



DECOMPRESSION CHAMBER—Dr. Raj Chhikara, UHCL professor of mathematics and statistics, helps Dr. Laura A. Thompson, post-doctoral ISSO research fellow, emerge from a decompression chamber. She and Dr. Chhikara are seeking ways to resolve the problem that humans experience in space when gas bubbles form in their blood. Typically during extravehicular activity (EVA), gas bubbles form in venous blood as a result of decompression.

A Limited Failure Population Model for Onset of Grade IV Venous Gas Emboli

22-ISSO

Abstract—

The presence of gas bubbles in venous blood (venous gas emboli, or VGE) is associated with an increased risk of decompression sickness (DCS) in hypobaric environments. A high grade of VGE (e.g., Grade IV) can be a precursor to serious DCS. We propose a limited failure population model for time to onset of Grade IV venous gas emboli (VGE) in astronauts in hypobaric environments. We fit a lognormal mixture survival model to interval- and right-censored data on time to Grade IV VGE that accounts for the possibility of a subset of individuals who are immune to experiencing the event (sometimes called the “cured” individuals). The model contains random subject effects to account for correlations between repeated measurements on a single individual. Model assessments and cross-validation indicate that this limited failure population mixture model is an improvement over a model that does not account for the potential of a fraction of cured individuals. Predictions from the best fitted mixture model indicate that the actual process is reasonably approximated by a limited failure population model.

HUMANS EXPOSED TO HYPOBARIC ENVIRONMENTS, such as astronauts during extravehicular activity (EVA), typically experience the formation of gas bubbles in venous blood as a result of the decompression. The reduction in pressure from the space shuttle or a space station to a pressurized space suit can cause nitrogen gas normally dissolved in body fluids and tissue to escape from solution too rapidly, causing the formation of bubbles in the tissue and blood. Prebreathing oxygen helps to eliminate nitrogen from tissues, and reduces the number of circulating bubbles while at altitude and during EVA.

Prebreathe procedures are evaluated prior to their use on space missions by using altitude chambers that produce low pressure conditions. In these tests, the existence of venous gas emboli (VGE) are typically monitored using Doppler detection, and the extent of bubble signals is measured in grades using the Spencer scale.¹ Grades range from Grade 0—“absence of bubble signals in cardiac cycles,” to Grade IV—“bubble signals are detected continuously throughout the monitoring period, overriding the amplitude of cardiac motion and blood flow signals.” The lower the grade, the lower the apparent number of bubbles. Conkin et al.,² describes the connection between VGE and DCS in more detail. The effectiveness of each prebreathe procedure can be measured by the percentage of cases of high bubble grade (Grade IV or above) produced in simulated low pressure conditions of an altitude chamber.

Previous research on statistical modeling of the time to onset of DCS in high altitude conditions has dealt with right-censored data and a single survival distribution for the entire population under study. The nature of DCS onset is such that the risk rises with time, reaches a maximum, then declines.³ As such, a commonly

used hazard model for time-to-onset of DCS is the log-logistic or log-normal hazard functions. Conkin, et al.³ used a log-logistic model for the time to onset of DCS symptoms, with covariates related to tissue ratio, which is a measure of nitrogen decompression stress, and exercise at altitude. They also described survival models for modeling the onset of VGE detection from Doppler ultrasound. Koti, Chhikara, and Spears⁴ used log-logistic and log-normal models for modeling time to onset of DCS, using heavily right-censored data from NASA's Hypobaric Decompression Sickness Databank (HDSD). Later, Chhikara, Spears, and English⁵ used Cox's proportional hazards model on the same data set.

Researchers have suggested that under certain circumstances, some individuals will never suffer serious DCS (or its reported precursor, Grade IVVGE) no matter how long they remain at high altitude.³ Thus, it may not be ideal to use an ordinary survival model for modeling the time to onset of DCS or its related causes. Limited failure population (LFP) models in survival analysis have frequently been applied in the biostatistical and medical literature where it is known or assumed that some fraction of the population (the "cured fraction") will never experience the event under study. Maller and Zhou⁶ provide a complete history of these models.

The survivor function for a limited failure population is

$$S_{pop}(t) = \pi S(t) + 1 - \pi, \quad (1)$$

where $S(t)$ is the survivor function for individuals who will experience the event, and π represents the probability of eventually experiencing the event, given enough time. Nice features of model (1) are that $S(\infty) = 0$, as for an ordinary survival distribution, and $S_{pop}(\infty) = 1 - \pi$, so that $1 - \pi$ denotes the cured fraction of the population.

We fit an LFP model for predicting the time to onset of Grade IV venous gas bubbles. We employ a databank from NASA consisting of test results from volunteer subjects undergoing monitoring for VGE under hypobaric conditions. All of the observations were either interval- or right-censored, and some individuals performed more than one test, providing multiple records in the data set. Measurements on certain explanatory variables known to be associated with DCS were also recorded for each subject.

Experimental Data and Testing Procedure

We analyzed a set of data taken from NASA's Hypobaric Decompression Sickness Databank.⁷ It comprises monitoring test results from human volunteer subjects undergoing denitrogenation test procedures prior to being exposed to low pressure. Exposure records are from 453 males and 96 females who participated in a total of 28 different tests from 1983 to 1998. However, the actual number of individuals tested was 238 (177 male) because some subjects participated in more than one test. The highest number of test results contributed by a single individual was 13. The median was two. The recorded data did not provide information on the order in which the tests were taken.

Explanatory variables included experimental variables and physical characteristics of the subjects. These variables and their summary statistics are given in Table 1. The importance of these variables in decompression sickness is well-documented.⁸⁻¹¹ The first variable (TR360) is a measure of decompression stress. It is the ratio of the partial pressure of nitrogen to ambient pressure prior to ascent. A theoretical compartment with a half-time of 360 minutes was used to model nitrogen elimination. The greater this

Table 1. Explanatory Variables Measured on Subjects

	TR360	SEX	AGE	NOADYN	ALTTIME
Minimum:	0.94	0.00	20.00	0.00	2.00
Mean:	1.57	0.83	31.85	0.85	4.04
Median:	1.68	1.00	30.00	1.00	3.00
Maximum:	1.89	1.00	54.00	1.00	6.00
SD:	0.26	0.38	7.17	0.36	1.34

ratio is above 1.0, the more quickly one would expect high bubble grade to be detected. The variable, NOADYN was an indicator for whether the test subject was ambulatory (NOADYN = 1) or lower body adynamic (NOADYN = 0) during the session. The variable ALTTIME was the prescheduled time at altitude (two, three, four, five, or six hours). The variable SEX was coded male = 1 and female = 0. Eighty-three percent of the 548 test records were contributed by males. However, 74.4 percent of the 238 singular individuals were male.

Model Fitting and Selection

For the Grade IV VGE data, the chosen parametric form for the survival distribution for the "non-cured" individuals was lognormal. The lognormal model was used because the form of its hazard function is consistent with the potential for experiencing Grade IV bubbles in the population of interest. Other parametric forms, such as the Weibull and Log-logistic, were also tried. But, the fit of these parametric forms was worse than for the lognormal. Thus, the log time to onset of Grade IV bubbles is assumed to be normally distributed with a mean that is a linear combination of several covariates and a constant scale parameter. The survival function is

$$S(t|\beta, \sigma, \mathbf{x}) = 1 - \frac{\log(t) - \mu(\mathbf{x})}{\sigma}, \quad (2)$$

where Φ is the cumulative distribution function for the standard normal distribution,

$$\mu(\mathbf{x}) = \beta_0 + \sum_{k=1}^p x_k \beta_k,$$

σ is the standard deviation of $\log(t)$, and $x_k, k = 1, \dots, p$ are values of p explanatory variables.

For an initial fit, we included all the covariates listed in Table 1, with the exception of ALTTIME because it reflects the dependent variable, time to Grade IVVGE. Also, its inclusion resulted in considerable under-prediction in the model. For computational purposes, the variable AGE was standardized, and is subsequently denoted as S.AGE. The covariates were all included in linear form. Some investigation into nonlinear forms yielded no reason to suspect non-linear effects of the covariates on the response variable or interactions among them. In addition, pairwise scatterplots between pairs of covariates did not reveal any noticeable relationships.

Random Effects Model for Dependent Observations

A random effects model handles repeated measurements by incorporating random subject effects into the linear form

Investigative Team

UHCL PI: Raj S. Chhikara, Ph.D., Professor
Department of Mathematics and Statistics
University of Houston-Clear Lake
2700 Bay Area Blvd
Houston, TX 77058
Phone: (281) 283-3726; Fax: (281) 283-3708
E-mail: chhikara@cl.uh.edu

UH PDAF: Laura A. Thompson, Ph.D.
ISSO Post-doctoral Research Fellow
Department of Mathematics and Statistics
University of Houston-Clear Lake
2700 Bay Area Blvd
Houston, TX 77058
Phone: (281) 283-3882; Fax: (281) 283-3708
E-mail: thompsonla@cl.uh.edu

NASA-JSC Co-PI: Michael R. Powell, Ph.D.
Human Adaptations and Countermeasures Office
NASA Johnson Space Center / SD3
2101 NASA Road 1
Houston, TX 77058-3696
Phone: (281) 483-5413; Fax: 281-244-5734
E-mail: mrpowell@ems.jsc.nasa.gov

NASA-JSC Co-PI: Johnny Conkin, Ph.D.
National Space Biomedical Research Institute
One Baylor Plaza, NA425
Baylor College of Medicine
6621 Fannin Street
Houston, TX 77030-3498
Phone: (281) 244-1121; Fax: (281) 483-3058
E-mail: jconkin@ems.jsc.nasa.gov

Previous Post-Doctoral Fellow
Kallapa M. Koti, Ph.D., June 1996-August 1998
Research Statistician, FDA, Washington, D.C.

Graduate Student:
Thomas English, M.A., Master's degree in
Statistics, UHCL (1998-2000)

representing the conditional mean of the log time to the onset of Grade IV VGE. The mean for the j th log onset time for the i th subject is modeled as

$$\mu_{ij}(\mathbf{x}) = \beta_0 + \sum_{k=1}^p x_{kij} \beta_k + b_i. \quad (3)$$

In this representation, b_i is the random effect for subject i , which is a realization from a normal distribution with mean zero and variance σ_b^2 . Because of the normality of $\log t_{ij}$, the correlation between any pair of log times for a given subject is $\rho = \sigma_b^2 / (\sigma^2 + \sigma_b^2)$.

A random effects model also allows for heterogeneity in log times across subjects. As such, it can capture variability not modeled already in the linear combination,

$$\beta_0 + \sum_{k=1}^p x_{kij} \beta_k + b_i.$$

A large value of σ_b^2 , the variance of subject effects, relative to σ^2 , the average variance among a single subject's observations, may indicate relatively high variability among observations made between individuals or could indicate a proportionately large amount of unmodeled effects.

If we denote the likelihood by $L(\beta, \sigma, \sigma_b, \mathbf{b} = (b_1, \dots, b_n))$, then maximum likelihood estimates of the unknown parameters, $(\beta, \sigma_b, \sigma)$ are obtained by maximizing, with respect to $(\beta, \sigma_b, \sigma)$, the likelihood,

$$L(\beta, \sigma, \sigma_b) = \int_i L(\beta, \sigma, b_i) N(b_i | \sigma_b) d\mathbf{b}$$

integrated over the distribution of the random effects, \mathbf{b} , where $N(b_i | \sigma_b)$ is the zero-mean Gaussian probability density function with standard deviation, σ_b , evaluated at b_i . Gauss-Hermite quadrature can be used to obtain the integrations.¹²

Another alternative for obtaining generalized maximum likelihood estimates is by a Bayesian estimation with diffuse priors on the unknown parameters. Markov Chain Monte Carlo (MCMC) sampling can make estimation much easier than other methods, particularly when the likelihood function is complicated. In MCMC sampling, repeated samples from the joint posterior distribution follow a first-order Markov chain. The joint posterior distribution of the parameters is proportional to the likelihood function, as a function of the parameters, times the joint prior distribution on the parameters. Estimates of expectations with respect to the posterior distribution can be obtained as Monte Carlo estimates using the samples from the chain.

An advantage of MCMC sampling with random effects models is that the random effects can be sampled from their conditional distribution, given the data and current parameter samples, and implicitly averaged. This methodology avoids direct integration required to get the integrated likelihood. We combined numerical integration and MCMC sampling in order to estimate parameters.

Construction of the Limited Failure Population Model

For our limited failure population, the cure rate is the probability of never experiencing Grade IV bubbles. We assume that the cure rate depends on certain explanatory variables, so that for a given individual, the probability of experiencing Grade IV bubbles is a

function of physical characteristics of the individual and also possibly of aspects of the test. For the observed data, the cure rate is zero for individuals who underwent tests that had interval-censored recorded times, and it is nonnegative otherwise.

For the i th observation from the j th individual, define the indicator variable

$$z_{ij} = \begin{cases} 1 & \text{if the observation will eventually experience the event} \\ 0 & \text{otherwise} \end{cases}$$

We make an assumption in this paper that the z_{ij} s are independent for all i and j , conditional on certain modeled covariates as discussed in the next section. This assumption is valid to the extent that all relevant covariates have been included in the model. Thus, we assume that dependence between any two z_{ij} s occurs only through shared covariate values.

Under the assumption that censoring times are independent of event times, and that observations from different *individuals* are independent of one another, we construct the likelihood as follows: For an interval-censored observation with covariates, \mathbf{x}_{ij} , the contribution to the likelihood is

$$P(t_{0ij} < T_{ij} \mid t_{1ij}, z_{ij} = 1 \mid \mathbf{x}_{ij}) = P(z_{ij} = 1 \mid \mathbf{x}_{ij}) \left(S(t_{0ij} \mid \mathbf{x}_{ij}) - S(t_{1ij} \mid \mathbf{x}_{ij}) \right),$$

for known left and right endpoints t_{0ij} and t_{1ij} . Similarly, for a Type I right-censored observation, the contribution to the likelihood is

$$P(t_{0ij} < T_{ij}, z_{ij} = 1 \mid \mathbf{x}_{ij}) + P(t_{0ij} < T_{ij}, z_{ij} = 0 \mid \mathbf{x}_{ij}) \\ = P(z_{ij} = 1 \mid \mathbf{x}_{ij}) S(t_{0ij} \mid \mathbf{x}_{ij}) + P(z_{ij} = 0 \mid \mathbf{x}_{ij}),$$

The contribution by right-censored observations reflects the mixture aspect of the model. A right-censored observation has a probability, $P(z_{ij} = 1 \mid \mathbf{x}_{ij})$, of experiencing the event, and probability $P(z_{ij} = 0 \mid \mathbf{x}_{ij})$ of never experiencing the event. The probability, $P(z_{ij} = 0 \mid \mathbf{x}_{ij})$, is the cure rate mentioned above.

The full likelihood is then

$$L = \prod_{i=1}^n \prod_j L_{ij} N(b_i \mid 0, \sigma_b) \quad (4)$$

where,

$$L_{ij} = P(z_{ij} = 1 \mid \mathbf{x}_{ij}) \left(S(t_{0ij} \mid \mathbf{x}_{ij}) - S(t_{1ij} \mid \mathbf{x}_{ij}) \right)^{\delta_{ij}} \\ \left[P(z_{ij} = 1 \mid \mathbf{x}_{ij}) S(t_{0ij} \mid \mathbf{x}_{ij}) + P(z_{ij} = 0 \mid \mathbf{x}_{ij}) \right]^{1-\delta_{ij}}$$

and

$$\delta_{ij} = \begin{cases} 1 & \text{if observation } i \text{ from individual } j \text{ was interval-censored} \\ 0 & \text{otherwise} \end{cases}$$

We modeled the probability of eventually experiencing Grade IV bubbles on a given test as a logistic function of the explanatory variables,

Table 2. Approximate Maximum Likelihood Estimates for Fitted Models

	Non-Mixture Model	LFP Mixture Model
-LogLH	552.237	550.48
AIC	559.237	558.48
Location Parameter Estimates:		
β_0	9.932 (1.123)	6.848 (1.157)
β_1 (SEX)	-0.873 (0.346)	
β_2 (S.AGE)	-0.236 (0.122)	-0.149 (0.120)
β_3 (TR360)	-3.392 (0.566)	-2.646 (0.572)
β_4 (NOADYN)	-1.347 (0.349)	-1.169 (0.323)
σ (scale)	1.297 (0.118)	0.956 (0.140)
Mixture Parameter Estimates		
α_1 (SEX)		1.916 (0.525)
α_2 (S.AGE)		
α_3 (TR360)		
α_4 (NOADYN)		
α_5 (ALTTIME)		-0.335 (0.100)
σ_b (sd of random effects)	1.079 (0.161)	0.900 (0.151)
ρ corr(log t_{ij} , log t_{ij} '')	0.409 (0.151)	0.530 (0.158)

$$P(z_{ij} = 1 \mid \mathbf{x}_{ij}) = \frac{\exp \left(\alpha_0 + \sum_{k=1}^p \alpha_k x_{ijk} \right)}{1 + \exp \left(\alpha_0 + \sum_{k=1}^p \alpha_k x_{ijk} \right)} \quad (5)$$

where the α_k are parameters relating the covariates to the cure rate. The survival function, given \mathbf{x} , is $S(t \mid \mathbf{x})$, as given by (2). With these forms for the cure rate and survival distribution, L_{ij} in Eq. (4) can now be expressed in terms of parameters α and β .

We refer to the parameters, as *location parameters* and to the parameters, in the cure rate, as *mixture parameters*. If the location parameters are all distinct from the mixture parameters, then the maximum number of distinct parameters is 13. However, a model that includes all 13 parameters was found to perform relatively poorly in terms of validation. So, the results from this model are not presented. Instead, certain reduced forms of this model are considered, as given next.

Estimation of the Parameters from Random Effects Models

We now fit an LFP random effects model to the Grade IV VGE data, and contrast its fit with that of a non-mixture model. The models were fit using the structures given above, with the addition of a random effect added to the mean log time to onset for each individual. We accomplished estimation in two steps. First, we used MCMC sampling to sample realizations of from their joint distribution. The WinBUGS software version 1.3¹³ was used for the computations. Independent normal priors

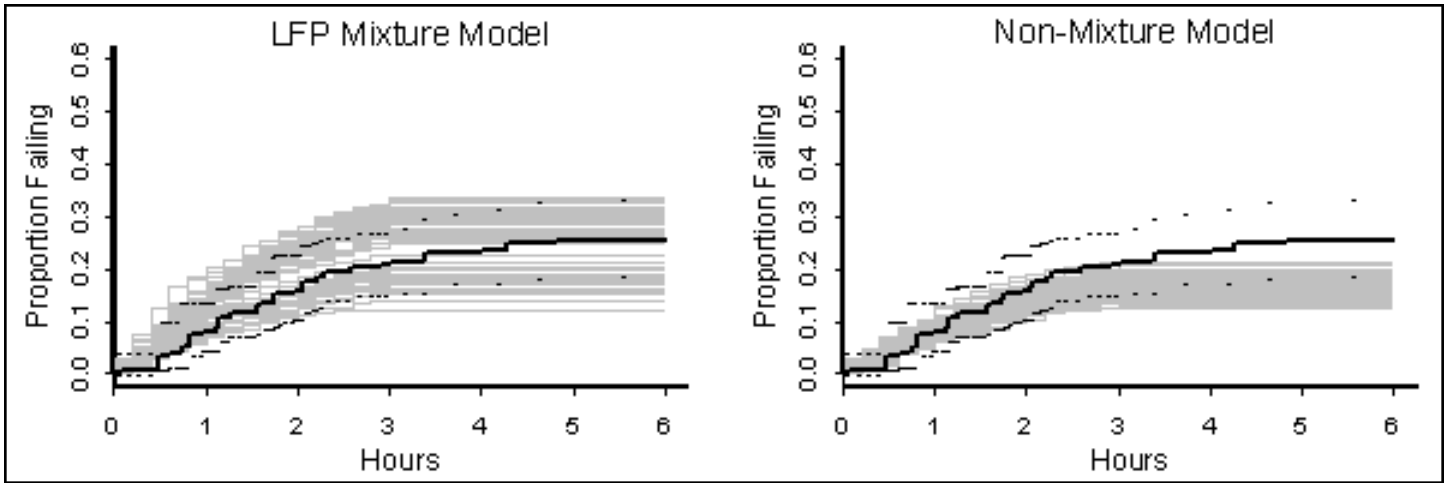


Figure 1. Goodness-of-fit graphs for two models examined: Non-mixture lognormal model and the LFP mixture lognormal model. The dark solid line represents the Turnball nonparametric probability estimate based on the observed data. The greyed lines represent the Turnball estimates based on 100 samples from the respective fitted model.

with mean 0 and variance 1,000 were chosen for the elements in $(\beta, \alpha, \sigma, \sigma_b, \mathbf{b})$. Results did not appear to be sensitive to the parametric form of the diffuse prior specification.

Table 2 gives the approximate maximum likelihood estimates of the coefficients for two models. The first model is a non-mixture model, and the second model is an LFP mixture model. Approximate standard errors in parentheses were obtained using the inverse of a finite difference approximation to the negative of the Hessian of the log likelihood evaluated at the point estimates.¹⁴ Researchers observed no evidence of nonidentifiability of parameters for either of the two models in Table 2.

The LFP model represented in Table 2 describes log survival time as dependent upon three of the explanatory variables: AGE, TR360 and NOADYN, and it models the cure rate as depending on two explanatory variables: SEX and ALTTIME (pre-scheduled time at altitude). Other configurations were tried, but none was uniformly better in terms of fit and accuracy of predictions.

Two sources of variability are estimated for each model given in Table 2. The parameter σ_b describes variability between individuals, and the parameter σ describes variability from measurement error not associated with between-individual variability. A relatively high estimate for σ_b may imply either large variability between individuals, or it could indicate unaccounted individuals for variability from covariates that were not modeled. The estimates of σ_b for both models are comparable to those of σ . One reason for this may be the small number of tests contributed by most subjects. For example, 42 percent of the 238 subjects contributed only one record, 30 percent contributed two records, and fewer than five percent contributed more than six records. Higher standard errors for σ_b versus σ also reflect the small number of tests by most subjects. The higher estimate for σ_b results in moderate estimates of correlation between two log responses on the same subject. These correlations appear in the last row of Table 2.

Goodness of Fit of Two Models

We describe a graphical goodness-of-fit approach that we have found useful in evaluating the model fits.

The graphical approach we used involves a parametric bootstrap procedure. We bootstrapped interval- and

right-censored observations from each of the fitted models with point estimates in Table 2, calculated the Turnball estimates of the probability of Grade IV bubbles for each bootstrapped sample and compared these nonparametric curves to the data-based Turnball estimate and 95 percent simultaneous confidence bands.¹⁵ The 95 percent simultaneous confidence bands may be considered to contain the likely probabilities of experiencing Grade IV VGE over time for our data or a similar set of data from the same population. To the extent that the confidence interval points cover curves from the parametric bootstrap, that model may be said to provide a relatively good fit to the data.

One hundred bootstrapped samples were drawn from each of the competing random effects models. After around 50 samples, the concentration of the resulting plotted Turnball estimates hardly changed such that 100 samples were deemed sufficient for graphical purposes. Figure 1 contains the resulting Turnball estimates of “failure” probabilities by time for the non-mixture model and the LFPmixture model described in Table 2. Each plot shows the 100 generated datasets for one of the models (the greyed lines), with the data-based curve based on the original data (the dark solid line) superimposed.

According to Fig. 1, the LFPmixture Model generates Turnball nonparametric maximum likelihood estimates that mostly fall within the 95 percent confidence intervals associated with the estimate computed from the original data. However, some of the generated curves that indicate a higher rate of Grade IV VGE up to three hours fall outside the intervals. The values of the coefficient estimates of the LFP model reveal that the model is consistent with either a high or low rate of Grade IV early on. All of the location estimates are negative, which means that any positive value for the associated covariate will reduce the mean (log) time to onset, increasing the presence of Grade IV VGE in the early hours. On the other hand, a higher value of ALTTIME lowers the probability of ever experiencing Grade IV VGE.

The plot containing the samples drawn from the non-mixture model shows that almost all the generated datasets fall below the original dataset after a time point of about two hours. In comparison with the LFPmixture model, the non-mixture model appears to support a lower occurrence rate in general. As all of the coefficient estimates in the location part have the same sign across the

two models, the lower event rate appears to be mainly attributed to the higher intercept value in the non-mixture model. Comparing the non-mixture model with the LFP model estimates in Table 2, we see that when a mixture portion is included in the latter, the intercept estimate drops by over three log hours. The higher intercept pulls the Grade IVVGE rate down, resulting in a model that appears to underpredict the rate of Grade IV VGE beginning just after two hours.

The generated data in both plots have no or few events after three hours. Although in the data set there is some incidence of Grade IVVGE after three hours, only eight such cases are reported. These occurrences are seen in the Turnbull estimate from the observed data, but neither the non-mixture nor LFP model captures these failures adequately. The eight cases in question were all male, seven with NOADYN = 1 (ambulatory), and six with ALTTIME = 6. A high ALTTIME reduces the predicted probability of eventually experiencing Grade IV VGE. Out of 50 generated sets of predicted event times for these eight cases, 41 sets have five or more cases with infinite times from the LFP model. Furthermore, for the non-mixture model, only two of the predicted times for the eight cases were less than six hours, the rest being right-censored. Neither model describes these cases well.

Based on the plots in Fig. 1, it seems more plausible that our data originated from a model similar to the LFP mixture model than from the non-mixture lognormal model because of the underprediction as reflected by the large gap in the plot for the non-mixture model that appears near the upper confidence bands, where there are no generated curves.

Predictive Accuracy of Models for Time to Grade IV VGE

To compare the predictive accuracy of the models, we used K -fold cross-validation. Briefly, cross-validation consists of leaving out K subsets of observations in turn, fitting the model to the remaining set of observations, then using the model to predict the values of the left-out subset. This process may be repeated M times to get M different groupings of K subsets. Cross-validation corrects for much of the positive bias that results from predicting observations when those observations were used to fit the model. We used an additional bias adjustment that corrects another “optimistic” bias; the procedure is called *K-fold Adjusted Cross-validation*.¹⁶

Predictions were obtained on log times to Grade IVVGE, conditional on the covariate values and censoring pattern. Error in prediction was assessed via a squared distance from the nearest interval endpoint if the predicted value fell outside of the recorded interval, and the error was zero otherwise. Predictions do not consider repeated observations conducted on the same individual.

The top half of Table 3 contains the results for two values of K ($K = 10, 548$), the number of subsets that partitioned the dataset. The choice $K = 10$ is recommended by the rule of thumb, $K = \min(10, n)$, where n is the size of the data set.¹⁶ The choice $K = 548$ entails leaving out each observation in turn. The table shows the results of the mean prediction error averaged over subsets, and averaged over 100 repetitions (for $K = 10$ only). The size of each subset is labeled as m . The two values for K give similar results and thus lead to the same conclusion: the LFP model predicts better than the non-mixture model on a new data set from the same population. Other choices of K led to very similar results as those seen in the table.

The bottom half of Table 3 contains results to get an idea of the sample size at which overfitting begins to display itself. We ran-

Table 3: Measure of Predictive Accuracy of Models (K-fold Cross-Validated Prediction Error)

	Non-mixture Model	LFP Model (Mixture)
Number of Groups (Size of Left-out group)		
$K = 10$ ($m = 55$)	0.900	0.348
$K = 548$ ($m = 1$)	0.902	0.347
Sample size used to fit model:		
$n = 25$	0.840	0.371
$n = 35$	0.875	0.370
$n = 50$	0.809	0.353

domly selected a small number of observations ($n = 25, 35$ or 50) from the data set and fit each model using that set. A sample size of 50 might be considered sufficient, whereas a sample size of 25 may not be sufficient. Then we computed the averaged prediction error using the left-out cases. We repeated this procedure 100 times for each model, then took the median of the resulting prediction errors. (The median was used instead of the mean because of outlying prediction errors). The median is displayed in the table. As expected, prediction generally gets better for both models as the sample size increases, and there does not appear to be problems with overfitting for either model. But, notably, the non-mixture model predicts poorly compared to the LFP model at all sample sizes.

Further Evaluation of the LFP Model

Based on the estimates from the LFP Model in Table 2, the estimated probability of remaining free from Grade IV VGE by a given time, t , is

$$\begin{aligned} \hat{P}(T > t) &= \hat{P}(z = 1)\hat{S}(t) + \hat{P}(z = 0)(1) \\ &= \hat{P}(z = 1|\mathbf{x}) \left[1 - \frac{\log t - \hat{\mu}(\mathbf{x})}{0.956} \right] + \hat{P}(z = 0|\mathbf{x}) \end{aligned} \quad (6)$$

where

$$\hat{\mu}(\mathbf{x}) = 6.848 - 0.149 S.AGE - 2.646 TR360 - 1.169 NOADYN,$$

and

$$\hat{P}(z = 1|\mathbf{x}) = \frac{\exp(1.916 SEX - 0.335 ALTTIME)}{1 + \exp(1.916 SEX - 0.335 ALTTIME)}.$$

The probability of experiencing Grade IV VGE by a given time, t , is the complement of Equation (6), i.e.,

$$\hat{P}(T \leq t) = \hat{P}(z = 1)\left(1 - \hat{S}(t)\right) = \hat{P}(z = 1|\mathbf{x}) \frac{\log t - \hat{\mu}(\mathbf{x})}{0.956}. \quad (7)$$

For calculating predictions, the random effect term, b_i , that appears in (3), is not used. Plots of Equation (7) are

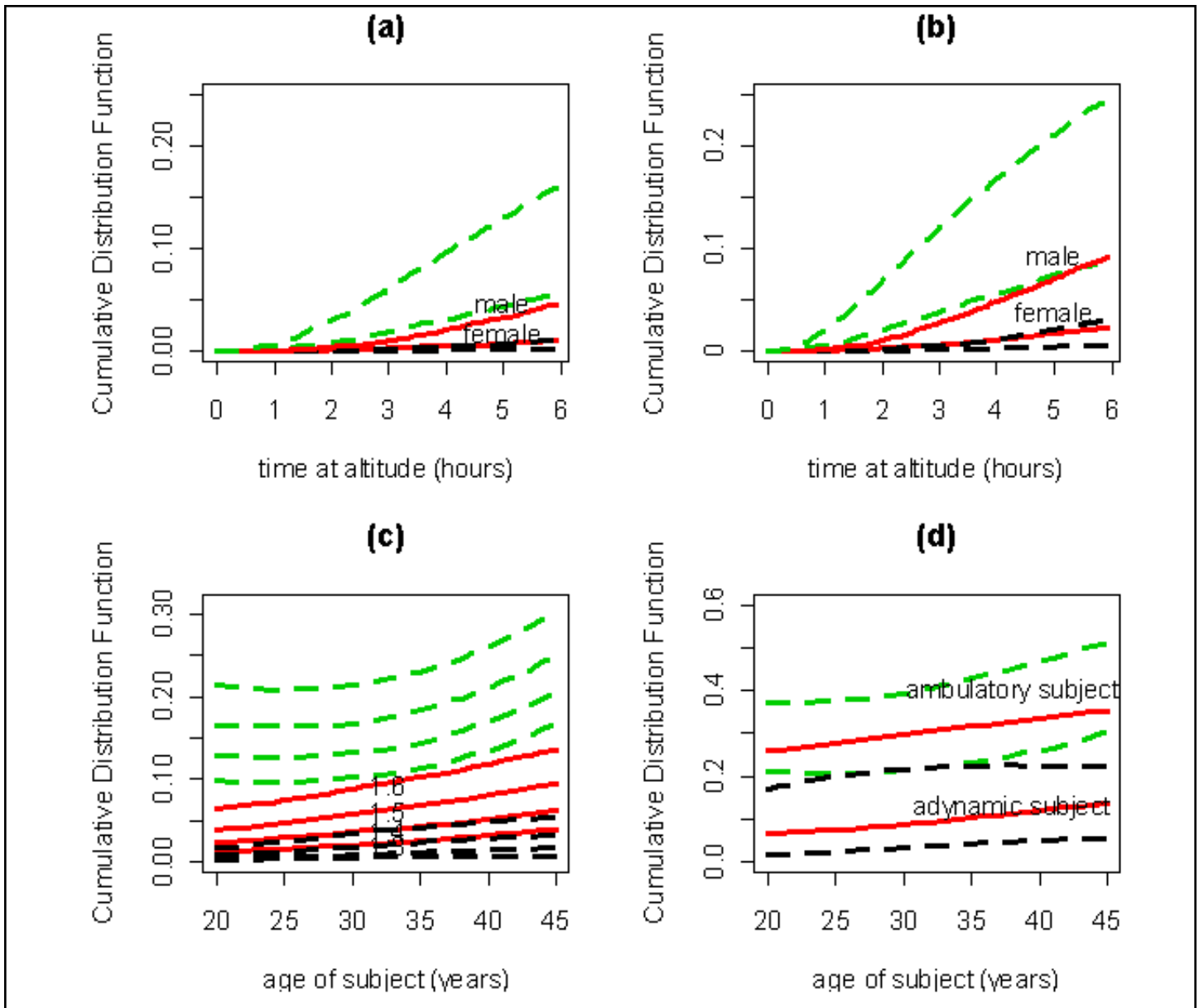


Figure 2. Estimated probabilities of Grade IV VGE, with 95 percent pointwise confidence intervals, as predicted by the LFP Model. Green dashed lines are upper points of the intervals; black dashed lines are the lower points. The solid red lines are point estimates. (See text for further explanation of individual panels.)

shown in Fig. 2 for several values of the explanatory variables, along with 95 percent pointwise confidence intervals obtained assuming asymptotic normality of the logit of the respective estimated probabilities, and then back-transforming the resulting interval to get a confidence interval on the probability itself. The confidence intervals are not symmetric, but are always contained within 0 and 1.¹⁵ The upper points of the intervals are represented by dashed green lines, and the lower points are represented by dashed black lines. For all four plots, the higher green dashed lines and the higher black dashed lines correspond to the higher solid lines [red on the web site], representing the point estimates.

Figures 2a and 2b compare probabilities by time at altitude for males and females ages 25 and 45, respectively, with a tissue ratio of 1.5, who are lower body adynamic, and who plan to spend a total of six hours at altitude (i.e., ALT-TIME = 6). Males have higher predicted probabilities of Grade IV bubbles

than do females at almost all hours at altitude. Figures 2c and 2d compare predicted probabilities by age for different tissue ratios and adynamic versus ambulatory individuals, respectively. For both graphs, only males are considered, staying a total of six hours at altitude. For Fig. 2c, the individual is considered to be lower body adynamic; for Fig. 2d, the individual has a tissue ratio of 1.5. Both of the plots give reasonable results in the direction expected. First, probability estimates increase with age. Also, at any age, the higher tissue ratios have higher estimated probabilities, as do ambulatory individuals, with tissue ratio held constant.

Analogous predictions from the non-mixture model (figures omitted) showed mean predictions that were in the same direction as those from the LFP model, except that confidence intervals were almost always slightly narrower for the non-mixture model. This was expected due to the smaller standard errors for the non-mixture model. Also, differences in mean predictions across tissue ratio and adynamia status became increasingly larger over age

for the non-mixture model. For the LFP model, differences in mean predictions did not increase as much over age, especially across adynamia status. Thus, in comparison with a non-mixture model, the LFP model predicted on average, similar likelihoods of Grade IVVGE, but that differences in predictions across tissue ratio and adynamia status did not increase over age. The empirical proportions of Grade IV incidence by age categories and adynamia status did not indicate an increasing gap in Grade IV incidence across adynamia status, as age increased. This corroborates the relative accuracy of predictions made by the LFP model, but not by the non-mixture model.

Conclusions

We have shown that a limited failure population model might suitably describe the onset of Grade IVVGE in volunteer subjects undergoing testing in altitude chambers. The LFP mixture model of Table 2 seems to be an appropriate candidate.

However, more flexible models might be considered. For example, the generalized-F distribution contains the log normal distribution as a special case. To apply this distribution to interval-censored data, however, may be practically quite complicated. A Bayesian approach might make the model fitting easier. A Cox model with frailties is another alternative model. At least one study⁵ has shown the superiority of the Cox model applied to onset of DCS, as opposed to parametric models such as the log normal or log-logistic. Whether this superiority remains when there is also interval censoring and dependent observations will be addressed in the future. We should note that a completely non-parametric model may not be possible due to having only a few observed occurrences of Grade IVVGE in the right tail of the survival distribution.

Finally, the data set we analyzed in this study contained many right-censored cases. We did not specifically address heavy censoring in our estimation procedures. As heavy right-censoring is common in data obtained from altitude chamber tests,²⁵ this is a potential concern and limitation in our analysis.

References

- ¹M. P. Spencer. "Decompression Limits for Compressed Air Determined by Ultrasonically Detected Blood Bubbles," *J. Applied Physiology* 40 (1976): 229-35.
- ²J. Conkin, M. R. Powell, P. D. Foster, and J. M. Waligora. "Information About Venous Gas Emboli Improves Prediction of Hypobaric Decompression Sickness," *Aviation, Space and Environ. Medicine* 69 (1998): 8-16.
- ³J. Conkin, P. D. Foster, M. R. Powell, and J. M. Waligora. "Relationship of the Time Course of Venous Gas Bubbles to Altitude Decompression Illness," *Undersea and Hyperbaric Medicine* 23.3 (1996): 141-51.
- ⁴K. Koti, R. S. Chhikara, and F. M. Spears. "Cox Regression Modeling and Analysis of Decompression Sickness Data," *1997-1998 Annual Report, ISSO, University of Houston* (1998): 20-29.
- ⁵R. S. Chhikara, F. M. Spears, and T. English. "Modeling and Analysis for Risk Assessment of Decompression Sickness," *1999-2000 Annual Report, ISSO, University of Houston* (2000): 66-67.
- ⁶R. Maller and X. Zhou. *Survival Analysis with Long-Term Survivors*. New York: Wiley, 1996.
- ⁷J. Conkin, S. Bedahl, and H. Van Liew. "A Computerized Databank of Decompression Sickness Incidence in Altitude

Chambers," *Aviation, Space and Environ. Medicine* 63 (1992): 819-24.

⁸D. Carturan, A. Boussuges, H. Burnet, J. Fondaral, P. Vanuxem, B. Gardette. "Circulating Venous Bubbles in Recreational Diving: Relationship with Age, Weight, Maximum Oxygen Uptake, and Body Fat Percentage," *International J. of Sports Medicine* 20 (1999): 410-14.

⁹J. Conkin and M. R. Powell. "Lower Body Adynamia as a Factor to Reduce the Risk of Hypobaric Decompression Sickness," *Aviation, Space and Environ. Medicine* 72 (2001): 202-14.

¹⁰Z. M. Sulaiman, A. A. Pilmanis, and R. B. O'Connor. "Relationship Between Age and Susceptibility to Altitude Decompression Sickness," *Aviation, Space and Environ. Medicine* 68 (1997): 695-98.

¹¹J. T. Webb, A. A. Pilmanis, K. M. Krause, and N. Kannan. "Gender and Altitude-Induced Decompression Sickness Susceptibility," 70th Annual Scientific Meeting of the Aerospace Medical Society, 1999.

¹²J. C. Naylor and A. F. M. Smith. "Applications of a Method for the Efficient Computation of Posterior Distributions," *Applied Statistics* 31 (1982): 214-25.

¹³D. J. Spiegelhalter, A. Thomas, and N. Best. *WinBUGS 1.3 User Manual*, <http://www.mrc-bsu.cam.ac.uk/bugs>, 2000.

¹⁴M. Tanner. *Tools for Statistical Inference*. Springer-Verlag, NY, 1996. 74

¹⁵W. Meeker and L. Escobar. *Statistical Methods for Reliability Data*. New York: Wiley, 1998.

¹⁶A. Davison and D. Hinkley. *Bootstrap Methods and their Application*. Cambridge: Cambridge University Press, 1997.

Publications

Chhikara, R. S. and L. A. Thompson. "Statistical Modeling of Survival Data on Decompression Sickness," *Bulletin of the International Statistical Institute, Contributed Papers Book 3* (August, 2001): 43-44.

Presentations

Chhikara, R. S. and L. A. Thompson. "Statistical Modeling of Survival Data on Decompression Sickness," 53rd Session of the Int'l Statistical Institute, Seoul, Korea, Aug. 22-29, 2001.