomplished by random sampling of the individuals of each haul from the second-stage sampling using a multinomial model parameterized with second-stage calculated age-probabilities. Two fixed subsample sizes (10 and 50 fish aged per haul) were explored during third-stage sampling, and represented ~175 and 626 fish, respectively, sampled on average (not all hauls have enough fish to obtain the fixed sample size). For first-stage sampling, the total number of fish in all hauls averaged 4000 individuals.

The survey age composition of hauls was estimated from the third-stage subsample by calculating the proportions-at-age (p_a):

$$\hat{p}_a = \frac{\sum_{i=1}^{n} M_i \cdot \hat{p}_{i,a}}{\sum_{i=1}^{n} M_i}$$

where M_i is the total number of fish in haul *i*, $p_{i,a}$ is the estimated proportion of age *a* in haul *i* from subsampling and *n* is the number of hauls (Aanes and Volstad, 2015). $p_{i,a}$ is calculated as $m_{i,a}/m_i$ where $m_{i,a}$ is the number of age *a* fish in haul *i* and m_i are the number sampled in the third stage from haul *i*.

To explore the impact of using the total number of fish caught in all hauls (*M*) or the third-stage total subsample size (*m*) as the expansion number to develop the numbers-at-age from the proportions-at-age, age frequencies of each age *a* were derived by:

$$\hat{M}_a = \hat{p}_a \cdot \sum_{i}^{n} M_i$$

and

$$\hat{m}_a = \hat{p}_a \cdot \sum_{i}^{n} m_i$$

Numbers-at-age were rounded to the nearest whole integer to specify discrete counts.

Performance

For each level of true *Z* and σ_R, 4000 population age frequencies were generated. If the number of fully recruited age classes was less than three, the catch age frequency was rejected and replaced with a catch age frequency from a new population. Intra-cluster correlation was introduced using the same 4000 simulated population age frequencies for each level of ρ.

The performance of each estimator was measured with percent bias (%*BIAS*) following Smith *et al.* (2012). For each estimator, %*BIAS* of a $\hat{Z}$ was calculated as

$$\%BIAS(\hat{Z}) = \frac{E(\hat{Z}) - Z}{Z} \cdot 100$$

and %*BIAS* of the standard error ($SE(\hat{Z})$) was calculated as

$$\%BIAS(SE(\hat{Z})) = \frac{E(\hat{SE}(\hat{Z})) - SE(\hat{Z})}{SE(\hat{Z})} \cdot 100$$

where $E()$ denotes expectation, *Z* is the true value and $SE(\hat{Z})$ is the true *SE* calculated as the standard deviation of the 4000 simulated estimates of *Z*. $E(\hat{Z})$ and $E(\hat{SE}(\hat{Z}))$ were approximated by

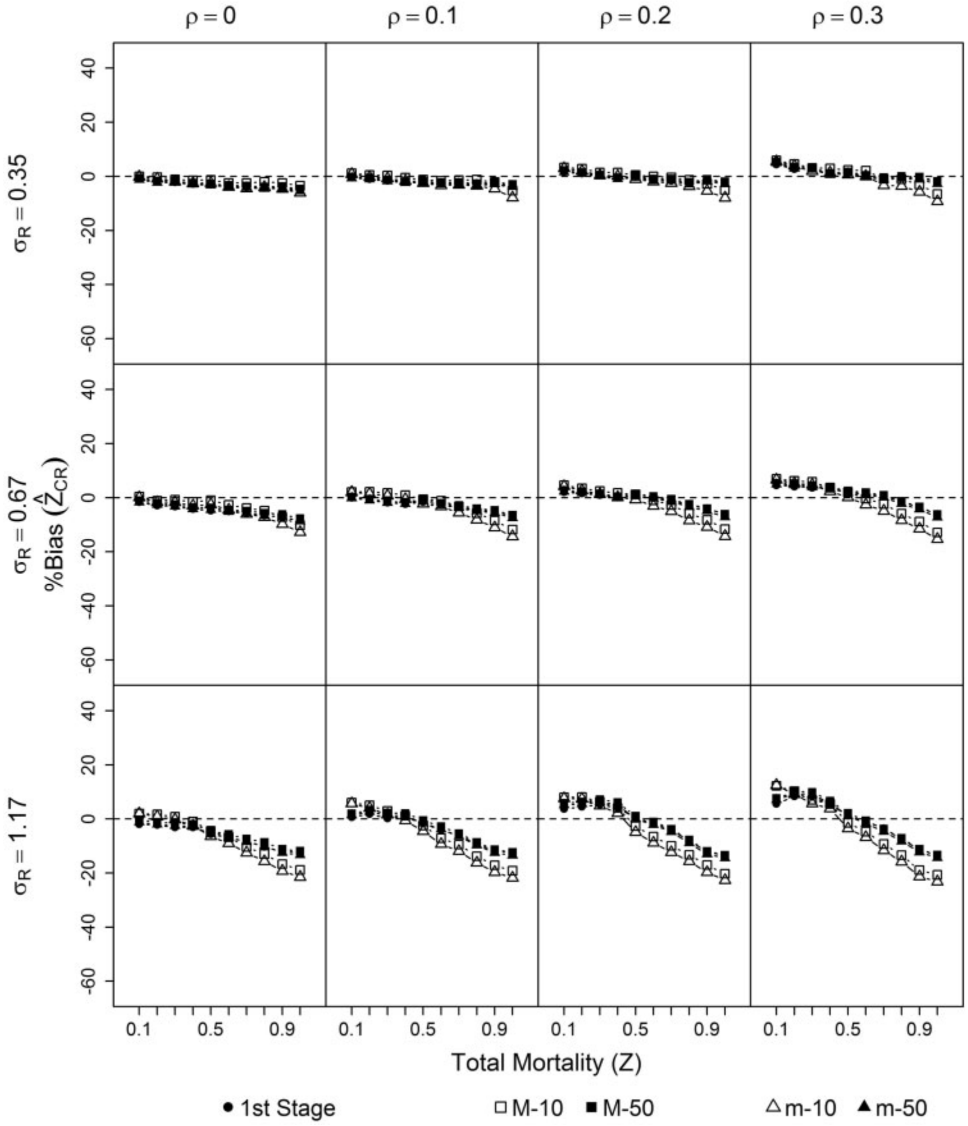

Figure 1. Percent bias (%Bias) of the Chapman–Robson total mortality rate (Z) estimator vs. the true total mortality under different levels of recruitment variation (σ_R), intra-cluster correlation (ρ), choice of using the total number caught (M), or the total subsample size (m) to derived age frequencies and third-stage subsample size (10 or 50 fish per haul).

averaging estimates of $\hat{Z}$ and $\hat{SE}(\hat{Z})$ over the simulation results (Smith *et al.*, 2012).

Results

Performance of Z estimators

The trends and magnitudes in %BIAS of $\hat{Z}_{CR}$ and $\hat{Z}_{GL}$ were similar across all scenarios (Figures 1 and 2). Under $\rho = 0$ (no intra-cluster correlation), %BIAS from first-stage sampling (all fish aged) was relatively low and negative, but it became more negative as true Z and σ_R increased (Figures 1 and 2). Compared with these baseline values, the impact of increasing ρ was to increase bias in the positive direction at low Z and to increase bias in the negative direction at high Z, although %BIAS in $\hat{Z}_{GL}$ became slightly more positive at low Z and less negative at high Z (Figures 1 and 2). As recruitment variation increased, these patterns became more pronounced. The effect of third-stage subsample size was to slightly shift bias in the positive direction at low Z and in the negative direction at high Z as subsample size decreased. The expansion number had little impact on bias (Figures 1 and 2). Overall, %BIAS under all scenarios was relatively narrow ($\hat{Z}_{CR}$ range: –23.2%, 12.5%; $\hat{Z}_{GL}$ range: –20.6%, 8.5%).

%BIAS of $\hat{Z}_{LM}$ and $\hat{Z}_{WL}$ was similar in trend but different in magnitude across all scenarios. Under the $\rho = 0$ scenarios, %BIAS for first-stage sampling was typically negative, but it became more negative as σ_R increased, particularly at low Z (Figures 3 and 4).

Compared with these baselines, the effect of increasing ρ was for %BIAS to become more negative at low Z and more positive at intermediate Z values as σ_R increased (Figures 3 and 4). The impact of third-stage sampling was to shift overall bias in the negative direction as subsample size decreased. When m was used in

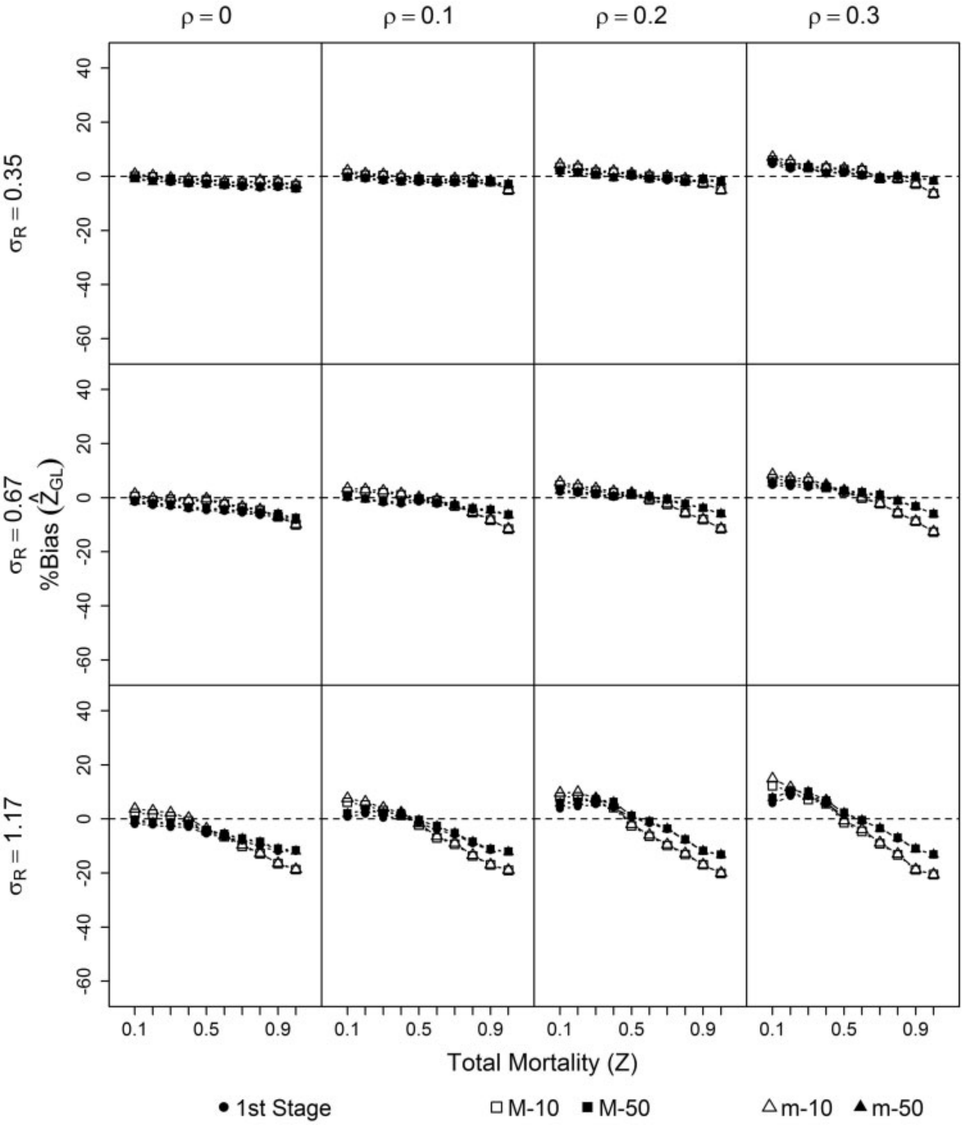

Figure 2. Percent bias (%Bias) of the generalized Poisson linear model estimator of total mortality rate (Z) vs. the true total mortality under different levels of recruitment variation (σ_R), intra-cluster correlation (ρ), choice of using the total number caught (M), or the total subsample size (m) to derived age frequencies and third-stage subsample size (10 or 50 fish per haul).

expansion, bias shifted in the negative direction, although less so for $\hat{Z}_{\mathrm{WL}}$ (Figures 3 and 4). Overall, the range of %BIAS under all scenarios was wide ($\hat{Z}_{\mathrm{LM}}$ range: –55.6%, 3.3%; $\hat{Z}_{\mathrm{WL}}$ range: –42.4%, 15.1%).

Under the $\rho = 0$ scenarios, the %BIAS of the $\hat{Z}_{\mathrm{JN}}$ estimator was mostly negative at low σ_R, but it became more positive at low true Z values and more negative at high Z values as σ_R increased (Figure 5). The effect of increasing ρ was to slightly increase bias in the negative direction (Figure 5). The impact of subsample size was for bias to become more negative at low sample size (10 fish per haul) (Figure 5). Overall, the range of %BIAS of $\hat{Z}_{\mathrm{JN}}$ under all scenarios was wide (range: –39.2%, 42.4%).

Under the $\rho = 0$ scenarios, %BIAS of $\hat{Z}_{\mathrm{RE}}$ for first-stage sampling was very low (range: –1.6%, 0.8%) in all σ_R scenarios and became only slightly negative as Z increased (Figure 6). As ρ increased, %BIAS of $\hat{Z}_{\mathrm{RE}}$ in all scenarios became more positive (Figure 6). The effect of third-stage subsampling was to increase bias in the positive direction as subsample size decreased and when M was used in expansion (Figure 6). Unexpectedly, the converse was true when m was used in expansion (Figure 6). Overall, the range of bias for $\hat{Z}_{\mathrm{RE}}$ under all scenarios was very wide (range: –3.6%, 139.1%).

Performance of standard error estimators

The performance of standard error estimators was poor in most scenarios as the true standard error was typically under-estimated (Supplementary Figures S1–S6). %BIAS was the most negative in baseline cases (first-stage sampling and all σ_Rs) under $\rho = 0$ and negative bias increased with increasing σ_R. SE($\hat{Z}_{\mathrm{RE}}$) was least biased negatively (range: –37.6, –22.6), followed by SE($\hat{Z}_{\mathrm{LM}}$) (range: –39.5, –11.5), SE($\hat{Z}_{\mathrm{WLM}}$) (range: –46.6, –31.0), SE($\hat{Z}_{\mathrm{GL}}$) (range: –53.8, –39.3), SE($\hat{Z}_{\mathrm{CR}}$) (range: –57.2, –39.9), and SE($\hat{Z}_{\mathrm{JN}}$) (range: –93.7, –78.1). In general, the effect of increasing ρ, decreasing subsample size and using m in expansion was for %BIAS to become

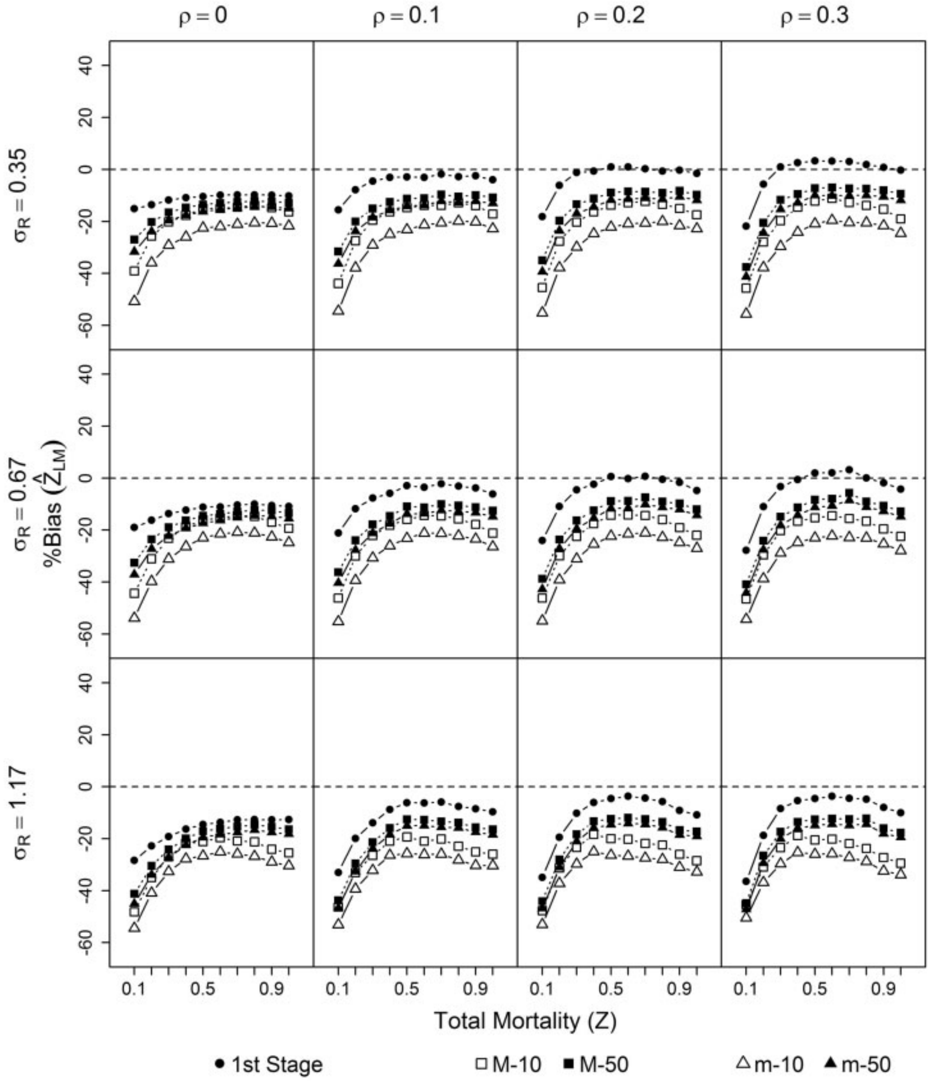

Figure 3. Percent bias (%Bias) of the standard linear model estimator of total mortality rate (Z) vs. the true total mortality under different levels of recruitment variation (σ_R), intra-cluster correlation (ρ), choice of using the total number caught (M), or the total subsample size (m) to derived age frequencies and third-stage subsample size (10 or 50 fish per haul).

less negative, although the trends in bias typically changed with increasing σ_R (Supplementary Figures S1–S6). Bias in SE($\hat{Z}_{\mathrm{CR}}$), SE($\hat{Z}_{\mathrm{GL}}$), SE($\hat{Z}_{\mathrm{WLM}}$), SE($\hat{Z}_{\mathrm{JN}}$), and SE($\hat{Z}_{\mathrm{RE}}$) was particularly sensitive to the expansion number used and third-stage subsample size, whereas bias in SE($\hat{Z}_{\mathrm{LM}}$) was particularly sensitive to subsample size regardless of the expansion number used (Supplementary Figures S1–S6). Plots of %BIAS in SE vs. %BIAS in Z showed that there were apparent relationships between bias for several estimators (Figure 7). %BIAS in SE became less negative as the bias in $\hat{Z}_{\mathrm{LM}}$, $\hat{Z}_{\mathrm{WLM}}$, and $\hat{Z}_{\mathrm{JN}}$ became more negative and as bias in $\hat{Z}_{\mathrm{RE}}$ became more positive (Figure 7).

Discussion

This study demonstrated that the performance of the six $\hat{Z}$ estimators examined herein are affected by ρ, subsample size in third-stage sampling, the choice of expansion number and the level of σ_R. In general, increasing ρ, small subsample size per haul, and use of total fish caught in all hauls to generate the age frequencies tended to increase bias (in negative or positive direction) in most $\hat{Z}$ estimators compared with the baseline estimates under the levels of σ_R examined, although to different degrees. Bias in $\hat{Z}_{\mathrm{CR}}$ and $\hat{Z}_{\mathrm{GL}}$ was influenced modestly by σ_R and ρ, but appeared relatively insensitive to the expansion number and subsample size. Bias in $\hat{Z}_{\mathrm{JN}}$ was very sensitive to σ_R and modestly sensitive to subsample size. Bias in $\hat{Z}_{\mathrm{LM}}$ and $\hat{Z}_{\mathrm{WLM}}$ was sensitive to increasing ρ and σ_R under first-stage sampling, but was most sensitive to decreasing sample size under third-stage sampling and expansion number. Increasing ρ, decreasing subsample size and using M to create age frequencies greatly increased bias in $\hat{Z}_{\mathrm{RE}}$.

Compared with work of other researchers, results of this study for scenarios under $\rho = 0$ and first-stage sampling were similar in magnitudes and trends in bias of Z estimators. Smith *et al.* (2012) and Millar (2015) under low ageing error scenarios also showed that bias in $\hat{Z}_{\mathrm{CR}}$ estimates was relatively low (under 22%), typically became more negative as Z increased, was moderately sensitive to increasing σ_R and as sample size decreased. Millar (2015)

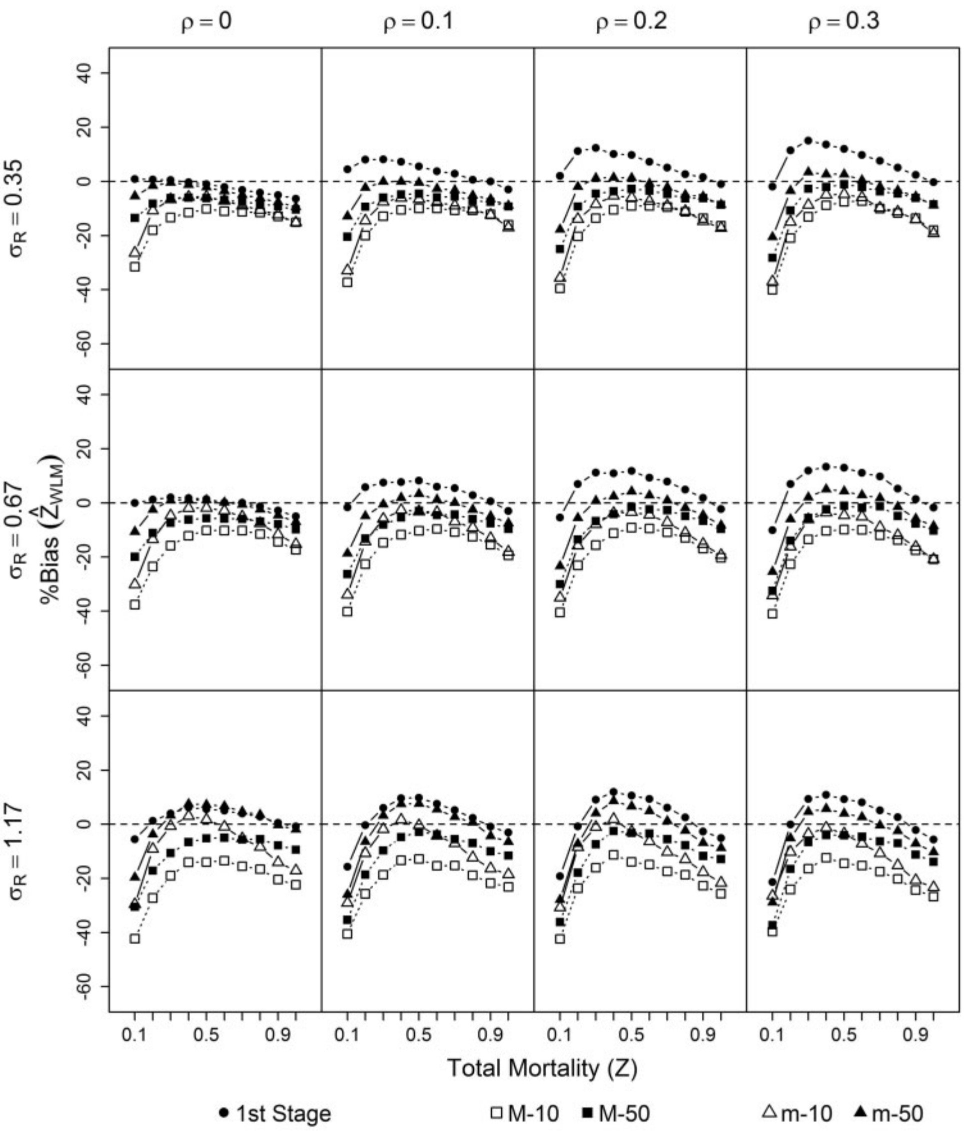

Figure 4. Percent bias (%Bias) of the weighted regression estimator of total mortality rate (Z) vs. the true total mortality under different levels of recruitment variation (σ_R), intra-cluster correlation (ρ), choice of using the total number caught (M), or the total subsample size (m) to derived age frequencies and third-stage subsample size (10 or 50 fish per haul).

found that the bias of $\hat{Z}_{\mathrm{GL}}$ was similar to $\hat{Z}_{\mathrm{CR}}$, but tended to be slightly less negative. For $\hat{Z}_{\mathrm{LM}}$, decreasing negative bias with increasing Z was also observed by Smith *et al.* (2012) for $Z > 0.3$ using the peak age criterion. For $\hat{Z}_{\mathrm{WL}}$, increasing negative bias with increasing Z was observed by Smith *et al.* (2012) for values of $Z > 0.3$. The emerging dome-shaped pattern in bias (high negative bias at low Z, low bias at intermediate Z and high negative bias at high Z) found in this study as σ_R increased and subsample size decreased was also observed by Millar (2015) across all low ageing error scenarios. Trends in bias of $\hat{Z}_{\mathrm{JN}}$ found in this study (declining bias with increasing Z) were similar to the trends in bias for the Heincke method described by Smith *et al.* (2012) but the magnitude shown in this study was much lower and more negative at intermediate and high Z values. The trends and magnitudes in bias of $\hat{Z}_{\mathrm{RE}}$ and sensitivity to σ_R were nearly identical to those observed in Millar (2015) for low ageing error scenarios.

The influence of ageing error on bias of the Z estimators was not examined like Millar (2015). Ageing error is likely a source of bias in reality, but the number of factors examined had to be restricted to a manageable size to make analysis and interpretation easier. Increased ageing error would likely increase estimator bias in the negative direction under the $\rho = 0$ scenarios, as found by Millar (2015), but the increased bias would be likely offset by the directional change in bias observed under the remaining scenarios.

Under the factors explored in this study, it was shown that estimates of standard errors were biased, although to different degrees. Under $\rho = 0$ and no sub-sampling, the baseline standard errors were under-estimated in all cases. Equivalent observations were noted by Smith *et al.* (2012) and Millar (2015) in their studies. Increasing ρ, lower subsample size and, depending on the estimator, the choice of expansion number caused bias to become less negative under the levels of σ_R examined compared with the

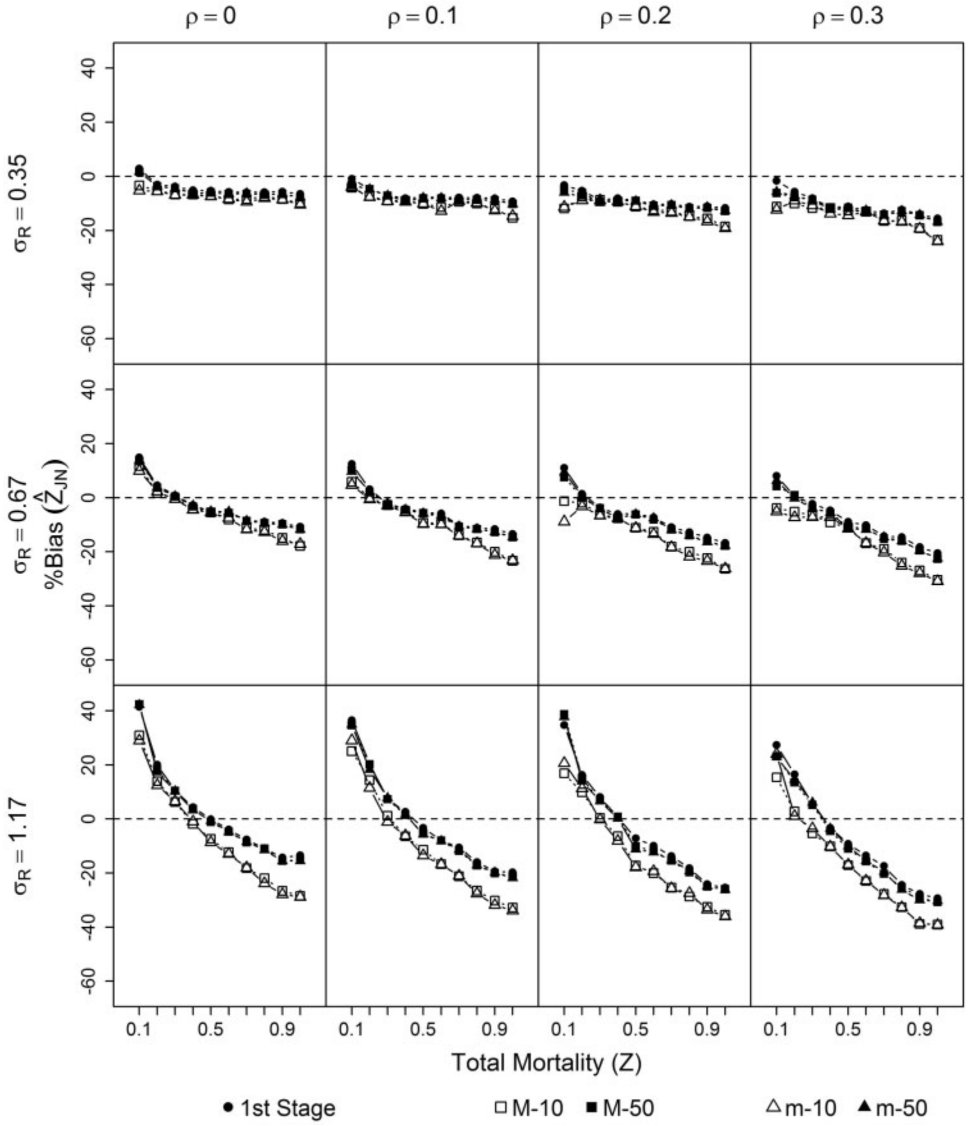

Figure 5. Percent bias (%Bias) of the modified Heincke estimator of total mortality rate (*Z*) vs. the true total mortality under different levels of recruitment variation (σ_R), intra-cluster correlation (ρ), choice of using the total number caught (*M*), or the total subsample size (*m*) to derived age frequencies and third-stage subsample size (10 or 50 fish per haul).

baseline estimates. One might interpret these findings as an improvement in the estimates of SE. However, it would be an incorrect conclusion because the declines in negative bias represent a change in the direction of bias, not an improvement. This feature is shown in Supplementary Figure S6 for SE($\hat{Z}_{\mathrm{RE}}$) where SE bias at high Z approached and then exceed 0% bias as ρ increased. If ρ continued to increase, bias would likely become even more positive given the observed trends.

Unfortunately, there was no single estimator that provided unbiased estimates of total mortality under simulated multi-stage cluster sampling. However, based on the range of bias in Z estimates and the apparent sensitivities to changes in the factors examined, there are two estimators that eclipsed the performance of the others: $\hat{Z}_{\mathrm{GL}}$ and $\hat{Z}_{\mathrm{CR}}$. These estimators produced the narrowest range of bias ($\hat{Z}_{\mathrm{GL}}$ range: –20.6%, 8.5%; $\hat{Z}_{\mathrm{CR}}$ range: –23.2%, 12.5%) under all scenarios (although $\hat{Z}_{\mathrm{GL}}$ performed slightly better), and were only modestly sensitive to changes in σ_R and ρ and rather insensitive to subsample size and expansion number. However, the standard errors produced by these estimators (and the remaining estimators) were quite biased as they were under-estimated in almost all cases. Therefore, standard errors should not be used in any type of statistical comparison because the type I error rate (probability of rejecting a true null hypothesis) will be greatly impacted and, in cases where SEs are under-estimated, significant differences will often be found where none exists (Nelson 2014).

The remaining estimators ($\hat{Z}_{\mathrm{JN}}$, $\hat{Z}_{\mathrm{LM}}$, or $\hat{Z}_{\mathrm{WLM}}$) should not be used because the former was very sensitive to recruitment variation and the latter two were moderately sensitive to all factors examined. In addition, the random-intercept Poisson log-linear mixed effects model should not be used when data are collected through multi-stage cluster sampling. If it were possible to collected fish individually and at random required under the model assumption, then $\hat{Z}_{\mathrm{RE}}$ would be the best estimator by far as shown

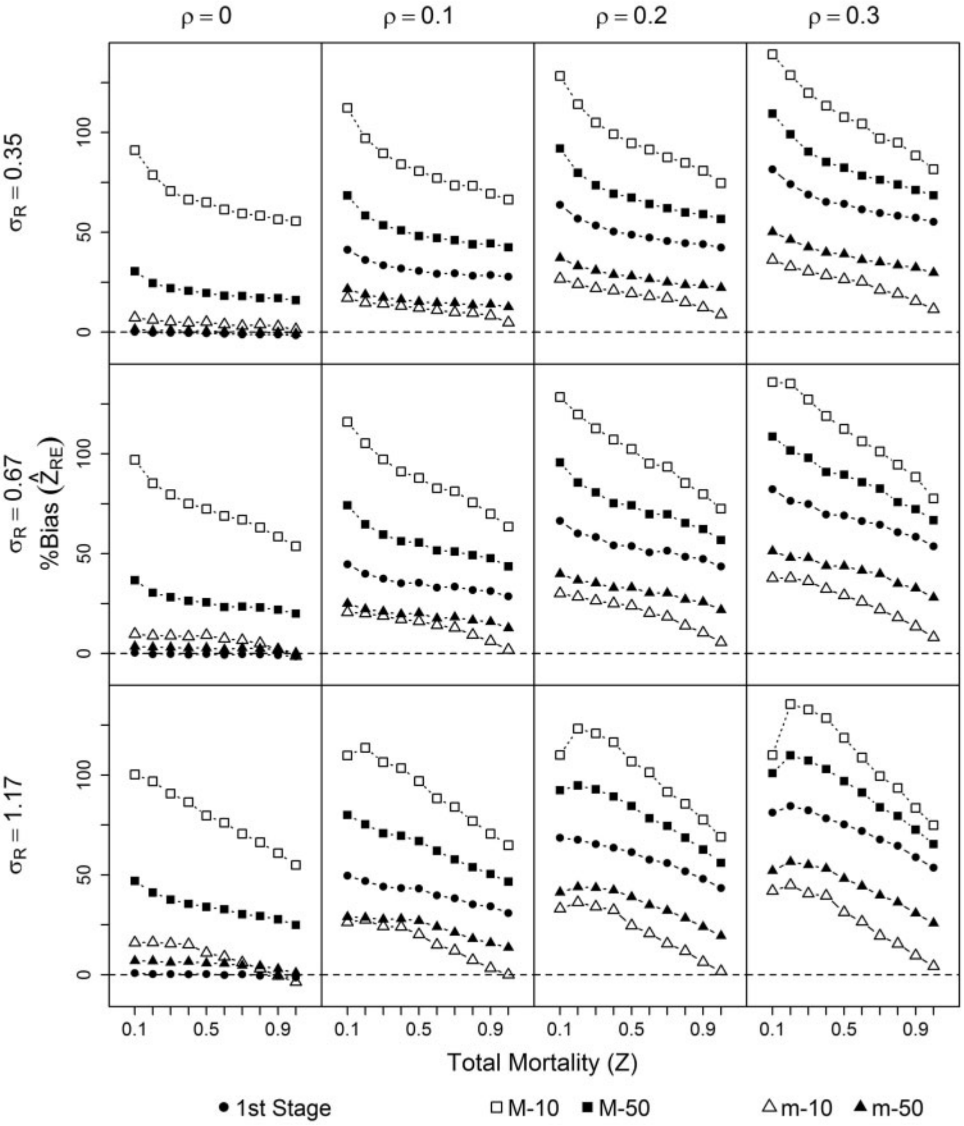

Figure 6. Percent bias (%Bias) of the random-intercept Poisson log-linear mixed model estimator of total mortality rate (Z) vs. the true total mortality under different levels of recruitment variation (σ_R), intra-cluster correlation (ρ), choice of using the total number caught (M), or the total subsample size (m) to derived age frequencies and third-stage subsample size (10 or 50 fish per haul).

by Millar (2015). However, when third-stage subsampling occurs, intra-cluster correlation is present, subsample size is low or the age frequencies are created by using the total number of fish caught in all hauls, the performance of the estimator degrades significantly. The degraded performance is likely the result of the signal in age-specific recruitment deviations being affected by the multi-stage random sampling because the proportions of older, low frequency age classes are over-represented in many cases. The random effects of those age classes are then over-estimated, which influences the steepness of the slope in numbers-at-age vs. age relationship, thus producing higher mortality rates. Surprisingly, the estimator performed better when the third-stage total sub-sample size was used to develop the age frequencies and 10 fish per haul were sub-sampled. This is likely an artefact of rounding the estimated numbers-at-age to whole digits, which eliminates the problematic, older, low frequency age classes that, when proportions are multiplied by small total subsample sizes, have fractional numbers <0.5.

This work showed that the least-biased estimators are the generalized Poisson linear model and the Chapman–Robson estimator, and these should be used to estimate total mortality when data are collected from fish trawl surveys. However, it should be kept in mind the only factor that can marginally reduce bias in $\hat{Z}_{GL}$ and $\hat{Z}_{CR}$ estimators, and is under control of the investigator, is sample size. As shown above, bias in $\hat{Z}_{GL}$ and $\hat{Z}_{CR}$ improves modestly as the subsample size increases, and this is directly related to the precision of the age composition estimates. In multi-stage cluster sampling, the best strategy for improving the precision of age composition estimates is to increase the number of hauls (or tows) that are made and sampled during a fish survey, not the number of fish sampled in each haul, because it is the variation among clusters that determines mostly the precision of

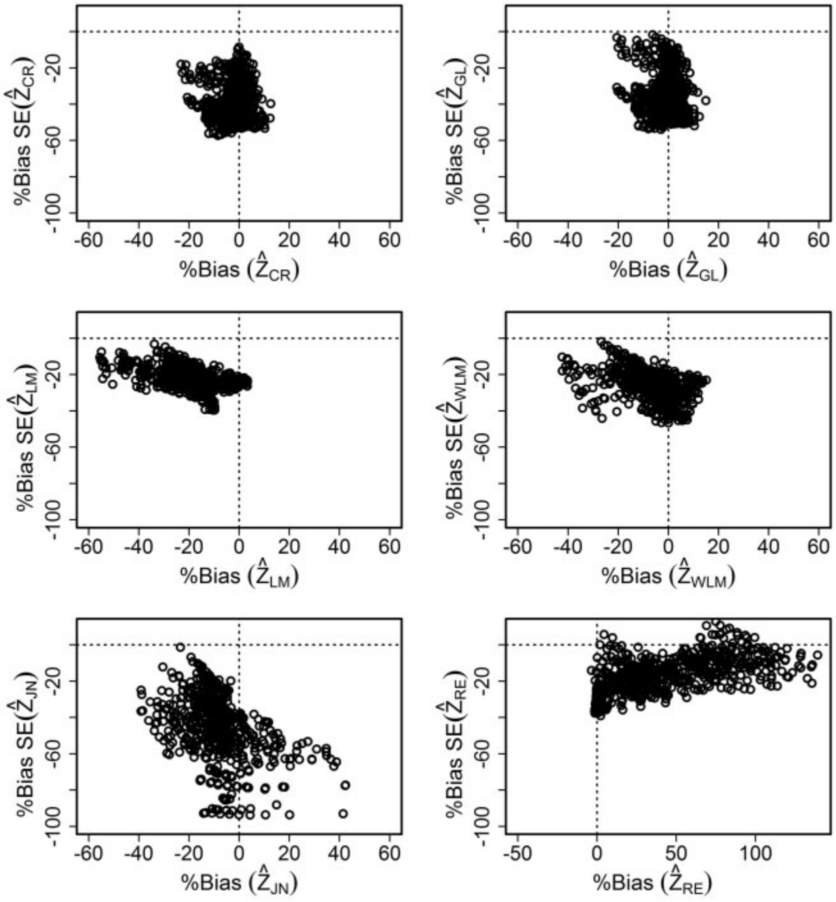

Figure 7. Percent bias of standard error estimators vs. percent bias of total mortality rates for the six methods examined.

estimates when intra-cluster correlation is high (Pennington and Volstad, 1994; Bogstad *et al.*, 1995; Aanes and Pennington, 2003). In fact, it appears the optimal sampling strategy is to measure fewer individuals from each haul (Bogstad *et al.*, 1995; Zhang and Cadrin, 2013; Aanes and Volstad, 2015). Therefore, the investigator must be aware that, if intra-cluster correlation is present and only a few hauls are available from which to sample, the precision of age composition estimates will be likely low (and bias in Z large), and increasing the number of fish sampled from each haul will not likely make a significant improvement in the bias of any total mortality estimator (see Supplementary Table S1).

None of the estimators examined herein was specifically formulated to incorporate intra-cluster correlation or the extra variation of multistage sampling. As stressed by one reviewer, the Poisson log-linear mixed modeling framework has the best potential for incorporating these sources of variability by modeling data from individual hauls and such configurations should be explored in the future. However, it should be noted that issues arise from these configurations (e.g. how to apply the “peak age” criterion to individual hauls) that will need to be assessed before such models are presented.

Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript.

Acknowledgements

Thanks to Micah Dean, Greg DeCelles, and two anonymous reviewers for helpful comments.

Funding

Funding for this study was provided, in part, by the U.S. Fish and Wildlife Service Sportfish Restoration Program Grant F-57-R.

References

ASMFC (Atlantic States Marine Fisheries Commission). 1994. Proceedings of the workshop on the collection and use of trawl survey data for fisheries management. ASMFC, Special Report 35, Arlington, Virginia.

ASMFC (Atlantic States Marine Fisheries Commission). 2017. River herring stock assessment update, volume I: coastwide summary. ASMFC, Stock Assessment Report, August 2017, Arlington, Virginia.

Aanes, S., and Pennington, M. 2003. On estimating the age composition of the commercial catch of Northeast Arctic cod from a sample of clusters. ICES Journal of Marine Science, 60: 297–303.

Aanes, S., and Volstad, J. H. 2015. Efficient statistical estimators and sampling strategies for estimating the age composition of fish. Canadian Journal of Fisheries and Aquatic Sciences, 72: 938–953.

Bogstad, B., Pennington, M., and Volstad, J. H. 1995. Cost-efficient survey designs for estimating consumption by fish. Fisheries Research, 23: 37–46.

Burnham, K. P., and Anderson, D. R. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd edn. Springer-Verlag, New York.

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

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

Heincke, F. 1913. Investigations on the plaice-general report: 1. Place fishery and protective measures, preliminary brief summary of the most important points of the report. Rapports et Proces-Verbaux des Reunions, Conseil International pour l'Exploration de la Mer 16.

Jensen, A. L. 1996. Ratio estimation of mortality using catch curves. Fisheries Research, 27: 61–67.

Lohr, S. L. 1999. Sampling: Design and Analysis. Duxbury Press, Boston.

Maceina, M. J., and Bettoli, P. W. 1998. Variation on largemouth bass recruitment in four mainstream impoundments of the Tennessee River. North American Journal of Fisheries Management, 18: 998–1003.

Millar, R. B. 2015. A better estimator of mortality rate from age-frequency data. Canadian Journal of Fisheries and Aquatic Sciences, 72: 364–375.

Nelson, G. A. 2002. Age, growth, mortality and distribution of pinfish (*Lagodon rhomboides*) in Tampa Bay and adjacent Gulf of Mexico water. United States Fishery Bulletin, 100: 582–592.

Nelson, G. A. 2014. Cluster sampling: a pervasive, yet little recognized survey design in fisheries research. Transactions of the American Fisheries Society, 143: 926–938.

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. 1996. Applied Linear Statistical Models, 3rd edn. WCB/McGraw-Hill, Boston, MA. 1408 p.

Pennington, M., and Volstad, J. H. 1994. Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. Biometric, 50: 725–732.

R Development Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available from: http://www.R-project.org./ (last accessed April 2018).

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Bulletin of the Fishery Research Board of Canada, 191: 1–382.

Seber, G. A. F. 2002. The Estimation of Animal Abundance and Related Parameters, 2nd edn. The Blackburn Press, Caldwell, New Jersey, 654 pp.

Smith, M. W., Then, A. Y., Wor, C., Ralph, G., Pollock, K. H., and Hoenig, J. M. 2012. Recommendations for catch-curve analysis. North American Journal of Fisheries Management, 32: 956–967.

Tvedebrink, T. 2013. Package *dirmult*: estimation in Dirichlet-Multinomial distribution. https://cran.r-project.org/web/packages/dirmult/dirmult.pdf (last accessed 6 May 2019).

Venables, W. N., and Ripley, B. D. 1999. Modern Applied Statistics with S-Plus. Springer-Verlag, New York Inc. 501 pp.

Wankowski, J. W., Hyduke, J. E., and William, S. 1988. Population age structure and mortality rates of jackass morwong (*Nemadactylus macropterus*) in Eastern Bass Strait, Australia. Fisheries Research, 6: 317–335.

Zhang, Y., and Cadrin, S. X. 2013. Estimating effective sample size for monitoring length distributions: a comparative study of Georges Bank groundfish. Transactions of the American Fisheries Society, 142: 59–67.

Handling editor: Stan Kotwicki