Parsimonious survival analysis models: Studying early attrition in the armed services by frailty and time-dependent survival analysis

Yuanzhang Li, Timothy Powers, and Margot Krauss

Introduction

Early attrition among new enlistees is a costly problem for the federal military. It is of interest to recruiters and planners to have an idea of how much attrition will take place, as well as when it will occur. It is also of interest to know what factors are related to likelihood of attrition so that the best prospects for retention can be sought. Survival analysis is one way of addressing these needs, and the current paper will examine different types of survival models for suitability to this task.

A popular approach to survival analysis problems with multiple covariates is Cox proportional hazards modeling. Also known as "semi-parametric" modeling, this approach does not require such strenuous assumptions as to the functional form of the survival curve as do parametric models. Two critical assumptions with regard to functional form are 1) a multiplicative relation between the underlying hazard function and the log-linear function of the covariates; and 2) the hazard associated with any particular combination of factor levels is proportional over time to that associated with any other combination of factor levels. Quite conveniently, this assumption holds true for many of the more commonly used parametric models, including the exponential and the Weibull.

However, while the assumption of proportional hazards is less restrictive than those needed for parametric modeling, even this lesser assumption is sometimes of dubious merit. In particular, the first assumption stated above outcome of interest may be related to factors beyond those included in the model, resulting at best in bias of effects estimates. Another possible violation of the proportional hazards assumption is that the effects of some factors on the outcome of interest might not be constant over time.

Frailty models were proposed by Vaupel et al (1979) to deal with the problem of underspecification or misspecification of covariates in a proportional hazards setting. In essence, these models allow for the survival among individuals with the same levels of predictor factors to have survival that differs according to a distribution, rather than simply by random error.

The second assumption listed above (time independence) might also not be true. For example, a military recruit with a pre-existing medical condition might be at increased risk of early attrition, although presumably more so at the beginning of training than once he/she has successfully undergone rigorous training. In such a case, the impact of the covariate is clearly dependent on time. Non-proportional hazards models are used to extend the proportional model to analyze such data. The user can specify arithmetic expressions to define covariates as functions of several variables and survival time.

The aim of this study is to develop more accurate depictions of early military attrition using these survival model variations.

Subjects and Methods

All first-time enlistees beginning active duty enlisted service in the Army during January 1998 - December 2001 were included in the analyses. Accession records on these individuals were linked with military personnel records to determine whether or not a subsequent early attrition occurred. In addition to the accession data, the demographic factors, such as sex, age, race, education, Armed Forces Qualification Test scores (AFQT), medical qualify, Body Mass Index (BMI), married status, the number of dependents as well as the geographic factor, the Military Entrance Processing Station (MEPS) are included in the model. These factors have been found in previous studies to be strongly related to likelihood of attrition.

Three types of survival models are used to relate attrition to these multiple factors:

Cox-Proportional Model-

$$h_i(t) = h_0(t)\exp(\beta_1 x_{i1} + \cdots + \beta_k x_{ik}) \qquad 1$$

where h is the hazard function, which is the instantaneous probability of failure given survival up to t and h_0 is the baseline hazard.

Frailty Model-

Unexplained variability, that not accounted for by including covariates, is known as overdispersion. Overdispersion is caused either by misspecification or omitted covariates, and makes the assumption of the proportional model invalid. The hazard is

$$h_i(t \mid z) = h_i(t) z_i$$

A frailty model attempts to account for variability that is not accounted for by the included covariates (overdispersion) by the following model:

$$h_i(t) = h_0(t) z_i \exp(\beta_1 x_{i1} + \cdots + \beta_k x_{ik}) \qquad 2.1$$

Here, z_i varies across the individuals. However, frailty models are also used to model survival times in the presence of group-specific random effects. Such models are termed "shared" frailty models, and depicted as follows:

$$h_{ij}(t_{ij}) = h_0(t_{ij}) z_j \exp(\beta_1 x_{ij1} + \cdots + \beta_k x_{ijk}) \qquad 2.2$$

In general, the random unobservable frailty effects are often assumed to follow either a gamma or inverse-Gaussian distribution. For the shared group-specific frailty, the model is constrained to be equal over those observations from a given group or panel.

Time-Dependent Proportional Model:

As discussed above, the hazard ratios for different combinations of factor levels may vary according the time. Hence we should consider a time-depend hazard model such as the one depicted below:

$$h_{ij}(t_{ij}) = h_0(t_{ij}) z_j \exp(\beta_1 x_{ij1}(t_{ij}) + \cdots + \beta_k x_{ijk}(t_{ij}) + f(t_{ij})\eta) \qquad 3.1$$

where f(t) is a function of the survival time t. In this study, we consider a special case of 3.1.

$$h_i(t) = h_0(t) \exp(\sum_{j=1}^{K} x_{ij} \beta_j [1 + \gamma_j \ln(t) + \eta_j \ln(t)^2]) \qquad 3.2$$

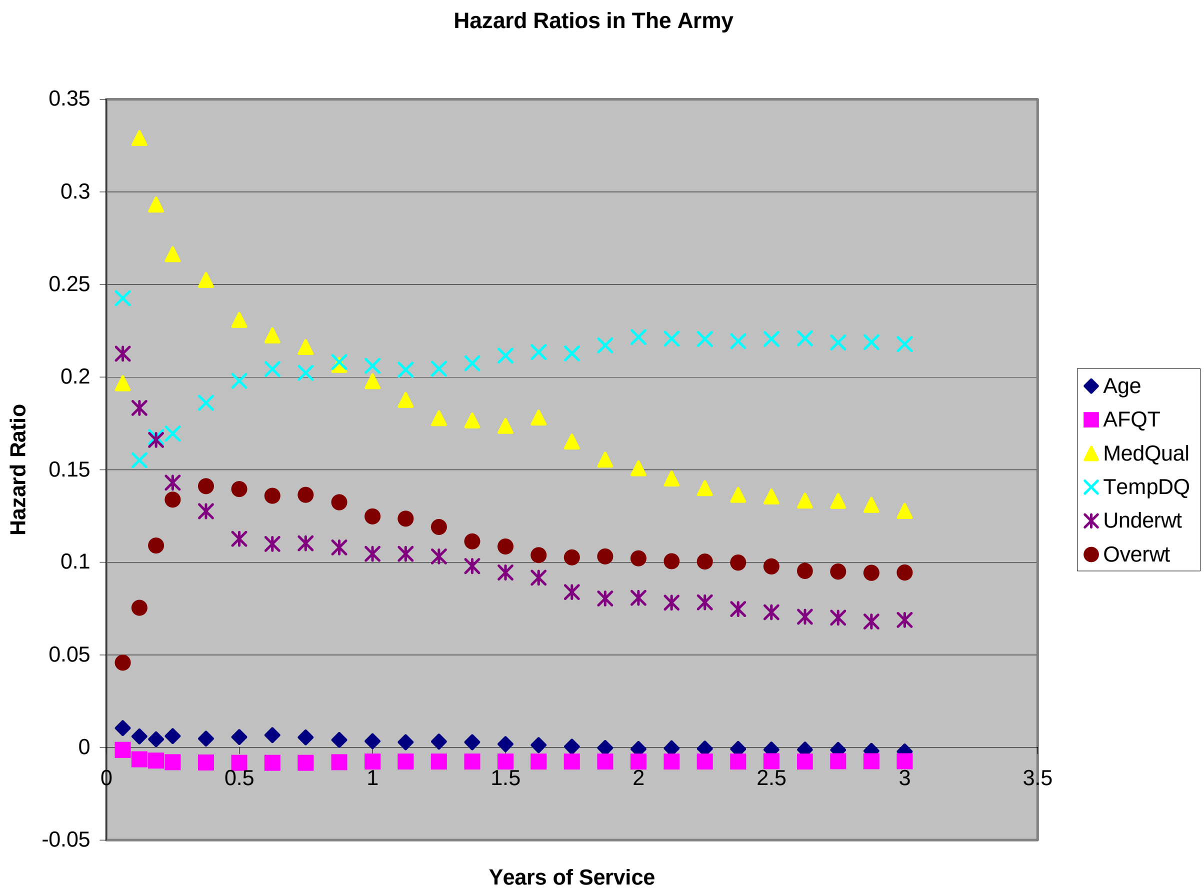
Note that the natural logarithm of survival time, ln(t), is used rather than survival time t, as suggested by many previous investigators. The square of ln(t) is used to measure non-linear effect of time.

Some special cases of model 3.2 should be noted. If $\eta_j=0$ for j=1, …, K, then model 3.2 becomes the time-dependent linear model. If $\gamma_j= \gamma$, $\eta_j= \eta$ for j=1, …, K, then model 3.2 becomes a special case frailty model. Finally, if $\gamma_j=\eta_j=0$ for j=1, …K, then model 3.2 becomes the Proportional Hazards Model.

A basic principal in model estimation is to include only those prediction factors that have some influence on the dependent variable. Hence, in the estimation process, we will delete all terms with non-significant γ or η. If none of these terms is significant, the proportional hazards model can be taken as the appropriate model.

Results

First we applied a proportional hazards model within different years of services to examine the effects of the various factors on hazard of attrition over the first three years of service. Figure 1 shows changes in some of the estimated factor effects over the examined time period. It is clear that the effects of several factors vary,over time, meaning that the proportional hazards assumption is not tenable. For example, the effect of being overweight increases over the first half-year of service, then diminishes gradually over the next 2 ½ years. Similar observations of changing effects can be seen for being underweight, being medically qualified, and being temporarily medically disqualified. The effects of age were a little higher for the first year of service, then relatively stable around the same value afterwards. The effect of AFQT score was relatively small and appeared more constant over time, with a minor wavering.



Hazard Ratios in The Army
Hazard Ratio
Years of Service
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
0
0.5
1
1.5
2
2.5
3
3.5
Age
AFQT
MedQual
TempDQ
Underwt
Overwt

Given that the proportional hazards assumption is not met by the Army attrition data, we proceed to consider a frailty model. We consider the frailty to be due to the geographic regions from which military applicants come, and the demographic distributions of applicants from the various regions. Table 1 shows the differences in factor effect estimates between the frailty model and the simpler Cox proportional hazards model.

It is seen that the inclusion of frailty has virtually no effect on factor effect estimates for attrition within the first year of service. However, the consideration of frailty does result in altered effect estimates at longer periods of service. For example, the effect of being black (and of being white, for that matter) on attrition within the first three years of service is estimated to be much less when frailty is considered than when it is not considered. Other factor effect estimates are seen to show such differences at the longer time range.

Table 1. Factor Coefficients With and Without Frailty Consideration

	Within 1 Year		Within 2 Years		Within 3 Years	
Factor	No Frailty	With Frailty	No Frailty	With Frailty	No Frailty	With Frailty
Age	0.0033	0.0033	-0.001	-0.001	-0.002	-0.002
AFQT	-0.0076	-0.0076	-0.008	-0.008	-0.007	-0.011
Black	0.0813	0.0811	0.092	0.094	0.106	0.083
Dependents	0.0374	0.0373	0.063	0.065	0.059	0.071
Less than HS	-0.0478	-0.0479	-0.063	-0.064	-0.049	-0.050
Married	-0.1533	-0.1528	-0.161	-0.167	-0.175	-0.178
MedQual	0.1977	0.1975	0.151	0.147	0.128	0.119
Single	-0.3551	-0.3539	-0.341	-0.360	-0.352	-0.373
White	0.4810	0.4811	0.463	0.470	0.448	0.376
Underwt	0.1045	0.1046	0.081	0.079	0.069	0.065
TempDQ	0.2060	0.2059	0.222	0.217	0.218	0.211
Overwt	0.1248	0.1248	0.102	0.104	0.095	0.128
Sex	0.6662	0.6667	0.628	0.605	0.601	0.479

The relation of several of the considered factors to attrition has been seen to differ according to the length of time served. We therefore consider a time-dependent model to account for this time dependency. The effect of being overweight showed a non-linear pattern, so a non-linear time effect is considered for it. The effects of sex, medical qualification status, temporary medical disqualification, being underweight, having less than a high school education, being black and being white showed a linear relation with time, and are modeled accordingly. Finally, the effects of age and AFQT score did not show a relation to time, therefore no time component is considered for these factors.

Table 2 shows hazard ratio estimates for three factors (sex, permanent medical disqualification, and temporary medical disqualification) from both the Cox proportional hazards model at the indicated time cut-points, and the Time-Dependent model. It is seen that the effect estimates for these three factors differ over time within each model, and across models. Further work is needed to account for factors influencing attrition that are not yet accounted for.

Table 2. Hazard Ratio Estimates for Selected Factors:
Cox Proportional Hazards Model vs. Time-Dependent Model

	Proportional Hazards Model			Time-Dependent Model		
Months	Sex	Perm't DQ	Temp DQ	Sex	Perm't DQ	Temp DQ
1.5	2.73	1.34	1.34	1.86	1.15	1.19
3	2.13	1.25	1.33	1.70	1.12	1.19
4.5	1.94	1.22	1.32	1.61	1.10	1.19
6	1.88	1.19	1.33	1.55	1.09	1.19
7.5	1.85	1.18	1.32	1.50	1.08	1.19
9	1.84	1.17	1.32	1.46	1.08	1.18
10.5	1.82	1.16	1.32	1.43	1.07	1.18
12	1.80	1.14	1.31	1.41	1.07	1.18
13.5	1.79	1.13	1.30	1.39	1.06	1.18
15	1.77	1.12	1.30	1.37	1.06	1.18
16.5	1.76	1.12	1.29	1.35	1.06	1.18
18	1.74	1.11	1.29	1.33	1.05	1.18
19.5	1.72	1.12	1.29	1.32	1.05	1.18
21	1.70	1.10	1.29	1.31	1.05	1.18
22.5	1.68	1.08	1.29	1.29	1.05	1.18
24	1.66	1.08	1.29	1.28	1.04	1.18
30	1.59	1.05	1.27	1.24	1.04	1.18
36	1.54	1.04	1.25	1.21	1.03	1.18

Conclusions

This study has found that the assumptions of the Cox proportional hazards model are not adequately met by Army attrition data. Several of the factors seen in many studies to be related to early attrition exhibit effects that are not constant over time in service, casting doubt on the proportional hazards model estimates. Consideration of a frailty model did not have an appreciable impact on factor effect estimates, although a time-dependent model did.

The time-dependent model considered here was derived in subjective fashion. Functional form of time-dependency for the various factors was decided by examining factor effects from Cox proportional hazards models applied at various time points. Future work will focus on a more objective way to develop the time-dependent model. Application to other services' attrition data will also be pursued.

Reference

Vaupel JW, Manton KG and Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography. 16(1979)3: 439-454.