

Quantifying migratory delay: a new application of survival analysis methods

Theodore Castro-Santos and Alex Haro

Abstract: Statistical techniques commonly used in fish passage research fail to adequately quantify delays incurred at obstacles, or the effects of modifications to those obstacles on passage rates. Analyses of telemetry data describing these effects can be misleading, particularly when passage route of some individuals is not established (e.g., because of mortality, tag failure, passage through unmonitored or alternate routes, etc.). Here, we demonstrate how event-time analysis, better known as survival analysis, can be used to quantify passage rates for any study that allows tracking of individuals through time, even when some individuals fail to pass the route or obstacle in question. We review two of the primary methods of event-time analysis (parametric and Cox's proportional hazards regression analyses) and use them in combination with logistic regression to provide unbiased estimates of delay incurred at a hydroelectric facility, as well as insights on factors affecting both rates of passage and route selection. Passage rate increased with increased depth of a surface bypass sluice gate and, among fish that passed through the turbines, with turbine flow. The data further indicate that risk of turbine passage increased with both delay and turbine flow.

Résumé : Les techniques statistiques couramment utilisées pour étudier le passage des poissons ne réussissent pas à quantifier adéquatement les délais face aux obstacles, ni à évaluer les effets des modifications de ces obstacles sur les taux de passage. Les analyses de données de télémétrie qui décrivent ces effets peuvent être faussées, particulièrement lorsque la voie de passage de certains individus ne peut être déterminée (e.g., à cause de la mortalité, de la perte des étiquettes, du passage par des routes non surveillées ou des routes de rechange, etc.). Nous démontrons comment l'analyse temporelle des événements (« event-time analysis »), mieux connue sous le nom d'analyse de survie, peut servir à quantifier les taux de passage dans toute étude qui permet de suivre des individus dans le temps, même lorsque certains ne suivent pas la route ou ne traversent pas les obstacles en question. Nous examinons deux des principales méthodes de l'analyse temporelle des événements (l'analyse de régression paramétrique et l'analyse de régression aléatoire proportionnelle de Cox) et les utilisons en combinaison avec la régression logistique pour obtenir des estimations non biaisées des délais encourus à un ouvrage hydroélectrique, de même que des informations sur les facteurs qui affectent à la fois les taux de passage et le choix de route. Les taux de passage augmentent en relation avec la profondeur d'une vanne de dérivation de surface et, chez les poissons qui passent par les turbines, en relation avec le débit de la turbine. Nos données montrent, de plus, que le risque associé au passage dans la turbine augmente tant avec le délai devant l'obstacle qu'avec le débit de la turbine.

[Traduit par la Rédaction]

Introduction

A growing body of research shows that delays to the migrations of anadromous fishes can cause dramatic reductions in adult recruitment and spawning success. For many species, the ability of juveniles to osmoregulate in both salt and fresh water (a prerequisite for successful transition to the marine environment) can only be maintained for a brief period (McCormick et al. 1998; Whalen et al. 1999). Similar time-dependent effects have been shown for thermal tolerance (Zydlewski and McCormick 1997a) and predation risk

(Hargreaves 1994; Venditti et al. 2000). Failure of migrants to reach the marine environment within the resulting "smolt window" reduces likelihood of survival. Adult migrants are also vulnerable to delay: freshwater spawning migrations are often powered exclusively by energy stores acquired at sea, and growing evidence suggests there is a trade-off between depletion of these stores and reproductive success (Glebe and Leggett 1981; Leonard and McCormick 1999; Hinch and Bratty 2000).

Concern over the delays incurred at dams and similar barriers has caused the U.S. National Marine Fisheries Service (NMFS) to call for operational changes at hydroelectric facilities to minimize this effect (NMFS 2000). How the effects of these changes should be quantified, however, remains unclear. Migration rate is affected by numerous environmental variables including flow, temperature, photoperiod, and previous experience of the migrants (Zabel and Anderson 1997; McCormick et al. 1998). Although data from various forms of telemetry can provide detailed information on migratory behavior near obstacles, current analytical methods fail to make full use of incomplete data, i.e., data from indi-

Received 8 August 2003. Accepted 9 April 2003. Published on the NRC Research Press Web site at <http://cjfas.nrc.ca> on 25 September 2003.
J17035

T. Castro-Santos¹ and A. Haro. United States Geological Survey, Biological Resources Division, Silvio O. Conte Anadromous Fish Research Center, One Migratory Way, P.O. Box 796, Turners Falls, MA 01376, U.S.A.

¹Corresponding author (email: TCastro_Santos@usgs.gov).

viduals that experience any form of tag failure, suffer predation, pass via undetermined routes, routes other than the route of interest, or fail to pass during the course of the study. Typically, such “censored” individuals are removed from the study and only those that pass are compared; alternatively, some analyses might require an assumption that the fish passed through one route or the other. Either approach introduces bias into the analysis and casts suspicion on the results (e.g., Nettles and Gloss 1987; Wilson et al. 1991; Johnson et al. 2000).

A further deficiency in standard analytical techniques is that they often fail to adequately account for covariates that change over time. Many variables that could potentially influence passage rate may not be constant during the pre-passage period, and depending on delay duration, fish may be exposed to multiple levels of these covariates.

In this paper, we demonstrate how the problems of incomplete data, alternate passage routes, and time-varying covariates can be overcome using statistical methods best known for their applications in biomedical research. Collectively known as “survival analysis” (Cox and Oakes 1984; Lee 1992; Hosmer and Lemeshow 1999), these methods were developed to describe the timing of events, incorporating data from individuals that are removed from studies or whose fate was not determined before ending a study. Although the name implies application to survival studies, we use it here to quantify passage rate with no direct inference on survival. Thus, to avoid confusion with actual survival studies, we will use the synonym “event-time analysis”.

Although these methods have broad application in ecological studies (e.g., Chambers and Leggett 1989) and have been used to estimate survival during the course of migration (Skalski et al. 1993; Lowther and Skalski 1997), they have yet to be applied to behavioral components of fish passage. Here, we present a novel application of event-time analysis and demonstrate its usefulness by quantifying effects of modifications at a hydroelectric facility on rates of passage. Results of a radiotelemetry study of migrating Atlantic salmon (*Salmo salar*) smolts were selected to demonstrate this application.

Methods

Rationale and techniques of event-time analysis

Because the theory and application of event-time analysis are unfamiliar to most fish passage researchers, we present a brief overview of event-time analysis techniques in the following subsections. Interested readers should consult Lee (1992), Allison (1995), and Hosmer and Lemeshow (1999) for more complete details. In each technique, a binary variable, δ , is used to denote whether the individual’s passage time was observed ($\delta = 1$) or not (censored observations; $\delta = 0$). This allows calculation of probability functions without attributing passage routes or times to censored individuals. The only limitation to the use of censored data is that censoring must not be informative, i.e., covariate effects should be the same for censored and uncensored observations.

In addition to censoring, the feature that best distinguishes event-time analysis from other parametric and nonparametric methods is its use of the hazard and survivorship functions ($h(t)$ and $S(t)$, respectively). The hazard function is the in-

stantaneous rate of passage for those individuals that have not yet passed, i.e.,

$$(1) \quad h(t) = \lim_{\Delta t \rightarrow 0} [P\{\text{an individual remaining at time } t \text{ passing in the interval } (t, t + \Delta t)\} / \Delta t]$$

The survivorship function is the complement of the cumulative distribution function and indicates the proportion of individuals remaining at time t .

$$(2) \quad S(t) = P\{\text{an individual passing after time } t\}$$

These relate to the more familiar probability density function (PDF, or $f(t)$)

$$(3) \quad f(t) = \lim_{\Delta t \rightarrow 0} [P\{\text{an individual passing in the interval } (t, t + \Delta t)\} / \Delta t]$$

in that $f(t) = h(t) \times S(t)$ (Lee 1992).

The utility of the hazard and survivorship functions becomes apparent when we consider the effects of censoring. Censored data preclude the usual approach of estimating mean and variance of passage times. The hazard function can be easily estimated, however, by dividing numbers of fish passing in an interval by the number of fish available to pass. Likewise, because censoring indicates that the fish has not yet passed, the last extant observation still contributes to the calculation of the survivorship function (see below).

The most straightforward approach to describing passage times and their associated probability functions is to construct a life table. This is done by breaking time down into meaningful, but not necessarily equal, intervals and calculating estimates of the above probability functions. Lee (1992) provides a clear and detailed description of this procedure. An alternative approach, developed by Kaplan and Meier (1958), is helpful when plotting data. Here, instead of being fixed, intervals are defined by the actual occurrence of events. At each time that a passage event occurs, the value of the survivorship function is estimated based on the cumulative product of the conditional proportion passing:

$$(4) \quad \hat{S}(t) = \prod_{i: t_i \leq t} \left[1 - \left(\frac{p_i}{n_i} \right) \right]$$

where p_i of n_i available individuals pass at each time t_i . Note that the t_i ’s refer only to uncensored observations, but both censored and complete observations are included in the denominator (n_i , or the risk set) as long as they remain available to pass. Life-table methods are best for constructing tables, Kaplan–Meier curves are best for plotting data. Non-parametric tests and predictions can be generated based on either life-table or Kaplan–Meier methods with similar results. However, when multiple covariate effects are present, these tests may be inappropriate, so we avoid them here.

Parametric models for event-time data with censoring

The influence of covariates on passage rate, as well as the shape of probability functions, can often be described by fitting models to the data and testing for fit. Covariate effects can be readily expressed as a linear model. An intuitive form is the accelerated failure time (AFT) model (Allison 1995; Hosmer and Lemeshow 1999):

$$(5) \quad \ln(T) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon$$

where passage time T is a random variable conditional on covariates x_1, \dots, x_k . The disturbance component ε determines the shape of the error distribution and thus the spread of the quantile estimates of passage time. Note that the “failure” term in AFT carries over from the survival analysis literature but refers here to passage events. Using this approach, covariate effects are multiplicative: a given quantile of passage time T changes by a factor of e^{β_j} , or increases by $100(e^{\beta_j} - 1)$ percent per unit increase in the covariate.

Most texts and computer programs fit these models using maximum likelihood (ML) estimation (Lee 1992). The likelihood, L , of a model is described by

$$(6) \quad L = \prod_{i=1}^n [f_i(t_i)]^{\delta_i} [S_i(t_i)]^{1-\delta_i}$$

where $f_i(t_i)$ and $S_i(t_i)$ are, respectively, the estimated PDF and survivorship functions of the fitted distribution for each individual, contingent on model covariates, and $\delta_i = 0$ for censored and 1 for uncensored observations. The likelihood function is maximized with respect to model covariates and parameters, usually by applying some version of the Newton–Raphson algorithm (Lee 1992).

The presence of censored data complicates the evaluation of model fit. Although Allison (1995) and Hosmer and Lemeshow (1999) provide a method for calculating a generalized R^2 statistic, both texts caution against its use, because its value is affected by the proportion of censored observations in the data set; indeed, no statistic can be calculated that quantifies the proportion of variance in a data set that is accounted for by a given model, because the variance of censored observations is not known. Therefore, alternative methods must be used for evaluating model fit. The log-likelihoods of ML-generated models can be used to test for differences in fit between nested distributions and models: a significant likelihood ratio indicates superior fit of the model with the greater likelihood. Akaike’s information criterion (AIC) can also be used to identify the best distribution or model and has the advantage of not requiring nesting (see Allison (1995) for a clear discussion of nested distributions and Burnham and Anderson (1998) for the theory and application of AIC).

A more general method for numerically testing whether the data follow a particular distribution was proposed by Hollander and Proschan (1979), summarized in Lee (1992): a statistic following the standard normal distribution is calculated by comparing predicted and observed values of the survivorship function. In a related approach, Cox–Snell residuals (defined as $e_i = -\ln(S(t_i|x_i))$, where x_i is the vector of covariate values for individual i and $S(t_i)$ is the estimated probability of that individual remaining until time t , based on the fitted model) are plotted against $-\ln(S'(t_i))$, where $S'(t_i)$ is the Kaplan–Meier estimate of the survivorship function. If the model adequately describes the data, then this plot yields a straight line with a slope of unity (Allison 1995).

Cox’s proportional hazards regression

More often than not, knowledge of underlying distributions is limited, or the complexity of the shape of those distributions precludes predictive modeling. Furthermore, the presence of covariates that change over time can complicate modeling efforts and may not even be possible using many standard software packages. The effect of various treatments on the hazard function can still be estimated, however, using a semiparametric, proportional hazards regression approach first described by Cox (1972). This model is based on the premise that the log of the hazard is a linear function of k covariates; the relationship between the hazard functions of two treatment groups i and j is described by

$$(7) \quad \ln(h_i(t) - h_j(t)) = \beta_1(x_{i1} - x_{j1}) + \beta_2(x_{i2} - x_{j2}) + \dots + \beta_k(x_{ik} - x_{jk})$$

The effects of covariates on the ratio of the two hazards are estimated by the coefficients β_1, \dots, β_k . This approach is similar to that described above for the parametric regression model, with the important distinction that where the parametric regression uses ML to model the effects of covariates on the actual time of the event, Cox’s proportional hazards regression uses partial likelihood (L_p) to describe their effects on the rate at which the event occurs. A general expression for the L_p of a proportional hazards model with fixed (i.e., not time-varying) covariates is

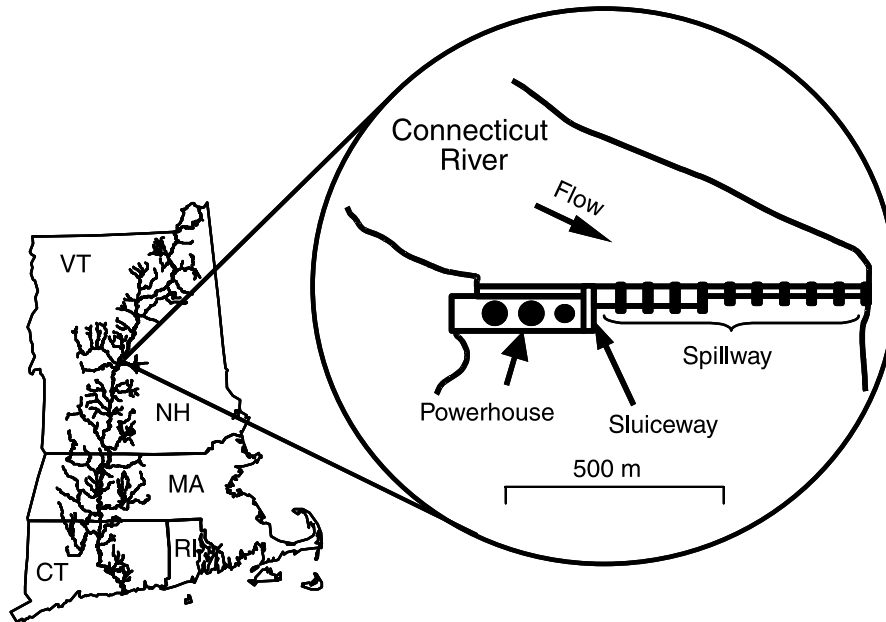
$$(8) \quad L_p = \prod_{i=1}^n \left[\frac{e^{x_i \beta}}{\sum_{j \in R(t_i)} e^{x_j \beta}} \right]^{\delta_i}$$

where $R(t_i)$ constitutes the risk set, all individuals available to pass at time t_i (equivalent to n_i in eq. 4) and $x \beta$ refers to the vector product of the covariates and their coefficients, either for the individual passing (i) or for each member of the risk set (j). As with ML estimation, the L_p is then maximized with respect to β using the Newton–Raphson algorithm. Note that by constructing the denominator in this way and including the censoring indicator δ_i , censored data are included in the analysis and contribute to the denominator until the last extant observation.

Because the L_p is based on the rank of time, rather than its actual value, combined, relative covariate effects on the hazard function can be tested without requiring the underlying probability to follow a particular distribution. In addition to significance, software packages may generate estimates of coefficients in eq. 7; their interpretation is simplified by using an alternate quantity, the hazard ratio, which equals e^{β_j} and indicates the proportional change in hazard per unit change in the covariate.

Another attractive feature of proportional hazards regression is that because it makes no assumptions about the underlying hazard function, inclusion of covariates that change over time is a simple process that is included as a standard feature in many software packages. Hazards are calculated at each event time based on the current risk set, regardless of whether individuals had previously been exposed to a different set of covariate values.

Fig. 1. Plan view of Wilder Station ($42^{\circ}40'N$, $72^{\circ}18'W$) showing location of forebay, powerhouse, and sluiceway. Inset shows location of facility on the Connecticut River in northeastern U.S.A. (VT, Vermont; NH, New Hampshire; MA, Massachusetts; CT, Connecticut; RI, Rhode Island).



Although Cox's proportional hazards regression is not fully parametric, it still requires certain assumptions about the data, primarily that the effects of covariates on hazard are constant over time. Deviation from this assumption can cause misleading results. The proportional hazards assumption can be tested by plotting Schoenfeld residuals (Schoenfeld 1982; Gramsch and Therneau 1994) against the log of time (any significant slope indicates that proportionality is time-dependent). Another residual, called the score residual, is useful for identifying influential and poorly fit observations (see Hosmer and Lemeshow (1999) for a thorough discussion of fit evaluation for proportional hazards models).

Competing risks

The binary approach to censoring applied in the preceding sections is complicated by the availability of multiple passage routes, a common feature of downstream passage studies. Before passage, fish are available to pass through all routes, i.e., they are part of the risk set. Once a route is selected, however, they no longer contribute to the passage rate of any route. In other words, an individual passing via a given route is effectively censored with respect to the other routes. This constitutes a competing risks situation for which event-time analysis methods are particularly well suited (Allison 1995).

When confronted with a competing risks situation, a multi-step approach is appropriate. First, all covariates except passage route are included in the model. This allows inclusion of individuals that were not observed passing through either route (i.e., censored observations) and provides the best estimate of overall passage rate. Next, separate models are developed for individuals passing through each route of interest by modifying the censoring variable. For each model, non-passers, as well as those that pass through alternate routes, are included, censored at time of passage; only those

that pass through the route of interest are noncensored. The advantage of this step is that it evaluates separately variables that affect the rates at which fish pass through each route. Although different results for competing passage routes suggest some underlying difference in covariate effects on passage rate that may be of substantial biological interest, the competing risks approach does not test for these differences explicitly; it is simply a means of quantifying covariate effects on passage rates through a particular route. This means that researchers can use the entire risk set when analyzing passage rates through any route, or combination of routes, and the approach can be applied to any of the methods detailed above, a potent tool for evaluating the effects of facility operations on passage rate through a particular route. The only assumption required is that fish that have not yet passed the dam are equally available to pass via all routes and should be included when evaluating the effects of covariates on the groups.

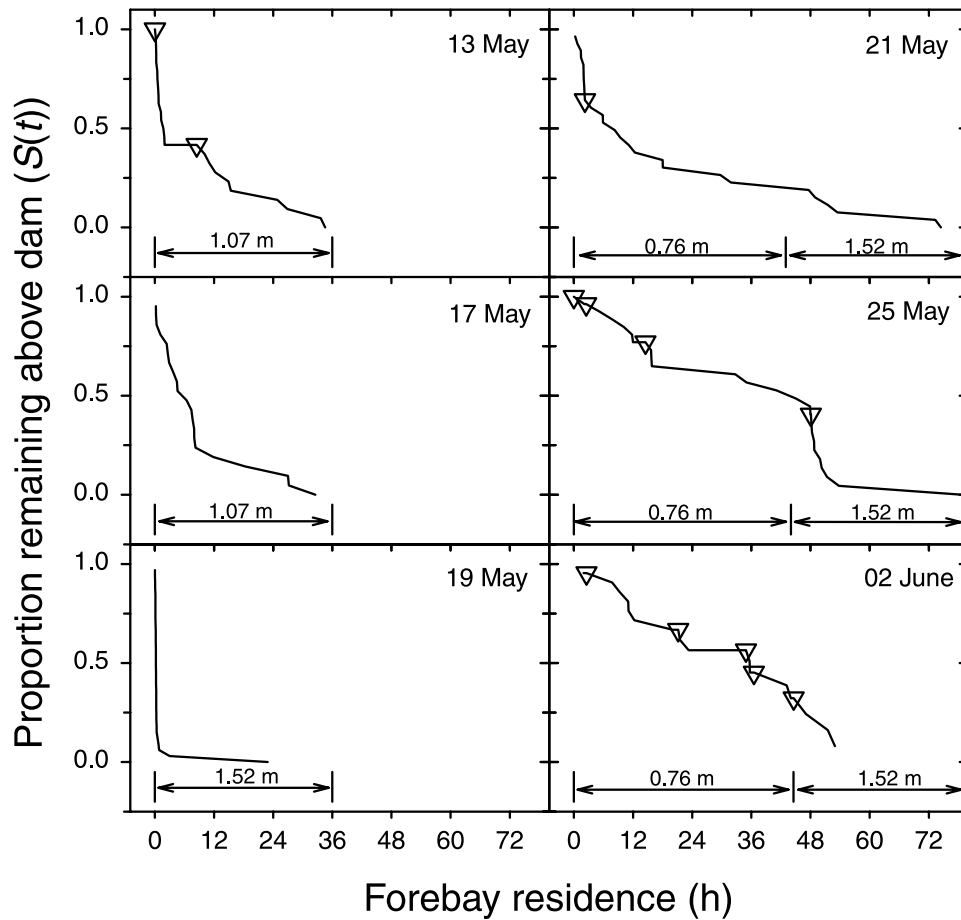
Logistic regression

Although the competing risks approach allows separate analysis of rates of passage through various routes, it does not directly quantify which variables most influence the likelihood of selecting one route over another. Logistic regression is a standard method for quantifying covariate effects on the likelihood of selecting one of two or more categorical variables (Hosmer and Lemeshow 1989). By including time as a covariate, the effect of delay on passage route selection can be tested directly. Censored individuals, because they are not observed to actually pass, are generally not included in such analyses (Allison 1995).

Data set

We selected a sample data set (RMC Environmental Services, Inc., currently Normandeau Associates, 917 Route 12,

Fig. 2. Kaplan–Meier curves describing forebay residence times of radio-tagged Atlantic salmon (*Salmo salar*) smolts passing Wilder Station during each of six releases. Open triangles indicate censored individuals. Gate depth is indicated in each panel, as well as shifts from 0.76 m to 1.52 m during the last three releases. Note the rapid passage on the 19 May release and increased passage rate following increased gate depth during the last three releases.



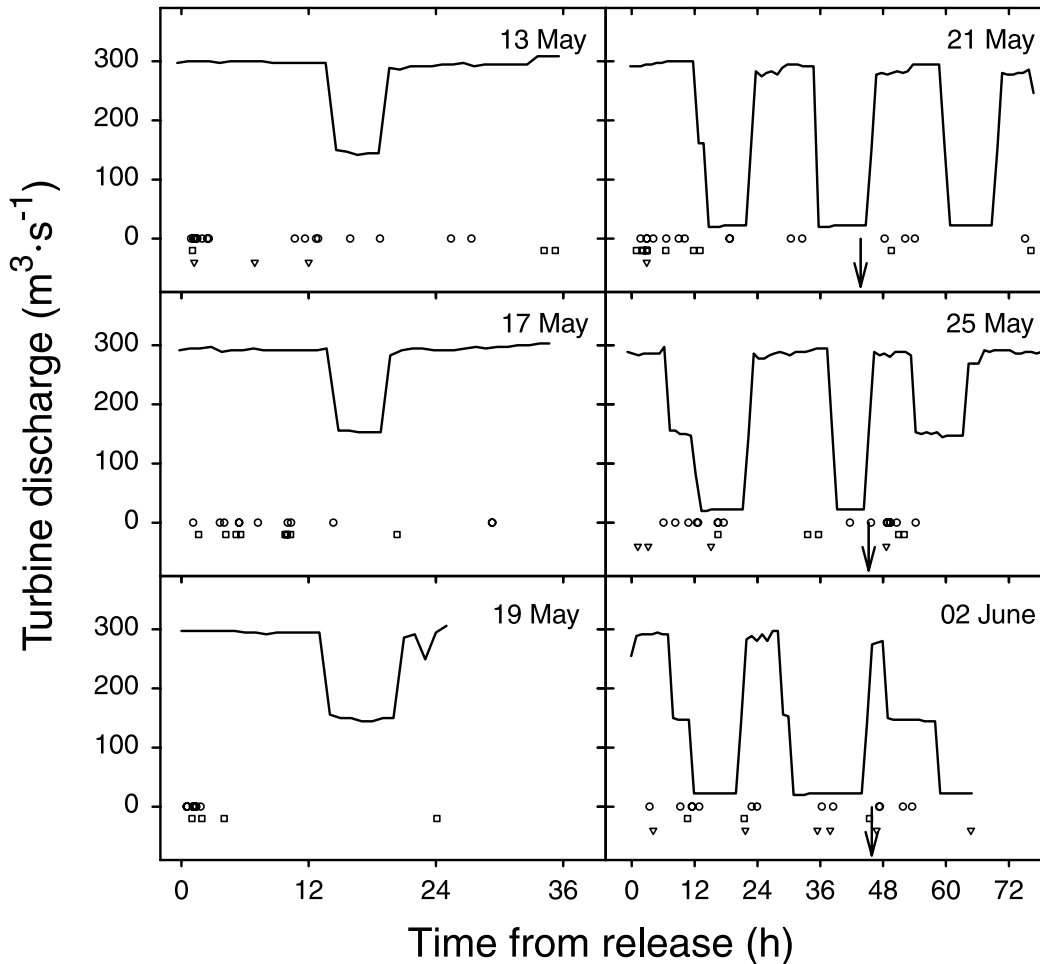
Suite 1, Westmoreland, NH 03467, U.S.A., unpublished data) to demonstrate the utility of each of the event-time analysis methods for describing the effects of operational modifications and other variables on delay. Here we define delay as the time elapsed between forebay entry and passage, i.e., forebay residence time. The study was conducted in 1994 at Wilder Station, a hydroelectric facility on the Connecticut River mainstem at river-kilometre 348, and was designed to test the effect of different depth settings of a bypass sluice gate on passage route selection. The ice-log sluice used for downstream smolt passage is located adjacent to the powerhouse (Fig. 1). A 3.0 m × 4.6 m skimmer gate regulates surface flows of 1.38 m³·s⁻¹ at 0.305-m depth to 18.7 m³·s⁻¹ at 1.83-m depth. When operated for smolt passage, the gate is normally set to 1.07-m depth, passing 8.77 m³·s⁻¹ down an 18.3-m-long sluice into the station tailrace. The powerhouse contains two 19-MW Kaplan turbines and one 3.2-MW generator, protected by trashracks with 15 cm horizontal and 48 cm vertical spacing. No spill occurred during this study, so all flow passed through the turbines or over the bypass sluice. These were the only two passage routes available to the smolts during this period.

Atlantic salmon smolts were obtained from two sources: a bypass sampler on the Connecticut River mainstem (“wild” fish; $n = 65$, fork length (FL) = 137–235 mm, $\bar{FL} = 179$ mm),

and the White River National Fish Hatchery in Bethel, Vermont (“hatchery” fish; $n = 93$, FL = 152–218 mm, $\bar{FL} = 190$ mm). Source of fish is referred to as “origin” and coded 0 and 1, respectively, for wild and hatchery smolts in analyses. Smolts were anaesthetized using MS-222 and radio-tagged using esophageal implants. Following a 24-h recovery period, fish were released 1 km upstream of Wilder Station. This was considered sufficient distance to prevent any predisposition on the part of the smolts to pass through one route over the other. Smolts were released in six groups of 21–33 individuals on 13, 17, 19, 21, and 25 May and 2 June 1994 (Fig. 2). Wild fish made up 48% (33–58%) of the first five releases; the last release consisted of hatchery fish only. Telemetry receivers were placed in such a way that smolts were detected when they entered the forebay of the project (which extended about 200 m upstream of the dam) and were monitored continuously during their forebay residence. A four-element YAGI antenna situated halfway down the sluice identified sluice passers, and antennas submerged at the entrance of each intake identified turbine passers. Time to passage (delay or residence time) was calculated from the time that fish first entered the forebay to ensure that only data from actively migrating fish were used.

Turbine flow was logged each hour and ranged from 19.8 to 308 m³·s⁻¹, $\bar{X} = 253$ m³·s⁻¹. Because fish that did not im-

Fig. 3. Turbine flow at Wilder Station during each of six releases. Data are presented as hours from each release. Points indicate passage time for sluice passers (circles) and turbine passers (squares); censored times are indicated by triangles. Arrows indicate time to increased gate depth in each of the last three releases.



mediately pass the station were often subjected to more than one level of turbine flow (Fig. 3), this was not included in the parametric models but was included as a time-dependent covariate in the proportional hazards models.

A similar complication arose with respect to sluice gate depth. This was set to 1.07 m for the first two releases, 1.52 m for the third release, and 0.76 m for the last three releases. For each of the last three releases, the sluice gate depth was increased to 1.52 m after 44–46 h (Fig. 2). To prevent the artificial association of greater gate depth with long passage delays, we censored data from these releases at the corresponding residence times for the parametric regression analysis but included the data with gate depth as a time-dependent covariate in the proportional hazards analysis.

Because fish could only pass through the turbines or over the sluice, passage route constituted a competing risks variable in this study. Separate models were generated for each route as well as for the combined data using fully parametric and Cox’s proportional hazards techniques. To test for differences between routes in covariate effects on passage rate, we ran the above tests, including route and its interactions with the other covariates.

We used SAS software (SAS 1999) to estimate covariate effects on passage time. We selected from among exponen-

tial, Weibull, lognormal, and generalized gamma distributions, testing for the best fit using likelihood ratio statistics and AIC. Where nested models were not significantly different from each other, we selected the most parsimonious model, i.e., the one with the fewest parameters. Adequacy of the parametric models was evaluated both numerically and graphically, using each of the methods detailed above. We evaluated adequacy of proportional hazards models using Schoenfeld and score residuals.

In addition to the above tests, we used logistic regression to test for covariate effects on likelihood of passing through the bypass sluice, including log of delay time as a covariate. All analyses were conducted using SAS software (SAS 1999).

Results

Of the 158 Atlantic salmon smolts used in this study, 14 (eight hatchery and six wild) had undetermined passage routes or failed to pass; these were included in the analyses, censored at their last extant observation. In all, 144 smolts passed the station by known routes: 106 over the bypass sluice and 38 through the turbines. Most fish entered the forebay shortly after release; mean ± standard deviation (SD) of postrelease delay was 1.27 ± 1.70 h, with all but three in-

Table 1. Results from parametric and Cox's proportional hazards regression.

Data set	Variables	Parametric			Proportional hazards		
		<i>N</i>	$\hat{\beta}$	<i>P</i> value	<i>N</i>	$\hat{\beta}$	<i>P</i> value
Combined data	Number passed	121			144		
	Number censored	37			14		
	Intercept		-787.190	0.018	—	—	
	Origin		-0.007	0.981	0.042	0.810	
	Release date		0.063	0.017	-0.019	0.323	
	Turbine flow (m ³ ·s ⁻¹)		—	—	0.001	0.611	
	Gate depth (m)		-5.459	<0.001	3.528	<0.001	
	Scale		1.589		—	—	
Sluice passers	Number passed	89			106		
	Number censored	69			52		
	Intercept		-499.125	0.183	—	—	
	Origin		0.163	0.609	-0.055	0.787	
	Release date		0.040	0.175	-0.025	0.289	
	Turbine flow (m ³ ·s ⁻¹)		—	—	-0.003	0.210	
	Gate depth (m)		-6.316	<0.001	3.972	<0.001	
	Scale		1.695		—	—	
Turbine passers	Number passed	32			38		
	Number censored	126			120		
	Intercept		-1460.460	0.034	—	—	
	Origin		-0.462	0.406	0.316	0.378	
	Release date		0.117	0.033	-0.029	0.433	
	Turbine flow (m ³ ·s ⁻¹)		—	—	0.009	0.024	
	Gate depth (m)		-2.562	0.039	1.820	0.068	
	Scale		2.185		—	—	

Note: The competing risks approach was applied to each passage route, where censored individuals include those passing through the alternate route. Parametric models are based on the lognormal distribution: coefficients ($\hat{\beta}$) indicate effect of each variable on the log of delay ($\ln(T)$); scale refers to the error term. Coefficients for the proportional hazards models indicate their effect on the log of the hazard ($\ln(h(t))$). Origin is coded 1 (hatchery) and 0 (wild).

individuals entering within 4 h. Kaplan–Meier survivorship curves of residence time ($S'(t)$; Kaplan and Meier 1958) are presented for each release group (Fig. 2).

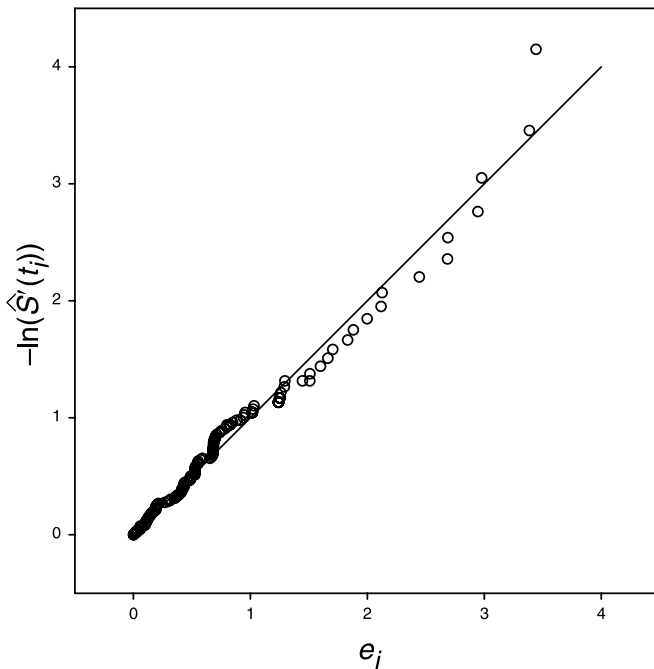
Results from both parametric and proportional hazards models (Table 1) should be interpreted with some caution, as there was significant collinearity among covariates. This effect was greatest between release date and both gate depth and turbine flow. Correlation coefficients were in all cases less than 0.45, however, and the effect of collinearity on the models should be small.

Among the parametric models, the generalized gamma distribution provided a better fit to the combined data than did the Weibull or exponential distributions (χ^2 ; 1 and 2 df, respectively; $P < 0.001$). However, the lognormal and gamma distributions provided nearly identical fits (χ^2 ; 1 df; $P = 0.81$). Based on these results, combined with the AIC values, the Hollander and Proschan test ($P = 0.20$), and analysis of Cox–Snell residuals (Fig. 4), we concluded that the lognormal distribution was the most appropriate and parsimonious of the distributions tested and that it adequately described the data. Under this parameterization, the scale variable is analogous to the error term under the standard normal distribution; location is estimated by $x\hat{\beta}$. The estimated scale values of 1.6–2.2 indicate that passage rate (i.e., hazard) follows an inverted U shape: initially low, it rapidly increases to a maximum value and then declines gradually over time (Meeker and Escobar 1998).

Coefficients of the parametric models indicate covariate effects on $\ln(T)$. These describe reduced delay with increased gate depth for combined passage data and for both passage routes, as indicated by a significant negative coefficient. The interpretation of this for the combined data, adjusting to centimetres, is $T = 100(e^{-0.055} - 1)$, or mean delay decreases by 5.4% for every centimetre of increased gate depth. The same transformation for sluice and turbine passers shows a 6.1% and 2.6% decrease in delay time per centimetre increase of gate depth, respectively. For the current data set, this implies that by increasing gate depth from the shallowest to deepest settings and setting all other covariates to their mean values, median delay declines from 19.9 to 0.3 h when both passage routes are available, and delay of the 90th percentile declines from 152.8 to 2.4 h (Table 2). The parametric approach also suggests that fish released later in the study passed more slowly than earlier releases (positive β), regardless of passage route.

Results of proportional hazards regression indicate that gate depth affected passage rate, particularly among sluice passers (Table 1). Faster passage rate at greater depths is indicated by a significant positive coefficient (greater hazard; note the contrast with the parametric approach). Adjusting to centimetres and transforming the data to risk ratios, we find here a 3.6% increase in passage rate associated with each centimetre of gate depth for combined data and a 4.0% increase for sluice passers.

Fig. 4. Graphical assessment for goodness-of-fit of the lognormal parametric regression model using the total data set (sluice and turbine passers combined). Cox–Snell residuals (e_i) are plotted against the negative log survivorship function ($-\ln(S'(t))$), calculated using the Kaplan–Meier method (circles). A line with slope of unity, an indicator that the model provides a reasonable fit, has been included for reference.



Turbine flow also affected passage rate, but only among fish that passed through the turbines. Here, the positive coefficient indicates a hazard ratio of 1.009. Over the range of turbine flows encountered in this study, this means that the rate of fish passage through the turbines was 2.6 times greater under the highest flow than under the lowest flow.

Residual analysis confirmed that the assumption of proportionality was met in these models. However, analysis of the score residuals suggested that the last fish to pass from the 19 May release was influencing the results, particularly with respect to the coefficients for gate depth. When this individual was removed from the analysis, the coefficients for total data and sluice passers increased to 4.7% and 5.3% per centimetre increase in gate depth, respectively. The only other notable effect of removing this data point was to increase the P value for the coefficient describing the gate depth effect on turbine passers to 0.121, casting further doubt on its importance.

Significance tests for different effects of gate depth and turbine flow on rates of sluice and turbine passage suggested that covariate effects did differ between the two routes, although neither the parametric nor the proportional hazards approach found a significant difference in passage rate by route (main effect), the interaction of route \times gate depth was strongly significant for both models ($P < 0.001$). In contrast, the interaction of route \times turbine flow was only marginally significant ($P = 0.063$; proportional hazards model only). It is important to emphasize here that these tests, because they include passage route, exclude the censored observations and should therefore be considered biased approximations.

Table 2. Predicted delay times from the parametric regression model (Table 1) for median and first and last deciles of Atlantic salmon smolts passing a hydroelectric facility by either route (“combined data”) under three different sluice gate settings.

Gate depth	Percent passed	Predicted delay (h)	95% bounds	
			Lower	Upper
0.76	10	2.60	1.83	3.71
	50	19.94	14.56	27.32
	90	152.78	101.16	230.77
1.07	10	0.49	0.37	0.66
	50	3.78	3.02	4.73
	90	28.93	20.67	40.50
1.52	10	0.04	0.02	0.07
	50	0.31	0.19	0.50
	90	2.38	1.41	4.03

Table 3. Logistic regression results describing the effect of covariates on passage route selection.

Source	$\hat{\beta}$	P value
Intercept	-4.0930	0.9936
Origin	-0.3732	0.3931
Release date	0.0004	0.9904
Turbine flow ($m^3 \cdot s^{-1}$)	-0.0077	0.0158
Gate depth (m)	0.4549	0.0339
ln(hours)	-0.2651	0.0234
N	144	

Note: Positive coefficients ($\hat{\beta}$) mean increased probability of sluice passage; origin was denoted as 1 (hatchery) and 0 (wild); gate depth and turbine flow indicate settings at time of passage. Fish that failed to pass were excluded from this analysis; sample size (N) refers to passers only.

The logistic model (Table 3) demonstrates that the probability of turbine passage increases with delay (note that data from censored individuals are omitted). Probability of turbine passage was likewise increased by greater turbine flow but was reduced by greater sluice gate depth.

Discussion

Although the results of this study inevitably have management implications, readers should bear in mind that our objective is to introduce a new technique and demonstrate its use. RMC (1994) correctly concluded that the gate depth setting of 1.52 m afforded more expedient downstream passage to emigrating smolts and the sluice is currently operated at this depth for fish passage. Therefore, the intent of this paper is not to call into question existing management decisions, but to present techniques that will lead to further and more complete investigations of bypass configurations and other passage evaluations. Moreover, it is important to recognize that although we define delay here as forebay residence time, the term implies a change relative to some minimum transit time. This value is unknown, but its identification and characterization should be an objective of future studies.

Our results do highlight some important differences among the event-time analysis techniques, as well as their relative strengths. Survivorship curves are simple to construct, and their significance is readily interpreted. Here, they show that although the first half of the fish pass fairly quickly, passage rate declines with time and the slowest fish take several days to pass the project. Reporting only median passage time masks this important feature (Venditti et al. 2000), a fact that should be of some concern to restoration efforts. The survivorship curves also provide graphical representation of the effects of sluice gate depth setting. Delay times were least for the 19 May release (gate depth of 1.52 m) and greatest for the last three releases (gate depth of 0.76 m). Those fish that delayed passage until after the gate was set to a greater depth show a correspondingly greater rate of passage at that time.

More censoring was observed for later release dates and shallower gate settings. However, there is no reason to believe that censored and complete observations differed with respect to their response to covariate values. With the exception of the competing risks models, the exact cause of censoring was not determined and may have been due to mortality, tag failure or expulsion, or undetected passage.

Although nonparametric methods for analyzing data with censored observations exist, adequate sample size and the presence of multiple, nonorthogonal covariates indicated the use of a regression approach in this analysis. Both the parametric and Cox's proportional hazard models simultaneously account for all covariate effects (Type III hypotheses; SAS 1999), but they differ in important ways. First, because the parametric models describe covariate effects on forebay residence time, we are able to use them to estimate time to passage of specific proportions of the population at defined covariate levels, a useful tool for managers and those interested in understanding population-level implications of delay. The proportional hazards models, by contrast, describe covariate effects on passage rate (hence the reversed sign of the regression coefficients: passage rate is inversely related to residence time) and cannot be used to directly estimate delay time (but see Hosmer and Lemeshow (1999) for methods by which indirect estimates can be extrapolated).

The parametric models can also be used to draw important inferences on the shape of the hazard and consequently the distribution of quantiles. Here, the numerical and graphical goodness-of-fit analyses both suggest that the lognormal models provide a reasonable fit to the data. This implies that the log delay times are approximately normally distributed (Allison 1995). The inverted U-shaped hazard function described by these models may have some biological significance, as it implies both initial delay (as might be expected from time required to locate the passage route) and reduced passage rates for fish that do not subsequently pass quickly. Thus, the risks of delay include increased likelihood of further delay, a pattern that one would predict if delay resulted in loss of migratory motivation (Meeker and Escobar 1998).

Although the Cox's proportional hazards models do not yield parametric descriptions of the hazard, their independence from specific distributions make them robust against fluctuating passage rate, e.g., resulting from diel migratory patterns, requiring only that covariate effects on hazard ratios be consistent across these patterns. With either approach,

we advise against extrapolating far beyond observed values of covariates or delay times.

A further distinction is that proportional hazards models allow ready computation of effects of covariates that change over time, whereas the parametric models assume that each fish is exposed to fixed covariate conditions. This feature has important implications for the interpretation of the effects of release date and gate depth, as well as turbine flow. Because the shallowest gate settings were applied at the end of the experiment, the experimental design was unbalanced, and variability caused by this factor is wrongly apportioned between gate depth and release date under the fully parametric model. By including gate depth as a time-dependent covariate, potentially confounding effects between the two variables are reduced. Under this less biased interpretation of the data, delay is appropriately attributed to gate depth and not to date.

Similarly, fluctuations in turbine flow precluded its inclusion in parametric models — any quantity used (e.g., initial, mean, or final turbine flow) would yield misleading results, because it would not reflect the range of flows to which individual fish were exposed with their associated passage rates. As with the sluice gate settings, minimum daily discharge was less for later releases, and the significant date factor in the parametric models may have arisen in part from the covariance between date and discharge. When turbine flow was included in the proportional hazards models as a time-dependent covariate, it was found to significantly increase the rate of turbine passage, but not sluice passage, and was nonsignificant for the total data.

Both turbine passage models, with their weak effects of turbine flow and gate depth, also illustrate one of the limitations of event-time analysis: because only 38 fish passed through the turbines, over 75% of the observations are censored. This increases the standard error of the estimates, and the power of these models decreases with increased censoring. Thus, the weak significance values for gate depth should be viewed with caution: it may be that increased gate depth does increase passage rate for fish that pass through the turbines (as indicated by the parametric model), but the proportional hazards model lacked sufficient power to detect this effect.

The competing risks condition of this study illustrates the utility of event-time analysis techniques in analyzing passage data. The experimental design reasonably allows one to ask the question: does gate depth affect the passage rate through the bypass sluice? Including turbine passers among the censored data provides the clearest, least-biased answer to this question. Conversely, by censoring sluice passers, we were able to detect the effect of turbine flow on rate of passage through the turbines, which was not apparent from the analysis of the total data. This approach, however, does not directly test for the significance of differences in passage time between the two routes. Indeed, in the presence of censored data, there is no available way to conduct such a test objectively. Here, we constructed a model on the non-censored data that included passage route and the interactions of route \times gate and route \times turbine flow. The results showed strong significance for the former test but marginal significance for the latter. Because data from censored fish are omitted, these results are of necessity biased; at best, the

resulting models tend to underestimate time to passage (eq. 6). They do, however, describe the data from those fish with observed passage times and routes and further support the results of the competing risks analysis.

Whether the increased passage rate with increased gate depth resulted directly from the associated increase in flow or whether it reflects behavioral avoidance of the surface is unclear from this study. Surprisingly, greater gate depth was also associated with increased turbine passage rate. This may be because increased flow over the sluice attracted fish to the powerhouse area or otherwise altered forebay hydraulics such that fish passed more quickly through both routes. The effect of turbine flow supports the view that hydraulics affect passage rate, but the results for sluice passers remain ambiguous. Future studies should attempt to simultaneously control for volume and depth of bypass flows, as well as the ratio of bypass flow to turbine flow, to improve orthogonality of these factors.

Each of the techniques described above has specific advantages and can provide unique information on passage rate. The most appropriate approach will depend on time resolution (Cox's proportional hazards regression can incorporate ordinal time data, whereas parametric regression requires a continuous time variable), shape of the hazard function, and research objectives. None of the above techniques, however, quantifies covariate effects on route selection as such. Logistic regression does just this and thereby complements event-time analysis. Bearing in mind that data from 14 individuals are missing, logistic regression reveals a significant time effect, with greater delays associated with increased risk of turbine passage. This result alone can be a powerful argument for trying to maximize passage rates. Also significant in the logistic model are effects of turbine flow and gate depth, with greater flow and shallower gate settings associated with increased risk of turbine passage.

Combining the logistic and event-time approach, we conclude that shallower gate depth not only increased delay, but also simultaneously (in part, because of the delay) increased the likelihood of turbine passage, particularly in the presence of high turbine flows. These results illustrate the complementary nature of event-time and logistic regression approaches: shallower sluice gate settings reduced passage rates, especially through the sluice. This, in turn, increased the time during which fish were exposed to the possibility of passing through both routes, thereby increasing the likelihood of turbine passage. Thus, by modifying operations to maximize the rate at which fish pass over the sluice, both delay and likelihood of turbine passage could be minimized simultaneously.

The bulk of current fish passage research work focuses on proportions of fish passing through various routes, primarily because this is thought to have the greatest relevance to survival and recruitment (Burnham et al. 1987; Skalski 1998; Skalski et al. 1998). Although the importance of delays to migration is not well understood, it is bound to vary by species, river system, and life history (McCormick et al. 1996; Zydlewski and McCormick 1997b; Zabel et al. 1998). Our understanding of the effect of delay is limited at the outset by our ability to quantify it. Event-time analysis provides a powerful set of tools for developing just such descriptions, as well as for evaluating effects of structural and operational

modifications on passage rates. Because they afford continuous monitoring of individuals, radiotelemetry and acoustic telemetry are particularly well suited to these analyses. Other forms of telemetry and monitoring (e.g., from passive integrated transponder (PIT) tags) may also be useful; however, it is important that time to passage or censoring is known. Because PIT tags tend to have relatively short read ranges (Prentice et al. 1990; Castro-Santos et al. 1996), it may not always be possible to identify censoring times using this technology, although the competing risks approach could still be applied to some data.

Although this paper focuses on the application of event-time analysis to a radiotelemetry study of downstream fish passage, the techniques have much broader potential. Analogous applications include quantifying attraction of upstream migrants to fishway entrances, monitoring progress up fishways (where height can be substituted for time as the dependent variable and successfully exiting the top of the fishway constitutes censoring), and quantifying timing of movements of other migratory species: in short, any application may be appropriate where censoring and competing risks confound the use of standard techniques.

Acknowledgments

The following individuals provided valuable assistance in the preparation of this report: James F. Jekel (Yale University, School of Public Health) provided early insight into the relevance of event-time analysis techniques to fish passage applications and Michael Sutherland (University of Massachusetts Statistical Consulting Services) and Paul Allison (University of Pennsylvania Sociology Department) provided guidance and instruction on event-time analysis techniques. We are particularly grateful to Timothy Brush and Brian Hanson (Normandean Associates, Westmoreland, N.H., formerly RMC Environmental Services) for providing the data set (courtesy of New England Power Company, now PG&E National Energy Group, Concord, N.H.), as well as extensive support and comments on the manuscript. David Hosmer and two anonymous reviewers provided careful reviews and many helpful recommendations for the final draft.

References

- Allison, P.D. 1995. Survival analysis using the SAS system: a practical guide. SAS Institute Inc., Cary, N.C.
- Burnham, K.P., and Anderson, D.R. 1998. Model selection and inference. A practical information-theoretic approach. Springer-Verlag New York Inc., New York.
- Burnham, K.P., Anderson, D.R., White, G.C., Brownie, C., and Pollock, K.H. 1987. Design and analysis methods for fish survival experiments based on release-recapture. American Fisheries Society, Bethesda, Md.
- Castro-Santos, T., Haro, A., and Walk, S. 1996. A passive integrated transponder (PIT) tagging system for monitoring fishways. *Fish. Res.* **28**: 253–261.
- Chambers, R.C., and Leggett, W.C. 1989. Event analysis applied to timing in marine fish ontogeny. *Can. J. Fish. Aquat. Sci.* **46**: 1633–1641.
- Cox, D.R. 1972. Regression models and life tables. *J. R. Statist. Soc.* **34**: 187–220.
- Cox, D.R., and Oakes, D. 1984. Analysis of survival data. Chapman and Hall, New York.

- Glebe, B.D., and Leggett, W.C. 1981. Latitudinal differences in energy allocation and use during the freshwater migrations of American shad (*Alosa sapidissima*) and their life history consequences. *Can. J. Fish. Aquat. Sci.* **38**: 806–820.
- Grampsch, P.M., and Therneau, T.M. 1994. Proportional hazards tests in diagnostics based on weighted residuals. *Biometrika*, **81**: 515–526.
- Hargreaves, N.B. 1994. Processes controlling behaviour and mortality of salmonids during the early sea life period in the ocean. *Nord. J. Freshw. Res.* **69**: 97.
- Hinch, S.G., and Bratty, J. 2000. Effects of swim speed and activity pattern on success of adult sockeye salmon migration through an area of difficult passage. *Trans. Am. Fish. Soc.* **129**: 598–606.
- Hollander, M., and Proschan, F. 1979. Testing to determine the underlying distribution using randomly censored data. *Biometrics*, **35**: 393–401.
- Hosmer, D.W., and Lemeshow, S. 1989. Applied logistic regression. John Wiley and Sons, Inc., New York.
- Hosmer, D.W., and Lemeshow, S. 1999. Applied survival analysis. John Wiley and Sons, Inc., New York.
- Johnson, G.E., Adams, N.S., Johnson, R.L., Rondorf, D.W., Dauble, D.D., and Barila, T.Y. 2000. Evaluation of the prototype surface bypass for salmonid smolts in spring 1996 and 1997 at Lower Granite Dam on the Snake River, Washington. *Trans. Am. Fish. Soc.* **129**: 381–397.
- Kaplan, E.L., and Meier, P. 1958. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**: 457–481.
- Lee, E.T. 1992. Statistical methods for survival data analysis. John Wiley & Sons, New York.
- Leonard, J.B.K., and McCormick, S.D. 1999. Effects of migration distance on whole-body and tissue-specific energy use in American shad (*Alosa sapidissima*). *Can. J. Fish. Aquat. Sci.* **56**: 1159–1171.
- Lowther, A.B., and Skalski, J. 1997. The design and analysis of salmonid tagging studies in the Columbia Basin. Vol. VII. A new model for estimating survival probabilities and residualization from a release–recapture study of fall chinook salmon (*Oncorhynchus tshawytscha*) smolts in the Snake River. U.S. Department of Energy, Bonneville Power Administration, Portland, Ore.
- McCormick, S.D., Shrimpton, J.M., and Zydlewski, J. 1996. Temperature effects on osmoregulatory physiology of juvenile anadromous fish. In *Global warming: implications for freshwater and marine fish*. Edited by C.M. Wood and D.G. McDonald. Cambridge University Press, Cambridge. pp. 279–301.
- McCormick, S.D., Hansen, L.P., Quinn, T.P., and Saunders, R.L. 1998. Movement, migration, and smolting of Atlantic salmon (*Salmo salar*). *Can. J. Fish. Aquat. Sci.* **55**: 77–92.
- Meeker, W.Q., and Escobar, L.A. 1998. Statistical methods for reliability data. John Wiley & Sons, New York.
- Nettles, D.C., and Gloss, S.P. 1987. Migration of landlocked Atlantic salmon smolts and effectiveness of a fish bypass structure at a small-scale hydroelectric facility. *N. Am. J. Fish. Manag.* **7**: 562–568.
- National Marine Fisheries Service (NMFS). 2000. Draft biological opinion: operation of the Federal Columbia River Power System including the Juvenile Fish Transportation Program and the Bureau of Reclamation's 31 projects, including the Entire Columbia Basin Project. National Marine Fisheries Service, Portland, Ore.
- Prentice, E.F., Flagg, T.A., and McCutcheon, S. 1990. Feasibility of using implantable passive integrated transponder (PIT) tags in salmonids. In *Fish marking techniques*. Edited by N.C. Parker, A.E. Giorgi, R.C. Heidinger, D.B. Jester, Jr., E.D. Prince, and G.A. Winans. American Fisheries Society, Bethesda, Md. pp. 317–322.
- RMC Environmental Services. 1994. Movement and behavior of radio-tagged Atlantic salmon smolts at Wilder Hydroelectric Station, Spring, 1994. Report prepared for New England Power Company, Westborough, Mass.
- SAS. 1999. SAS. SAS Institute Inc., Cary, N.C.
- Shoenfeld, D. 1982. Partial residuals for the proportional hazards regression model. *Biometrika*, **69**: 239–241.
- Skalski, J.R. 1998. Estimating season-wide survival rates of out-migrating salmon smolts in the Snake River, Washington. *Can. J. Fish. Aquat. Sci.* **55**: 761–769.
- Skalski, J.R., Hoffmann, A., and Smith, S.G. 1993. Development of survival relationships using concomitant variables measured from individual smolt implanted with PIT-tags. U.S. Department of Energy, Bonneville Power Administration, Portland, Ore.
- Skalski, J.R., Smith, S.G., Iwamoto, R.N., Williams, J.G., and Hoffmann, A. 1998. Use of passive integrated transponder tags to estimate survival of migrant juvenile salmonids in the Snake and Columbia rivers. *Can. J. Fish. Aquat. Sci.* **55**: 1484–1493.
- Venditti, D.A., Rondorf, D.W., and Kraut, J.M. 2000. Migratory behavior and forebay delay of radio-tagged juvenile fall chinook salmon in a lower Snake River impoundment. *N. Am. J. Fish. Manag.* **20**: 41–52.
- Whalen, K.G., Parrish, D.L., and McCormick, S.D. 1999. Migration timing of Atlantic salmon smolts relative to environmental and physiological factors. *Trans. Am. Fish. Soc.* **128**: 289–301.
- Wilson, J.W., Giorgi, A.E., and Stuehrenberg, L.C. 1991. A method for estimating spill effectiveness for passing juvenile salmon and its application at Lower Granite Dam on the Snake River. *Can. J. Fish. Aquat. Sci.* **48**: 1872–1876.
- Zabel, R.W., and Anderson, J.J. 1997. A model of the travel time of migrating juvenile salmon, with an application to Snake River spring chinook salmon. *N. Am. J. Fish. Manag.* **17**: 93–100.
- Zabel, R.W., Anderson, J.J., and Shaw, P.A. 1998. A multiple-reach model describing the migratory behavior of Snake River yearling chinook salmon (*Oncorhynchus tshawytscha*). *Can. J. Fish. Aquat. Sci.* **55**: 658–667.
- Zydlewski, J., and McCormick, S.D. 1997a. The ontogeny of salinity tolerance in the American shad, *Alosa sapidissima*. *Can. J. Fish. Aquat. Sci.* **54**: 182–189.
- Zydlewski, J., and McCormick, S.D. 1997b. The loss of hyperosmoregulatory ability in migrating juvenile American shad *Alosa sapidissima*. *Can. J. Fish. Aquat. Sci.* **54**: 2377–2387.