Survival Analysis Using Cox Proportional Hazards Modeling For Single And Multiple Event Time Data

Tyler Smith, MS; Besa Smith, MPH; and Margaret AK Ryan, MD, MPH

Department of Defense Center for Deployment Health Research,

Naval Health Research Center, San Diego, CA

Abstract

Survival analysis techniques are often used in clinical and epidemiologic research to model time until event data. Using SAS® system's PROC PHREG, Cox regression can be employed to model time until event while simultaneously adjusting for influential covariates and accounting for problems such as attrition, delayed entry, and temporal biases. Furthermore, by extending the techniques for single event modeling, the researcher can model time until multiple events. In this real data example, PROC PHREG with the baseline option was instrumental in handling attrition of subjects over a long study period and producing probability of hospitalization curves as a function of time. In this paper, the reader will gain insight into survival analysis techniques used to model time until single and multiple hospitalizations using PROC PHREG and tools available through SAS®

Introduction

Survival analysis pertains to a statistical approach designed to take into account the amount of time an experimental unit contributes to a study period, or the study of time between entry into observation and a subsequent event. Originally, the event of interest was death and the analysis consisted of following the subject until death. The use of survival analysis today is primarily in the medical and biological sciences, however these techniques are also widely used in the social sciences, econometrics, and engineering. Events or outcomes are defined by a transition from one discrete state to another at an instantaneous moment in time. Examples include time until onset of disease, time until stockmarket crash, time until equipment failure, and so on.

Although the origin of survival analysis goes back to mortality tables from centuries ago, this type of analysis was not well developed until World War II. A new era of survival analysis emerged that was stimulated by interest in reliability (or failure time) of military equipment. At the end of the war, the use of these newly developed statistical methods quickly spread through private industry as customers became more demanding of safer, more reliable products. As the use of survival analysis grew, researchers began to develop nonparametric and semiparametric approaches to fill in gaps left by parametric methods. These methods became popular over other parametric methods due to the relatively robust model and the ability of the researcher to be blind to the exact underlying distribution of survival times.

Survival analysis has become a popular tool used in clinical trials where it is well suited for work dealing with incomplete data. Medical intervention follow-up studies are plagued with late arrivals and early departure of subjects. Survival analysis techniques allow for a study to start without all experimental units enrolled and to end before all experimental units have experienced an event. This is extremely important because even in the most well developed studies, there will be subjects who choose to quit participating, who move too far away to follow, who will die from some unrelated event, or will simply not have an event before the end of the observation period. The researcher is no longer forced to withdraw the experimental unit and all associated data from the study. Instead, censoring techniques enable researchers to analyze incomplete data due to delayed entry or withdrawal from the study. This is important in allowing each experimental unit to contribute all of the information possible to the model for the amount of time the researcher is able to observe the unit.

The recent strides in the application of survival analysis techniques have been a direct result of the availability of software packages and high performance computers which are now able to run the difficult and computationally intensive algorithms used in these types of analyses relatively quickly and efficiently.

Basic Tools of Survival Analysis

First recall that time is continuous, and that the probability of an event at a single point of a continuous distribution is zero. Our first challenge is to define the probability of these events over a distribution. This is best described by graphing the distribution of event times. To ensure the

reader will start with the same fundamental tools of survival analysis, a brief descriptive section of these important concepts will follow. A more detailed description of the probability density function (pdf), the cumulative distribution function (cdf), the hazard function, and the survival function, can be found in any intermediate level statistical textbook.

So that the reader will be able to look for certain relationships, it is important to note the one-to-one relationship that these four functions possess. The pdf can be obtained by taking the derivative of the cdf and likewise, the cdf can be obtained by taking the integral of the pdf. The survival function is simply 1 minus the cdf, and the hazard function is simply the pdf divided by the survival function. It will be these relationships later that will allow us to calculate the cdf from the survival function estimates that the SAS procedure PROC PHREG will output.

The Cumulative Distribution Function

The cumulative distribution function is very useful in describing the continuous probability distribution of a random variable, such as time, in survival analysis. The cdf of a random variable T, denoted $F_T(t)$, is defined by $F_T(t) = P_T(T < t)$. This is interpreted as a function that will give the probability that the variable T will be less than or equal to any value t that we choose. Several properties of a distribution function F(t) can be listed as a consequence of the knowledge of probabilities. Because F(t) has the probability $0 < F(t) < 1$, then F(t) is a nondecreasing function of t, and as t approaches ∞, F(t) approaches 1.

The Probability Density Function

The probability density function is also very useful in describing the continuous probability distribution of a random variable. The pdf of a random variable T, denoted $f_T(t)$, is defined by $f_T(t) = d\,F_T(t)\,/\,dt$. That is, the pdf is the derivative or slope of the cdf. Every continuous random variable has its own density function, the probability $P(a < T < b)$ is the area under the curve between times a and b.

The Survival Function

Let $T > 0$ have a pdf f(t) and cdf F(t). Then the survival function takes on the following form:
$$S(t) = P\{T > t\}$$
$$= 1 - F(t)$$

The survival function gives the probability of surviving or being event-free beyond time t. Because S(t) is a probability, it is positive and ranges from 0 to 1. It is

defined as S(0) = 1 and as t approaches ∞, S(t) approaches 0. The Kaplan-Meier estimator, or product limit estimator, is the estimator used by most software packages because of the simplistic step idea. The Kaplan-Meier estimator incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored times. The survival curve describes the relationship between the probability of survival and time.

The Hazard Function

The hazard function h(t) is given by the following:

$$h(t) = P\{ t < T < (t + \Delta) \mid T > t\}$$

$$= f(t) / (1 - F(t))$$

$$= f(t) / S(t)$$

The hazard function describes the concept of the risk of an outcome (e.g., death, failure, hospitalization) in an interval after time t, conditional on the subject having survived to time t. It is the probability that an individual dies somewhere between t and $(t + \Delta)$, divided by the probability that the individual survived beyond time t. The hazard function seems to be more intuitive to use in survival analysis than the pdf because it attempts to quantify the instantaneous risk that an event will take place at time t given that the subject survived to time t.
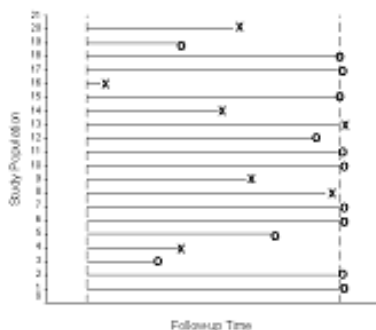
Incomplete Data

Observation time has two components that must be carefully defined in the beginning of any survival analysis. There is a beginning point of the study where time=0 and a reason or cause for the observation of time to end. For example, in a complete observation cancer study, observation of survival time may begin on the day a subject is diagnosed with the cancer and end when that subject dies as a result of the cancer. This subject is what is called an uncensored subject, resulting from the event occurring within the time period of observation. Complete observation time data like this example are desired but not realistic in most studies. There is always a good possibility that the patient might recover completely or the patient might die due to an entirely unrelated cause. In other words, the study cannot go on indefinitely waiting for an event from a participant, and unforeseen things happen to study participants that make them unavailable for observation. The censoring of study participants therefore deals with the problems of incomplete observations of time due to factors not related to the study design. **Note:** This differs from truncation where observations of time are incomplete due to a selection process inherent to the study design.

Left and Right Censoring

The most common form of incomplete data is right censored. This occurs when there is a defined time (t=0) where the observation of time is started for all subjects involved in the study. A right censored subject's time terminates before the outcome of interest is observed. For example, a subject could move out of town, die of an unexpected cause, or could simply choose not to participate in the study any longer. Right censoring techniques allow subjects to contribute to the model until they are no longer able to contribute (end of the study, or withdrawal), or they have an event. Conversely, an observation is left censored if the event of interest has already occurred when observation of time begins. For the purposes of this study we focused on right censoring.

The following graph shows a simple study design where the observation times start at a consistent point in time (t=0). The X's represent events and the O's represent censored observations. Notice that all observations are classified with an event, censored at time of loss to follow-up, or censored at the end of the study period. Some subjects have events early in the study period and others have events at the end of the study period. Likewise some subjects leave early, but most do not have an event during the entire study and are simply right censored at the end. There is no need for left censoring or truncation techniques in this simple example.

Figure1.

Time Dependencies

In some situations the researcher may find that the dynamic nature of a variable causes changes in value over the observation time. In other instances the researcher may find that certain trends affect the probability of the event of interest over time. There are easy ways to test and account for these temporal biases within PROC PHREG but be careful if you have a large number of observations as the computation of the subsequent partial likelihood is very taxing and time consuming. An easier way to see if

there are existing temporal biases is to look at the plots of the cumulative distributions of the probability of event. If there is a steep increase or decrease in the cumulative probability, it may suggest more investigation is needed. It is important to note here that when a time dependent variable is introduced into the model, the ratios of the hazards will not remain steady. This only affects the model structure. We will still be doing a Cox regression but instead the model used is called the extended Cox model.

The Study

The objective of this study was to investigate the effect of exposure to Kuwaiti oil well fire smoke and subsequent health events by comparing the postwar hospitalization experiences of various Gulf War exposure groups.

Demographic data

Demographic data available for analysis included social security numbers (for linking purposes only), gender, date of birth, race, ethnicity, home state, marital status, primary military occupation, military rank, length of military service, deployment status, salary, date of separation from military service, military service branch, and exposure to oil well fire smoke status.

Hospitalization Data

Data describing hospitalization experiences were captured from all United States Department of Defense military treatment facilities for the period of October 1, 1988, through December 31, 1999. The actual observation period varied by study. Removal of personnel with diagnoses of interest prior to the start of the study follow-up period was completed. These data included date of admission in a hospital and up to eight discharge diagnoses associated with the admission to the hospital. Additionally, a pre-exposure period covariate (coded as yes or no) was used to reflect a hospital admission during the 12 months prior to the start of the exposure period. Note: the exposure period was the year from August 1, 1990, to August 1, 1991. Diagnoses were coded according to the *International Classification of Diseases, Ninth Revision* (ICD-9). For these analyses, we scanned for the specific 3,4, or 5-digit component of the ICD-9 diagnoses.

Observation Time

The focus of each study was to see if certain estimates of exposure had any influence on the targeted disease outcomes. For each subject, hospitalizations (if any) were scanned in chronological order and diagnostic fields were

scanned in numerical order for the ICD-9-CM codes of interest. Only the first hospitalization meeting the outcome criteria was counted for each subject. Subjects were classified as having an event if they were hospitalized in any Department of Defense hospital worldwide with the targeted diagnoses, and as censored otherwise. Observation time start and end dates varied between studies, but the same methods were used to calculate total time from the start date of follow-up until event, separation from military service, or the end of the study period, whichever occurred first. Subjects were allowed to leave the study and found to follow a random early departure distribution. Delayed entry and events occurring before the start date of the study were not allowed and therefore not a concern. Right censoring was needed to allow for the early departure of subjects from the military active duty status (see figure 1).

Simple Survival Curves Using PROC LIFETEST

We should take a moment to mention another popular nonparametric approach used when the researcher simply wants to know if there are differences in the survival curves and what these differences look like graphically.

```
proc lifetest data=analydat plots=(s)
graphics;
 model inhosp*censor(0);
  strata exposure;
  test sex age marital;
  symbol1 v=none color=black line=1;
  symbol2 v=none color=black line=2;
run;
```

This procedure will produce survival function estimates and plot them, as well as give the Wilcoxon and Log Rank statistics for testing the equality of the survival function estimates of the two or more strata being investigated. The test statement within the procedure will test the association of the covariates with the survival time. The last two lines of code in the procedure will visibly give the graph different lines with no symbols differentiating them.

Why We Used Cox's Proportional Hazards Regression

Cox's proportional hazards modeling was chosen to investigate the effect of exposure to oil well fire smoke on time until hospitalization, while simultaneously adjusting for other possibly influential variables. Other attractive features of Cox modeling include: the relative risk type measure of association, no parametric assumptions, the use of the partial likelihood function, and the creation of survival function estimates.

Relative Risk

The simple interpretation of the measures of association given by the Cox model as "relative risk" type ratios is very desirable in explaining the risk of event for certain categories of covariates or exposures of interest. For example, when a two-level (dichotomous) covariate with a value of 0=no and 1=yes is observed, the hazard ratio becomes e^{β} where β is the parameter estimate from the regression. If the value of the coefficient is $\beta = 1.099$, then $e^{1.099} = 3$. The measure is simply saying that the subjects labeled with a 1 (yes) are three times more likely to have an event than the subjects labeled with a 0 (no). In this way we have a measure of association that gives insight into the difference between our exposure categories and those not being exposed.

No Parametric Assumptions

Another attractive feature of Cox regression is not having to choose the density function of a parametric distribution. This means that Cox's semiparametric modeling allows for no assumptions to be made about the parametric distribution of the survival times, making the method considerably more robust. Instead, the researcher must only validate the assumption that the hazards are proportional over time. The proportional hazards assumption refers to the fact that the hazard functions are multiplicatively related. That is, their ratio is assumed constant over the survival time, thereby not allowing a temporal bias to become influential on the endpoint.

Use of the Partial Likelihood Function

The Cox model has the flexibility to introduce time-dependent explanatory variables and handle censoring of survival times due to its use of the partial likelihood function. This was important to our study in that any temporal biases due to differences in hospitalization practices for different strata of the significant covariates over the years of study needed to be identified so that they might be handled properly. This ensured that any differences in hospitalization experiences between the exposed and nonexposed would not be coming from temporal differences over the study period.

Survival Function Estimates

With the SAS option BASELINE, a SAS dataset containing survival function estimates stratified by exposure category levels can be created and output. These estimates correspond to the means of the explanatory variables for each stratum.

Analysis

Univariate Analyses

Using PROC FREQ and PROC UNIVARIATE, an initial univariate analysis of the demographic, exposure, and deployment variables crossed with hospitalization experience was carried out to determine possible significant explanatory variables to be included in the model runs. An exploratory model analysis was then performed to explore the relations between the variables while simultaneously adjusting for all other variables that had influences on the outcome of interest. After investigation of confounding, all variables with p-values of 0.15 or less were considered possible confounders and were retained for the model analysis. Additionally, the distributions of attrition were checked to see if attrition rates differed for the categories of exposure over the study period.

Multivariable Cox Modeling Approach

Dummy variables were created for reference cell coding of the categorical variables. These were necessary for the output of measures of association using the reference category of choice. Starting with a saturated model, PROC PHREG was run using a manual backward stepwise model building approach. This created a final model with statistically significant effects of explanatory variables on survival times while controlling for possible confounding of exposure effects.

SAS Programming

```
PROC PHREG data=analydat;
 model inhosp*censor(0)=expose1-expose6
 pwhsp status1 sex1 age1-age3 ms1
 paygr1-paygr2 oc_cat1-oc_cat9 ccep
                / rl ties=efron ;
title1 'Cox regression with exposure
 status in the model';
run;
```

The options used in this survival analysis procedure are described below:

DATA=ANALYDAT names the input data set for the survival analysis.

RL requests for each explanatory variable, the 95% (the default alpha level because the ALPHA= option is not invoked) confidence limits for the hazard ratios.

TIES=EFRON gives the researcher the approximations to the EXACT method without using the tremendous CPU it takes to run the EXACT method. Both the EFRON and the BRESLOW methods do reasonably well at approximating the EXACT when there are not a lot of ties. If there are a lot of ties, then the BRESLOW

approximation of the EXACT will be very poor. If the time scale is not continuous and is therefore discrete, the option TIES=DISCRETE should be used.

Stratification By Exposure Status

These data were then stratified by exposure and the models were run with the exposure flag covariate withdrawn from the model. This allowed for inspection of confounding between exposure status and covariates. Running these separate models also allowed for the computation of survival function estimates using the BASELINE function in PROC PHREG. The survival curves (which are really step functions, however there are such numerous events that they appear continuous) were now available to compute the cumulative distribution function for the separate exposure categories.

Time Dependent Covariates

After the final model of significant explanatory variables was created, it was necessary to validate the proportional hazards assumption. If the researcher believes that there may be a time dependency from a certain variable then simply add x1time to the list of independent variables and the following below the model statement.

x1time=x1*(t)

Where t is the time variable and x1 is the suspected time dependent variable.

If the interaction term is found to be insignificant we can conclude that the proportional hazards assumption holds. This was necessary to ensure that there was no adverse effect from time-dependent covariates creating different rates for different subjects, thus making the ratios of their hazards nonconstant.

Survival Function Estimates By Exposure Category

The following is the code used after the dataset ANALYDAT was stratified into exposed or nonexposed. This produced the survival function estimates by exposure while simultaneously checking to see if there were any interactions between the covariates and the exposure status.

```
PROC PHREG data=expose1;
 model inhosp*censor(0)=pwhsp status1
   sex1 age1-age3 ms1 paygr1-paygr2
   oc_cat1-oc_cat9 ccep
   /rl ties=efron ;
   baseline out=survs survival=s;
run;
```

The different options used in this survival analysis procedure are described below:

BASELINE without the COVARIATES= option produces the survival function estimates corresponding to the means of the explanatory variables for each stratum.

OUT=SURVS names the data set output by the BASELINE option.

SURVIVAL=S tells SAS to produce the survival function estimates in the output data set.

A simple calculation of 1-Survival function estimates in SURVS, obtained from running the BASELINE option, produced the cumulative distribution functions. This allowed investigation of the cumulative probability estimates of hospitalization over time and the ability to scan for differences in hospitalization experiences between the exposure categories. This investigation helped determine whether or not the proportional hazards assumption had been violated.

Probability of Hospitalization Over Time

Figure 2: What the cumulative distribution function might look like if there were a violation of the proportional hazards assumption. Note the sharp increase in probability of hospitalization beginning right before the third year and lasting for approximately 1 year. After this one year period the top curve then levels off and becomes parallel with the bottom curve once again.
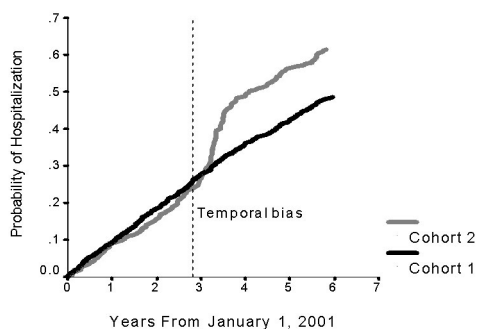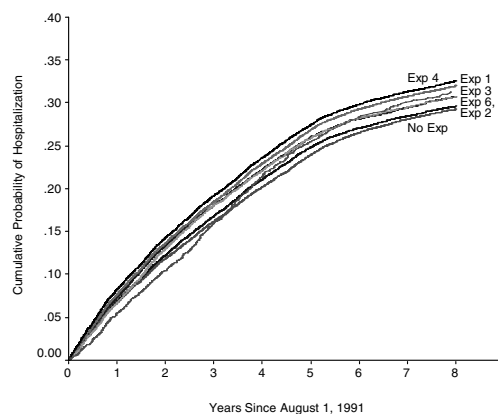
Figure 2.

Figure 3: The stratified cumulative distribution functions of postwar hospitalization for any cause by level of exposure to smoke from Kuwaiti oil well fires. There was no violation of the assumption of proportional hazards but there did happen to be an observed significant difference in

the hospitalization experiences between the 7 exposure levels.

Figure 3.

Computing the Generalized R^2

It may be helpful to know if SAS computes the $\mathbf{R^2}$ value and what it is for this particular model. If the researcher desires, the $\mathbf{R^2}$ value can be computed easily from the output of a regression, although it is not an option of PROC PHREG. Simply compute

$$R^2 = 1 - \exp(LR^2/n)$$

Where LR is the Likelihood-ratio chi-square statistic for testing the null hypothesis that all variables included in the model have coefficients of 0, and n is the number of observations. The researcher needs to take extreme caution when comparing the R^2 values of Cox regression models. Remember from linear regression analysis, R^2 can be artificially increased by simply adding explanatory variables to the regression model (ie; more variables does not equal a better model necessarily). Also, the above computation does not give the proportion of variance of the dependent variable explained by the independent variables as it would in linear regression, but does give a measure of how associated the independent variables are with the dependent variable.

Residual Analysis

A residual analysis is very important especially if the sample size is relatively small. Add the following after a model statement to output the Martingale and deviance residuals:

```
baseline out=survs survival=s
xbeta=xbet resmart=marting resdev=rdev;
```

Then a simple plot of the residuals against the linear predictor scores will give the researcher an idea of the fit or lack of fit of the model to individual observations.

```
PROC GPLOT data=survs;
  plot (marting rdev) * xbet / vref=0;
  symbol1 value=circle;
```

Multiple Hospitalization Analysis

The next question that we must answer is whether those exposed are hospitalized more often than those who were not exposed. That is, have we wasted information by only modeling time until first hospitalization? There are many methods which can be used for this type of an analysis, however, the researcher first must decide if they are modeling ordered or unordered outcomes. Since we are investigating multiple hospitalizations for the same broad diagnostic category or same unique diagnosis, we will focus on ordered outcomes. Three popular marginal regression models are the independent increment model (mutual independence of the observations within a subject), the WLW model (treating the ordered outcome data set with an unordered competing risks approach), and the conditional model (assuming an individual can't be at risk for outcome 2 if outcome 1 has not happened to the individual). After ordering the admission dates in chronological order, follow-up time periods for multiple hospitalization time modeling were calculated from the start of follow up, and subsequent hospital admission dates, until first or subsequent hospitalization, separation from service, or the end of follow-up, whichever occurred first.

Results

Complete exposure, demographic, and deployment data were available for approximately 400,000 veterans. Using the univariate comparisons for any cause hospitalizations, the following variables were selected for the subsequent model analyses: gender, age group, marital status, race/ethnicity, military occupational category, military pay grade, salary, service branch, pre-exposure period hospitalization, and oil well fire smoke exposure status. Home state was not shown to significantly affect the endpoints in any models and was not identified to be a potential confounder, so was therefore removed from further modeling. Salary and length of service were dropped from analyses due to colinearity with age.

The adjusted risk of any cause hospitalization for three of the six oil well fire smoke exposed groups of veterans was significantly less than the risk for the nonexposed group, although 2 of the 3 upper limits of the 95% confidence intervals were within .01 of including 1.0. The corresponding exposure group cumulative probability of hospitalization plots remained very stable and nearly parallel over the 8-year follow-up period (Figure 3). However, the Cox model did reveal some consistently better predictors of any cause hospitalization. These included female gender (RR=1.53; 95% CI=1.49, 1.57), pre exposure period hospitalization (RR=1.60; 95% CI=1.56, 1.63), enlisted pay grade (RR=1.50; 95% CI=1.46, 1.55), and health care workers (RR=1.32; 95% CI=1.27, 1.37). There were no noticeable temporal biases and therefore no need for time-dependent variables.

Summary

The Cox proportional hazard model's robust nature allows us to closely approximate the results for the correct parametric model when the true distribution is unknown or in question. Using the SAS® system procedure PROC PHREG, Cox's proportional hazards modeling was used to compare the hospitalization experiences of several levels of an exposure variable with those who were not exposed. Additionally, by extending these methods, investigation of multiple hospitalizations to the same individual was possible. The results of this study did not support the theory that Gulf War veterans were at increased risk of hospitalization due to exposure to oil-well-fire smoke.

Conclusions

The SAS® system's PROC PHREG with right censoring capabilities and the baseline option is a powerful tool for handling early departure causing incomplete data for subjects during the study period. It is also useful for computing data sets of survival function estimates, which can be used in a simple equation to produce estimates of cumulative probability of hospitalization over time. These graphs can be investigated for temporal bias of the covariates in the model. Additionally, the extension to multiple hospitalization helps to answer whether those who were exposed may be hospitalized more often or may have chronic hospitalizations at increased rates over those not exposed.

Bibliography

Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York: Springer-Verlag; 2000.

Hosmer JR. DW, Lemeshow S. *Applied Survival Analysis; Regression Modeling of Time to Event Data.* New York: John Wiley & Sons; 1999

Kleinbaum DG, *Survival Analysis: A self-Learning Text.* New York: Springer-Verlag; 1996

SAS Institute Inc., SAS/STAT® *User's Guide, Version 6, Fourth Edition, Volume 1,* Cary, NC: SAS Institute Inc., 1989. 943 pp.

SAS Institute Inc., SAS/STAT® *User's Guide, Version 6, Fourth Edition, Volume 2,* Cary, NC: SAS Institute Inc., 1989. 846 pp.

SAS Institute Inc. SAS/STAT® *Software: Changes and Enhancements through Release 6.11*. Cary, NC: SAS Institute Inc., 1996. 1104 pp.

Allison, Paul D., *Survival Analysis Using the SAS® system: A Practical Guide*, Cary, NC: SAS Institute Inc., 1995. 292 pp.

Smith TC, Heller JM, Hooper TI, Gackstetter GD, Gray GC. *Are Gulf War veterans experiencing illness due to exposure to smoke from Kuwaiti oil well fires? Examination of Department of Defense hospitalization data.* Amer J of Epidemiol; 2002; 155:908-17

Smith TC, Gray GC, Knoke JD. *Is systemic lupus erythematosus, amyotrophic lateral sclerosis, or fibromyalgia associated with Persian Gulf War service? An examination of Department of Defense hospitalization data.* Amer J of Epidemiol; 2000; Vol 151 No 11.

Gray GC, Smith TC, Knoke JD, Heller JM. *The postwar hospitalization experience among Gulf War veterans exposed to chemical munitions destruction at Khamisiyah, Iraq.* Amer J of Epidemiol; 1999; Vol 150 No 5.

Acknowledgments

About The Authors and Contact Information

Tyler has used SAS for 12 years including work as an undergraduate in Mathematics and Statistics, a graduate at the University of Kentucky Department of Statistics, and senior statistician with the DoD. Responsibilities include mathematical modeling, analysis, management, and documentation of large data, culminating in more than 20 peer reviewed journal manuscripts in major scientific journals. Invitations to speak include the International Biometrics Society Meetings, WUSS99-02, SUGI2000-01, SUGI2003, and the San Diego SAS users group 1999-00 fall meetings.

Tyler C. Smith, MS
Statistician, Henry Jackson Foundation
Department of Defense Center for Deployment Health Research, at the Naval Health Research Center, San Diego
(619) 553-7593
SMITH@nhrc.navy.mil

Besa Smith has used SAS for 6 years including work as a graduate at the San Diego State University Graduate School of Public Health in Biostatistics, and currently as a senior biostatistician with the DoD Center for Deployment Health Research. Her responsibilities include management of large military data sets, mathematical modeling and statistical analyses. She has been section chair and an invited speaker for multiple WUSS conferences and the San Diego area's SAS group meetings.

Besa Smith, MPH
Biostatistician, Henry Jackson Foundation
Department of Defense Center for Deployment Health Research, at the Naval Health Research Center, San Diego
(619) 553-7603
besa@nhrc.navy.mil