

2016

The logic of comparative life history studies for estimating key parameters, with a focus on natural mortality rate

JM Hoenig

Virginia Institute of Marine Science

AYH Then

Virginia Institute of Marine Science

EA Babcock

NG Hall

DA Hewitt

See next page for additional authors

Follow this and additional works at: <https://scholarworks.wm.edu/vimsarticles>



Part of the [Aquaculture and Fisheries Commons](#)

Recommended Citation

Hoenig, JM; Then, AYH; Babcock, EA; Hall, NG; Hewitt, DA; and Hesp, SA, "The logic of comparative life history studies for estimating key parameters, with a focus on natural mortality rate" (2016). *VIMS Articles*. 789.

<https://scholarworks.wm.edu/vimsarticles/789>

Authors

JM Hoenig, AYH Then, EA Babcock, NG Hall, DA Hewitt, and SA Hesp



Quo Vadimus

The logic of comparative life history studies for estimating key parameters, with a focus on natural mortality rate

John M. Hoenig^{1,*}, Amy Y.-H. Then^{1,2}, Elizabeth A. Babcock³, Norman G. Hall^{4,5}, David A. Hewitt⁶, and Sybrand A. Hesp⁵

¹Virginia Institute of Marine Science, College of William & Mary, PO Box 1346, Gloucester Point, VA 23062, USA

²Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, 50603, Malaysia

³University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149, USA

⁴Centre for Fish and Fisheries Research, Murdoch University, 90 South Street, Murdoch, Western Australia 6150, Australia

⁵Department of Fisheries, Western Australian Fisheries and Marine Research Laboratories, PO Box 20, North Beach, Perth, Western Australia 6920, Australia

⁶U.S. Geological Survey, Western Fisheries Research Center, 2795 Anderson Avenue Suite 106, Klamath Falls Field Station, Klamath Falls, OR 97603, USA

*Corresponding author: tel: 18046847125; e-mail: hoenig@vims.edu

Hoenig, J. M., Then, A. Y.-H., Babcock, E. A., Hall, N. G., Hewitt, D. A., and Hesp, S. A. The logic of comparative life history studies for estimating key parameters, with a focus on natural mortality rate. – ICES Journal of Marine Science, 73: 2453–2467.

Received 7 September 2015; revised 11 March 2016; accepted 4 April 2016; advance access publication 21 June 2016.

There are a number of key parameters in population dynamics that are difficult to estimate, such as natural mortality rate, intrinsic rate of population growth, and stock-recruitment relationships. Often, these parameters of a stock are, or can be, estimated indirectly on the basis of comparative life history studies. That is, the relationship between a difficult to estimate parameter and life history correlates is examined over a wide variety of species in order to develop predictive equations. The form of these equations may be derived from life history theory or simply be suggested by exploratory data analysis. Similarly, population characteristics such as potential yield can be estimated by making use of a relationship between the population parameter and bio-chemico-physical characteristics of the ecosystem. Surprisingly, little work has been done to evaluate how well these indirect estimators work and, in fact, there is little guidance on how to conduct comparative life history studies and how to evaluate them. We consider five issues arising in such studies: (i) the parameters of interest may be ill-defined idealizations of the real world, (ii) true values of the parameters are not known for any species, (iii) selecting data based on the quality of the estimates can introduce a host of problems, (iv) the estimates that are available for comparison constitute a non-random sample of species from an ill-defined population of species of interest, and (v) the hierarchical nature of the data (e.g. stocks within species within genera within families, etc., with multiple observations at each level) warrants consideration. We discuss how these issues can be handled and how they shape the kinds of questions that can be asked of a database of life history studies.

Keywords: biological reference points, data selection bias, empirical relationships, F_{msy} , hierarchical Bayesian models, indirect methods, intrinsic rate of population growth, life history correlates, mixed effects models, steepness parameter, stock-recruit relationships.

Introduction

The models used by resource assessment biologists, ecosystem modellers and other applied scientists frequently require values of certain key parameters that are difficult to estimate reliably and precisely. In these cases, it is natural to examine similar situations

for guidance on possible values of the parameters. Such guidance can be derived from observations from similar locations, species, time periods, observation systems (e.g. fisheries), and so forth. Indeed, even when an estimate of a parameter is believed to be reliable and precise, it is prudent to check its reasonableness by

comparing it to estimates in the realm of experience. For example, Hewitt *et al.* (2007) estimated the natural mortality rate of blue crabs (*Callinectes sapidus*) from field data and compared the results to values obtained from several methods based on comparative life history studies.

Sometimes, a parameter can be estimated within a population dynamics model but the estimates may be imprecise and highly correlated with those of other parameters. In this case, it may be of interest to provide additional information about the parameter to the modelling process. In a Bayesian analysis, auxiliary information can be used to develop a prior distribution for the parameter which is incorporated into the estimation scheme; the prior distribution may be developed on the basis of comparative life history data (see Hamel, 2015). Another approach is to assess several stocks simultaneously, allowing the parameter to be estimated as a compromise between what the data say about an individual stock and what other stocks say about the parameter value (Punt *et al.*, 2011). In essence, a penalty is imposed for departure from a shared value; the amount of penalty decreases as the information about the particular species of interest increases.

The problem of obtaining values for these difficult to estimate parameters can be of tremendous importance. For example, the instantaneous natural mortality rate, M , enters into almost all aspects of fishery stock assessment but can be difficult to estimate. Consequently, a number of indirect methods based on life history correlations have found widespread use. One, due to Pauly (1980), has been cited over 2400 times and another, due to Hoenig (1983), has been cited almost 1100 times according to Google Scholar (<http://scholar.google.com/>, accessed 5 March 2016).

Another example of a widely used parameter that is difficult to estimate is the intrinsic rate of population increase (maximum per capita rate of population growth). This parameter occurs in surplus production models of yield, is directly related to the fishing mortality rate that gives maximum sustainable yield (F_{msy}), and occurs in methods for calculating allowable biological catch and for formulating stock rebuilding plans. It also occurs in models of population dynamics of low fecundity species and in Lotka-Volterra and similar models of predator-prey interactions. Although this parameter can be estimated in a surplus production model, the estimates tend to be imprecise especially when observations are available over only a limited range of population size. The intrinsic rate of population increase can also be estimated by observing population growth at low population density, or by observing growth rate at two or more population densities and extrapolating down to the depleted state (Gedamke *et al.*, 2007), among other methods. Reliable, precise estimates are difficult to obtain. The intrinsic rate of increase is inversely related to adult body mass, as evidenced in organisms ranging from viruses, bacteria and protozoa up to large cetaceans (Blueweiss *et al.*, 1978). Pauly (1984) demonstrated that this relationship also holds over a much narrower range of body sizes (fish and cetaceans, Figure 1). It is also inversely related to generation time and to reproductive output per generation (Heron, 1972; Figure 2).

A third example is the characterization of stock-recruitment relationships. These relationships are used to determine the maximum sustainable yield and the fishing mortality generating such yield and thus to derive an allowable biological catch, to forecast catches, in population viability analysis (where the importance of density-dependent mortality is key) and in management strategy evaluations based on simulating population responses to exploitation. They are also crucial for interpreting trends in population

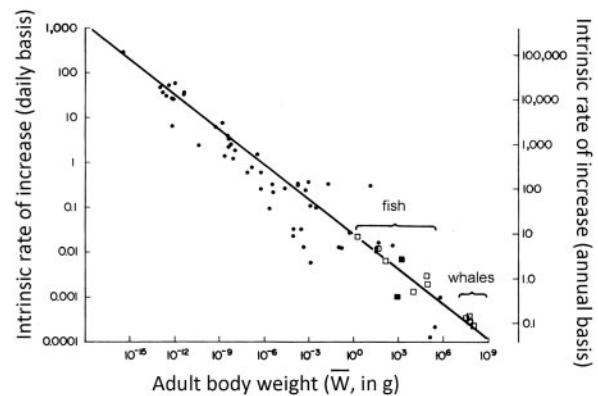


Figure 1. Regression of the logarithm of the intrinsic rate of population increase on the logarithm of adult body weight. From Pauly (1984).

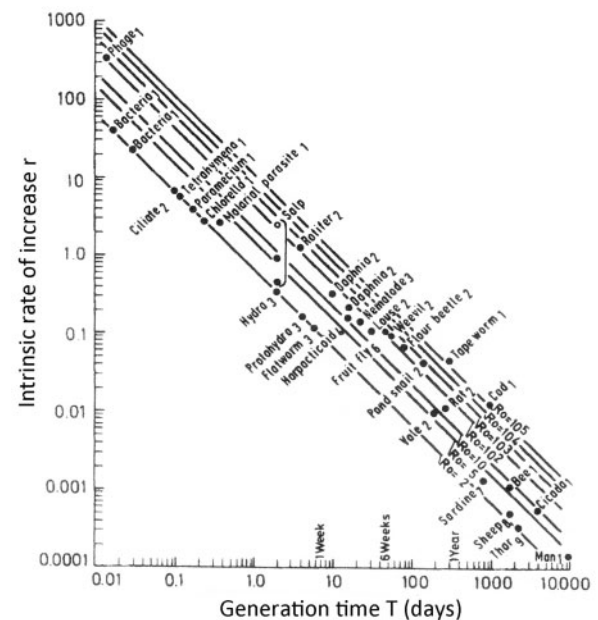


Figure 2. Regression of the logarithm of the intrinsic rate of population increase (per day) on the logarithm of the generation time for several levels of reproductive output per generation, R_0 . The parallel lines represent values of R_0 from 2 to 10^5 . From Heron (1972).

abundance resulting from historical fishing activity. Stock-recruitment data typically display apparently weak structural relationships although this may be due to high measurement error of both stock and recruitment (Walters and Ludwig, 1981). A range of stock sizes must be observed in order to determine how recruitment varies as a function of stock size. Because of the difficulty in estimating stock-recruitment curves and derived parameters (such as steepness, slope at the origin, and unfished equilibrium) there is interest in looking for patterns and general properties in collections of parameter estimates (Myers, 2001; Dorn, 2002). To this end Myers *et al.* (1995) assembled data on over 700 fish stocks.

Another important problem in fisheries science is the prediction of potential fish yields on a per area basis. This can be

accomplished by modelling the relationship between fish production and biological, physical, chemical or other predictors of fish productivity. A widely used empirical model for freshwater fishes in northern North America is the morphoedaphic index which predicts total fish yield based on mean depth of a lake and total dissolved solids (Ryder, 1965, 1982; Ryder *et al.*, 1974). Fryer and Iles (1972, cited in Jones, 1982) also showed a decline in annual fish yields per unit area with depth in African lakes. There are many additional freshwater examples (e.g. SPOF, 1982; Welcomme, 1985; Lester *et al.*, 2004). In the marine context, for fisheries on the continental shelf, Marten and Polovina (1982) showed a negative relationship between maximum sustainable yields of some multispecies demersal fisheries and depth and between some tropical pelagic fisheries and productivity. Although these studies related a key parameter to environmental characteristics, as opposed to relating it to life history characteristics of the species, the approach uses the same logic as comparative life history studies and the issues raised in this article are pertinent to this kind of study as well.

There are other uses of comparative life history studies beyond obtaining parameter estimates for newly considered species or stocks. Such studies can provide insights into ecological, physiological, evolutionary, and other processes (Charnov, 1993; Myers and Mertz, 1998; Thorson *et al.*, 2013). For example, Liermann and Hilborn (1997) applied Bayesian hierarchical meta-analysis to stock-recruit data to evaluate the evidence for the existence of depensation (Allee effect) in fish populations. These aspects are important but beyond the scope of this paper which focuses on prediction of parameter values.

Given the importance of these relationships it is surprising how little work has been done to establish the logic behind comparative life history studies and establish methods and procedures for characterizing relationships and evaluating their performance in predicting parameter values in new situations. Reporting the findings of a 2009 workshop on estimation of natural mortality rate, Brodziak *et al.* (2011) called for a thorough evaluation of natural mortality estimators that are based on comparative life history studies. As the use of comparative studies is widely used in studies of natural systems, it is important that key issues associated with the development and use of such studies are identified, and appropriate methods are developed to evaluate the performance of the relationships derived from those studies.

Indirect (Empirical) vs. Direct Methods

Life history parameter estimation methods may be characterized as direct or indirect, depending on the source of the information utilized by the estimator. Hewitt *et al.* (2007) defined direct methods as those that use just data pertaining to the species (or stock) of interest. Examples would be the estimation of mortality rate in an unexploited population from the age composition or from mark-recapture data. In contrast, an indirect method makes use of a relationship across a variety of species between the parameter of interest (e.g. M) and one or more life history or environmental variables. An example would be the regression of M in over 200 fish stocks versus the corresponding oldest known age for the stock (Then *et al.*, 2015); the regression line can then be used to predict M in a species for which just the oldest age is known (Figure 3). Thus, the term indirect method pertains to an estimator based on comparison of life histories across species. An alternative and equivalent term would be empirical method, e.g. Cubillos *et al.* (1999).

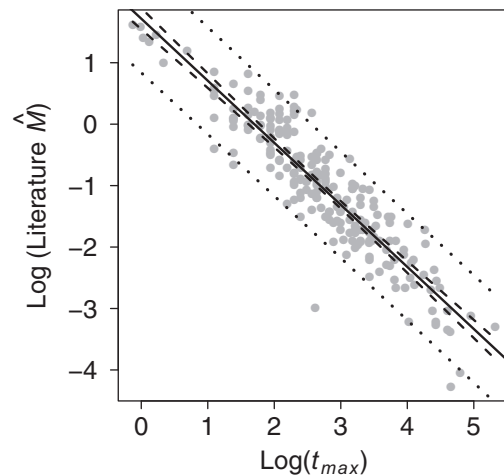


Figure 3. Regression of the logarithm of natural mortality rate M on the logarithm of maximum age t_{max} . Also shown are the 95% confidence interval (dashed lines) and 95% prediction interval (dotted lines). Data on 215 species of fish from Then *et al.* (2015).

The distinction between direct and indirect can be subtle. For example, various researchers have used relationships of the form $M = c/t_{max}$ where c is a constant (i.e. a parameter) and t_{max} is the oldest known age. It has been suggested that t_{max} corresponds to the age at which a cohort has been reduced to 5% of its initial size; this leads to a value for c of ~ 3.0 . This is a direct estimator for M according to the terminology of Hewitt *et al.* (2007) (though it has little to recommend it because the 5% criterion, and thus the value of $c = 3.0$, is arbitrary and the results will vary systematically with the choice of the percentage). On the other hand, Hewitt and Hoenig (2005) showed that a value of 4.22 was optimal for c in the sense of minimizing the sum of squared differences between direct estimates of M collated by Hoenig (1983) from the literature and predicted values from c/t_{max} [an updated analysis of this relationship appears in Then *et al.* (2015)]. Because this involved data from a variety of species the method makes use of a comparative life history study and the estimator is indirect.

The choice of model to fit to empirical data may well come from theoretical considerations. For example, Charnov *et al.* (2013) derived on theoretical grounds an equation for predicting fish weight-specific values of natural mortality rate from growth and maturity parameters:

$$M = 0.41(W/W_a)^{-1/2} A W_a^{-1/3}$$

where $A = 3 K W_\infty^{-1/3}$, W_a is the mass of a fish at first reproduction, K and W_∞ are parameters of the von Bertalanffy growth equation for weight, and weight is assumed to be proportional to the cube of length. Predictions from this model could be compared with empirical studies of weight-specific natural mortality if such estimates were available. In doing so, one would want to distinguish between within-species and among-species departures of empirical observations from theoretical predictions. For example, all sizes of all species may have observations above the theoretical predictions suggesting that the coefficient 0.41 might need adjusting. Another possibility is for all species to have a similar pattern to the residuals over weight; yet another

possibility is that some species have positive residuals for all weights while others have negative residuals. The different patterns of residuals suggest ways in which the theory may need revision. This underscores the need for a logical approach to evaluating comparative life history studies. However, at present it may be difficult to obtain enough weight-specific estimates of natural mortality to perform this evaluation. An alternative might be to compare weighted averages of the weight-dependent mortalities predicted by theory for lightly fished stocks with the average values estimated from catch curve or tagging data for those stocks (e.g. as reported by Lorenzen, 1996).

Key issues and answerable questions

The issues surrounding comparative life history studies for prediction include the following.

- (i) The parameters of interest may be ill-defined idealizations of the real world
- (ii) True values of the parameter(s) are not known for any species
- (iii) Selecting data based on the quality of the estimates can introduce a host of problems
- (iv) The estimates that are available for comparison constitute a non-random sample of species from an ill-defined population of species of interest; and, in general, we wish to make predictions for a particular species, not a randomly selected species
- (v) The hierarchical nature of the data (e.g. stocks within species within genera within families, etc., with multiple observations at each level) warrants consideration

These issues are not trivial but neither are they insurmountable. Their solution depends on establishing a logical framework to define the multiple questions of interest, synthesize life history information, and devise appropriate methodology for deriving estimators, evaluating their performance, and comparing competing estimators.

The questions we consider in this article, in the context of the above five issues, are:

- (i) How well does a given method predict parameter values reported in the literature? Would the performance of the estimator benefit substantially from an increase in sample size (number of life histories considered)?
- (ii) In what circumstances does a method work? Does it work for subsets of the data? Does it withstand the test of time?
- (iii) How do we devise a fair comparison of methods for estimating a parameter from life history information?
- (iv) What can we say about the performance of an estimator for a particular prediction?

We focus on how comparative life history studies for the purpose of estimating parameters should be done: their logic, the questions that can be answered and the methodology appropriate to the various questions. We focus primarily on the problem of predicting M because this problem is important and of widespread interest, there are diverse approaches in the literature to using life history information to estimate this parameter, and there is much

work to be done to evaluate and improve the methods that have been proposed.

Dealing with the key issues

Issue 1. The parameter(s) of interest may be ill-defined idealizations of the real world

Life history parameters may vary by year, time within a year, location, component of the population (e.g. age, sex, reproductive condition), and so forth. Yet, it is often useful to treat a parameter as a constant because the variability of the parameter may not be of practical significance over a wide segment of the population or treating the parameter as variable may render the problem of parameter estimation intractable, or both. For example, M likely varies with the age (or size) of the animal, over time in response to climatic and other changes in environmental conditions and changes in population density of both the species of interest and its predators, by habitat, by sex, and so forth. Assessment models in which mortality rates vary over time based on the density of predators are sometimes used in fisheries stock assessment, particularly for forage species (e.g. Tyrrell *et al.*, 2008). Yet it has generally been found useful to assume for most species that, after a certain age, M can be treated as a constant (see Johnson *et al.*, (2015) for an example of extensive simulations to address this question). As another example, the generation of surplus production (i.e. potential growth of population biomass) may also vary by year, season, location, and age composition yet a single value of the intrinsic rate of increase is usually estimated. George E.P. Box famously remarked: 'All models are wrong, but some are useful'. The point is that, while studies of particular, well-studied species may be able to estimate changes in life history parameters across time and space, comparative life history studies often treat the parameter of interest as being a constant for each species, and such a simplistic treatment of life history often has utility.

Issue 2. True values of the parameters are not known for any species

It is not possible to answer the questions 'How well does a given indirect method estimate a parameter?' and 'Which indirect method best estimates a parameter?' if the true value of the parameter is not known for any species. (It is possible, using mathematical analysis or simulation, to evaluate the performance of an estimator under specific (i.e. hypothesized) conditions. However, it is rarely if ever possible to know the true conditions of a population, i.e. whether the assumptions of the analysis are met.) Some would argue that some parameter estimates are quite precise and are unbiased as evidenced by the standard errors of the estimates and from careful consideration and testing of assumptions. We do not dispute there are some high quality estimates (though establishing such is the case is not easy). But, we argue that a precise, unbiased estimate of a parameter such as M would represent a snapshot of a single place, time, and set of conditions and that a large series of repeated studies might well demonstrate variability, and thus uncertainty, about the true (average) parameter value.

We can, however, address the questions 'How well does an indirect method reproduce what is known about a parameter?' and 'Which indirect method best reproduces what we do know about a parameter?' The key is to view what is known about the parameter as the collection of estimates that are available in the

scientific literature plus those estimates that can be derived from unpublished reports and databases available from various governmental and scientific organizations. These, for all their faults and limitations, are the values used in practice (for assessment and management of resources, for ecosystem modelling, etc.). Hence, because this collection of estimates is of obvious value, it is reasonable to ask how similar information can be collected more easily and to ask if the indirect-method surrogates could provide estimates similar to those of more advanced, i.e. direct methods. It should be borne in mind that the indirect estimate may be better than a direct estimate if the direct estimate is poorly done. Indeed, indirect estimates are often used to judge whether a direct estimate is reasonable (e.g. a high estimate of M does not seem consistent with a high longevity). This is an important point to which we return below when we discuss Answerable Question A: How well does a given method predict parameter values in the literature?

Maunder and Wong (2011), based on Pascual and Iribarne (1993), commented that estimates of M based on relationships with life history parameters are too imprecise to be useful in stock assessments. Kenchington (2014) also struggled with the issue of what constitutes reliable information about natural mortality and seemingly drew the opposite conclusion. He criticized ‘regression estimators’—i.e. those indirect methods that relate direct estimates of M to easier to estimate life history parameters—because of their reliance on the estimates of M in the literature which he deems uncertain. But, the values of M in the literature are used for managing fisheries so there is a de facto general acceptance that these estimates constitute valuable information; it is reasonable to use the estimates to develop predictors using regression analyses.

The objection that parameter estimates in the literature (the dependent variable) are subject to uncertainty is not a persuasive argument against regression analysis. A regression analysis explicitly incorporates uncertainty in the dependent variable in the model as an error term. The error term can be specified in various ways to accommodate distributional assumptions, dependence on other factors, etc. Furthermore, statistical theory provides methods for evaluating performance; if the estimates of M in the literature are too uncertain to be useful one would expect that the performance of the various models relating M to life history variables would be poor. Yet high correlations have been obtained in regression analyses relating life history parameters to each other in general, and to M in particular.

This is not to say there are not additional factors to consider such as the quality of individual estimates and the representativeness of the available estimates. But, these issues can be resolved, as described below. In summary, the information in the scientific literature constitutes what is known about M and it is worthwhile seeing which empirical methods can best reproduce the collection of literature values.

Issue 3. Selecting data based on the quality of the estimates can introduce a host of problems

The issue of quality control involves more than a trade-off between accepting more data of lesser quality—or of poorly known quality—vs. accepting less data that are of higher quality. Rejection of some studies can introduce selection bias. For example, when predicting M , elimination of studies using length-based methods (as opposed to age-based methods or tagging

studies) to estimate M may eliminate studies of tropical species of fish disproportionately. Elimination of poorly studied species may eliminate low-valued fish which may be from lower trophic levels than species of greater commercial value (see ‘Issue 4’ below for a discussion of selection bias). An additional consideration beyond bias is that elimination of select studies reduces the likelihood that the data considered span the range of variability in life history patterns; this, in turn, has two consequences. First, it reduces the ability to estimate the parameters of a structural model. That is, one cannot measure (well) the change in γ per unit change in x if x does not vary (much). Second, elimination of studies limits the ability to test the applicability of derived relationships across a range of conditions, (e.g. taxonomic groups, ecological guilds, habitats).

From a statistical viewpoint, there are two aspects of the estimates in the data to consider—variance and bias. Suppose that all estimates of the key parameter (dependent variable), y , are unbiased estimates but the variance varies (say, according to sample size used in the individual studies). If one wished to construct a regression estimate, and one had estimates of the variance for each y value, one could perform a weighted regression. Failure to assign proper weight to the data points does not induce bias in the estimated regression coefficients. However, the estimates of the regression coefficients are not of minimum variance and the confidence intervals for the regression coefficients may not be reliable. Of more concern is that some of the estimates of the key parameter in the database may come from biased estimators and it may be difficult to quantify or identify the bias.

Clearly, it is important to inspect comparative life history data for quality control purposes. It may be possible to develop objective criteria for accepting or rejecting studies (e.g. Gislason *et al.*, 2010) but the degree of stringency is a subjective judgment. Because researchers may differ in the criteria they apply, the conclusions will only be widely accepted if they are robust to choice of criteria for filtering the data. If all data are made freely available, interested researchers can apply their own selection criteria to investigate the sensitivity of the results to selection criteria.

Therefore, we believe that a necessary condition for a comparative life history study is to make all data considered available for scrutiny and filtering by other researchers. It is best if all studies considered, including those that were rejected, are included, and reasons for rejection are documented. However, few studies document data sources that were rejected. The study by Zhang and Megrey (2006) provides a good example of the importance of documenting data. They published the data for 91 species that they used in their study of an estimator of M based on maximum age (t_{\max}) and growth parameters. Examination of their data reveals that one third of their supposedly independent, direct estimates of M are, in fact, calculated as $3/t_{\max}$. (The authors noted that many of their estimates came from the study by Ault *et al.* (1998), and that study was not clear about the origins of the M estimates.)

Issue 4. The estimates that are available for comparison constitute a non-random sample of species from an ill-defined population of species of interest

Consider the prediction of M of fishes based on life history traits such as growth parameters, water temperature, longevity, and age or size of maturation. What is the population of interest? That is, for which species do we wish to make predictions? Is it the

population of commercially and recreationally exploited species? All coastal, estuarine, and freshwater fishes? What about pelagic and hadal species? Clearly, different researchers will be interested in different groups of fishes. For most comparative studies, the population of interest has not been defined.

Some comparative life history studies have focused on a well-defined taxonomic group. For example, Bayliff (1967) examined the relationship between longevity and M in the fish family Engraulidae and Ohsumi (1979) examined this relationship among the cetaceans. Although authors rarely have a random sample of the species in the group of interest (or data for every species in the group) it may not be hard to imagine that the results are applicable to the entire group in some cases.

In general, however, comparative life history studies are conducted using whatever data are available and thus the data do not constitute a random sample of species from the population of interest. It is then difficult to know if the results can be validly extrapolated to a larger population of species.

If a comparative life history study gives rise to a general relationship that holds across all taxonomic groups and all environmental conditions then the definition of the target population is irrelevant. One way to evaluate the representativeness of the results of the comparative study is to look at how the resulting model predicts values of the key parameter for subsets of the data defined by taxonomic position, ecological role, habitat, etc. (e.g. Beverton and Holt, 1959). Analysis of residuals is a way to detect problems with the assumption that the model being fitted has some universal applicability over a wide range of species and conditions. For example, Pauly (1980) recognized this; he found that his model for predicting M needed adjustment for polar fishes and clupeoids but generally appeared to be widely applicable based on tests of universality (sign tests on the residuals). A caveat is that the greater the number of tests of universality that is conducted, the more exceptions to the prediction rule will be found because of pure sampling error (when 20 statistical tests are done at $\alpha = 0.05$, the expected number of significant results is one when none of the null hypotheses is false). Thus, one needs to be careful about concluding that separate estimators are needed for select groups. Nonetheless, examination of the performance of an estimator over subgroups is a logical way to assure that lack of random sampling of species from the population of interest is not a fatal flaw of the study. We expand on this idea in the next paragraph.

The non-random selection of species likely introduces bias into the estimated regression coefficients because the probability of a species being sampled is likely to be correlated with the response variable (Nakagawa and Freckleton, 2008). Consider the regression of $\log(\text{natural mortality rate})$ on $\log(\text{maximum age})$. Then *et al.* (2015) had values of M up to 5.07 y^{-1} , corresponding to small fish (as short as 49 mm) with short lifespans (as short as 38 weeks). Suppose they wished to evaluate the relationship between M and maximum age for species of economic importance. They might restrict their attention to species for which M (the response variable) is estimated to be less than 1.0 y^{-1} . This would cause selection bias (Figure 4a). On the other hand, if they restricted attention to species with a maximum age (i.e. explanatory variable) >2 years, this would not cause selection bias (Figure 4b). (Lest the reader get the impression that comparative studies are hopelessly biased, we point out that Figure 4a represents an extreme in selection bias with one group having a sampling probability of 1.0 and the other having a sampling probability of zero.

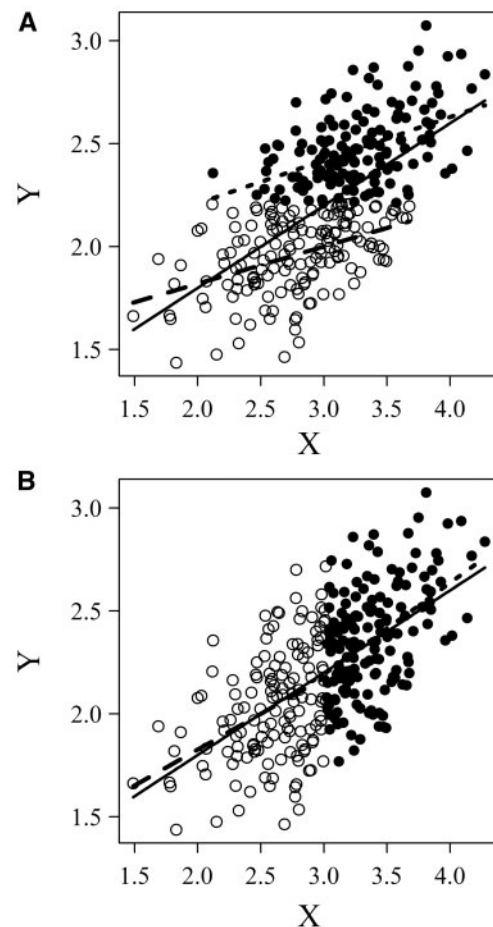


Figure 4. Regression models fitted to partial data sets where data are missing (a) depending on the value of Y or (b) depending on the value of X, after Nakagawa and Freckleton (2008). The bold line is the true regression, the thin line is the regression calculated from only the solid points, and the dashed line is the regression calculated from only the open points.

The bias would be less if, say, one group had a sampling probability of 0.7 and the other had 0.3.) Now suppose that species with high values of M are less likely to be sampled (i.e. studied) because most species for which mortality has been estimated are important to either recreational or commercial fisheries, and fisheries tend to focus on the larger species within taxonomic groups which tend to have lower M and higher longevities. Thus, selection is not on the basis of M (response variable) or longevity (explanatory variability) but on some other (unspecified) variable such as economic importance. The selection is on the basis of a variable (or variables) that is (are) correlated with both the response and explanatory variable. This causes statistical bias. An example of selection biased by the value of the response of interest is given by Best (1993) who reported that observations on rate of population increase of baleen whales at low population size were restricted to populations that had increased enough to make it feasible to monitor the populations.

Statistical methods exist that could be used to extract unbiased estimates of regression coefficients when the probability of a species being sampled (studied) varies by species. This missing data

problem is statistically similar to the problem of estimating species richness from field surveys when the species have different probabilities of being detected (given that the species is present in the sampled site), and detection probability may be correlated with taxonomic group, body size, or habitat preferences (Dorazio *et al.*, 2006; Kery and Royle, 2008; Royle and Dorazio, 2008). A similar problem exists in mark-recapture studies in which the probability of an animal being captured varies with body size, sex or other characteristics (Royle, 2008, 2009), and in surveys of animal abundance that expand in area over time (Schmidt *et al.*, 2009). In all of these cases, the missing response data (e.g. species occurrence) can be imputed either as a random effect or as a function of variables that have been observed. The probability of detection (of a species or individual) can also be modelled as a function of covariates (Royle, 2009). Although the bias caused by non-random sampling has been addressed in a meta-analysis of fishery management systems (Melnychuk *et al.*, 2012; Thorson *et al.*, 2013), to our knowledge, such methods have not been used for comparative life history studies. It may be possible to reduce the bias in such studies by modelling the probability of a species being studied as a function of the species' importance to fisheries, or variables such as size, region, latitude, family, habitat type, or trophic level.

For example, we extracted all the species for which both t_{\max} and L_{\max} were available from Fishbase (Froese and Pauly, 2015), and used this dataset to evaluate the propensity of each species to have an M value in the database of Then *et al.* (2015). Whether a species was studied (i.e. was included in the database of Then *et al.*) was modelled with logistic regression with two fixed effects explanatory variables of whether or not the species was fished (as reported in Fishbase) and L_{\max} , and with a random effect of family. Each species' propensity to be studied was calculated as the predicted probability of being studied from this model. To evaluate whether propensity to be studied was related to the M values, indicating that the missing data were not missing at random, we ran a linear regression of M against propensity to be studied for those species included in the Then *et al.* database. The results indicate that fish species with a higher propensity to be studied have a lower natural mortality rate (solid regression line in Figure 5); furthermore, the variability in estimates of M decreases with increasing propensity score. Species that are unfished (circles) have a lower propensity score and are extremely variable in their natural mortality rates compared with species that are fished (triangles). If species with low propensity to be sampled are eliminated from consideration then there does not appear to be a relationship between natural mortality rate and propensity (dashed regression line). Estimates of the parameter c in the model $M = c/T_{\max}$ are similar when computed with and without the low-propensity data indicating that the amount of low propensity data is not sufficient to have a large effect on the model for estimating M . However, application of the model to a new species with a low propensity score is likely to give poor results; thus, one should check the propensity score before estimating M .

Issue 5. The hierarchical nature of the data (e.g. stocks within species within genera within families, etc., with multiple observations at each level) warrants consideration

Often, multiple sets of estimates of parameters will be available for some species, e.g. commercially or recreationally important species and species that are easy to study. Including multiple

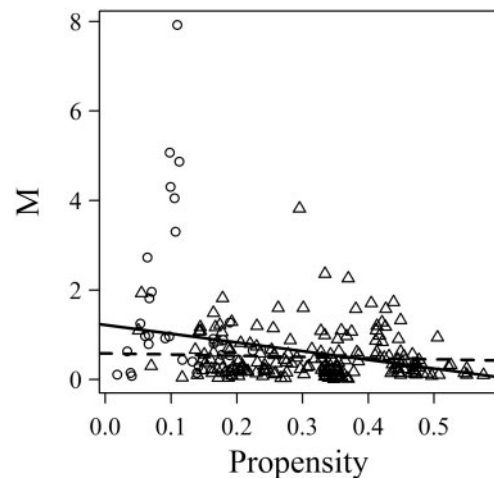


Figure 5. Natural mortality rate plotted against propensity to be studied for the species in the database of Then *et al.* (2015). Propensity scores are computed from the mixed effects logistic regression of whether or not the species is in the database of Then *et al.* (response variable) on the fixed effect explanatory variables L_{\max} and whether or not the species is fished, and on the random effect for family. Circles represent unfished species and triangles represent fished species. The solid line is a regression across all of the data, and the dashed line is a regression including only data with a propensity > 0.15 .

estimates for some species weights the results towards those species; this may affect the representativeness of the results to a wider population of interest unless a hierarchical model is used to partition the variance between measurement error and variation among populations of the same species. Also, to the extent that the multiple estimates *within a species* have less variability than the between-species variability one will be misled into thinking one has better predictive ability than is warranted unless a hierarchical model is used.

Nonetheless, there is value in compiling multiple estimates for species because this can provide insight into the fineness (scale) of inference that can be made. For instance, in comparative studies of M , Pauly (1980) had a total of nine estimates of M and corresponding parameters for the Atlantic cod (*Gadus morhua*) while Hoening (1983) included eight estimates for the cisco (*Coregonus artedii*). For a single species, the range of actual variability in life history correlates is likely to be less than that among different species, genera, families, etc. (although there are a number of examples of freshwater fishes exhibiting high variability in growth and other parameters; see Donald and Alger, 1989; Shuter *et al.*, 1998; McDermid *et al.*, 2010; Helser and Lai, 2004). Hence, it is of interest to see if the range of variability observed in life history parameters, say t_{\max} , within a taxonomic group can explain the range in the M estimates for that group. Ideally, the range of M estimates for a taxon can be accounted for by the variation in the explanatory life history traits instead of measurement errors. That is, it is of interest to see if an indirect method can use subtle variation in life history traits to predict the variability in M among members of a restricted taxonomic group such as an individual species.

The same considerations pertain to higher order taxonomic groups (genera, families, etc.). Well-studied groups will receive more weight in the development of estimators unless techniques

which partition variability into measurement error and variability associated with taxonomic group are used. Two approaches are available: mixed effects models and hierarchical Bayesian models. We will give examples of the latter but we note that the former can accomplish the same goals except in a classical rather than Bayesian framework.

Hierarchical models have been used to evaluate stock-recruit relationships (Myers, 2001; Dorn, 2002), and occasionally for growth (Helser and Lai, 2004; Zhang *et al.*, 2009) and mortality studies (Zhou *et al.*, 2012). Such models generally find important differences among taxonomic groups in life history parameters and the relationships between them. Thorson *et al.* (2013) provide an excellent review of the use of hierarchical models for meta-analysis in fisheries science.

Hierarchical models treat the differences between sample units, such as populations or species, as random effects, drawn from a distribution with an estimated variance (Royle and Dorazio, 2008; Thorson *et al.*, 2013). For models of life history parameters, random effects can be estimated at multiple taxonomic levels, such as population, species, genus, family, and order. Additional explanatory variables such as environmental conditions or location can also be included in the model as fixed effects at an appropriate taxonomic level. Errors in variables and multivariate correlation structure among parameters can be incorporated into the hierarchical framework (Helser and Lai, 2004; Zhou *et al.*, 2012; Thorson *et al.*, 2014). As a practical matter, the dataset will not be extensive enough to model all the variance components for many of the parameters we wish to predict. Nevertheless, it is often possible to estimate random effects for at least the higher taxonomic levels such as order and family (Zhou *et al.*, 2012). Random effects at the level of species or population can also be useful. Because the variance among taxonomic groups (e.g. species) and the sampling error variance (e.g. variation among studies of the same species) can be distinguished (at least in theory) in the hierarchical model, the model can appropriately weight data from groups that have different numbers of studies. This potentially allows all available data to be included in the model (after screening for data quality and differences in estimation methodology).

In effect, a model with random effects at some taxonomic level is intermediate between the two extremes of pooling all the data for regression and regressing data from each taxonomic group separately. A hierarchical model will estimate separate regression coefficients for each group, but the values at the group level will be drawn from a shared distribution. Thus, the coefficients for each group tend to shrink toward the mean coefficients. Gelman and Hill (2007) refer to this effect as ‘partial pooling’. Taxonomic groups with a larger sample size will experience less pooling (shrinkage towards the mean) than will groups with a smaller sample. This partial pooling is useful because it allows less studied groups to borrow information from well-studied groups [what Punt *et al.* (2011) call the ‘Robin Hood’ approach]. On the other hand, for a very well-studied taxonomic group, an independent model may be preferable to a hierarchical model if borrowing information from other groups is not necessary or desirable (it might not be desirable because it introduces stronger assumptions into the analysis).

In the case that there is significant variation among taxonomic groups in their life history parameters and the relationships among them, a hierarchical model may produce more accurate predictions of life history parameters for unstudied species within

taxonomic groups that are included in the dataset that was used to fit the model (Thorson *et al.*, 2013). This is illustrated using simulation of Bayesian hierarchical models in Appendix 1. Also, because hierarchical models partition variance at multiple taxonomic levels, they can give more accurate estimates of the variance associated with a prediction at any given taxonomic level.

To expand upon this idea, we first review the distinction between prediction intervals and confidence intervals. In a regression of an independent variable y on a predictor variable x , there is a cloud of points and a best fitting line (in some sense). The line describes the average value of y for each value of x . Of course, not all data points constitute average cases, i.e. there are differences among cases (stocks, species, etc.). This has two implications. First, there is uncertainty where the average lies, and this uncertainty is described by the confidence interval. Second, there is the fact that individual observations are not the same as the average because of individual variability. The uncertainty in where a new observation will lie, given the value of x , is described by a prediction interval, which is wider than the confidence interval since it contains the uncertainty in the mean and the deviation of individual observations from the mean (see section on answerable question D below for a further discussion of this).

Thus, e.g. if a model includes family- and species-specific random effects, the expected value of the mean for an unstudied species in a studied family will include the family effect. In both the cases of single and of multiple observations per species, the prediction interval contains the sampling error (among-studies variability) and the among-species variability within a family. In the case of no multiple studies of species, the sampling error and among-species variability are lumped together and are not separately estimable, so the prediction interval would be about the same for both the case of one study per species and multiple studies for some species. However, the confidence interval for the expected value of a new species in the case with multiple studies per species would not include the among-studies variance, so it would be narrower than if there is only one study for each species. When making a prediction for a new species in a studied family, given multiple studies per species, it makes sense to characterize the uncertainty in the prediction by using the confidence interval because it includes the variability among species but not the variability among studies of the same species. Having multiple studies of species would be necessary to separate the variances to make this possible. If a prior distribution is needed for a parameter, the confidence interval of the mean prediction would probably be reasonable for an unstudied species, because it would include individual variation among species, but not sampling error. See Appendix 1 for details.

For cross-species comparisons of life history parameters, non-hierarchical regression models have been more commonly used than hierarchical models, perhaps in part because existing datasets have not been considered sufficient to estimate variances at multiple hierarchical levels. Nevertheless, considering that some life history traits do appear to vary taxonomically (Zhou *et al.*, 2012), it would be worth developing hierarchical models where data allow, and also comparing the results of hierarchical models to the more commonly used regression approaches. There may be very little difference between regression and hierarchical models in the prediction for a species from an unstudied taxon. For unstudied taxa, the hierarchical model may not have any benefit to justify the extra model complexity. The benefit of the hierarchical model is to give better predictions for an unstudied species from a studied taxon.

As an example, consider the dataset on natural mortality rates compiled by [Then et al. \(2015\)](#) available at http://www.vims.edu/research/departments/fisheries/programs/mort_db/index.php. Bayesian estimates for the model $\hat{M} = cK$, where K is the von Bertalanffy growth coefficient and c is a parameter to estimate, were computed using uninformative priors for c and for residual variance. Next, models were fitted with either family or order treated as a normally distributed random effect with an estimated variance. Additionally, a two-level hierarchical model with family and order was fitted. In this model, order effects are random deviations from the mean slope parameter c , and family effects are random deviations from the mean of the corresponding order effect. The model with just family effect performed best based on its having the lowest DIC (Deviance Information Criterion) value. The point estimate for a new species in a new family from the model with family effects, and the point estimate from the model without family, were very similar. However, the point estimates for the species in studied families differed by family ([Figure 6](#)). Thus, the model with family effects may improve prediction for new species in a studied family, especially if the number of studied species in the family is high.

Answerable questions

(A) *How well does a given method predict literature values?*

What is the appropriate measure of performance (absolute or relative error)?

The focus of this article is on predicting values of a parameter for a new (previously not considered) species rather than on modeling the structural relationship between the expected (average) value of a variable of interest and possible explanatory variables. Therefore, we seek a way to evaluate predictive ability.

One way to do this is to look at the average ‘discrepancy’ between predictions and ‘observed values’ where an ‘observed value’ is an independent (direct) estimate of the parameter and ‘discrepancy’ can be the absolute difference between a predicted value and an observed value or the squared difference between the prediction and the observed value (e.g. root mean squared error). For example, consider the estimator of M developed by [Hoenig](#)

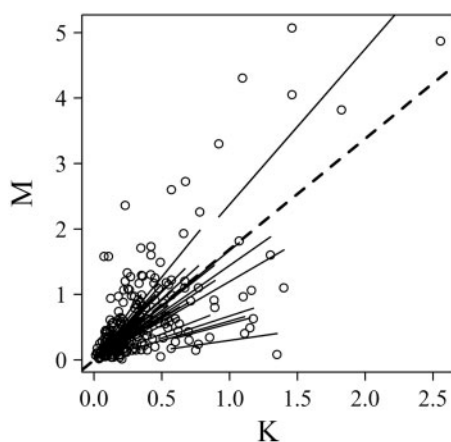


Figure 6. Original data from [Then et al. \(2015\)](#) (points) showing the relationship between natural mortality rate, M , and von Bertalanffy growth coefficient (K), predicted values by family (solid lines) for a hierarchical model with a random effect of family, and predicted values from a non-hierarchical model (dashed line).

(1983) in which the logarithms of independent estimates of M were regressed on the logarithms of the maximum ages known for 134 stocks of fish, mollusks and cetaceans. We could look at the average squared deviation of the data points (independent estimates of M) from the predictions (regression line). However, it is well known that this assessment is too optimistic because the same data that were used to fit the model (the training data) were used to evaluate the predictive performance (the checking data).

Therefore, it is common to divide the data into two groups and use the data in one group to fit the model and the data in the other group to evaluate the model. Typically, k -fold cross-validation prediction error is computed where the data are randomly divided into k folds (groups) and the model is fitted k times, each time leaving out one fold. The omitted fold then serves as the independent, checking data for evaluation of predictive ability. Typical values for k range from 10 to n (the number of data points) with n -fold cross-validation consisting of a ‘leave one out’ scheme.

There are three key points about k -fold cross-validation. The first is that it does not say how well we can predict the parameter, e.g. M , because the true values of the parameter are unknown. Rather, it tells us how well we can predict (reproduce) the existing estimates. It is not clear if the prediction error (the discrepancy) is due to the prediction being bad or if it is due to the estimate being predicted being bad. In essence, this procedure is getting at the question: If, instead of doing time consuming, expensive or tedious studies to estimate a parameter directly, we rely on a simpler, indirect method, how close can we come to the direct method (on average)?

The second point about cross-validation is that it does not estimate exactly what is desired ([Hastie et al., 2009](#)). We want it to estimate the conditional prediction error but it estimates the unconditional prediction error. The conditional prediction error quantifies the expected error when the existing model (conditioned on the comparative life history data used to fit the model) is applied to a randomly selected new set of life history parameters. The unconditional prediction error refers to: if we were to collect other data sets like the one we have at hand, and use them to develop indirect estimators, how well would the new estimators predict a new direct estimate. Bootstrapping is an alternative to cross-validation but it, too, does a better job at estimating the unconditional than the conditional error rate ([Hastie et al., 2009](#)). Nonetheless, these techniques provide the best indication of prediction error and are valuable.

The third point is that random sampling is not operative in the selection of species to develop the estimator nor is it operative in the selection of the new species to which the estimator is to be applied. Thus, the usual statistical theory governing properties of the estimator may not hold. For example, as the sample size of species used to develop an estimator increases the regression coefficients may not converge to stable values—this might arise if the underlying model is not universally applicable to all species groups. In this case, formal modelling of the probability of a species being sampled as a function of, e.g. body size or importance to fisheries, may help to mitigate this problem, given sufficient data.

(B) *In what circumstances does a method work? Does it work for subsets of the data? Does it withstand the test of time?*

As we have seen, studying a universally applicable relationship avoids problems with defining populations and selecting data. It

is thus important to assess the applicability of any relationship among life history parameters. Techniques such as analysis of residuals should be applied to multiple subsets of the data reflecting taxonomy, ecology, physiology and other factors that might be relevant. The possibility that measurements depend on methodology should also be examined. For example, improvements in methods for determining the age of fish have led to increases in maximum known age for many species. If methodological changes are significant, then apparent relationships among life history parameters may not be stable over time. In performing these tests it is important to remember that the more tests that are performed, the greater the chances of finding spurious differences among groups.

(C) How do we devise a fair comparison of methods for estimating a parameter?

It is important to compare estimators on as level a playing field as possible. We believe parameter estimators should be tested (compared) on the same set of species. In this way, outliers in the response variable will affect all methods, and the non-random collection of species for which data are available will at least be the same for all methods. The disadvantage of this approach is that the sample size is reduced because you can only use species for which all life history characters needed for all methods are known. It may be desirable to perform pairwise comparisons of estimators so that sample size is reduced minimally.

(D) What can we say about the performance of an estimator for a particular prediction?

Although there are many uses for comparative life history studies, the focus of this paper is on the prediction of parameter values in new situations. Researchers then want to know how well an estimator is likely to perform for the particular stock or species of interest rather than how well it works on average. As noted above, cross-validation tells us about how well an indirect estimator can replicate values from direct estimators, not how close to true parameter values we can come.

One might think that standard regression theory can provide guidance. Consider the regression of the logarithm of M on the logarithm of maximum known age in Figure 3. Clearly there is a relationship between the dependent and independent variables. For what follows, we will assume there is no error in the independent variable, no uncertainty in the model being linear, and the residual errors are independent and identically distributed normal random variables. Also shown in Figure 3 are 95% confidence bands for the regression and 95% prediction intervals for a new prediction. The question is: why don't the data points fall on the line and, thus, why doesn't a particular data point of interest fall on the line?

There are two extreme cases to consider. First, suppose that the true values of $\log(M)$ fall on a (true, but unknown) regression line. The confidence interval describes in some sense the uncertainty in the regression line and thus in the prediction of the M for a new situation. By implication, for this case, direct estimates of M should be replaced by the predicted values from the regression because the regression estimates are better than the direct estimates upon which the regression is based.

At the other extreme, suppose there is no measurement error—all of the data points are correct. The 95% prediction interval describes where 95% of new (future) observations should

fall and hence gives a range of plausible values (for this case) for M given a value of maximum age.

In reality, species with a given longevity vary in M and there is uncertainty about every estimate of M . Hence, the uncertainty in a new prediction is probably more than what is indicated by the confidence band and less than what is indicated by the prediction interval. (Theoretically, we could obtain multiple estimates of M for each of several species (by conducting multiple studies) to examine the very interesting question of the extent to which departures from the regression line are due to measurement error versus individual variability in the functional relationship between longevity and mortality.) Given data, at least some of this can be addressed with a hierarchical model. Any variability that can be attributed to random effect differences among taxonomic groups will not be attributed to sampling error (Appendix 1).

There may be more than one empirical method for estimating a parameter. Thus, it may be tempting to average predictions or to develop an estimator based on multiple predictor variables. To date, attempts to do this for the estimation of natural mortality rate do not appear to have met with much success (Then *et al.*, 2015; see also Hamel, 2015).

Often, one's interest is focused on a particular stock. It is then natural to judge a prediction for the stock in the context of auxiliary information such as taxonomic position, habitat or ecological role. Such considerations can be accommodated in hierarchical models, such as in Figure 6 where it is seen that separate relationships can be developed for individual taxonomic groups (but, because all data are considered in a hierarchical model, the data from each group lends strength to, i.e. influences, the estimated relationship for every other group). However, if a new model is developed because one is dissatisfied with a particular prediction from a standard (established) model one loses objectivity and the prediction interval for the particular estimate from the new model becomes hard to interpret.

Conclusions and Recommendations

Comparative life history studies, at least in an informal sense, are a basic part of routine stock assessment activities. That is, scientists routinely judge the reasonableness of parameter estimates in terms of the realm of experience with similar stocks. In a more formal sense, regression estimators, e.g. to estimate natural mortality rate, are used in many if not most stock assessments.

Perhaps the biggest obstacle to more widespread development and use of comparative life history studies is a widespread misunderstanding of what these studies accomplish. The average discrepancy between a regression prediction and a direct estimate is often taken as a measure of the (in)accuracy of the regression model. Thus, there are statements in the scientific literature that empirical estimators are highly imprecise, perhaps too imprecise to be used for stock assessment purposes. However, this is only true if the direct estimates are correct and the error is entirely or largely in the regression model. But, this is an unwarranted and unverified assumption. Certainly, an indirect regression estimate may be preferable to a poor direct estimate. And, it is certainly possible that a regression estimator provides more accurate and precise results than can be found in the collection of individual estimates in the literature.

We are now moving beyond simple regression analysis and using hierarchical models, mixed effect models, empirical Bayes estimators and data-mining techniques. The field of comparative life history studies is likely to grow because this approach

provides information needed for fundamental tasks of stock assessment.

We offer the following guidelines for conducting comparative life history studies.

- (i) Outline the structure of the available data including fixed and random effects variables and hierarchies, e.g. taxonomic, that are likely to affect the predictions of a life history parameter.
- (ii) Based on available data and a list of all potentially sampleable units, determine what units are being excluded from sampling; this is a fundamental step in determining the population about which inferences are to be made, e.g. all fishes versus fishes in certain habitats or taxonomic positions.
- (iii) Document and make publicly available all data encountered, including data that are rejected as unsuitable, as well as the data screening procedures adopted; explore the consequences of using various subsets of the data
- (iv) Consider the effects of hierarchies in the data on predictions
- (v) Examine the possibility of data selection bias, both in the literature and through quality control procedures, and its consequences

Acknowledgements

This study was funded by an NMFS Stock Assessment Improvement Grant awarded to the Southeast Fisheries Science Centre and was supported by Virginia Sea Grant and by the Malaysian Ministry of Higher Education—University of Malaya scholarship awarded to A.Y.-H.T. The work of E.A.B. was supported in part by a NOAA grant through the Cooperative Institute for Marine and Atmospheric Studies at the University of Miami (Award Number NA150AR4320064). Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government. We thank Andre Punt and an anonymous reviewer for helpful comments. This article is Contribution No. 3548 of the Virginia Institute of Marine Science, College of William & Mary.

References

- Ault, J. S., Bohnsack, J. A., and Meester, G. A. 1998. A retrospective (1979–1996) multispecies assessment of coral reef fish stocks in the Florida Keys. *Fishery Bulletin*, 96: 395–414.
- 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.
- Best, P. B. 1993. Increase rates in severely depleted stocks of baleen whales. *ICES Journal of Marine Science*, 50: 169–186.
- 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. *Ciba Foundation Colloquia on Ageing*, 5: 142–180.
- Blueweiss, L., Fox, H., Kudzma, V., Nakashima, D., Peters, R., and Sams, S. 1978. Relationships between body size and some life-history parameters. *Oecologia*, 37: 257–272.
- Brodziak, J., Ianelli, J., Lorenzen, K., and Methot, R.D. (Eds). 2011. Estimating natural mortality rate in stock assessment applications. NOAA Technical Memorandum NFMS-F/SPO-119.
- Charnov, E. L. 1993. Life History Invariants: Some Explorations of Symmetry in Evolutionary Ecology, Vol. 6. Oxford University Press, Oxford. 184 p.
- Charnov, E. L., Gislason, H., and Pope, J. G. 2013. Evolutionary assembly rules for fish life histories. *Fish and Fisheries*, 14: 213–224.
- Cubillos, L. A., Alarcon, R., and Brante, A. 1999. Empirical estimates of natural mortality for the Chilean hake (*Merluccius gayi*): evaluation of precision. *Fisheries Research*, 42: 147–153.
- Donald, D. B., and Alger, D. J. 1989. Evaluation of exploitation as a means of improving growth in a stunted population of brook trout. *North American Journal of Fisheries Management*, 9: 177–183.
- Dorazio, R. M., Royle, J. A., Soderstrom, B., and Glimskar, A. 2006. Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology*, 87: 842–854.
- Dorn, M. W. 2002. Advice on West Coast rockfish harvest rates from Bayesian meta-analysis of stock-recruit relationships. *North American Journal of Fisheries Management*, 22: 280–300.
- Froese, R., and Pauly D. 2015. FishBase. World Wide Web electronic publication. <http://www.fishbase.org>
- Fryer, G., and Iles, T. D. 1972. The Cichlid Fishes of the Great Lakes of Africa. Oliver and Boyd, Edinburgh. 641 p.
- Gedamke, T., Hoenig, J. M., Musick, J. A., DuPaul, W. D., and Gruber, S. H. 2007. Using demographic models to determine intrinsic rate of increase and sustainable fishing for elasmobranchs: pitfalls, advances and applications. *North American Journal of Fisheries Management*, 27: 605–618.
- Gelman, A., and Hill, J. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York. 625 p.
- 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.
- Hamel, O. S. 2015. A method for calculating a meta-analytical prior for the natural mortality rate using multiple life history correlates. *ICES Journal of Marine Science*, 72: 62–69.
- Hastie, T., Tibshirani, R., and Friedman, J. 2009. *The Elements of Statistical Learning*, 2nd edn. Springer, Heidelberg. 739 p.
- Helser, T. E., and Lai, H. L. 2004. A Bayesian hierarchical meta-analysis of fish growth: with an example for North American largemouth bass, *Micropterus salmoides*. *Ecological Modelling*, 178: 399–416.
- Heron, A. C. 1972. Population ecology of a colonizing species: the pelagic tunicate *Thalia democratica*. II. Population growth rate. *Oecologia*, 10: 294–312.
- Hewitt, D. A., and Hoenig, J. M. 2005. Comparison of two approaches for estimating natural mortality based on longevity. *Fishery Bulletin*, 103: 443–437.
- 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. 1983. Empirical use of longevity data to estimate mortality rates. *Fishery Bulletin*, 82: 898–903.
- Johnson, K. F., Monnahan, C. C., McGilliard, C. R., Vert-pre, K. A., Anderson, S. C., Cunningham, C. J., Hurtado-Ferro, F., *et al.* 2015. Time-varying natural mortality in fisheries stock assessment models: identifying a default approach. *ICES Journal of Marine Science*, 72: 137–150.
- Jones, R. 1982. Ecosystems, food chains and fish yields. *In Theory and Management of Tropical Fisheries*, pp. 195–239. Ed. by D. Pauly, and G. I. Murphy. ICLARM Conference Proceedings 9. International Center for Living Aquatic Resources Management, Manila, Philippines, and Division of Fisheries Research, Commonwealth Scientific and Industrial Research Organization, Cronulla, Australia. 360 p.

- Kenchington, T. J. 2014. Natural mortality estimators for information-limited fisheries. *Fish and Fisheries*, 15: 533–562.
- Kery, M., and Royle, J. A. 2008. Hierarchical Bayes estimation of species richness and occupancy in spatially replicated surveys. *Journal of Applied Ecology*, 45: 589–598.
- Lester, N. P., Dextrase, A. J., Kushneriuk, R. S., Rawson, M. R., and Ryan, P. A. 2004. Light and temperature: key factors affecting walleye abundance and production. *Transactions of the American Fisheries Society*, 133: 588–605.
- Liermann, M., and Hilborn, R. 1997. Depensation in fish stocks: a hierarchical meta-analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 15: 1976–1984.
- 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.
- Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. 2014. *The BUGS Book: A Practical Introduction to Bayesian Analysis*. CRC Press, Boca Raton, Florida. 381 p.
- Marten, G. G., and Polovina, J. J. 1982. A comparative study of fish yields from various tropical ecosystems. In *Theory and Management of Tropical Fisheries*, pp. 255–289. Ed. by D. Pauly, and G. I. Murphy. ICLARM Conference Proceedings 9. International Center for Living Aquatic Resources Management, Manila, Philippines, and Division of Fisheries Research, Commonwealth Scientific and Industrial Research Organization, Cronulla, Australia. 360 p.
- 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.
- McDermid, J. L., Sloan, W. N., Wilson, C. C., and Shuter, B. J. 2010. Early life history variation among hatchery- and wild-origin lake trout reared in a hatchery environment. *Transactions of the American Fisheries Society*, 139: 21–28.
- Melnichuk, M. C., Essington, T. E., Branch, T. A., Heppell, S. S., Jensen, O. P., Link, J. S., Martell, S. J. D., et al. 2012. Can catch share fisheries better track management targets? *Fish and Fisheries*, 13: 267–290.
- Myers, R. A. 2001. Stock and recruitment: generalizations about maximum reproductive rate, density dependence, and variability using meta-analytic approaches. *ICES Journal of Marine Science*, 58: 937–951.
- Myers, R. A., Bridson, J., and Barrowman, N. J. 1995. Summary of worldwide stock and recruitment data. *Canadian Technical Report on Fisheries and Aquatic Sciences*, 2024: 327.
- Myers, R. A., and Mertz, G. 1998. The limits of exploitation: a precautionary approach. *Ecological Applications*, 8: S165–S169.
- Nakagawa, S., and Freckleton, R. P. 2008. Missing inaction: the dangers of ignoring missing data. *Trends in Ecology and Evolution*, 23: 592–596.
- 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.
- Pascual, M. A., and Iribarne, O. O. 1993. How good are empirical predictions of natural mortality? *Fisheries Research*, 16: 17–24.
- 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. 1984. Fish population dynamics in tropical waters: a manual for use with programmable calculators. ICLARM Studies and Reviews 8. International Center for Living Aquatic Resources Management, Manila.
- Punt, A. E., Smith, D. C., and Smith, A. D. M. 2011. Among-stock comparisons for improving stock assessments of data-poor stocks: the “Robin Hood” approach. *ICES Journal of Marine Science*, 68: 972–981.
- Royle, J. A. 2008. Modeling individual effects in the Cormack–Jolly–Seber model: a state–space formulation. *Biometrics*, 64: 364–370.
- Royle, J. 2009. Analysis of capture-recapture models with individual covariates using data augmentation. *Biometrics*, 65: 267–274.
- Royle, J. A., and Dorazio, R. M. 2008. *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*. Academic Press, Amsterdam. 444 pp.
- Ryder, R. A. 1965. A method for estimating the potential fish production of north-temperate lakes. *Transactions of the American Fisheries Society*, 94: 214–218.
- Ryder, R. A. 1982. The morphoedaphic index—use, abuse, and fundamental concept. *Transactions of the American Fisheries Society*, 111: 154–164.
- Ryder, R. A., Kerr, S. R., Loftus, K. H., and Regier, H. A. 1974. The morphoedaphic Index, a fish yield estimator — review and evaluation. *Journal of the Fisheries Research Board of Canada*, 31: 663–688.
- Schmidt, J. H., Lindberg, M. S., Johnson, D. S., Conant, B., and King, J. 2009. Evidence of Alaskan Trumpeter Swan Population Growth Using Bayesian Hierarchical Models. *Journal of Wildlife Management*, 73: 720–727.
- Shuter, B. J., Jones, M. L., Korver, R. M., and Lester, N. P. 1998. A general life history based model for regional management of fish stocks: the inland lake trout (*Salvelinus namaycush*) fisheries of Ontario. *Canadian Journal of Fisheries and Aquatic Sciences*, 55: 2161–2177.
- SPOF (Strategic Planning for Ontario Fisheries). 1982. Partitioning yields estimated from the morphoedaphic index into individual species yields: report of SPOF Working Group Number 12. Ontario Ministry of Natural Resources, Fisheries Branch. 71 p.
- Then, A. Y., Hoenig, J. M., Hall, N. G., and Hewitt, D. A. 2015. Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. *ICES Journal of Marine Science*, 72: 82–92.
- Thorson, J. T., Cope, J. M., Kleisner, K. M., Samhoury, J. F., Shelton, A. O., and Ward, E. J. 2013. Giants’ shoulders 15 years later: lessons, challenges and guidelines in fisheries meta-analysis. *Fish and Fisheries*, 16: 342–361.
- Tyrrell, M. C., Link, J. S., Moustahfid, H., and Overholtz, W. J. 2008. Evaluating the effect of predation mortality on forage species population dynamics in the Northeast US continental shelf ecosystem using multispecies virtual population analysis. *ICES Journal of Marine Science*, 65: 1689–1700.
- Walters, C. J., and Ludwig, D. 1981. Effects of measurement errors on the assessment of stock-recruitment relationships. *Canadian Journal of Fisheries and Aquatic Sciences*, 38: 704–710.
- Welcomme, R.L. 1985. River fisheries. *FAO Fisheries Technical Paper*, (262): 330 p.
- Zhang, C. I., and Megrey, B. A. 2006. A revised Alverson and Carney model for estimating the instantaneous rate of natural mortality. *Transactions of the American Fisheries Society*, 135: 620–633.
- Zhang, Z. N., Lessard, J., and Campbell, A. 2009. Use of Bayesian hierarchical models to estimate northern abalone, *Haliotis kamtschatkana*, growth parameters from tag-recapture data. *Fisheries Research*, 95: 289–295.
- Zhou, S., Yin, S., Thorson, J. T., Smith, A. D. M., and Fuller, M. 2012. Linking fishing mortality reference points to life history traits: an empirical study. *Canadian Journal of Fisheries and Aquatic Sciences*, 69: 1292–1301.

Appendix 1. Hierarchical modelling of life history relationships

Methods

To illustrate the effect of using hierarchical models versus simple regression models, we generated a simulated dataset that included variation in regression parameters between species and between taxonomic groups (Figure A1). Hierarchical and non-hierarchical models were fitted to these data and used to predict life history parameter Y (say, natural mortality) for a new species with a known value of the life history parameter X (say, inverse longevity). We considered the case where the new species was from a taxonomic group already studied and where it was from a taxonomic group not included in the analysis.

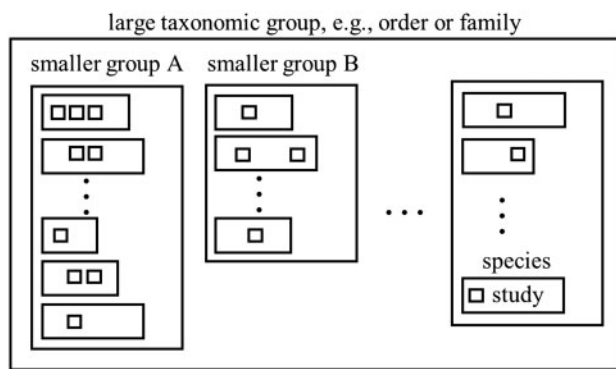


Figure A1. Hierarchical structure of simulated data. There are studies (smallest boxes) within species nested within a higher taxonomic group such as genus or family or order nested within an even higher taxonomic group. From the perspective of predicting the dependent variable Y for a new species, the key question is whether there are significant differences in Y associated with the group identity.

Data were generated for n studies of Y with the following model:

$$Y_i = \beta_{0,g(i)} + \beta_{1,g(i)}X_i + \alpha_{sp(i)} + \epsilon_i \tag{A1}$$

where i denotes the study, $\beta_{0,g(i)}$ is the intercept term and $\beta_{1,g(i)}$ is the slope for taxonomic group g (a family or genus, say) corresponding to study i , $\alpha_{sp(i)}$ is a species effect for the species in study i , and ϵ_i is a normally distributed random error [$\epsilon_i \sim \text{Normal}(0, \sigma_{\text{error}}^2)$], which, for simplicity, is the same for all studies of all species (Table A1). The intercepts, slopes and species effects are random effects drawn from independent normal distributions with specified means (μ_{β_0} , μ_{β_1} , and 0) and variances ($\sigma_{\beta_0}^2$, $\sigma_{\beta_1}^2$, and σ_{α}^2), respectively, with the random effect across species having the same distribution for all taxonomic groups. The simulated dataset was generated with 25 species in 5 taxonomic groups, and 4 studies per species ($n = 500$), so that the data were informative to estimate all of the parameters.

The models fitted to the data were: (i) a simple linear regression, (ii) a one-level hierarchical model with species as a random effect, (iii) a two-level hierarchical model including both species and group effects, and (iv) a one-level hierarchical model with a random effect of group. All models were fitted using a Bayesian statistical framework with uninformative priors. However, similar results could be obtained using linear mixed models in a classical statistical framework. The regression model treated all data points equally and estimated a single intercept, slope, and error variance:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{A2}$$

The one-level hierarchical model was the same as the regression, except that it also included a normally distributed random effect of species ($\alpha_{sp(i)}$) with a mean of zero and an estimated variance [$\alpha_{sp} \sim \text{Normal}(0, \sigma_{\alpha}^2)$]:

$$Y_i = \beta_0 + \beta_1 X_i + \alpha_{sp(i)} + \epsilon_i. \tag{A3}$$

Table A1. Values of the parameters used to generate the simulated data ('true') and the values estimated by the three models

Parameter	Description	True	Regression	Species effect	Two-level	Group effect
σ_{error}	Error SD	0.050	0.144	0.053	0.054	0.055
σ_{α}	SD in intercept between species	0.030	NA	0.142	0.020	NA
σ_{β_0}	SD in intercept between groups	0.070	NA	NA	0.120	0.125
σ_{β_1}	SD in slope between groups	0.600	NA	NA	0.797	0.833
μ_{β_0}	mean intercept between groups	0.040	0.023	0.023	0.052	0.058
$\beta_{0,1}$	intercept of group 1	0.099	NA	NA	0.114	0.114
$\beta_{0,2}$	intercept of group 2	0.031	NA	NA	0.030	0.030
$\beta_{0,3}$	intercept of group 3	-0.026	NA	NA	-0.054	-0.054
$\beta_{0,4}$	intercept of group 4	0.024	NA	NA	0.023	0.023
$\beta_{0,5}$	intercept of group 5	0.081	NA	NA	0.147	0.147
μ_{β_1}	mean slope between groups	3.000	3.237	3.242	2.961	2.898
$\beta_{1,1}$	slope of group 1	2.051	NA	NA	1.997	1.659
$\beta_{1,2}$	slope of group 2	2.940	NA	NA	2.888	2.888
$\beta_{1,3}$	slope of group 3	2.835	NA	NA	3.032	3.065
$\beta_{1,4}$	slope of group 4	2.561	NA	NA	2.655	2.611
$\beta_{1,5}$	slope of group 5	4.674	NA	NA	4.225	4.273
α_1	species effect for species 1	0.017	NA	-0.032	0.000	NA
α_2	species effect for species 2	0.002	NA	0.036	0.011	NA
α_3	species effect for species 3	-0.046	NA	-0.086	-0.017	NA
α_4	species effect for species 4	0.013	NA	0.056	0.011	NA
α_5	species effect for species 5	-0.008	NA	-0.067	-0.004	NA

Only the first 5 of the 25 species effects are shown.

The two-level hierarchical model had the same structure that was used to simulate the data (Equation A1), with random effects of species, and random effects for the intercept $[\beta_{0,g} \sim \text{Normal}(0, \sigma_{\beta_0}^2)]$ and slope $[\beta_{1,g} \sim \text{Normal}(0, \sigma_{\beta_1}^2)]$ of the regression at the group level:

$$Y_i = \beta_{0,g(i)} + \beta_{1,g(i)}X_i + \alpha_{sp(i)} + \epsilon_i \quad (\text{A4})$$

The last model included a random effect of group on both the slope and the intercept, but no species effect:

$$Y_i = \beta_{0,g(i)} + \beta_{1,g(i)}X_i + \epsilon_i \quad (\text{A5})$$

To evaluate the implications of these different model structures on the prediction of Y for a new species, we calculated predictions from all the models for a hypothetical unstudied species with an arbitrary value of X . At that value of X , we calculated, with each model, both the credible interval of the mean of Y (i.e. the range from the 2.5 percentile to the 97.5 percentile of the posterior distribution of the mean of Y at the specified value of X) and the prediction interval (i.e. the range from the 2.5 percentile to the 97.5 percentile of the posterior distribution of a hypothetical new data point at the same X value; this is wider than the credible interval of the mean because it includes observation error). For the regression model and the one-level hierarchical model with a species effect, the taxonomic group of the new species does not influence the prediction. For the models with random effects of group, we predicted values for all the groups that were included in the dataset, and for a species from a new taxonomic group. For a species within one of the studied groups, the expected value of Y would be calculated as:

$$Y_i = \beta_{0,g(i)} + \beta_{1,g(i)}X_i + \alpha_{sp,new} \quad (\text{A6})$$

for the two-level model, using the values of $\beta_{0,g(i)}$ and $\beta_{1,g(i)}$ for that group. Because the value of $\alpha_{sp,new}$ for a new species is not known, values from this parameter are randomly drawn from the estimated distribution of species random effects $[\alpha_{sp} \sim \text{Normal}(0, \sigma_{\alpha}^2)]$, so that the estimated posterior distribution of Y_i includes the uncertainty in this parameter. For a species in a new taxonomic group, $\beta_{0,g(i)}$ and $\beta_{1,g(i)}$ would also be drawn from their respective distributions, further increasing the uncertainty in Y_i .

Simulations were conducted in R, and models were fit using the Bayesian software JAGS (Lunn *et al.*, 2013). The JAGS software uses the MCMC algorithm to estimate the marginal posterior distribution of each estimated parameter or prediction. The values of all the predictions are calculated at each iteration of the MCMC algorithm, so that the uncertainty in all the estimated parameters and the expected variances at the level of species and group are integrated to give the posterior distribution of the predictions.

Results

The simulated data had quite different relationships between X and Y in the different taxonomic groups (Figure A2a, Table A1). Because of this, the predicted value of Y for a new species would be highly dependent on its taxonomic group, even if the underlying values of the regression parameters for each group were

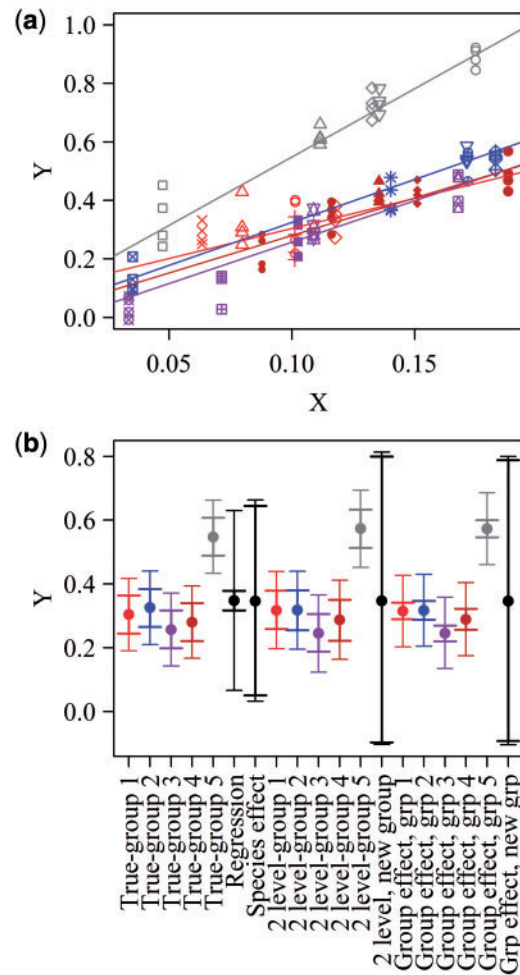


Figure A2. (a) Simulated data for life history parameter (Y) predicted by life history parameter (X), with both the slope and the intercept varying by taxonomic group. Plotting symbols indicate the species, lines are the true relationship between X and Y for each group. The top line is group 5, followed by (at $X = 0.1$) groups 2, 1, 4 and 3. (b) Predicted values of Y for a new species with $X = 0.1$. Inner bounds are the credible interval for the mean value of Y for a new species. Outer bounds are the prediction interval for a new study of a new species (including measurement error). Central value is the posterior median. The values shown at the left are the 'true' values calculated from the known parameter values for each group. On the right are the predictions from a simple linear regression, a one-level hierarchical model with a species random effect, a two-level hierarchical model with both species and group random effects, and a one-level hierarchical model with a group effect but no species effect. For models with a group effect, predictions are shown for a new species in each studied group, and for a new species in an unstudied group.

known perfectly (Figure A2b, values labeled 'true' where, among the five groups shown, Group 5 is much higher than the others). When a simple linear regression was used to fit these data, the variation between groups and between species was interpreted as observation error (Figure A2b, Table A1). Thus, the credible interval of the mean Y at $X = 0.1$ is quite narrow, but the prediction interval is broad (Figure A2b: Regression) reflecting the fact that the regression method does not utilize the information in the group identity of the new species to predict Y . On the other hand,

a one-level hierarchical model with a species random effect interprets most of the variation in the data as differences between species rather than sampling error (Figure A2b: Species effect). Thus, although the one-level hierarchical model and the regression model give very similar point estimates of Y and similar prediction intervals, the credible intervals of the mean are very different.

The two-level hierarchical model interprets much of the variation in the data as differences between groups, so that both the variance between species and the sampling error variance are smaller (Table A1, Figure A2b, 2-level). Therefore, when the two-level hierarchical model is used to predict Y values for a new species in one of the five studied groups, both the credible interval of the mean and the prediction interval are narrow. On the other hand, when the two-level hierarchical model is used to predict a value for a new species in an unstudied group, the credible interval of the mean and the prediction interval are broad, because they include the variation between groups.

Finally, when a one-level hierarchical model with a group effect but no species effect is used, the model interprets differences between species within each group as being part of the differences between groups (Figure A2b: Group effect). Thus, the credible interval for the mean of a new species in one of the studied groups is narrow, and the prediction interval is broad, compared with the results from the model with both group and species effects. The credible interval and prediction interval for a new species in an unstudied group are fairly similar for any of the hierarchical models, because all the variation between groups and species is included in the prediction interval.

Discussion

Which of these models is preferable (in general) for prediction depends on whether there is significant variation between taxonomic groups in the relationship between X and Y , and also on the research question being asked. Models with different hierarchical structures differ from each other mainly in how they partition the variance in the data into observation error variance versus random effect variance at the level of various groups. If the researchers are only interested in producing a point estimate of Y for a particular value of X for an unstudied species, the models may give very similar results. In the case that there is no variation between taxonomic groups, any of these model structures, including the simple regression, will provide very similar point estimates of Y at any particular value of X . In the case where

taxonomic groups vary, the models also produce very similar point estimates if the prediction is for a species in an unstudied group. The point estimates only differ for predictions within the studied groups, when there are differences between the groups. In that case, a hierarchical model that includes a group random effect can improve predictions for a new species in the studied groups. Interestingly, including a species random effect has a negligible influence on the point estimate of Y for a new species in the studied groups. Species effects may be difficult to estimate, since it is necessary to have multiple studies for some species in order to distinguish species effects from observation error. However, our results indicate that it may not be necessary to estimate species effects, if only point estimates are needed.

If researchers want credible intervals or prediction intervals as well as point estimates, or if the partitioning of variance is of interest, then the choice of hierarchical structure requires more thought. In our simulated data example, there were differences between groups and between species within groups. Without hierarchical structure, the regression model overestimated the observation error variance. Thus, for the regression, the credible intervals for the mean Y at any value of X are narrower than they should be. The model that had only a species random effect provided a fairly accurate estimate of the observation error variance, but greatly overestimated the variance between species, because it was including all the differences between groups in the species effect. The two-level hierarchical model was able to estimate all of the variance components, although it did somewhat overestimate the difference between groups and underestimate the difference between species (But, this observation pertains to the specific dataset that was simulated and may not hold in general). The model with only group effects was very similar to the two-level hierarchical model. This implies that, as expected, if there is hierarchical structure in the data, then a model that is able to estimate the important elements of the structure will produce better estimates of the variance components. In our simulated data, the differences between groups were the most important hierarchical component.

In an actual study of life history relationships among the fishes, there may be important differences at multiple taxonomic levels (order, family, genus, species, population) and data may not exist to estimate random effects at more than one or two levels. These simulation results seem to imply that capturing the differences between groups at one or two levels can improve prediction, even if some taxonomic levels have to be left out due to data constraints.