Vitality Models Found Useful in Modeling Tag-Failure Times in Acoustic-tag Survival Studies

John R. Skalski[1] *and Steven L. Whitlock*

Columbia Basin Research

School of Aquatic and Fishery Sciences, University of Washington

1325 Fourth Ave, Suite 1515, Seattle, Washington, 98101

[1] Corresponding author: skalski@uw.edu

Abstract

Acoustic telemetry studies often rely on the assumption that premature tag failure does not affect the validity of inferences. However, in some cases this assumption is possibly or likely invalid and it is necessary to apply a correction to estimation procedures. The question of which approaches and specific models are best suited to modeling acoustic tag failures has received little research attention. In this short communication, we present a meta-analysis of 42 acoustic tag-life studies, originally used to correct survival studies involving outmigrating juvenile salmonids in the Columbia/Snake river basin. We compare the performance of nine alternative parametric models including common failure-time/survival models and vitality models of Li and Anderson (2009 and 2013), which characterize demographic heterogeneity in the mortality of populations. The tag-life studies used acoustic tags from three different tag manufacturers, had expected lifetimes between 12 and 61 days, and had dry weights ranging from 0.22 to 1.65 grams. In 57% of the cases, the vitality models of Li and Anderson (2009 and 2013) fit the tag failure-times best. The vitality models were also the second-best choices in 17% of the cases. Together, the vitality models, log-logistic, (19%), and gamma models (14%) accounted for 90% of the models selected. Unlike more traditional failure-time models, the vitality models are capable of characterizing both the early onset of tag failure due to manufacturing errors and the anticipated battery life. We provide further guidance on appropriate sample sizes (50–100 tags) and procedures to be considered when applying precise tag-life corrections in release-recapture survival studies.

Keywords: acoustic tags, fish survival, tag life, failure time analysis, release-recapture

Background

Acoustic telemetry is a powerful tool for studying fish movement and survival (Jellyman 2009, Crossin et al. 2017). While many studies reasonably assume that tags do not fail during the study period (Melnychuk 2010, Plumb et al. 2016), there are other studies with design limitations related to the size of the organism, duration of the study, and detection capability that make a degree of tag failure within a study unavoidable (Adams et al. 2012, Steig and Holbrook et al. 2009). Under these circumstances, it becomes necessary to correct for tag life expectancies in order to make reliable inferences (Cowen and Schwarz 2005, Townsend et al. 2006).

Correcting for premature acoustic tag failures is particularly critical in estimating the survival of outmigrating juvenile salmonids at dams in major rivers (Harnish et al. 2012, Skalski et al. 2014, Skalski et al 2016a). Often in these studies, investigators apply single (Cormack 1964) and multistate (Buchanan et al. 2013, Perry et al. 2013) release-recapture models to estimate perceived survival, the joint probability of the fish and tag being alive from one detection point to another over time. These perceived estimates of survival are negatively biased in the presence of post-release tag failure (Arnason and Mills 1981), unless information on tag life or failure times is available for correction.

The severity of bias from post-release tag failure is dictated by the temporal overlap between detections at the farthest interrogation site and the tag-failure distribution. However, even minor overlap may be consequential when estimates are required to meet specific standards. For example, federal hydropower operations in the Columbia River Basin must comply with minimum juvenile salmonid passage survival standards, which stipulate both a particular survival threshold and a minimum level of precision for estimates (Skalski et al. 2016b). The pattern of tag failure is known to vary across manufacturing lots, so it is advisable to conduct concurrent tag-life studies, in which a sample of tags is activated alongside active tags used in the survival study (Albert et al. 2010). These sampled tags are monitored by a hydrophone to measure the time until failure, and a model is then fitted that represents the expected tag life and is used to correct survival estimates (Townsend et al. 2006).

Some studies have modeled tag failure using nonparametric approaches (Cowen and Schwarz 2005, Holbrook et al. 2013), while Townsend et al. (2006) recommended a parametric approach to modeling the failure-time data, because if a parametric model is found that fits the empirical data the precision of the tag-life corrected survival estimates is improved. There is a suite of traditional failure-time distributions to select from when performing tag-life corrections including gamma, Gompertz (1825), log-logistic, lognormal (Elandt-Johnson and Johnson 1980: 62–63), and Weibull (1939). Alternative models vary in flexibility and how well they fit failure-time data based on the number of parameters and the assumption of how risk of failure changes through time.

A seemingly unlikely source for further model consideration comes from the study of population demographics and animal survival. Li and Anderson (2009, 2013) modeled death times as a survival process that depends on two components, a vitality-dependent process intrinsic to the individual and a vitality-independent process associated with accidental death. These two processes are analogous to the propensity of battery failure and accidental manufacturing errors in modeling tag life. Because tag lots often have a mixture of these two sources of failure, the 4-parameter versions of these models (“Vitality (2009)” and “Vitality (2013)” hereafter) have the potential to better model tag-failure times where simpler models cannot capture the complexity of the survival process.

Here, in this short communication, we took the opportunity to evaluate the fit of alternative failure-time models to 42 different acoustic-tag life studies all conducted using the same protocol between 2002–2018. Our purpose was to thoroughly examine the relative performance of these models so as to provide guidance to investigators on the best candidate models and strategies for incorporating tag-life corrections into release-recapture survival studies of fish.

Methods

We first describe the nine alternative models, then our procedure for evaluating goodness-of-fit (GOF) and ranking the performance of models for each study. We have limited our model descriptions to their general characteristics and relationships. Additional details on the conventional failure-time/survival models that we evaluated may be found in Lee and Wang (2003). The structure and motivation of the two 4-parameter Vitality models is thoroughly outlined in the Li and Anderson (2009) and Li and Anderson (2013).

Tag-Failure Models

The survival function begins with a value of 1 (i.e., 100%) at $t = 0$ and declines as a function of time. A survival function $S(t)$ can be formed from any positive continuous probability distribution via its cumulative distribution function

$$S(t) = 1 - F(t)$$

where $F(t)$ is the cumulative distribution function where,

$$F(t) = \int_{0}^{\infty} f(t)dt$$

and where $f(t)$ is the density function. For reference, the hazard function is defined as

$$h(t) = \frac{f(t)}{S(t)}$$

and is also known as the instantaneous failure rate and characterizes the risk of failure over time (Lee and Wang 2003). The shape of the hazard function is often useful in selecting a failure-time model to a specific failure-time process (Table 1).

Table 1. Brief description of hazard functions for the nine failure-time models fit to the 42 sets of tag-life data plus the exponential function.

Failure-time functions	Hazard Function Description
Exponential	Constant over time
Weibull(2)	Linear or log-linear monotonic increase/decrease
Weibull(3)	Linear or log-linear monotonic increase/decrease with a shift that represents a guaranteed (“failure free”) period
Log-normal	Monotonic decrease or dome-shaped
Log−logistic	Monotonic decrease or dome-shaped. Allows for steeper decrease from apex than the log-normal
Gompertz	Constant initial hazard, which can remain constant or then have a quadratic increase/decrease
Gamma	Monotonic increase or decrease to a hazard rate of 1
Generalized Gamma	Monotonic increase or decrease to a specific hazard rate of α
Vitality 2009	Initial hazard rate followed by a shift in rate
Vitality 2013	Initial hazard rate followed by a shift in rate

Perhaps the simplest parametric failure-time model is the exponential model, with survival function

$$S(t) = e^{(-\lambda t)}$$

where the hazard rate is constant and defined by λ. Acoustic tag-failure rate is not uniform over time; thus, the exponential model is a generally a poor choice for this application. We excluded the exponential model from our analysis for this reason. Nonetheless, the exponential model is an appropriate starting point as it forms the basis of more complex failure-time models. The exponential distribution is a special case of the 2-parameter Weibull distribution, with survival function

$$S(t) = e^{-(\lambda t)^{\gamma}} \tag{1}$$

which in turn is a special case of the 3-parameter Weibull distribution with survival function (Weibull 1939, Elandt-Johnson and Johnson 1980: 62)

$$S(t) = e^{-\left(\frac{t-\gamma}{\lambda}\right)^{\beta}} \tag{2}$$

with shape (β), scale (λ), and shift (γ) parameters. The shift parameter describes the endpoint of an initial "failure free" portion of the curve.

Other common survival models that we considered were the 2-parameter gamma

$$S(t) = \int_{t}^{\infty} \frac{\lambda}{\Gamma(\gamma)} (\lambda t)^{\gamma-1} e^{-\lambda t} dt \tag{3}$$

with γ shape and λ scale parameters and the 3-parameter generalized gamma (Stacy 1962, Khodabin and Ahmadabadi 2010)

$$S(t) = \int_{t}^{\infty} \frac{\alpha\lambda^{\alpha\gamma}}{\Gamma(\gamma)} t^{\alpha\gamma-1} e^{\lfloor -(\lambda t)^{\alpha} \rfloor} dt \tag{4}$$

which includes α, an intercept parameter. The hazard function of the 2-parameter gamma decreases or increases to 1, whereas the generalized gamma approaches the value of α. The exponential, gamma, and Weibull distributions are special cases of the generalized gamma distribution.

The fifth distribution that we evaluated was the 2-parameter Gompertz distribution (Gompertz 1825), which is an extension of the exponential model that assumes the hazard rate increases exponentially with time or age. The survival function for the Gompertz model is

$$S(t) = e^{\left\{-\frac{e^{\lambda}}{\gamma}(e^{\gamma t}-1)\right\}} \tag{5}$$

where parameters λ and γ describe the intercept and slope of a log-linear regression equation for the hazard rate, respectively.

We considered the 2-parameter lognormal survival model that has a dome-shaped hazard function

$$S(t) = \frac{1}{\sigma\sqrt{2\pi}}\int_t^{\infty}\frac{1}{x}e^{\left(-\frac{1}{2\sigma^2}[\log(t)-\mu]^2\right)}dx, \tag{6}$$

with σ shape and μ scale parameters. The 2-parameter log-logistic has a similar shaped hazard function to the lognormal but allows for steeper declines from the apex

$$S(t) = \frac{1}{1+\gamma t^{\lambda}} \tag{7}$$

with γ shape and λ scale parameters.

The final two survival functions we examined were the 4-parameter vitality models. The Vitality (2009) model assumes a normal distribution of initial vitality across a batch of tags and a stochastic decline toward zero vitality. The survival function of the Vitality (2009) is defined as

$$S(t) = 1 - \left(\Phi\left(\frac{1-rt}{\sqrt{u^2+s^2t}}\right) - e^{\left(\frac{2u^2r^2}{s^4}+\frac{2r}{s^2}\right)}\Phi\left(\frac{2u^2r+rt+1}{\sqrt{u^2+s^2t}}\right)\right)^{e^{-kt}} \tag{8}$$

where Φ = cumulative normal distribution

r = wear rate,

s = standard deviation in wear rate,

k = rate of accidental failure,

u = standard deviation in quality of original components.

The Vitality (2013) model has a slightly different parameterization that assumes the same stochastic decline in tag vitality but combined with a Poisson process of challenge events of varying difficulty across a tag's lifetime. The survival function is

$$S(t) = \frac{t^{-3/2} e^{(1-rt)^2/2s^2 t}}{s\sqrt{2\pi}\left(\Phi\left(\frac{1-rt}{s\sqrt{t}}\right) - e^{2r/s^2}\Phi\left(-\frac{1+rt}{s\sqrt{t}}\right)\right)} + \lambda e^{-(1-rt)/\beta} \tag{9}$$

where

r = rate of vitality loss (intrinsic),

s = spread of initial and evolving vitalities (intrinsic),

λ = frequency of challenges during life (extrinsic),

β = magnitude of challenges (extrinsic).

We hypothesized that one or both vitality models would tend to fit acoustic-tag failure times well, as they allow for early onset of random tag failure due to manufacturing as well as systemic battery failure later on. The random failure component, in addition to battery discharge, gives the vitality model additional flexibility to fit data not found in other models.

Tag-Life Studies

The 42 different tag life studies were all performed with the same study procedures. Tags were systematically sampled from the tag lots used in salmon smolt survival studies conducted within the Columbia River Basin, 2002–2018. Within each test, tags were activated and monitored with hydrophones continuously until complete failure of all tags. The tags were submerged in ambient water the same temperature as the tagged fish encountered during the survival studies. Failure times were recorded to the minute. The failure time analyses used the time-to-failure measured in days and fractions of days.

The various acoustic tags analyzed were manufactured by Advanced Telemetry Systems, Hydroacoustic Technology Incorporated, and Lotek, with 16, 25, and 5 separate tag-life evaluations, respectively. Mean tag lives ranged from 12 to 61 days and sample sizes ranged from 38 to 125 tags per study. Tag sizes ranged from 0.22 to 1.65 grams dry weight. Tags were set to emit acoustic pulses between 1 and 6 times per minute, depending on the specific needs of the study.

Model Fitting and Comparison

The failure time data from the different tag-life studies conducted between 2002–2018 were fit to the nine alternative failure-time models within the R programming language and free software environment (https://www.r-project.org). For the more conventional survival analysis models (1–7), we used model-fitting routines in the "FAdist" and "flexsurv" R packages (Aucoin 2015, Jackson 2016). We fitted the two vitality models using routines available in the "Vitality" R package (Passolt et al. 2018).

Because of the diversity of models that we examined and the fact that many of the distributions involved were non-nested, we had to devise new metrics for assessing GOF and ranking model performance. The exponential 2- and 3-parameter Weibull models and gamma and generalized gamma models are nested and as such can be compared using likelihood ratio tests (Kotz, Johnson, and Read 1983: 647–650). However, the Gompertz, log-normal, log-logistic, and vitality models are not nested among themselves or the others. Also, in this situation, Akaike information criterion (Burnham and Anderson 2002: 61–64) cannot be used because the approach requires the alternative models share the same distribution.

Instead, we compared the various model fits to the empirical survival function using the nonparametric Kaplan-Meier (1958) product-limit method. The Kaplan-Meier (K-M) method estimates the survival function as

$$\hat{S}(t) = \prod_{t(i)<t} \left(\frac{n-i}{n-i+1} \right) \tag{10}$$

where

n = sample size,

i = number of failures before time t.

Relative GOF of the alternative parametric models was measured by the average squared deviation between the empirical K-M and the fitted model values (Figure 1) of $S(t)$ across the n observed failure times, i.e.,

$$GOF = \frac{\sum_{i=1}^{n} \left(\hat{S}(t_i) - S(t_i) \right)^2}{(n - p - 1)} \tag{11}$$

where

$\hat{S}(t_i)$ = survival value for parametric model at time t for the ith failure ($i = 1, \cdots, n$),

$S(t_i)$ = K-M survival value at time t for the ith tag failure ($i = 1, \cdots, n$),

p = number of fitted model parameters.

The number of parameters (p) serves as a penalty function for the number of estimated model parameters. The GOF was modeled after the mean square error for regression. The tag-failure model with the smallest GOF value was selected as the most appropriated.

We also performed lack-of-fit tests as characterized by the K-M nonparametric curve (10). The test statistic for the Kolmogorov-Smirnov (K-S) test is the absolute value of the largest discrepancy between $\hat{S}(t_i)$ and $S(t_i)$ anywhere along the fitted curve, i.e.,

$$D = \underset{i=1,\cdots,n}{\mathrm{MAX}} \left| \hat{S}(t_i) - S(t_i) \right| . \tag{12}$$

Whereas the traditional K-S test assumes the theoretical distribution being tested and its parameters are *a priori* specified, in our case, they were estimated from the data. Therefore, we used the approach of Lilliefors (1967), where the test distribution under the null hypothesis was simulated from the fitted model via parametric bootstrap. For each test performed, a random sample size n was drawn from the fitted survival functions and the value D calculated. This simulation process was replicated 50,000 times to create a distribution (D_{sim}) under the null hypothesis to which the observed statistic (D_0) was compared. This number of simulations was selected to guarantee a precision of ± 0.004 in the estimated P-values ($\sqrt{z_{0.975}^2 \cdot 0.5 \cdot 0.5/50000}$). Estimated P-values for the Lilliefors tests are reported in the supporting information, based on a $\alpha = 0.05$ rejection criterion. Whereas GOF provided a measure of relative goodness-of-fit to compare alternative models, the K-S test assessed whether there is a significant lack-of-fit of the selected model (i.e., H_o: model fits vs. H_a: model does not fit). By construct, the GOF and the D_0 of the K-S test are positively correlated.

In addition to testing and comparing the fit of models with the procedures described above, we also judged the most competitive models based on how well they appeared to fit empirical data in the initial portion of the curve preceding the primary decline in tag-life associated with battery failure, which we term the "shoulder" of the curve. Evaluating fit in this portion of the curve was valuable because it represented the portion of the function most likely to overlap with the last few detections in the survival studies.

Results

In 24 of the 42 cases (57%), a vitality model (2009 or 2013) was selected as the best fit among the nine alternative parametric failure-time models evaluated. The log-logistic model was the second most common (19%) choice, followed by the gamma or generalized gamma (17%), Gompertz (5%), 3-parameter Weibull (2%), and log-normal (2%). In numerous cases there were little differences in GOF between first, second, or even third choices of survival models. The two versions of the vitality model (2009 and 2013) were found to be top-ranking with equal frequency (12 data sets each), suggesting that no one version was clearly superior from the standpoint of model fit. Both version of the gamma model also performed equally well (3 data sets each).

The vitality models often outperformed other candidates because they could account for both early failures defining the shoulder of the function and the later precipitous decline, presumably due to battery failure. The log-logistic model fit these initial failures better than the remaining candidate models, although their survival functions were almost always positioned above those of the vitality models in the shoulder of the curve.

In all cases, the top-ranking survival model according to GOF was not rejected by the K-S test of lack-of-fit ($P < 0.05$). However, we found the K-S test to be insensitive to lack-of-fit. The K-S test rejected a fitted model only 60 times out of 378 (15.9% = 42 data sets x 9 candidate models) despite visually obvious cases of lack-of-fit. Therefore, non-rejection of a K-S test should not be the sole criterion for model selection. Nevertheless, we found a strong inverse relationship between the natural log of GOF value and the P-value of the K-S tests ($r = -0.79$, $P < 0.001$). With P-values ranging from 0 to 1, P-values near 1 indicated smaller discrepancy between observed and fitted values of the failure-time data. Using the K-S maximum P-value as a criterion for model selection, the vitality models were again selected in 57% of the cases studied, followed again by the log-logistic model at 19%.

In addition to being most frequently top-ranking, the vitality models also demonstrated considerable flexibility in the shape of the survival curves. We found that many of the tag-life datasets could be categorized as having a particular shape to which one of the conventional failure model was best suited. For example, gamma models tended to be top-ranking for data sets with survival curves resembling a half-normal distribution. Although vitality models were not always top-ranking for these cases, they consistently provided a fit that was competitive with the top-ranking model (Figure 2).

Discussion

Tagging studies with the objectives of describing fish movement and life history often do not include tag life studies as part of the investigation. Such studies are designed based on the anticipated life expectancy of the tags and the temporal requirements of the investigation (Adams et al. 2012). On the other hand, fish survival studies based on regulatory requirements with mandated survival thresholds will generally need to include formal tag-life studies (Skalski et al 2016a). Without the ancillary tag-life information, perceived survival estimates calculated by classic release-recapture models will be negatively biased by the presence of tag failure (Townsend et al. 2006, Holbrook et al. 2009). When actual fish survival is close to the regulatory thresholds, even small bias corrections can be consequential. For example, the compliance threshold for yearling Chinook salmon (*Oncorhynchus tshawytscha*) smolt survival through a hydroelectric project (i.e., reservoir plus dam) in the mid-Columbia River is typically ≥ 93%, with a 95% confidence interval on the survival estimate of ± 3% (Skalski and Bickford 2014, Skalski et al. 2012, Skalski et al. 2016b). Here even small tag life corrections of less than a percentage point can be important.

We found clear evidence to support the use of the vitality models for tag-life correction on the basis that these models were top-ranking in terms of GOF for the majority of data sets and exhibited a variety of survival function shapes that matched empirical tag-life data sets. The vitality models even tended to perform well when applied to the minority of tag-life data sets, which were discontinuous or had other aberrant characteristics. We do not recommend that model selection be based solely on the non-rejection of the K-S lack-of-fit test, as the test is rather conservative in the range of sample sizes (38 to 125) we evaluated. We instead recommend that investigators evaluate the GOF of their tag-life data to a suite of alternative survivorship models using both ocular and numerical evaluations of model fit. Among these, the alternative models should include vitality, log-logistic, and the gamma families of models.

We found the Vitality (2009) model to be preferable to the Vitality (2013) model. The GOF measure did not suggest clear dominance of one version of the vitality model over another, our determination is instead based on what we viewed as a stronger theoretical foundation for Vitality (2009) as well as post hoc comparisons of the two models across all datasets. We found the tag-failure process to be more analogous to the Vitality (2009) model, which assumes early failures as a result of a variability in initial vitalities in the population followed by a stochastic decline. While the Vitality (2013) has some similar properties, it further assumes that individuals encounter challenges of varying magnitude over a lifetime, which is not particularly representative of the process that acoustic tags undergo. Our second reason for favoring the Vitality (2009) model was that the survival curve for this model was less frequently above the K-M estimates in the shoulder of the curve than its counterpart.

In our experience, the shoulder of the survivorship curve is where most of the tag-life correction occurs and therefore should be estimated with greatest accuracy. A common reason for the poor fit of many models was that the curve descended too early, “cutting off” the shoulder present in the empirical data. Proceeding with a model misspecified in this manner would result in an overcorrection of survival estimates. Poor fit in the shoulder of some the tag-life data was partly what motivated our experimentation with the vitality class of models. This shortcoming was common for all models that we compared with the exception of the vitality models and to a lesser extent the log-logistic model. In fact, Weibull, lognormal, and gamma models only properly fit tag-life datasets without any early outlying failures. The Gompertz model was somewhat exceptional in that it was competitive with the vitality models for 6 out of 42 (14%) in which the initial decline in tag-life was relatively steep.

While providing tag-life corrected survival estimates are within the reach of all investigators, it remains the responsibility of individual studies to determine the appropriateness of collecting this expensive auxiliary information. It must be acknowledged that tag-life studies are costly and it is important that we discuss some of tradeoffs involved in choosing tag-life studies. Cost considerations occur at two levels. First, there is a question of a whether unique a tag-life study is warranted for a particular survival study. Second, there is a question of the number of tags that should be used. There are situations in which a single tag-life study may be applied to multiple survival studies, however, it may be necessary to adjust the tag-life correction protocol if the studies to which they are applied have dissimilar release schedules. We discuss this situation in greater detail below. With respect to the second consideration, tags cost approximately $200 each, resulting in a total expenditure of between \$10,000–\$20,000, if 50–100 tags are used. In our experience with juvenile salmonid acoustic-tag studies, sample sizes for tag-life studies should range between 50–100 tags. With 50 tags, the standard errors of the survival estimates are typically increased at the second and third decimal place. With 100 tags, the standard errors are changed at the third or fourth decimal place with the incorporation of the variability in tag-life data. Admittedly, not all studies warrant the degree of precision in survival estimates that the tag-life data sets in this study were intended to inform. However, it is worth noting that the lower the sample size the greater the chance that none of tags sampled for the tag-life study will possess defects that are actually present in tag population, in which case the early-failure process will not be incorporated into the correction.

Another important consideration when applying tag-life corrections is whether it is appropriate to perform a censored analysis of the tag-life data. There are at least two scenarios where a right-censored tag-life analysis may be useful and appropriate. The first scenario occurs when the tag-life study is stopped/truncated before the last tag failure. In this case, a right-truncated failure-time analysis is essential. Let T be the time of truncation, then the maximum likelihood estimates of the truncated model are based on the likelihood

$$L \propto \prod_{i=1}^{r} f(t_i) \cdot S(T)^{n-r}$$

where r is the number of tags that failed on or before the truncation time T. A second truncation scenario can occur when observed fish travel times are relatively short compared to observed tag-failure times and it is more accurate and easier to model tag-failure times to some truncation point beyond the longest travel time. This truncation strategy is when traditional failure-time distributions have difficulty fitting both the shoulder and tail of the failure-time curve. When inferences near the tail of the failure-time distributions are unnecessary, a truncated right-tailed analysis may do a better job fitting the shoulder of the survivorship curve where travel times are likely more relevant.

Ideally, the duration of the survival studies should be timed to be completed while still in the left-hand shoulder of the tag-life curves. Should the duration of the survival study coincide with the right-hand cascade of tag failures, tag life corrections will be greater and consequences to precision more profound. In the case where the duration of the survival study exceeds the tag-life curve, tag life corrections will be underestimated and the survival estimates will remain negatively biased to an unknown extent. Consequently, despite the mathematical ability to account for tag failure, it remains important to coordinate duration of the field study with tag selection and function. Harnish et al. (2012) discussed an issue of tag-life correction, unforeseen by Townsend et al. (2006). In their case, tag failures occurred so immediately and severely that it caused apparent negative bias in the distribution of arrival times. As a result, the recommended procedure for tag life correction applied too conservative of a correction.

This paper describes a meta-analysis of the performance of various models in fitting tag-life data sets and draws on extensive experience related to the application of tag-life correction to juvenile salmonid survival studies. We direct investigators to the freeware Program ATLAS (Active Tag Life Adjusted Survival), which can be used to interactively examine alterative tag-life models (i.e., vitality, Weibull), perform truncated tag-life analyses, and obtain tag-life corrected fish survival estimates (http://www.cbr.washington.edu/analysis/apps/atlas).

List of Abbreviations

GOF	goodness-of-fit
K-M	Kaplan-Meir product limit method
K-S	Kolmogorov-Smirnov test of lack-of-fit

Declarations

Ethics approval and consent to participate

Not Applicable.

Consent for publication

The manuscript does contain any individual person's data, details, images, or videos.

Availability of data and materials

Plots of the data and fitted survivorship curves, results of model selection (GOF), and K-S tests available in supplemental data provide.

Competing interests

The authors declare that they have no competing interests.

Funding

This publication was produced with the financial support of the Bonneville Power Administration, project 1989-107-00, contract #76910 REL 10.

Authors' Contributions

Both authors contributed equally to the paper.

Acknowledgements

We wish to thank Public Utility District No. 1 of Chelan County and Public Utility District No. 2 of Grant County of Washington State and the Portland, Oregon, and Walla Walla, Washington, districts of the U.S. Army Corps of Engineer for providing the tag-life data used in this paper.

References

Arnason, AN, Mills, KH. Bias and loss of precision due to tag log in Jolly-Seber estimates for mark-recapture experiments. Biometrics. 1981; 61:657–664.

Adams, NS, Beeman, JW, Eiler, JH. Telemetry techniques: a user guide for fisheries research. Bethesda: American Fisheries Society; 2012.

Albert, G, Skalski, JR, Pevin, C, Langeslay, M, Smith, S, Counihan, TD, Perry, RW, Bickford, S. Guidelines for conducting smolt survival studies in the Columbia River. United States Geological Survey. 2010.

Aucoin F. FAdist: distributions that are sometimes used in hydrology. R package version 2.2. 2015. https://CRAN.R-project.org/package=FAdist.

Buchanan, RA, Skalski, JR, Brandes, PR, Fuller, A. Route use and survival of juvenile Chinook salmon through the San Joaquin River Delta. North American Journal of Fisheries Management. 2013;33:216–229.

Burnham, K.P, Anderson, DR. Model selection and multimodel inference, 2nd ed. New York: Springer; 2002.

Cormack, RM. Estimates of survival from sighting of marked animals. Biometrika. 1964;51:429–438.

Cowen, LL, Schwartz, CJ. Capture-recapture studies using radio telemetry with premature radio-tag failure. Biometrics. 2005;61:657–664.

Crossin, GT, Heupel, MR, Holbrook, CM, Hussey, NE, Lowerre-Barbieri, SK, Nguyen, VM, Raby, GD, Cooke, SJ. Acoustic telemetry and fisheries management. Ecological Applications. 2017;27(4):1031–1049.

Elandt-Johnson, RC, Johnson, NL. Survival models and data analysis. New York: John Wiley & Sons; 1980.

Gompertz, B.: On the nature of the function expressive of the law of human morality, and on a new mode of determining life contingencies. London: Royal Society; 1825;513–585.

Harnish, RA, Johnson, GE, McMichael, GA, Hughes, MS, Ebberts, BD. Effect of Migration Pathway on Travel Time and Survival of Acoustic-Tagged Juvenile Salmonids in the Columbia River Estuary. Transactions of the American Fisheries Society. 2012;141(2):507–519.

Holbrook, CM, Perry, RW, Adams, NS. Distribution and joint-tag survival of juvenile Chinook salmon migrating through the Sacramento–San Joaquin River Delta, California, 2008. United States Geological Survey. 2009.

Holbrook, CM, Perry, RW, Brandes, PL, Adams, NS. Adjusting survival estimates for premature transmitter failure: a case study from the Sacramento-San Joaquin Delta. Environ Biol Fish. 2013;96(2):165–173.

Jackson, CH. Flexsurv: a platform for parametric survival modeling in R. Journal of Statistical Software. 2016;70(8).

Jellyman, D. 2009. A review of radio and acoustic telemetry studies of freshwater fish in New Zealand. Marine and Freshwater Research. 2009;60(4):321–327.

Kaplan, EL, Meier, P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association. 1958;53:457–481.

Khodabin, M, Ahmadabadi, A. Some properties of generalized gamma distribution. Mathematical Science. 2010;4(1):9–28.

Kotz, S, Johnson, NL, Read, CB. Encyclopedia of statistical sciences, vol 4. New York: John Wiley & Sons; 1983.

Lee, ET, Wang, JW. Statistical methods for survival data analysis, 3rd edn. New York: John Wiley & Sons; 1992.

Li, T, Anderson, JJ. The vitality model: A way to understand population survival and demographic heterogeneity. Theoretical Population Biology. 2009;76(2):118–131.

Li, T, Anderson, JJ. Shaping human mortality patterns through intrinsic and extrinsic vitality processes. Demographic Research. 2013;28:341–372.

Lilliefors, HW. On the Kolmorgorov-Smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association. 1967;62:399–402.

Melnychuk, MC. Estimation of survival and detection probabilities for multiple tagged salmon stocks with nested migration routes, using a large-scale telemetry array. Mar. Freshwater Res. 2010;60(12):1231–1243.

Passolt G, Anderson, JJ, Li, T, Salinger, DH, Sharrow, DJ. Vitality: fitting routines for the vitality family of mortality models. R package version 1.3. 2018. https://CRAN.R-project.org/package=vitality.

Perry, RW, Brandes, PL, Burau, JR, Klimley, AP, MacFarlane, B, Michel, C, Skalski, JR. Sensitivity of population-level survival to migration routes used by juvenile Chinook salmon to negotiate the Sacramento-San Joaquin River Delta. Environmental Biology of Fishes. 2013;96:381–392.

Plumb, JM, Adams, NS, Perry, RW, Holbrook, CM, Romine, JG, Blake, AR, and Burau, JR. Diel Activity Patterns of Juvenile Late Fall-run Chinook Salmon with Implications for Operation of a Gated Water Diversion in the Sacramento–San Joaquin River Delta. River Research and Applications. 2016;32(4):711–720.

Skalski, JR, Steig, T, Hemstrom, SL. Assessing compliance with fish survival standards: a case study at Rock Island Dam, Washington. Environmental Science & Policy. 2012;18:45–51.

Skalski, JR, and Bickford, S. Decadal compliance with the no-net-impact survival standards at Wells Hydroelectric Project, Columbia River, Washington. Northwest Science. 2014;88(2):120–128.

Skalski, JR, Eppard, MB, Ploskey, GR, Weiland, MA, Carlson, TJ, Townsend, RL. Assessment of Subyearling Chinook Salmon Survival through the Federal Hydropower Projects in the Main-Stem Columbia River. North American Journal of Fisheries Management. 2014;34(4):741–752.

Skalski, JR, Weiland, MA, Ploskey, GR, Woodley, CM, Eppard, MB, Johnson, GE, Carlson, GE, Townsend, RL. Establishing and using survival criteria to ensure the rigor and robustness of survival compliance testing at hydroelectric dams. Environment Systems and Decisions. 2016a; 36:404–420.

Skalski, JR, Weiland, MA, Ham, KD, Ploskey, GR, McMichael, GA, Colotelo, AH, Carlson, TJ, Woodley, CM, Eppard, MB, Hockersmith. EE. Status after five years of survival compliance testing in the Federal Columbia River Power System (FCRPS). North American Journal of Fisheries Management.2016b;36:720–730.

Steig, T, and CM Holbrook. Use of acoustic telemetry to evaluate survival and behavior of juvenile salmonids at hydroelectric dams: a case study from Rocky Reach Dam, Columbia River, USA. In: Adams, NS, Beeman, JW, Eiler, JH, editors. Telemetry Techniques. Bethesda: American Fisheries Society; 2012. p. 361–387.

Stacy, EW: Quasimaximum likelihood estimates for two-parameter gamma distribution. IBM Journal of Research and Development. 1962;17:115–124.

Townsend, RL, Skalski, JR, Dillingham, P, Steig, TW. Correcting bias in survival estimation resulting from tag failure in acoustic and radiotelemetry studies. Journal of Agricultural, Biological, and Environmental Statistics. 2006;11:183–196.

Weibull, W. A statistical theory of the strength of materials. Ingeniors Vetenskaps Akakemien Handlingar. 1939;151:293–297.

Figure 1. Illustration of Kaplan-Meier (K-M) nonparametric and a fitted parametric survival function and an observed deviation in survival values at the time of a failure event. The deviation in survival values is calculated at each time step in the K-M curve used in the GOF assessment (11)

Figure 2. Four examples of tag failure datasets with curves representing top-ranking survival models based on our GOF measure versus the Vitality 2009 model. Each panel describes an example with the Kaplan-Meier empirical estimates, the fit from the 2009 Vitality model and the top-ranked model according to our GOF measure (a) log-logistic (b) 3-parameter Gamma (c) Gompertz (d) and 2-parameter Gamma.