ICES Journal of
Marine Science

Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species

Amy Y. Then[1,2]*, John M. Hoenig[1], Norman G. Hall[3,4], and David A. Hewitt[5]

[1]*Virginia Institute of Marine Science, College of William & Mary, PO Box 1346, Gloucester Point, VA 23062, USA*
[2]*Institute of Biological Sciences, Faculty of Science, University of Malaya, Lembah Pantai, 50603 Kuala Lumpur, Malaysia*
[3]*Centre for Fish and Fisheries Research, Murdoch University, South Street, Murdoch, Western Australia 6150, Australia*
[4]*Department of Fisheries, Western Australian Fisheries and Marine Research Laboratories, PO Box 20, North Beach, Perth, Western Australia 6920, Australia*
[5]*US Geological Survey, Western Fisheries Research Center, Klamath Falls, OR, USA*

Corresponding author: tel: +60128977292; fax: +60379674178; e-mail: amythen79@gmail.com, amy_then@um.edu.my

Many methods have been developed in the last 70 years to predict the natural mortality rate, M, of a stock based on empirical evidence from comparative life history studies. These indirect or empirical methods are used in most stock assessments to (i) obtain estimates of M in the absence of direct information, (ii) check on the reasonableness of a direct estimate of M, (iii) examine the range of plausible M estimates for the stock under consideration, and (iv) define prior distributions for Bayesian analyses. The two most cited empirical methods have appeared in the literature over 2500 times to date. Despite the importance of these methods, there is no consensus in the literature on how well these methods work in terms of prediction error or how their performance may be ranked. We evaluate estimators based on various combinations of maximum age (t_{max}), growth parameters, and water temperature by seeing how well they reproduce >200 independent, direct estimates of M. We use tenfold cross-validation to estimate the prediction error of the estimators and to rank their performance. With updated and carefully reviewed data, we conclude that a t_{max}-based estimator performs the best among all estimators evaluated. The t_{max}-based estimators in turn perform better than the Alverson–Carney method based on t_{max} and the von Bertalanffy K coefficient, Pauly's method based on growth parameters and water temperature and methods based just on K. It is possible to combine two independent methods by computing a weighted mean but the improvement over the t_{max}-based methods is slight. Based on cross-validation prediction error, model residual patterns, model parsimony, and biological considerations, we recommend the use of a t_{max}-based estimator ($M = 4.899t_{max}^{-0.916}$, prediction error = 0.32) when possible and a growth-based method ($M = 4.118K^{0.73}L_\infty^{-0.33}$, prediction error = 0.6) otherwise.

Keywords: Alverson and Carney, data limited, data poor situations, fish mortality, Hoenig, indirect estimators of M, Jensen, natural mortality, Pauly, prior distribution.

Introduction

One of the most influential stock assessment parameters, natural mortality rate (M), is generally believed to be difficult to estimate reliably and directly. By direct, we refer to estimation of M using information strictly pertaining to the species or stock of interest. Five examples are (i) measuring total mortality in an unexploited stock, (ii) relating total mortality to the amount of fishing and extrapolating to zero fishing effort, (iii) measuring both total mortality and exploitation rates and solving for components of mortality (e.g. Hewitt *et al.*, 2007), (iv) mark-recapture and telemetry studies (e.g. Hoenig *et al.*, 1998; Knip *et al.*, 2012), and (v) estimating M internally in an integrated stock assessment model (see Maunder and Punt, 2013). Direct estimation methods of M are often data intensive, thus limiting their application to relatively data-rich stocks.

A host of methods have been developed in the last 70 years to estimate M from surrogate life history information. These

life history correlates include maximum age t_{max} (Tanaka, 1960; Bayliff, 1967; Ohsumi, 1979; Hoenig, 1983), von Bertalanffy growth coefficient K (Beverton and Holt, 1959; Ralston, 1987; Charnov, 1993; Jensen, 1996) as well as composites of these variables—for example, the von Bertalanffy asymptotic size L_∞ or W_∞, growth coefficient K, and water temperature T (Pauly, 1980), and both t_{max} and K (Alverson and Carney, 1975). Ecological theory and empirical evidence provides strong basis for prediction of M from surrogate information not only for fish stocks but also for other animals and even plants (Hoenig, 1983; McCoy and Gillooly, 2008).

We use the term indirect or empirical to categorize this suite of methods since their derivation relies on comparative life history studies to borrow strength from many species (e.g. Pauly, 1980; Hoenig, 1983). Some models for estimating M have been derived based on theoretical ecological considerations; these models constitute empirical methods by our definition if they rely on data to estimate one or more unknown parameters (e.g. Alverson and Carney, 1975; Gunderson and Dygert, 1988).

Although these empirical methods are often perceived as being less reliable than their data-rich counterparts, a consensus is that empirical methods are useful and very important particularly in a data-poor setting (e.g. Brodziak et al., 2011). Empirical methods are routinely applied in stock assessments, both for data-poor and data-rich stocks, in the following ways: (i) obtain point estimates of M in the absence of direct information, (ii) examine the reasonableness of a directly estimated value of M, (iii) obtain a range of plausible values for M for the stock by applying a suite of empirical methods, and (iv) define prior distributions of M in Bayesian analyses.

A large body of evidence suggests that M varies over age and size (e.g. Peterson and Wroblewski, 1984; McGurk, 1986; Lorenzen, 1996; Gislason et al., 2010). Nonetheless, most fisheries scientists would agree that a single value for M can provide a useful representation of mortality over much of the exploitable lifespan of a species. Simulation studies have indicated that the assumption of a constant M in stock assessments is still very useful even when the simulated populations are subject to age- and time-varying M dynamics (Deroba and Schueller, 2013; Johnson et al., this issue). We proceed on the assumption that it is worthwhile to seek better ways to predict a single (constant) value of M for a stock.

Among the empirical estimators of M, the Pauly (1980) and Hoenig (1983) log-transformed linear regression equations are the two most widely applied estimators. The former regression was fitted to 175 fish stocks (113 unique species; see Griffiths and Harrod, 2007), whereas the latter was derived based on a total of 130 stocks (51 species of fish, 11 of molluscs, and 13 of cetaceans). As of August 5, 2014, the number of citations for Pauly (1980) equalled 2129 while Hoenig (1983) had been cited 936 times (Google Scholar http://scholar.google.com). The widespread use of these methods can be attributed to (i) simplicity and ease of application, (ii) perceived reliability given the amount of empirical information used to "train" their derivation, and (iii) minimal data required to apply them.

Given the importance of empirical estimators, it is surprising there has not been a comprehensive study to compare their performance. Kenchington (2013) described 30 estimators and minor variants and applied them to 12 species of fish and one invertebrate but, with a small sample size, it was difficult to draw general conclusions regarding the performance of the methods. Various studies that employed a suite of these empirical estimators have noted a wide disparity in resulting estimates [e.g. empirical M estimates

ranged from 0.134 to 0.706 yr^{-1} for the tropical lutjanid *Lutjanus quinquelineatus* (Newman et al., 1996), 0.003−0.14 yr^{-1} for *Sebastes variabilis* (Malecha et al., 2007), 0.3−2.2 yr^{-1} for the blue crab (Hewitt et al., 2007)]. There is no consensus on why M estimates from different empirical predictors vary so widely or which should be preferred. In stock assessments, it is a common practice to utilize multiple estimates of M as a means of characterizing the uncertainty of M; this assumes implicitly that all the estimates are equally reliable and independent (but see Hamel, this issue).

A major obstacle to such a comparative endeavour is the lack of a comprehensive dataset with all necessary estimates of life history variables used in the empirical estimators of interest. In addition to a comprehensive review of predictive ability, it is perhaps of even greater value to improve upon the existing empirical estimators with an updated extensive dataset of M and life history estimates.

The Gunderson and Dygert (1988) and the Pauly (1980) estimators have been "updated" with larger and presumably better datasets [the former by Gunderson (1997) and the latter by Jensen (2001) and Griffiths and Harrod (2007)]. Punt et al. (2005, Table 14, p. 38) presented updated regression estimators for both Pauly (1980) and Hoenig (1983). Beverton and Holt (1959) and Beverton (1963) noted a relationship between M and K. Charnov (1993) termed this a "Beverton and Holt life history invariant" relationship; he fitted the model $M = a/K$ which we refer to as the one-parameter, K-based method. Jensen (1996) updated this relationship and later proposed a two-parameter K method using a subset of Pauly's data (Jensen, 2001). However, it is unclear if the updated estimators predict M rates of fish significantly better than the original formulations of the estimators.

The focus of this paper is on improving the point estimation of M for teleosts and elasmobranchs. The challenge to address appropriately the questions outlined above is the fact that we do not know the true M of any fish. This raises the question: How might we objectively evaluate and rank the predictive ability of these estimators? The approach we took in this paper is to ask which estimator best reproduces what we know about M, which is the collection of direct estimates of M in the scientific literature and what we can estimate from both published and unpublished data provided by various fisheries scientists around the world.

Hence, the goal of this paper is to compare and rank the predictive abilities of four major empirical estimation approaches for M, namely the K-based methods and those of Alverson and Carney (1975), Pauly (1980), and Hoenig (1983), and as well as variants of these estimators. We also explored the possibility of weighting independent estimators to improve M prediction. Specifically, this paper seeks to answer the following questions: (i) How well do the estimators perform in predicting M estimates not used to train the original equation? (ii) When evaluated on a common dataset, how do the various methods rank in terms of prediction error? (iii) What are the updated model parameters for the preferred estimators when evaluated with a larger, better dataset? and (iv) Would a weighted combination of estimators improve prediction of M?

In conjunction with this research, we have compiled a dataset of direct M estimates and best-matched life history parameters of >200 unique fish species with documentation of the methods used to estimate M as well as the ageing methods (where applicable). The need for a well-documented database of M rates of species has been outlined in Brodziak et al. (2011) based on recommendations from a national workshop on estimation of mortality rates.

One basic problem with using compilations of parameter estimates is that the quality of individual values can be highly variable

and the quality is often unknown or controversial. One investigator's inclusion of a particular stock may be rejected by another investigator [e.g. Gislason *et al.* (2010) included data for a serranid but Kenchington (2013) rejected those data]. This is an inherent and unavoidable aspect of comparative life history studies. Gislason *et al.* (2010) developed and applied a set of quality control criteria to select usable data. This seemingly logical approach can potentially introduce biases because the accepted species may be the best studied species and these might not be typical of the wider collection of species about which one wishes to make inferences about natural mortality. Gislason *et al.* (2010) also highlighted assumptions that had to be made and other limitations in collating their dataset and subsequent analyses.

We have eliminated obviously erroneous data and will make the full, documented dataset available to the public in July of 2015 (hosted at http://bit.ly/vims_mort). This will allow researchers to filter the data according to whatever criteria they deem appropriate and explore whether the results are robust to choice of data selection criteria.

Methods
Data compilation and quality control
We included in the dataset fish stocks that are of commercial and sport value as well as those that are currently of no economic importance. We utilized existing compilations of estimates of *M* (e.g. Pauly, 1980; Hoenig, 1982; McGurk, 1986; Gislason *et al.*, 2010) as well as extensive literature searches of published materials (journal articles, stock assessment reports, grey literature). We verified that each *M* estimate was derived from a direct method by examining the original sources of the *M* estimates. Similarly, the original source document was consulted to verify the supporting life history information. We also derived new estimates of natural mortality rate using our own data and data supplied by colleagues. Exploitation level of species was determined primarily based on the judgment or assessments made by the authors of the age-growth or mortality studies. Catch curve-based estimates for stocks that were clearly heavily exploited according to any source document were not included.

A noteworthy mention is the heavy reliance of the Pauly (1980) and Hoenig (1982) datasets on the mortality compilation of Beverton and Holt (1959). We examined the original sources of these *M* estimates and found many of the sources did not themselves provide direct *M* estimates. Beverton and Holt (1959) most likely estimated *M* as well as the von Bertalanffy parameters based on the data (mostly some form of numbers-at-age and mean length-at-age data) in these source papers. For some of these stocks, we re-estimated the *M* and the von Bertalanffy parameters primarily to validate the published estimates in Beverton and Holt (1959); for catch-at-age data, *M* was re-estimated using the Chapman and Robson (1960) estimator based on the recommendations given in Smith *et al.* (2012). In general, the selected age of full recruitment corresponded to the age with peak catch plus one, with the exception of short-lived species (at least three age groups were used to estimate *M*) and the cases in which the age of full recruitment was specified by the source authors.

Where available, *M* (yr^{-1}), L_∞ (mm), *K* (yr^{-1}), t_{max} (yr), and *T* (°C) estimates were extracted from the same study. Otherwise, we searched the literature to find the set of parameters that best matched the *M* estimates in terms of location and timing of study. When a range of values were given (say *M* estimates that were based on three different catch curve methods), we used the mean

value. When the *M* estimate was pooled for both sexes but only sex-specific estimates were given for L_∞ and *K*, the sex-specific values were averaged. Mean *T* for a stock was either obtained from the same mortality-growth study, similar stock area (where *T* was available for another species in the same area), FishBase (http://www.fishbase.org), or estimated loosely based on the general criteria outlined in Pauly (1980). For certain species, no estimates were available of von Bertalanffy parameters (due to perceived linearity in growth by authors) or t_{max} (due to lack of ageing work). We did not attempt to "borrow" parameter estimates for these, but left them out of the analysis. Length measurements were not standardized; they were comprised primarily of total (36%) and fork length measurements (32%) with 32% unspecified or other types.

One of the assumptions of fitting a linear regression is that the observations themselves are independent of each other. The inclusion of multiple *M* estimates for a single species (stocks from multiple locations or separate male and female estimates) very likely violates that assumption. For instance, Pauly (1980) had a total of nine estimates of *M* and corresponding parameters for the Atlantic cod while Hoenig (1982) included eight estimates for the cisco. To tackle this issue, one could consider modelling the hierarchical structure, or use only a single observation for a given species. We elected to use the latter approach because most genera and species were represented by single observations, thus making it difficult to estimate components of variability. We selected the best set of estimates for a species based primarily on the ageing method, validation of the ages, sample size as well as the author's and our evaluation of the reliability of the estimates. An exception to this is the lake whitefish (*Coregonus clupeaformis*) where we included both the "normal" (*M* = 0.15) and "dwarf" (*M* = 1.3) forms whose differences in mortality and growth were vast (data from Beverton and Holt (1959))—we considered these forms as two distinct "species" in the context of this analysis.

Evaluation and updating of estimators
We evaluated the approaches of Pauly (1980), Hoenig (1983), and Alverson and Carney (1975) and the *K*-based estimators and various variants arising from differences in model formulation and fitting procedures (Table 1). Of note is a one-parameter estimator based on t_{max} given by $M_{est} = a/t_{max}$, where M_{est} denotes a new prediction of *M* and *a* is a coefficient that has been estimated (by least squares) from a compilation of estimates of *M* and t_{max} from the literature (Bayliff, 1967; Ohsumi, 1979; Hewitt and Hoenig, 2005). This model corresponds to setting the slope to -1 in the model formulation of Hoenig (1983). Some investigators have fixed the value of *a* at an arbitrary value rather than estimating it but there is little justification for this (see Hewitt and Hoenig, 2005) and such estimators are not considered here. Evaluation of taxa-specific estimators [e.g. Serranidae and Lutjanidae (Ralston, 1987; Pauly and Binohlan, 1996) and Engraulidae (Bayliff, 1967)] is also not considered here. The full list of all the variants of the estimators evaluated, including bias-correction of the Hoenig (1983) estimator, is recorded in Then and Hoenig (2014).

We evaluated the predictive performance of the estimators based on three metrics: (i) "Historical Performance"—how well do the original estimators predict estimates of *M* that were not used to train the original estimators? (ii) "Approach Evaluation"—how well do the different estimators compare with each other when refitted to, and evaluated on, a common dataset? and (3) "Updated Performance" –when updated using the fullest applicable dataset,

Table 1. The estimation approaches for predicting natural mortality, *M*, investigated in this study.

Model name	Formula	Fitting method	Model used empirically by
t_{max}^a			
One-parameter t_{max}	$M = a/t_{max}$	nls	Tauchi (1956), Tanaka (1960), Bayliff (1967), Ohsumi (1979)
Hoenig$_{lm}$	$\log(M) = a + b\log(t_{max})$	ls	Hoenig (1983)
Hoenig$_{gm}$	$\log(M) = a + b\log(t_{max})$	gm	Hoenig (1983)
Hoenig$_{nls}$	$M = at_{max}^b$	nls	This study
K^b			
One-parameter K	$M = aK$	ls	Beverton and Holt (1959), Beverton (1963), Charnov (1993), Jensen (1996)
Two-parameter K	$M = a + bK$	ls	Ralston (1987), Jensen (2001)
K, L_∞, and T			
Pauly$_{lm}$ [c]	$\log(M) = a + b\log(K) + c\log(L_\infty) + d\log(T)$	ls	Pauly (1980), Pauly and Binohlan (1996)
Pauly$_{lm\text{-}T}$ [c]	$\log(M) = a + b\log(K) + c\log(L_\infty)$	ls	This study
Pauly$_{nls}$	$M = aK^bL_\infty^cT^d$	nls	This study
Pauly$_{nls\text{-}T}$	$M = aK^bL_\infty^c$	nls	This study
K and t_{max}			
Alverson–Carney	$M = 3K/(e^{aKt_{max}} - 1)$	nls	Alverson and Carney (1975)
Composites[d]			
Weighted M	$M = pM_{Estimator1} + (1 - p)M_{Estimator2}$	NA	This study

ls, least squares; nls, non-linear least squares; gm, geometric mean or functional regression, as described by Ricker (1975); NA, not applicable.
[a]One-parameter t_{max} refers to the estimator, where *a* is an estimated parameter (not a value assumed for theoretical reasons). Tauchi (1956) apparently proposed that *M* should be proportional to the reciprocal of t_{max} (see Ohsumi, 1979). Tanaka (1960) fitted a line to *M* vs. $1/t_{max}$ for five species but did not propose an actual empirical estimator as far as we could tell.
[b]Beverton (1963) and Beverton and Holt (1959) plotted *M* vs. *K* and established that a relationship exists but did not fit a model.
[c]Although the original model formulation in Pauly (1980) was in base 10, we examined variants of the Pauly model in natural log scale for consistency.
[d]The combination of Estimator1 (with weighting of *p*) and Estimator2 (with weighting of $1 - p$) examined for the weighted *M* estimator consists of possible pairs of the estimators listed above that do not have predictor variables in common.

how well do the estimators reproduce the *M* estimates? The three metrics differed in the subsets of data used in the evaluation exercise.

The main purpose of the Historical Performance analysis is to determine how well the original estimators have served stock assessment scientists. For the Historical Performance analysis, we applied the original formulas to a new (independent) dataset and calculated the root-mean-square prediction error (RMSE), which is given as $\sqrt{\sum_{i=1}^{n}(M_{est,i} - M_{lit,i})^2/n}$, where $M_{est,i}$ is the predicted *M* for species *i* from the estimator and $M_{lit,i}$ is the literature-derived estimate for species *i*, for *i* = 1,...,*n* species. We evaluated the published log–log regression equations of Pauly (1980) and Hoenig (1983). We did not evaluate the historical one-parameter t_{max} or the two-parameter *K* models since the same regression form has multiple coefficient estimates by different authors (see Table 1). Alverson and Carney (1975) estimated the constant in their equation based on 63 data points, which were not provided in their paper. Jensen's (1996) one-parameter *K*-based equation was derived from the substitution of a derived equation into another without provision of published data. Hence, we evaluate the Historical Performance of the Alverson and Carney (1975) and Jensen (1996) estimators based on the fullest dataset recognizing that a small portion of the data may occur in both the derivation and the evaluation.

The Approach Evaluation was conducted using the common dataset (215 species) with complete sets of parameter estimates available for all estimators. For both the Approach Evaluation and the evaluation of Updated Performance, we also explored fitting the Pauly and Hoenig models as power functions using non-linear least squares, thus modelling *M* directly, in addition to evaluating the original log-transformed linear regression formulation (Table 1). This corresponded to fitting the model $M = aK^bL_\infty^cT^d + \varepsilon$ for

Pauly and $M = at_{max}^b + \varepsilon$ for Hoenig, where *a*, *b*, *c*, and *d* are parameters to be estimated and ε is a random error. To distinguish these estimators from the log–log linear versions, we denoted them as Pauly$_{nls}$ and Hoenig$_{nls}$, respectively. For the Pauly models, we fitted the full model as well as one excluding the temperature variable *T* to examine the importance of *T* in prediction of *M*. We used the Wald likelihood ratio test to compare the nested models.

For the tests of the Approach Evaluation and the Updated Performance, we used the tenfold cross-validation technique (Hastie *et al.*, 2009). We used the prediction RMSE as the overall metric of performance (from now on referred to as cross-validation prediction error, CVPE) and also presented as measures of model fit the coefficient of determination (unadjusted r^2) and the mean absolute difference (MAD) between the literature estimates and estimator-derived *M* predictions. In addition, we took into consideration the behaviour of resulting model residuals. For the log–log linear models of Hoenig$_{lm}$ and Pauly$_{lm}$, the predicted *M* estimates used for the model evaluation and calculation of CVPE were the exponentiated log(M_{est}).

We calculated the coefficient of variation (CV) for the log(M_{est}) for both the Hoenig$_{lm}$ and Pauly$_{lm}$ models based on the current dataset. Specifically, we fitted both models to obtain the residual standard error (SE). For the lognormal distribution, the CV can be estimated as $\sqrt{e^{SE^2} - 1}$, where SE is the SE of the log-transformed M_{est} (Johnson *et al.*, 1994). This allows us to compare the CV of the Hoenig$_{lm}$ and Pauly$_{lm}$ models calculated as 0.53 and 0.61, respectively, in MacCall (2009).

We investigated the possibility of combining the best estimators (with lowest CVPE) using a weighting scheme to create the best linear unbiased combination of estimators. This was based on the fact that for two normally distributed, unbiased and independent

estimators, the variance of the optimally weighted combined mean will be lower than the individual estimators' variances. The updated Hoenig and Pauly estimators could be assumed as independent since there is no overlap in the variables used in each estimator and their M estimates could thus be combined as a weighted mean. The form of the weighted estimator of M is given as $pM_{\text{Estimator1}} + (1 - p)M_{\text{Estimator2}}$ and the weighting factor p was calculated from $p = \text{Var}_{\text{Estimator2}}/(\text{Var}_{\text{Estimator1}} + \text{Var}_{\text{Estimator2}})$. We examined the independence of the pairs of estimators via residual and correlation plots.

We used 5000 bootstrap samples to estimate the SE and 95% bias-corrected and accelerated (BCa) confidence intervals (CI) of the model parameters for the best set of updated estimators. All analyses and plotting were conducted using the R statistical programming language (R Development Core Team, 2012). The least-squares fitting was done using the lm routine while the non-linear least-squares fitting was performed with nls. The Wald likelihood ratio test was conducted using the lmtest package (Zeileis and Hothorn, 2002). Geometric mean regressions were computed from the output of the lm routine using the procedure described by Ricker (1975). The bootstrapping was implemented with the boot package (Davison and Hinkley, 1997; Canty and Ripley, 2011).

Results

Dataset description

Our "common" dataset with complete parameter estimates for K, L_∞, $t_{\max}$, and T comprises 215 unique fish species from 22 orders, 66 families, and 143 genera with direct M estimates ranging from 0.014 to 5.07 yr^{-1}. Most of the M estimates were derived from age-based catch curve analyses (79%), 5% from length-based catch curves, 8% from tagging, and 2% from the regression of total mortality rate vs. effort (or variants). Of the 229 species in the full dataset, three species did not have corresponding $t_{\max}$ estimates while 12 were missing corresponding von Bertalanffy growth estimates. These additional species were included in the "Updated Performance" evaluations.

The dataset comprises primarily marine (81%) and temperate (43%) species, but a broad range of reproductive strategies and habitat associations are represented as well. Ninety-eight per cent of the stocks are teleosts. The dataset is largely weighted by stocks from the North American, Australian, and European continents and their surrounding water bodies. The von Bertalanffy growth K estimates ranged from 0.012 to 2.56 yr^{-1}, L_∞ from 49 to 3164 mm, and $t_{\max}$ from 38 weeks to 205 years. Mean T of stocks ranged from 4 to 30°C after accounting for physiological adjustments (as described in Pauly, 1980). The strongest linear predictor for M (in natural log scale) is the log-transformed $t_{\max}$ ($r = -0.92$; Figure 1). This is followed by the log-transformed K ($r = 0.61$; Figure 1). The literature $\log(M)$ estimates appeared to be fairly normally distributed, suggesting an underlying lognormal distribution for M (Figure 1). When age-based M estimates from the literature were excluded, the correlation between $\log(M)$ and $\log(t_{\max})$ remained high ($r = -0.81$, $n = 46$). The correlation between M and K is stronger than the log-transformed counterpart ($r = 0.68$), and M/K ratios ranged from 0.06 to 21.2.

Evaluation of estimators and their performance

In all, 8 of 82 stocks from Hoenig (1982) and 14 of 175 entries from Pauly (1980) with exact matching of all variables were included in the present dataset. It should be noted that the datasets of Hoenig
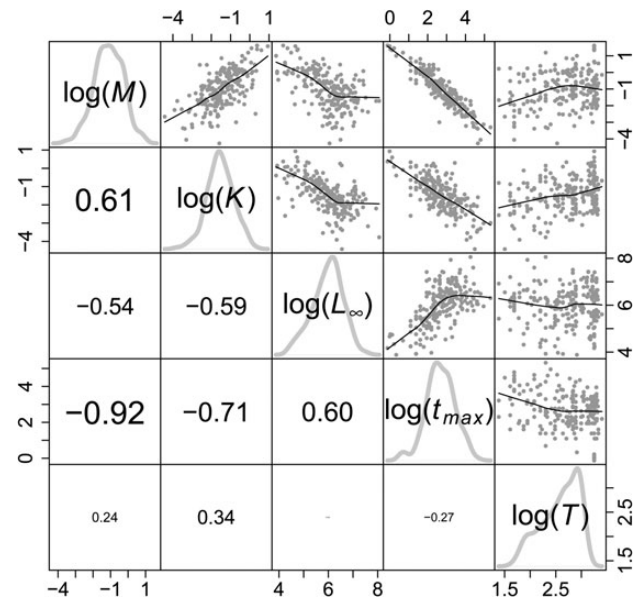

Figure 1. Scatterplot of pairs of log-transformed variables in the upper half of the panel, with locally weighted scatterplot smoothing (LOWESS) lines added (smoothing parameter $f = 2/3$). Variables: natural mortality rate M, maximum age $t_{\max}$, von Bertalanffy growth parameters K and L_∞, and mean temperature T. Kernel density plots of the log-transformed variables are shown in the diagonal panels. Correlation coefficients (r) for variable pairs are shown in the lower half of the panel, where the font size corresponds to the magnitude of the r values.

(1982) and Pauly (1980) were not combed through exhaustively, some of the source papers could not be located, age and growth studies for some of the species have been updated, and for species with multiple estimates only one estimate was included based on our criterion. The inclusion of only a small fraction of the Hoenig (1982) and Pauly (1980) stocks in the present dataset is not necessarily a reflection of the quality of their datasets.

The Historical Performance evaluation of the estimators in Table 1 based on the RMSE indicated that the Hoenig (1983) estimators performed the best, followed by the Alverson and Carney (1975), Pauly$_{\text{lm}}$ (1980), and Jensen (1996) one-parameter K-based estimators (Table 2). The Hoenig$_{\text{gm}}$ method performed slightly better than that of Hoenig$_{\text{lm}}$. The original Hoenig (1983) equation appeared to give lower estimates of M relative to the original Pauly (1980) equation for stocks with literature M estimates <0.9 yr^{-1}; as the literature M estimates increase, the Hoenig (1983) equation predicted higher M estimates than the Pauly (1980) equation (Figure 2a).

For the Approach Evaluation of individual estimators, the Hoenig$_{\text{nls}}$ method performed the best among the $t_{\max}$-based estimators, followed closely by the one-parameter $t_{\max}$, Hoenig$_{\text{lm}}$, and distantly by Hoenig$_{\text{gm}}$ (Table 2). Among the estimators that are not based solely on $t_{\max}$, the Alverson–Carney approach performed the best, followed by the Pauly$_{\text{nls}}$, Pauly$_{\text{nls-}T}$, two-parameter K, one-parameter K, Pauly$_{\text{lm}}$, and Pauly$_{\text{lm-}T}$ approaches. The K-based estimators and all the Pauly-type approaches are comparable with each other performance-wise. With the exception of Hoenig$_{\text{gm}}$, the $t_{\max}$-based estimators were clearly better than the other types of estimators evaluated, with approximately twice the predictive ability over the K-based and Pauly-type approaches. It was also evident that the

Table 2. Cross-validation prediction error (CVPE) of empirical estimators of natural mortality, *M*, evaluated in terms of three tests: Historical Performance, Approach Evaluation, and Updated Estimators.

Models	Historical		Approach (n = 215)	Updated			
	RMSE	*n*	CVPE	r^2	MAD	*n*	CVPE
t_{max}							
One-parameter t_{max}	NA	–	0.305	0.89	0.19	226	0.317
Hoenig$_{lm}$	0.36	219	0.328	0.89	0.19	226	0.329
Hoenig$_{gm}$	0.34	219	0.510	0.88	0.24	226	0.519
Hoenig$_{nls}$	NA	–	0.281	0.89	0.19	226	0.323
K							
One-parameter K	0.59	218	0.582	0.46	0.37	218	0.593
Two-parameter K	NA	–	0.580	0.46	0.38	218	0.591
K, L_∞, and T							
Pauly$_{lm}$	0.56	203	0.605	0.51	0.34	215	0.605
Pauly$_{lm-T}$	NA	–	0.610	0.49	0.35	218	0.627
Pauly$_{nls}$	NA	–	0.577	0.53	0.35	215	0.577
Pauly$_{nls-T}$	NA	–	0.578	0.50	0.36	218	0.597
K and t_{max}							
Alverson–Carney	0.42	215	0.414	0.81	0.26	215	0.414
Weighted M							
Hoenig$_{lm}$ _Pauly$_{lm}$	NA	–	0.298	0.86	0.19	215	0.298

The MAD and the coefficient of determination (unadjusted r^2) of predicted and literature *M* estimates are presented as well for the updated estimators. *n* denotes sample size. NA denotes not applicable. See Table 1 for definition of models. The Approach Evaluation column provides a fair test of the approaches. The Updated Estimators column provides an evaluation of each estimator based on the most data for training and checking the estimator.
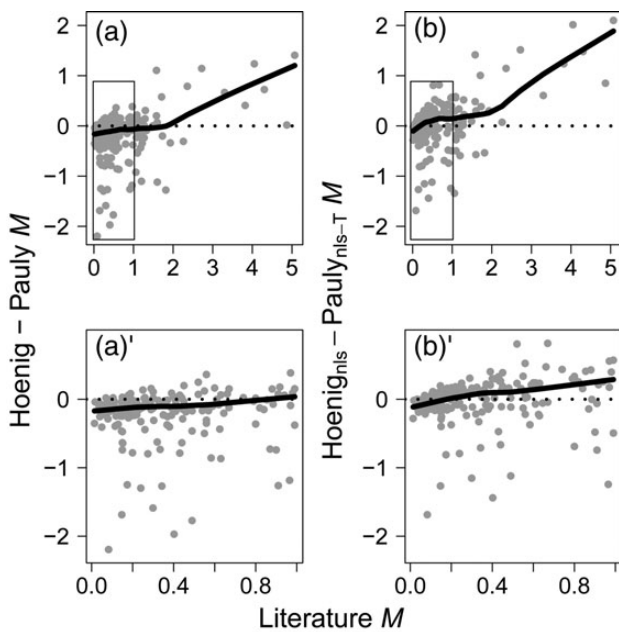
Figure 2. Difference in the predicted *M* estimates between (a) the original Hoenig (1983) and Pauly (1980) estimators as well as (b) the updated Hoenig$_{nls}$ and the Pauly$_{nls-T}$ estimators, plotted against literature *M* estimates (*n* = 215). (a)′ and (b)′: The enlargement of the boxed regions in (a) and (b), respectively. LOWESS lines (solid) shown in each panel (smoothing parameter *f* = 2/3).

non-linear least-squares fitted estimators of Hoenig and Pauly were better than the log–log regression counterparts, based on the lower CVPE. The CVPE values were very similar for both the Pauly$_{lm}$ and
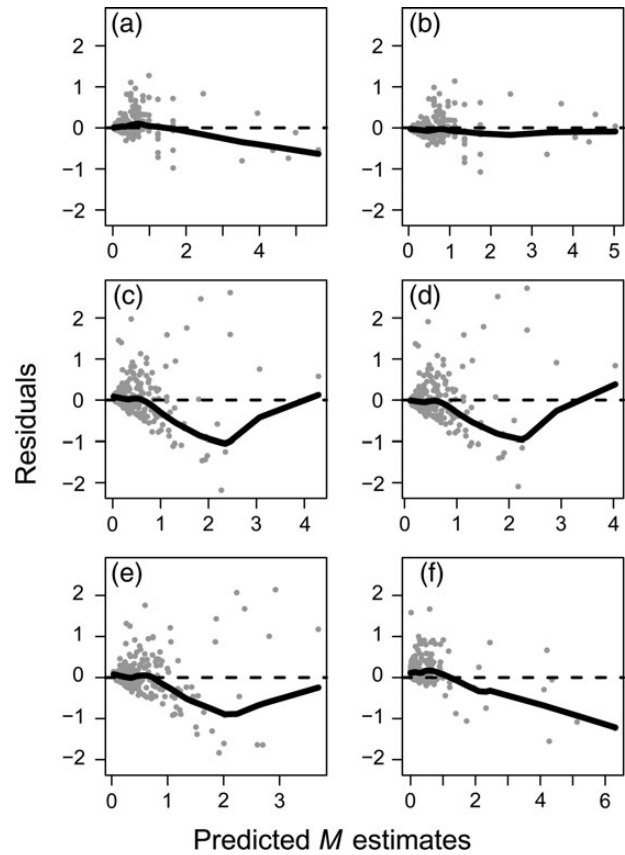
Figure 3. Model residuals (literature—predicted *M* estimates) for the updated estimators of (a) one-parameter t_{max}, (b) Hoenig$_{nls}$, (c) one-parameter K, (d) two-parameter K, (e) Pauly$_{nls-T}$, and (f) Alverson–Carney, based on the common dataset (*n* = 215). LOWESS lines (solid) shown in each panel (smoothing parameter *f* = 2/3).

Pauly$_{nls}$ approaches with the inclusion and exclusion of *T*, suggesting that *T* was not a useful variable in predicting *M* rates.

Model residuals appeared to show no strong trend for the best two t_{max}-based estimators (Figure 3). Most of the residual plots exhibit negative residuals for stocks with predicted values of $M > 1$. Although the variances of the residuals are heteroscedastic, one should bear in mind that these are raw residuals that are not scaled to the magnitude of the literature *M* estimates. The Alverson–Carney estimator had a number of large positive residuals at low predicted *M* values (Figure 3f). The one-parameter t_{max} model showed slightly better residuals than the Hoenig$_{nls}$ model at low predicted *M* values. The *K*-based estimators exhibited very similar residual patterns as the Pauly$_{nls-T}$ method (Figure 3c–e), suggesting that the *K* variable common to all is primarily responsible for the residual patterns. Visually, the Pauly$_{nls-T}$ model generally showed better looking residuals than the one-parameter *K* model (cf. Figure 3c and e).

The estimated residual SE from the Hoenig$_{lm}$ model in this study is 0.43, yielding an estimated CV of 0.45 which is slightly lower than the value of 0.53 for the Hoenig (1983) dataset (MacCall, 2009). For the Pauly$_{lm}$ model, the estimated residual SE is 0.83 which corresponds to CV of 0.997. This is considerably higher than the CV of 0.61 for the Pauly (1980) dataset (MacCall, 2009).

Table 3. Selected updated estimators based on the fullest dataset (sample size *n*).

Models	Updated equations	*n*	Parameter	Model SE	Bootstrap SE	BCa 95% CI
t_{max}						
One-parameter t_{max}	$M_{est} = 5.109/t_{max}$	226	Scaling	0.10	0.22	(4.72, 5.57)
Hoenig$_{lm}$	$\log(M_{est}) = 1.717 - 1.01 \log(t_{max})$	226	Intercept	0.08	0.08	(1.57, 1.88)
			Log (t_{max}) coef.	0.03	0.03	(−1.07, −0.96)
Hoenig$_{nls}$	$M_{est} = 4.899 t_{max}^{-0.916}$	226	Scaling	0.11	0.33	(4.37, 5.65)
			t_{max} exp.	0.02	0.04	(−1.01, −0.84)
K						
One-parameter K	$M_{est} = 1.692K$	218	K coef.	0.08	0.16	(1.37, 2.01)
Two-parameter K	$M_{est} = 0.098 + 1.55K$	218	Intercept	0.06	0.06	(−0.02, 0.22)
			K coef.	0.11	0.24	(1.08, 2.01)
K, L_∞, and t_{max}						
Pauly$_{nls-T}$	$M_{est} = 4.118 K^{0.73} L_\infty^{-0.33}$	218	Scaling	0.80	2.11	(1.89, 9.29)
			K exp.	0.08	0.18	(0.32, 1.00)
			L_∞ exp.	0.08	0.15	(−0.60, −0.04)

Model and bootstrap-based estimates of SE and the bootstrap 95% CI for the model parameter estimates are presented. coef., coefficient; exp., exponent. All length measurements are in millimetres.
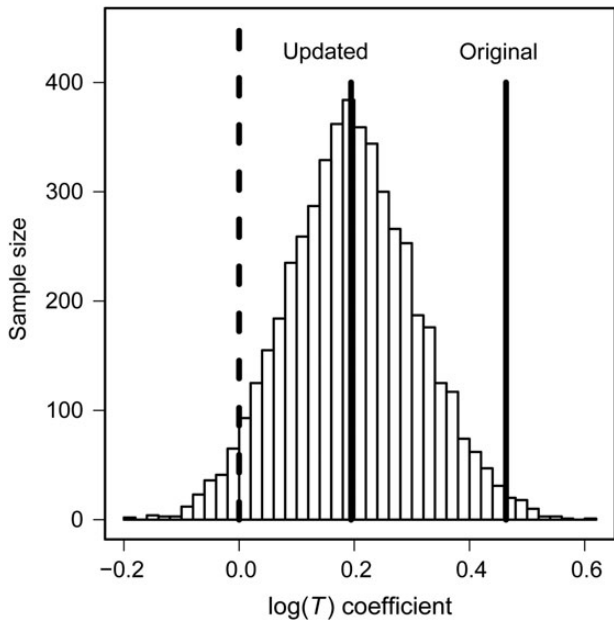

Figure 4. 5000 bootstrap estimates of the coefficient for log-transformed temperature *T* coefficient in the Pauly$_{lm}$ (log − log regression) model. Log(*T*) coefficient estimates for the original Pauly (1980) (=0.4634) and the updated Pauly$_{lm}$ (=0.196) models are shown for comparison. Coefficient estimates < 0 constitute 4% of the 5000 bootstrap estimates.

After the models were updated with the fullest dataset available, the rank of the estimators performance-wise remained almost unchanged from the Approach Evaluation results (Table 3). Among the t_{max}-based estimators, the updated one-parameter t_{max} performed slightly better in terms of CVPE than the Hoenig methods, except the Hoenig$_{gm}$ estimator, which was clearly inferior to the other t_{max}-based methods (Table 2). The t_{max}-based estimators were followed in performance by the updated Alverson− Carney, Pauly$_{nls}$, two-parameter K, one-parameter K, Pauly$_{nls-T}$, and the Pauly$_{lm}$ estimators. MAD estimates ranged from 0.19 for the three best t_{max}-based estimators to 0.38 for the two-parameter K method. Even with the K variable in addition to t_{max}, the

updated Alverson and Carney model performed worse than the solely t_{max}-based estimators, based on the CVPE and examination of the residuals. Hence, no further results will be presented on this estimator.

The CVPE for the updated Pauly methods indicates that inclusion of T did not improve M prediction. Convergence of the Pauly$_{nls}$ model with T fitted to bootstrap samples was particularly problematic. Bootstrap estimates of the $\log(T)$ coefficient for the Pauly$_{lm}$ model showed high variability and suggested ambiguity of the role of T in predicting M since negative estimates occurred (Figure 4). The Pauly$_{lm}$ model was not significantly better than the Pauly$_{lm-T}$ model ($\chi^2 = 2.86$; d.f. = 1; *p*-value = 0.09). The effect of T in prediction of M is dampened in the present updated Pauly$_{lm}$ model ($\log(T)$ coefficient = 0.196) relative to the original Pauly (1980) estimator (=0.463); the effect is even less evident when the physiological adjustment of temperature was not performed for fish with mean T of < 4°C ($\log(T)$ coefficient = 0.136). Hence, we focus only on the Pauly models excluding T in the predictive equation since T does not appear very informative.

Among the different pairs of estimators assumed to be independent, the weighted mean from the combination of the updated Hoenig$_{lm}$ and Pauly$_{lm}$ yielded the lowest CVPE (=0.298), where $M_{est} = 0.77 M_{Hoenig_{lm}} + 0.23 M_{Pauly_{lm}}$. Correlation between residuals of the two updated models was very weak ($r^2 = 0.003$), thus providing some justification for assuming independence and thereby combining the two estimators. However, this best combination estimator offered no advantage in predictive improvement relative to the updated estimators of one-parameter t_{max}, Hoenig$_{nls}$, and Hoenig$_{lm}$ and we present no further results on this estimation approach.

Discussion

Based on the CVPE alone (Table 2), it was difficult to ascertain the single best t_{max}-based estimator and the best K-based estimator. However, the updated two-parameter K equation ($M_{est} = 0.098 + 1.55K$) imposes a lower bound on M which is rather high such that M_{est} cannot be < 0.098. Based on available M estimates of long-lived species, it is biologically possible for M to be considerably < 0.098 and the two-parameter K equation will overestimate M substantially for these species. For this reason, we do not recommend the two-parameter K estimator. The L_∞ exponent in the Pauly$_{nls-T}$

model appeared to be estimated precisely (Table 3); however, inclusion of the L_∞ variable in the model did little to improve prediction performance over the one-parameter K model.

Based on the individual CVPE, model residuals, biological considerations, and principle of model parsimony, we recommend the updated Hoenig$_{nls}$ estimator, $M_{est} = 4.899t_{max}^{-0.916}$ (CVPE = 0.32, $n = 226$ species). In situations where an estimate of t_{max} is not available, we recommend the updated Pauly$_{nls-T}$ estimator, $M_{est} = 4.118K^{0.73}L_\infty^{-0.33}$ (CVPE = 0.60, $n = 218$ species). The estimated coefficient of 5.109 from the updated one-parameter t_{max} model implies that the t_{max} estimates from this study corresponds empirically to the age where $\sim$0.6% of the population survives.

The use of non-linear least-squares fitting to estimate the parameters of select estimators, such as the Hoenig$_{nls}$ and Pauly$_{nls}$ models, implies that the errors in M are additive. The log$-$log formulations of Hoenig and Pauly are likely more appropriate because they imply a multiplicative error structure in the original scale. However, non-linear regression when the residual variance is heteroscedastic does not result in biased estimates though the estimates are not of minimum variance. We proceed with the recommendation of these non-linear models because they performed the best among all models tested.

In examining the patterns of M prediction of the two preferred approaches, the updated Hoenig$_{nls}$ model generally yielded lower estimates of M than the Pauly$_{nls-T}$ estimator for stocks that appear to have low M rates ($<$0.2 yr^{-1} according to the literature) but higher M estimates than the Pauly$_{nls-T}$ method for stocks that experience higher M rates (literature $M > 0.2$ yr^{-1}; Figure 2b). It would appear that relative to the differences in the M estimates between original Hoenig$_{lm}$ and Pauly$_{lm}$ (Figure 2a), the differences in M estimates between the two preferred updated estimators has shifted upwards such that one may expect considerably higher positive discrepancies for stocks with high literature M ($>$0.8 yr^{-1}) and less negative discrepancies for stocks with literature $M < 0.8$ yr^{-1} (Figure 2b).

In this paper, we definitively ranked and quantified the predictive performance of the most widely used empirical approaches for estimating M using cross-validation on a dataset with over 200 fish species of varying life histories. In addition, we reported the precision of the parameter estimates for the preferred estimators using robust methods. The t_{max}-based estimators are unequivocally better in predicting the literature M values than the K-based estimators evaluated; use of a composite estimator offers no clear advantage over using estimators based on just t_{max}. Notably, the performance of the one-parameter t_{max}, Hoenig$_{nls}$, and Hoenig$_{lm}$ estimators was similar within the group of t_{max}-based estimators and the actual ranking may change slightly with addition of new data; the same can be said of the Pauly and the K-based methods as well. There was consistency in predictive performance across the three metrics for evaluation, with agreement based on the CVPE and also on the r^2 and MAD values for each estimator. We updated the best performing t_{max}-based estimators as well as the preferred non t_{max}-based estimators and the precision of the associated parameter estimates—the non t_{max}-based estimators are useful alternatives when one lacks age estimates.

Results from our study are generally corroborated by others. Punt et al. (2005) noted in a study of applying information from data-rich stocks to data-poor ones that Hoenig's (1983) method appears more reliable than Pauly's (1980) method for the purpose of estimating M. It is worth noting that the correlation between log(M) and log(t_{max}) was lower in their study ($r^2 = 0.6$) than in ours ($r^2 = 0.9$). MacCall (2009) examined the SE and coefficient

of variation of estimates of log(M) from both Pauly and Hoenig models and concluded that the latter is the better model. Although based on only 13 species, Kenchington (2013) also showed that the t_{max}-based estimators performed better than the other estimators reviewed.

Among the variables examined in this study (t_{max}, K, L_∞, and T), the most informative set to estimate M empirically for a given fish species is the observed t_{max} alone. This may be of no real surprise given that M and t_{max} are functionally related by theory. What may be of surprise is that the inclusion of information additional to t_{max} did not improve M prediction, as observed from the Alverson–Carney model which incorporated both K and t_{max}. This was also observed from the weighted Hoenig$_{lm}$—Pauly$_{lm}$ estimator; inclusion of the three-parameter Pauly$_{nls-T}$ model in the combination offered no additional prediction improvement than the stand-alone Hoenig$_{nls}$ estimator. Despite uncertainties and various issues surrounding t_{max} estimates (discussed below), our evaluation of various empirical estimators shows clearly that the observed t_{max} is both the best and a sufficient predictor of M.

Although various authors have highlighted the issue of sample size in t_{max} estimates (e.g. Brodziak et al., 2011; Maunder and Wong, 2011; Kenchington, 2013), we did not attempt to "correct" or adjust the observed t_{max} for sample size. Hoenig (1983) noted that maximum age tends to go up as the logarithm of the sample size and Beverton (1992) noted that the accuracy of t_{max} appeared to be less dependent on sample size than might be thought. Kenchington (2013) offered two versions of t_{max}-based estimators which purportedly incorporate effective sample size and the age of full recruitment and claimed that they outperformed other t_{max}-based estimators. The main issues with the Kenchington estimators, in addition to the small sample size in his evaluation, are as follows: (i) Hoenig (unpublished proof on file) demonstrated that the derivations of the estimators are mathematically faulty, (ii) the assumption of the exponential (or geometric) model underlying the Kenchington estimators is critical and it is unlikely fish stocks conform closely to that assumption, and (iii) no guidelines were provided for computing the effective sample size to "correct" the t_{max} estimate. Kritzer et al. (2001) demonstrated that a sample size of $\sim$200 is generally sufficient to obtain a reliable estimate of t_{max}, which is likely a lot less than the sample size of ageing for many exploited stocks.

The M/K ratio underlying the one-parameter K-based estimator, first discussed by Beverton and Holt (1959) as being relatively invariant within specific taxonomic groups, has since been reviewed by various authors with mixed conclusions. Beverton (1992) revisited the work and found M/K ratios ranging from 0.2 to 2.5 across four main taxa, with narrower range in individual taxa. Charnov (1993) reported M/K ratios ranging from 1.6 to 2.1 based on Pauly (1980)'s dataset. Frisk et al. (2001) found the M/K ratio to differ significantly between teleosts and elasmobranchs. Vøllestad et al. (1993) found that the originally developed M/K invariant did not hold for 29 brown trout stocks in Norway. Similarly, Purchase et al. (2006) found little evidence of life history invariance for walleye based on 435 stocks in Ontario. These results suggest that the M/K "invariance" may not necessarily hold even within a species and certainly not universally across all taxa, thus limiting its application in predicting M rates.

The mean annual temperature variable T as defined in Pauly (1980) is not a strong predictor of M in our analyses. Griffiths and Harrod (2007) re-examined Pauly's estimator with a larger dataset derived from FishBase and found that, using path analysis, the

effect of temperature on *M* was only significant indirectly. They also showed that temperature was significant in predicting *M* of perciform fish but not for the non-perciforms. In addition to taxa-specific differences, another possible reason temperature was not important in our analyses is that the mean *T* was not well estimated. Perhaps a more rigorous definition of *T* would produce different results.

The process of rigorous quality control and documentation of the dataset is central to any comparative life history study and meta-analysis. One recurring issue that we have encountered when cross-checking existing mortality compilations is the lack of reproducibility of estimates when referencing the original data sources cited due to (i) the lack of clear documentation on the sources of the *M* estimates and the methods by which they are derived, especially in stock assessment reports; (ii) careless cross-referencing of sources and "borrowing of parameters"; and (iii) continual referencing of a source paper that appeared to be obsolete. We encountered cases where parameter estimates that were purportedly for a stock were actually taken from a closely related species from an entirely different geographical location. Such practices compromise the quality of a dataset as a useful resource for future investigations. The challenge to maintain an updated and well-documented database of *M* estimates put forth by Brodziak *et al.* (2011) is one that we have undertaken for this study.

While we strived to develop and adhere to strict guidelines in building our dataset, we also recognize that utmost stringency would severely truncate the dataset and consequently reduce the usefulness of the resulting analyses in terms of generality across fish of different life history strategies. In determining what constitutes a study of acceptable quality, extracting desired estimates, pooling data, and rejecting some estimates, some degree of subjectivity was unavoidable. Various workers may wish to apply alternative data selection criteria. Hence, the existing dataset should become living documentation that should be reviewed critically and updated constantly as improved estimates are made available. Our database will be maintained on the Internet by an oversight group committed to long-term maintenance of the data.

Given the results from this paper, we make specific recommendations for best practices when estimating *M* empirically, particularly for data-poor stocks:

(i) We recommend the use of the updated Hoenig$_{nls}$ estimator ($M_{est} = 4.899t_{max}^{-0.916}$) when a t_{max} estimate is available. This model performed slightly better than the other t_{max}-based estimators, and exhibited better looking residuals. The use of the Alverson–Carney estimator has no additional advantage over the estimators based solely on t_{max} and is not recommended.

(ii) When t_{max} is not available, we recommend the use of the updated Pauly$_{nls-T}$ estimator ($M_{est} = 4.118K^{0.73}L_{\infty}^{-0.333}$). Although the updated one-parameter *K* estimator has slightly lower CVPE and may be preferred to the Pauly$_{nls-T}$ estimator based on parsimony, the latter estimator showed better behaved residuals.

(iii) The updated Pauly$_{nls}$ (with *T* included) is not preferred based on the following: (a) the precision of the temperature coefficient in the updated Pauly models is poor, and bootstrapping revealed the coefficient could be positive or negative; (b) eliminating temperature from the model provides a more parsimonious model; and (c) the effects of temperature in predicting

M appeared to be taxa-specific based on the study of Griffiths and Harrod (2007).

(iv) The practice of simple averaging of multiple *M* estimates derived from various empirical estimators is not advocated since this study clearly showed that (a) the empirical estimators are not all equally reliable and (b) some of the derived *M* estimates are not independent of each other. The uncertainty of the different estimators evaluated in this study, and their mutual dependencies, should be given due consideration in applications such as the development of prior distributions for *M* for Bayesian analyses.

(v) Compilers of stock assessment reports and future meta-analytic research involving compilations of *M* and life history parameter estimates should be diligent in recording the sources of data and estimates used especially in relation to the method of derivation of the estimate of *M*.

(vi) Previous stock assessments that used an empirical estimator of natural mortality rate may need to be revisited if the estimator performed poorly in this study.

Acknowledgements

References

Alverson, D. L., and Carney, M. J. 1975. A graphic review of the growth and decay of population cohorts. Journal du Conseil International pour l'Exploration de la Mer, 36: 133–143.

Bayliff, W. H. 1967. Growth, mortality, and exploitation of the Engraulidae, with special reference to the anchoveta, *Cetengraulis mysticetus*, and the colorado, *Anchoa naso*, in the eastern Pacific Ocean. Inter-American Tropical Tuna Commission Bulletin, 12: 365–432.

Beverton, R. J. H. 1963. Maturation, growth and mortality of Clupeid and Engraulid stocks in relation to fishing. Rapports et Procès-Verbaux des Réunions du Conseil Permanent International pour l'Exploration de la Mer, 154: 44–67.

Beverton, R. J. H. 1992. Patterns of reproductive strategy parameters in some marine teleost fishes. Journal of Fish Biology, 41: 137–160.

Beverton, R. J. H., and Holt, S. J. 1959. A review of the lifespans and mortality rates of fish in nature, and their relation to growth and other physiological characteristics. *In* Ciba Foundation Symposium – the Lifespan of Animals (Colloquia on Ageing), 5, pp. 142–180. Ed. by G. E. W. Wolstenholme, and M. O'Conner. John Wiley & Sons, Ltd, Chichester, UK. 324 pp.

Brodziak, J., Ianelli, J., Lorenzen, K., and Methot, R. D., Jr (Eds) 2011. Estimating natural mortality in stock assessment applications. NOAA Technical Memorandum NMFS-F/SPO-119. 38 pp.

Canty, A., and Ripley, B. 2011. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-2.

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

Charnov, E. L. 1993. Life History Invariants: Some Explorations of Symmetry in Evolutionary Ecology. Oxford University Press, Oxford. 167 pp.

Davison, A. C., and Hinkley, D. V. 1997. Bootstrap Methods and Their Application. Cambridge University Press, Cambridge. 582 pp.

Deroba, J. J., and Schueller, A. M. 2013. Performance of stock assessments with misspecified age-and time-varying natural mortality. Fisheries Research, 146: 27–40.

Frisk, M. G., Miller, T. J., and Fogarty, M. J. 2001. Estimation and analysis of biological parameters in elasmobranch fishes: a comparative life history study. Canadian Journal of Fisheries and Aquatic Sciences, 58: 969–981.

Gislason, H., Daan, N., Rice, J. C., and Pope, J. G. 2010. Size, growth, temperature and the natural mortality of marine fish. Fish and Fisheries, 11: 149–158.

Griffiths, D., and Harrod, C. 2007. Natural mortality, growth parameters, and environmental temperature in fishes revisited. Canadian Journal of Fisheries and Aquatic Sciences, 64: 249–255.

Gunderson, D. R. 1997. Trade-off between reproductive effort and adult survival in oviparous and viviparous fishes. Canadian Journal of Fisheries and Aquatic Sciences, 54: 990–998.

Gunderson, D. R., and Dygert, P. H. 1988. Reproductive effort as a predictor of natural mortality rate. Journal du Conseil International pour l'Exploration de la Mer, 44: 200–209.

Hamel, O. S. 2014. A method for calculating a meta-analytical prior for the natural mortality rate using multiple life-history correlates. ICES Journal of Marine Science, this issue. doi:10.1093/icesjms/fsu131.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edn. Springer, New York. 745 pp.

Hewitt, D. A., and Hoenig, J. M. 2005. Comparison of two approaches for estimating natural mortality based on longevity. Fishery Bulletin, 103: 433–437.

Hewitt, D. A., Lambert, D. M., Hoenig, J. M., Lipcius, R. N., Bunnell, D. B., and Miller, T. J. 2007. Direct and indirect estimates of natural mortality for Chesapeake Bay blue crab. Transactions of the American Fisheries Society, 136: 1030–1040.

Hoenig, J. M. 1982. Compilation of mortality and longevity estimates for fish, mollusks, and cetaceans with a bibliography of comparative life history studies. University of Rhode Island, Graduate School of Oceanography, Technical Report Reference No. 82-2. 14 pp.

Hoenig, J. M. 1983. Empirical use of longevity data to estimate mortality rates. Fishery Bulletin, 82: 898–903.

Hoenig, J. M., Barrowman, N. J., Hearn, W. S., and Pollock, K. H. 1998. Multiyear tagging studies incorporating fishing effort data. Canadian Journal of Fisheries and Aquatic Sciences, 55: 1466–1476.

Jensen, A. L. 1996. Beverton and Holt life history invariants result from optimal trade-off of reproduction and survival. Canadian Journal of Fisheries and Aquatic Sciences, 53: 820–822.

Jensen, A. L. 2001. Comparison of theoretical derivations, simple linear regressions, multiple linear regression and principal components for analysis of fish mortality, growth and environmental temperature data. Environmetrics, 12: 591–598.

Johnson, N. L., Kotz, S., and Balakrishnan, N. 1994. Continuous Univariate Distributions, John Wiley, New York.

Johnson, K., Monnahan, C., McGilliard, C., Vertpre, K., Anderson, S., Cunningham, C., Hurtado-Ferro, F., et al. 2014. Time-varying natural mortality in fisheries stock assessment models: identifying a default approach. ICES Journal of Marine Science, this issue. doi: 10.1093/icesjms/fsu055.

Kenchington, T. J. 2013. Natural mortality estimators for information-limited fisheries. Fish and Fisheries. doi: 10.1111/faf.12027.

Knip, D. M., Heupel, M. R., and Simpfendorfer, C. A. 2012. Mortality rates for two shark species occupying a shared coastal environment. Fisheries Research, 125: 184–189.

Kritzer, J. P., Davies, C. R., and Mapstone, B. D. 2001. Characterizing fish populations: effects of sample size and population structure on the precision of demographic parameter estimates. Canadian Journal of Fisheries and Aquatic Sciences, 58: 1557–1568.

Lorenzen, K. 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. Journal of Fish Biology, 49: 627–642.

MacCall, A. D. 2009. Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations. ICES Journal of Marine Science, 66: 2267–2271.

Malecha, P. W., Hanselman, D. H., and Heifetz, J. 2007. Growth and mortality of rockfishes (Scorpaenidae) from Alaska waters. NOAA Technical Memorandum, NMFS-AFSC-172. 61 pp.

Maunder, M. N., and Punt, A. E. 2013. A review of integrated analysis in fisheries stock assessment. Fisheries Research, 142: 61–74.

Maunder, M. N., and Wong, R. A. 2011. Approaches for estimating natural mortality: application to summer flounder (*Paralichthys dentatus*) in the U.S. mid-Atlantic. Fisheries Research, 111: 92–99.

McCoy, M. W., and Gillooly, J. F. 2008. Predicting natural mortality rates of plants and animals. Ecology Letters, 11: 710–716.

McGurk, M. D. 1986. Natural mortality of marine pelagic fish eggs and larvae: role of spatial patchiness. Marine Ecology Progress Series, 34: 227–242.

Newman, S. J., Williams, D. M., and Russ, G. R. 1996. Age validation, growth and mortality rates of the tropical snappers (Pisces: Lutjanidae) *Lutjanus adetii* (Castelnau, 1873) and *L. quinquelineatus* (Bloch, 1790) from the central Great Barrier Reef, Australia. Marine and Freshwater Research, 47: 575–584.

Ohsumi, S. 1979. Interspecies relationships among some biological parameters in cetaceans and estimation of the natural mortality coefficient of the Southern Hemisphere minke whale. Report of the International Whaling Commission, 29: 397–406.

Pauly, D. 1980. On the interrelationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. Journal du Conseil International pour l'Exploration de la Mer, 39: 175–192.

Pauly, D., and Binohlan, C. 1996. FishBase and AUXIM as tools for comparing life-history patterns, growth and natural mortality of fish: applications to snappers and groupers. *In* Biology, Fisheries and Culture of Tropical Groupers and Snappers, pp. 218–243. Ed. by F. Arreguín-Sánchez, J. L. Munro, M. C. Balgos, and D. Pauly. ICLARM Conference Proceedings 48. 449 pp.

Peterson, I., and Wroblewski, J. S. 1984. Mortality rate of fishes in the pelagic ecosystem. Canadian Journal of Fisheries and Aquatic Sciences, 41: 1117–1120.

Punt, A. E., Smith, D. C., and Koopman, M. T. 2005. Using information for "data-rich" species to inform assessments of "data-poor" species through Bayesian stock assessment methods. Final report to Fisheries Research and Development Corporation Project No. 2002/094, Primary Industries Research Victoria, Queenscliff. 259 pp.

Purchase, C. F., Hutchings, J. A., and Morgan, G. E. 2006. The biological and statistical significance of life-history invariants in walleye (*Sander vitreus*). Evolutionary Ecology Research, 8: 295–308.

Ralston, S. 1987. Mortality rates of snappers and groupers. *In* Tropical Snappers and Groupers: Biology and Fisheries Management, pp. 375–404. Ed. by J. J. Polovina, and S. Ralston. Westview Press, Boulder.

R Development Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org/ (last accessed 11 June 2014).

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

Tanaka, S. 1960. Studies on the dynamics and the management of fish populations. Bulletin of Tokai Regional Fisheries Research Laboratory, 28: 1–200.

Tauchi, M. 1956. Management Theory of Fishery Resources. Faculty of Agriculture, University of Kyoto. 15 pp. (cited in Ohsumi, 1979).

Then, A. Y., and Hoenig, J. M. 2014. Results of evaluating the performance of empirical estimators of natural mortality rate. Virginia Institute of Marine Science, Data Report 62, Gloucester Point, Virginia. http://web.vims.edu/library/GreyLit/VIMS/dr062.pdf (last accessed 14 July 2014).

Vøllestad, L. A., L'Abée-Lund, J. H., and Sægrov, H. 1993. Dimensionless numbers and life history variation in Brown Trout. Evolutionary Ecology, 7: 207–218.

Zeileis, A., and Hothorn, T. 2002. Diagnostic checking in regression relationships. R News, 2: 7–10. http://CRAN.R-project.org/doc/Rnews/ (last accessed 17 February 2014).

Handling editor: Ernesto Jardim