Survival heterogeneity review

Gordon A. Fox and Bruce E. Kendall

February 4, 2026

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.

1   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.

The models used for studying survival have some special characteristics. Survival processes generally do not lead to symmetric distributions of survival time, as they have an abrupt beginning, such as birth or the start of an experiment. Consider 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 I), survival times are distributed exponentially. As a consequence, survival models do not rely on the normal distribution, and the population mean (or marginal) survival is not generally a good descriptor for the entire population. In fact, 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. Moreover, 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. Because of the censorship and the skewed distribution of time-to-death, models based on the normal distribution are not generally used in survival analysis.

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 (Therneau and Grambsch, 2000; Kalbfleisch and Prentice, 2002; Lebreton et al., 1992). There are many forms of each. They have their own pros and cons in any application, and these are discussed elsewhere (Fox, 2001) (XX, bunch more cites). Here we focus on the issues (and pros and cons) related to heterogeneity in populations.

1.1   Regression models

The AFT model, in its simplest form, is

$$\log T_i = x_i^{'}\beta + \sigma W_i$$

where T_i is the event time for individual i, x_i is a matrix of covariates and $x_i^{'}$ its transpose, β is a vector of regressors, σ is a dispersion factor, and W is the error, typically distributed as Weibull or log-logistic. Less commonly some other distributions are used, including the gamma, Gompertz, and lognormal.

The Cox PH model is

$$h_i(t) = h_0(t) \exp \left[\Sigma_{j=1}^{r} \beta_j X_{ij} \right]$$

where $h_i(t)$ is the hazard for individual i at time t, h_0 is the baseline hazard, and the terms in the summation are covariates and regression terms, as in the AFT model.

The terms in these two models differ in sign — a positive effect of a covariate in the AFT model delays death, while the analogous term in a Cox PH model increases the hazard and thus tends to advance death.

1.2   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, Iachine, et al., 2001; Elbers and Ridder, 1982; Heckman and Singer, 1984; Balan and Putter, 2020)). For example, in a Cox proportional hazards (PH) model, one cannot estimate individual effects without specifying their parametric distribution, and some covariates.

1.3   Omitted covariates

We often 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 the Cox PH 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 PH model (Gail et al., 1984; Struthers and Kalbfleisch, 1986).

	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.

Table 1 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. A minute's thought should make it clear why the Cox model is so sensitive to missing covariates. Since the hazard under this model has no specified form, without further information, an early (or late) death might be equally due to luck, or to individual heterogeneity.

1.4   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 covariates

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 some of 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.

1.5   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.

The meaning of "random" here is important. It does not imply that the factor occurs in some truly random sense, nor that it is sampled at random. This is similar to random effects in a GLMM, although some papers still repeat these fallacies. Different authors use the term "frailty model" differently. Some refer to the form the frailty term takes within the statistical model, while others use "frailty model" to mean the entire statistical model that includes a frailty term.

If a frailty term is used, the AFT model becomes

$$\log T_i = x_i^{'}\beta + \epsilon_i + \sigma W_i$$

where ϵ is the frailty term; because it is additive in this model for the log of time-to-event, it acts multiplicatively on the untransformed time. With the addition of frailty, the Cox PH model becomes

$$h_i(t) = Z_i h_0(t) \exp\left[\Sigma_{j=1}^r \beta_j X_{ij}\right]$$

where the Z is the frailty. Thus the AFT and Cox PH models are on different scales, and they model, respectively, the log time to death and the hazard (the instantaneous risk of death). The terms differ in sign — a positive effect of a covariate in the AFT model delays death, while the analogous term in a Cox PH model increases the hazard and thus tends to advance death.

Frailty terms are most often taken to be gamma-distributed (typically with mean 0), because this turns out to be mathematically convenient. Other distributions are used, including the inverse gaussian (Wald) and positive stable distributions. There has been little study of how to choose the frailty distribution; Balan and Putter (2020) suggest that the choice is typically made by convenience.

Under the **shared frailty** model (Balan and Putter, 2020; Hougaard, 2000), the frailty term is the same for all members of a cluster — for example, a family or a site. The frailty term then represents the portion of frailty that is shared (say, by family members); frailty at other hierarchical levels is taken up by the underlying hazard function. The **correlated frailty** model (Yashin, Vaupel, et al., 1995) relaxes the assumption that the frailty is identical for all members of a cluster, and can be estimated using the `frailtypack` library in R (Rondeau et al., 2012). There are numerous other types of frailty models, including time-varying (Unkel et al., 2014) models. However, there is currently no standard software that can estimate these models.

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 PH 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 PH models, and by (Gorfine and Zucker, 2023) for accelerated failure time models.

2   Todo

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

- **Time-varying covariates**

- **Discrete time. Loglog**

- **Mark-recapture data**

- **What's done in IPMs**

- **Incorporate passages below**

- **Empirical**

2.1   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.

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.

3   Notes made earlier, to be incorporated or tossed

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.
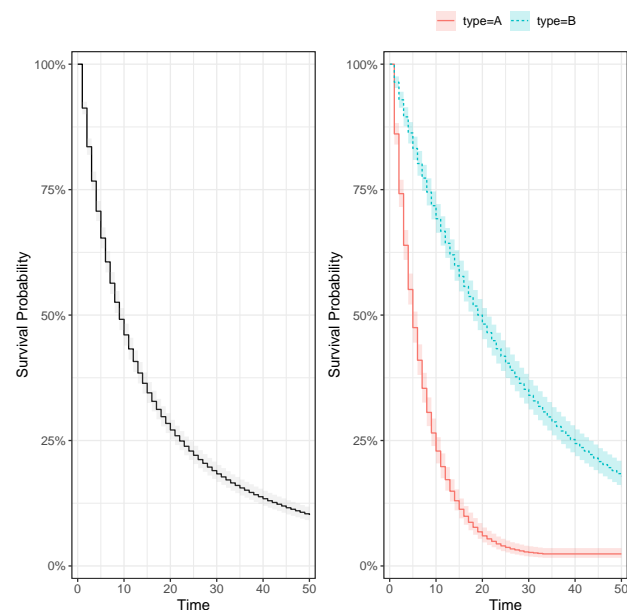

Figure 1: Survival in a population composed of two types, one of which has superior survival. Survival estimated for the entire population, ignoring the heterogeneity (left) does not represent that of any individual.

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.

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.

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.

Fox, G.A. 2001. Failure-time analysis: studying times to events and rates at which events occur. *Design and analysis of ecological experiments* **2**: 235–266.

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.

Hougaard, P. 2000. *Analysis of Multivariate Survival Data*. Ed. by K. Dietz, M. Gail, K. Krickeberg, A. Tsiatis, and J. Samet. Statistics for Biology and Health. New York, NY: Springer New York.

Kalbfleisch, J.D. and Prentice, R.L. 2002. *The statistical analysis of failure time data*. Hoboken, NJ: Wiley.

Lebreton, J.D., Burnham, K.P., Clobert, J., and Anderson, D.R. 1992. Modelling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs* **62**: 67–118.

Rondeau, V., Mazroui, Y., and Gonzalez, J.R. 2012. Frailtypack: An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal Of Statistical Software* **47**: 1–28.

Struthers, C.A. and Kalbfleisch, J.D. 1986. Misspecified proportional hazard models. *Biometrika* **73**: 363–369.

Therneau, T.M. and Grambsch, P.M. 2000. *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer New York.

Unkel, S., Farrington, C.P., Whitaker, H.J., and Pebody, R. 2014. Time Varying Frailty Models and the Estimation of Heterogeneities in Transmission of Infectious Diseases. *Journal of the Royal Statistical Society Series C: Applied Statistics* **63**: 141–158.

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.

Yashin, A.I., Vaupel, J.W., and Iachine, I.A. 1995. Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies* **5**: 145–159, 183.