

Survival heterogeneity review

Gordon A. Fox and Bruce E. Kendall

January 27, 2026

This is a working draft. Hell, it's not even that yet ...

1 Initial questions

- Why models?
- Why covariates?
- But there are many possible covariates, most of which can't be measured in practice.
- What if you ignore some, or just don't have enough data? This is one reason for using frailties
- What are some examples of usefully analyzed data? And what do these show?

2 Introduction

Individuals vary in their survival probabilities. Not only because of variation in age, size, or stage, but because of their histories of nutrition, parasites, or disease, their varied access to resources or exposure to toxicants, competition, and their varied positions in social hierarchies. There may also be genetic variation that contributes to varied survival propensities. Consequently, there is a rich store of biological questions to be revealed through the study of survival. But it also presents some knotty conceptual and technical issues. We introduce these problems first, and then review results from published empirical studies.

3 Estimating survival and its problems

A key problem is that the survival probability of an individual is not measurable. We can measure its death date, but an early death may be due to bad luck, not an inherently poor survival probability. We can only estimate survival probabilities for aggregates of individuals. Doing so requires modeling, based on biological insight.

An additional problem is that survival processes generally do not lead to symmetric distributions of survival time. Intuition on this point can be had by considering the three types of survival curves in Deevey (1947). If there are many late or early deaths (types I or III), obviously the survival times are not distributed normally. If mortality is constant (type II), survival times are distributed exponentially. Add to this the facts that survival processes have an abrupt beginning, such as birth or the start of an experiment, and that most data sets involve some censorship (time of death known only as an inequality, e.g., because it was after the end of the study, between two observations, or caused by the researcher. As a consequence, the population mean (or marginal) survival is not generally a good descriptor for the entire population. The mean or marginal survival estimate may not describe survival for any individual. If there is substantial heterogeneity, these estimates may be far off, and the analysis may miss some important biology.

Given this, how can we use models to estimate survival? There are many useful approaches. The most widely used are proportional hazard models (including the Cox proportional hazards model and parametric hazard models), accelerated failure time models (AFT), and mark-recapture estimates. There are many forms of each. They have their own pros and cons in any application, and these are discussed elsewhere (XX, bunch of cites). Here we focus on issues (and pros and cons) related to heterogeneity **rewrite**:

- **Identifiability.** Individuals die once. It may seem intuitively impossible to estimate any sort of individual propensity for mortality. But it turns out that all that is needed are (a) a named parametric distribution of these propensities, the expected value of which is finite, and (b) some covariates in the model ((Yashin et al., 2001; Elbers and Ridder, 1982; Heckman and Singer, 1984; Balan and Putter, 2020)). For example, in a Cox proportional hazards model, one cannot estimate individual effects without specifying their parametric distribution, and some covariates.
- **Omitted covariates.** We may have data on some important predictors but not others. For example, we may have data about individual types of microsites, but we rarely have data about variation in gene loci that may affect survival. In ordinary least-squares regression with independent regressors, omitting covariates still allows for consistent estimates of the remaining covariates. This is not always true for survival models. Under a proportional hazard model, omitting a covariate causes underestimation of the magnitude of remaining covariates (Bretagnolle and Huber-Carol, 1988). Indeed, this appears to have been the cause of the so-called “obesity paradox” in health research, in which some scientists concluded erroneously that obesity has a protective effect on those with some problems like cardiovascular diseases. The intuitive explanation is simple (Balan and Putter, 2020): with two independent factors that both increase risk, those most at risk are those with large values of both. They tend to die early. Under a model with only one of these factors, over time a spurious negative correlation develops between the factors. The overall effect that is estimated is consequently too small for the included factor. At least for some distributions, omitted covariates do not have as large effects on AFT models as for the Cox proportional hazards model (Gail et al., 1984; Struthers and Kalbfleisch, 1986).

Table 1 on the following page gives results of a simulation comparing the estimates of regression slopes under both Cox and accelerated failure time models, with and without a covariate, and with and without a frailty term. The same qualitative results occur if $\beta_1 = \beta_2$. These results show that the AFT model is robust to omitted covariates, but the Cox model is less so. Adding a frailty term can help with this problem in the Cox model.

- **Frailty.** Including a random term – usually called “frailty” – can overcome some of the problem caused by omitted covariates. Frailty terms are random effects; in some respects they bear a similarity to random effects in the familiar generalized linear models. The term “frailty model” is used in at least two ways in the literature: it may refer to the form of the frailty term within a statistical model, or to the entire statistical model itself.
- **Measurability.** An important conceptual distinction is between measurable and nonmeasurable predictors. We can measure, at least in principle, the height of an organism. On the other hand, while we can posit that there are predictors that make some individuals more or less prone to die (these may be, e.g., genetic, site-related, or socially induced), in practice it may be often be impossible to measure these. Similarly, while we may find that some families live longer than others, characterizing the genetic and environmental factors that cause this is typically beyond our grasp in practice. Some of these unmeasurable effects can be absorbed into frailty terms.
- **Levels of random variation.** Heterogeneity may occur at many different hierarchical levels (see Section X), and we expect that it is often likely to occur simultaneously at multiple levels. While this may be the case, it is not always practical to estimate such models. Sample size, of course, may be limiting, but there is a thornier problem: at present, there is little software that can estimate random effects for survival models at multiple levels. The `coxme` library in R does allow for mixed effects for the Cox proportional hazards model, and the `frailtyHL` library can fit models with multiple random levels for both Cox and accelerated failure time models. Models including shared frailty within specified groups can be estimated using the `streg` function in the Stan statistical modeling system. As of this writing, there is little experience with these approaches in ecology or evolutionary biology. Larger compendia of software for frailty modeling are provided in (Balan and Putter, 2020) for proportional hazard models, and by (Gorfine and Zucker, 2023) for accelerated failure time models.

Frailty: what it is and isn’t. Not really an estimate of heterogeneity, b/c of model dependence.

There are many other types of survival models that may have potential uses in population biology. These include parametric hazard models

	Mean($\hat{\beta}_1$)	SD($\hat{\beta}_1$)
Cox (1,2)	1.00169	0.032309
Cox (1)	0.65322	0.028168
Cox (1, f)	0.86420	0.054916
AFT (1,2)	-1.00007	0.024453
AFT (1)	-0.98488	0.035053

Table 1: Summary of results of simulations including both covariates or only one. Each of 200 runs used a sample population of 2000, with two independent covariates, each $\sim \mathcal{N}(0, 1)$, with $\beta_1 = 1, \beta_2 = -1$, and an exponential baseline hazard. 50 time steps were used, with uniform censoring on (20, 50). The table gives the mean and standard deviation of $\hat{\beta}_1$ for 200 runs of the simulation. We did not add a frailty term to the AFT models because the regressions were indistinguishable without them.

Some individuals die quickly: did they have poor survival probabilities, or bad luck? In individual cases, we generally can't say, but appropriate sampling and modeling can address this problem. Similarly, survival is usually affected by many things, some of which we can readily identify and measure, like family, site, or presence of a disease. Accounting for these is necessary, both to make appropriate estimates about the survival process, and to understand some of the causes underlying variation in survival.

Mark-recapture data

3.1 Predictors in survival models

There's a bestiary of names – covariates, factors, random factors, independent variables, and so on. For the present, just call them all “predictors.”

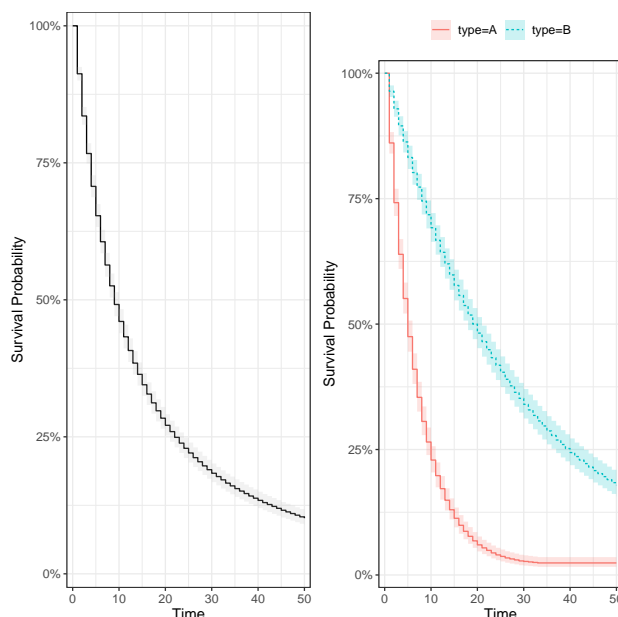
- Because sometimes that's where a lot of biological interest lies. E.g., we may want to know the effect of predator density on prey survival. Sometimes these have been manipulated experimentally.
- Because these factors induce variation in survival within the population.
- Because otherwise we may delude ourselves by e.g., just taking the mean survival.

How is it that we may be deluding ourselves by ignoring underlying variation? Consider a simple example: there are two types in a population. Both have constant survival probabilities, but one (perhaps in poor microsites) survives each interval at $p \approx 0.86$, while the better survivors have $p \approx 0.96$. The simulated survival process is shown in Figure 1.

```
source("SimpleSurvHetSimulation.R")
```

This example (Figure 1) may seem extreme, but it provides several important conclusions. Most important, the cumulative survival probability for the pooled population does not reflect that for any individual. Moreover, reliance simply on the survival estimate for the pooled population would miss the biological processes underlying the difference. Ignoring the predictors (in this case, type) can lead to misestimation. For example, if the initial proportions of the two types were different (in this example they are equal), the right-hand figure would be unchanged, but the left-hand figure (for the pooled population) would be different.

An important conceptual distinction is between measurable and nonmeasurable predictors. We can



measure, at least in principle, the height of an organism. On the other hand, while we can posit that there are predictors that make an individual more or less prone to some disease (these may be, e.g., genetic, site-related, or socially induced), in practice it may be often be impossible to measure these. Similarly, while we may find that some families live longer than others, characterizing the genetic and environmental factors that cause this is typically beyond our grasp in practice.

4 Frailty

5 Notes made earlier, to be incorporated or tossed

By its nature, estimating survival probabilities, and heterogeneity in them, presents some conceptual and technical challenges. Individuals have single sample - there's no replication. For individuals, we can observe date of death, but not the survival probability. And the process is inherently stochastic. So for individuals without any covariates, we can't distinguish between a high probability of survival and good luck, or vice versa.

But there are ways of estimating meaningful quantities. They all involve some combination of model-based estimation, measurement of meaningful covariates, and aggregation of individuals.

There are 3 main types of models:

- AFT
- Hazard-based regression
- Logistic regression
- Mark/recapture approaches

They each have strengths and weaknesses; for an introduction, see XXX.

One difficulty is that there are relatively few ecologists or evolutionists who are deeply familiar with the statistical methods used for analyzing survival data. Deep assumption in much of ecol/evol that GLM and its relatives covers most of the statistics needed. Most of the survival literature is in biostatistics and human demography; these are approachable, but the problems and the jargon are somewhat different.

Frailty models do not provide an estimate of the amount of heterogeneity, unless one can defend certain assumptions. This is because the variance of the random effect depends on the model specified and on the questions asked. CHECK ON THIS the random terms can be inflated/deflated by other terms in model. Obviously, its magnitude is meaningful in a qualitative sense.

There are multiple frailty models, and multiple senses of the word. Individual frailty, shared frailty, correlated frailty are the most common. It seems likely there will be more.

Frailty distributions

Value of studying covariates which you *may* treat as random.

Study of heterogeneity in survival presents additional challenges.

***** Individuals vary in their survival probabilities. Not only because of variation in age, size, or stage, but because of their histories of nutrition, parasites, or disease, their varied access to resources or exposure to toxicants, and their varied positions in social hierarchies. There may also be genetic variation that contributes to varied survival propensities.

This said, distinguishing between individual variation and stochasticity can be a knotty problem. Some individuals die quickly: did they have poor survival probabilities, or bad luck? In individual cases, we generally can't say, but appropriate sampling and modeling can address this problem.

There are several types of regression models used for survival data. These are somewhat specialized for survival data because survival times are not normally distributed, and the data are typically censored. By censored, we mean that we know only inequalities about the data (an individual survived at least as long as $\tau < x$, or between two values ($x_1 \leq \tau \leq x_2$), or less than some value ($\tau < x_2$). Censorship and non-normality usually make it impossible to use well-known approaches like GLM. Mark-recapture data

But there are powerful methods for survival analysis, mainly developed in biostatistics and industrial reliability testing.

Empirical studies: scrub-jays.

References

- Balan, T.A. and Putter, H. 2020. A tutorial on frailty models. *Statistical Methods in Medical Research* **29**: 3424–3454.
- Bretagnolle, J. and Huber-Carol, C. 1988. Effects of omitting covariates in Cox’s model for survival data. *Scandinavian Journal of Statistics* **15**: 125–138.
- Deevey, E.S. 1947. Life tables for natural populations of animals. *The Quarterly Review of Biology* **22**: 283–314.
- Elbers, C. and Ridder, G. 1982. True and spurious duration dependence: the identifiability of the proportional hazard model. *The Review of Economic Studies* **49**: 403–409.
- Gail, M.H., Wieand, S., and Piantadosi, S. 1984. Biased Estimates of Treatment Effect in Randomized Experiments with Nonlinear Regressions and Omitted Covariates. *Biometrika* **71**: 431–444.
- Gorfine, M. and Zucker, D.M. 2023. Shared Frailty Methods for Complex Survival Data: A Review of Recent Advances. *Annual Review of Statistics and Its Application* **10**: 51–73.
- Heckman, J. and Singer, B. 1984. The identifiability of the proportional hazard model. *The Review of Economic Studies* **51**: 231–241.
- Struthers, C.A. and Kalbfleisch, J.D. 1986. Misspecified proportional hazard models. *Biometrika* **73**: 363–369.
- Yashin, A.I., Iachine, I.A., Begun, A.Z., and Vaupel, J.W. 2001. *Hidden frailty: myths and reality*. Research Report. Odense, Denmark: Department of Statistics & Demography, University of Odense.