Modeling Grade IV Gas Emboli Using a Limited Failure Population Model with Random Effects

Laura A. Thompson, Ph.D., M.A
University of Houston – Clear Lake
School of Natural and Applied Sciences

Johnny Conkin, Ph.D., M.S.
National Space Biomedical Research Institute

Raj S. Chhikara, Ph.D., M.S.
University of Houston – Clear Lake
School of Natural and Applied Sciences

Michael R. Powell, Ph.D., M.S.
Human Adaptations and Countermeasures Office
NASA/Johnson Space Center

THE NASA STI PROGRAM OFFICE . . . IN PROFILE

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the lead center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA's counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or cosponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and mission, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results . . . even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at *http://www.sti.nasa.gov*

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA Access Help Desk at (301) 621-0134

- Telephone the NASA Access Help Desk at (301) 621-0390

- Write to:
NASA Access Help Desk
NASA Center for AeroSpace Information
7121 Standard
Hanover, MD 21076-1320

Modeling Grade IV Gas Emboli Using a Limited Failure Population Model with Random Effects

Laura A. Thompson, Ph.D., M.A
University of Houston – Clear Lake
School of Natural and Applied Sciences

Johnny Conkin, Ph.D., M.S.
National Space Biomedical Research Institute

Raj S. Chhikara, Ph.D., M.S.
University of Houston – Clear Lake
School of Natural and Applied Sciences

Michael R. Powell, Ph.D., M.S.
Human Adaptations and Countermeasures Office
NASA/Johnson Space Center

Acknowledgements

Available from:

NASA Center for AeroSpace Information
7121 Standard
Hanover, MD 21076-1320

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161

This report is also available in electronic form at http://techreports.larc.nasa.gov/cgi-bin/NTRS

Contents

Tables

Figures

Acronyms and Nomenclature

AIC   Aikaike's Information Criterion

ALTTIME pre-scheduled time at altitude

CDF   cumulative distribution function

DCS   decompression sickness

EVA   extravehicular activity

GIV   Grade IV

HDSD   hypobaric decompression sickness databank

ISSO   Institute for Space Systems Operations

LBA   lower body adynamia
LFP   limited failure population
LogLH   log likelihood

MCMC   Markov chain Monte Carlo
MLE   maximum likelihood estimate

NOADYN adynamia status

PFO   patent foramen ovale

TR360   tissue ratio in the 360-minute half-time compartment

VGE   venous gas emboli

Abstract

Venous gas emboli (VGE) (gas bubbles in venous blood) are associated with an increased risk of decompression sickness (DCS) in hypobaric environments. A high grade of VGE can be a precursor to serious DCS. In this paper, we model time to Grade IV VGE considering a subset of individuals assumed to be immune from experiencing VGE. Our data contain monitoring test results from subjects undergoing up to 13 denitrogenation test procedures prior to exposure to a hypobaric environment. The onset time of Grade IV VGE is recorded as contained within certain time intervals. We fit a parametric (lognormal) mixture survival model to the interval- and right-censored data to account for the possibility of a subset of "cured" individuals who are immune to the event. Our 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. We also evaluated some alternative mixture models. Predictions from the best fitted mixture model indicate that the actual process is reasonably approximated by a limited failure population model.

1. Introduction

Humans exposed to hypobaric environments, such as astronauts performing an extravehicular activity (EVA), typically experience the formation of gas bubbles in venous blood as a result of decompression. The reduction in pressure from the space shuttle or a space station to a pressurized space suit can cause nitrogen gas, which is normally dissolved in body fluids and tissue, to escape from solution too rapidly, resulting in the formation of bubbles in tissue and blood. The movement of gas bubbles into venous blood is called venous gas emboli (VGE). In most cases, the lungs filter out the bubbles prior to any danger of their being passed to arterial blood, where they may block the pulmonary artery or lodge in the brain. In some cases, however, especially those cases where an individual has a patent foramen ovale (PFO), there is the chance of bubbles traversing to arterial blood. Briefly, a PFO is an opening in the heart between the right and left arterial chambers that allows the passage of bubbles into arterial circulation and critical tissues. The presence of these circulating gas bubbles in the blood stream could contribute to serious hypobaric decompression sickness (DCS). For a further description of the process, see Bove (1998).

Types of DCS are distinguished by seriousness. The symptoms of Type I DCS are milder and include joint pain, or "the bends," in the elbows, knees, or shoulders. The symptoms of Type II DCS ("serious" DCS) are neurological (e.g., memory loss, unconsciousness, stroke), cardiac, or pulmonary (e.g., deep chest pain caused by bubbles blocking the pulmonary artery); Type II DCS can be fatal if it is not treated immediately. This is particularly problematic for astronauts performing an EVA because of the lack of quick rescue capability. To reduce the risk of the occurrence of VGE, astronauts typically breathe 100% oxygen prior to EVAs. Prebreathing oxygen helps to eliminate nitrogen from tissues, and reduces the number of circulating bubbles while at altitude and during the EVA.

Prebreathe procedures are evaluated in altitude chambers prior to their use on spaceflight missions by tests that produce low-pressure conditions. In these tests, the existence of VGE is typically monitored using Doppler detection, and the extent of bubble signals is measured in grades using the Spencer scale (Spencer, 1976). Spencer scale grades range from Grade 0 (i.e., the "absence of bubble signals in cardiac cycles" to Grade IV (i.e., "bubble signals are detected continuously throughout the monitoring period, overriding the amplitude of cardiac motion and blood flow signals"). The lower the grade in the scale, the lower the apparent number of bubbles. Conkin et al. (1998) describe 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 (e.g., in an altitude chamber). Grade IV bubbles are of concern because they are associated with an increased risk of Type II DCS, especially for individuals with a PFO. So, the choice of a particular denitrogenation procedure for use prior to an EVA will depend on the effectiveness of the procedure in avoiding Grade IV bubbles and, thus, potentially serious DCS.

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 parametric form of the survival distribution for the entire population under study. The nature of DCS onset is such that risk rises with time, reaches a maximum, then declines (Conkin et al., 1996). 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. (1996) used a log-logistic model for the time to onset of DCS symptoms, with covariates related to tissue ratio (a measure of nitrogen decompression stress) and exercise at altitude. The authors also described survival models for modeling the onset of VGE detection from Doppler ultrasound. Kumar and Powell (1994) modeled log time to onset of DCS as a parametric function of the explanatory variables, tissue ratio, and presence of circulating microbubbles in venous blood. Kannan and Raychaudhuri (1998) fit both parametric (log-logistic) and semi-parametric (Cox) models to DCS data from individuals in hypobaric chamber tests. For each individual, these authors

considered the time to onset of DCS symptoms as well as several explanatory variables such as exercise at altitude, amount of time spent prebreathing oxygen prior to exposure, maximum bubble grade using the Spencer scale, and pressure at altitude. In addition, Koti et al. (1998) used log-logistic and log-normal models for modeling time to onset of DCS, using heavily right-censored data from the NASA hypobaric decompression sickness databank (HDSD). Chhikara et al. (2000) later used Cox's proportional hazards model on the same dataset.

Limited Failure Population Models

Conkin et al. (1996) suggested that under certain circumstances, there are some individuals who will never get Grade IV VGE no matter how long they remain at high altitude. 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 often 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 Zhou (1996) provide a complete history of these models.

The application of a survival model for an LFP typically involves a mixture model, which can be a standard mixture of several different models or a nonstandard mixture model where one component is degenerate. For example, in clinical settings where the endpoint under study is death due to disease, the survival function can be expressed as a mixture model with different component survival distributions for death due to different causes, including the one under study, as well as "normal mortality". The resulting model is a standard finite mixture model in which mixture components may or may not depend on covariates, and the survival distributions for each competing risk may or may not be completely parametric. Examples of this type of LFP mixture model appear in Gordon (1990), Kuk and Chen (1992), and Larson and Dinse (1985).

In situations where the endpoint under study occurs by only one known means (e.g., paralysis due to excess radiation or onset of hypoxia due to low atmospheric pressure) or where competing causes of the endpoint are not observable (e.g., relapse of cancer caused by overgrowth of any number of cancer cells), the resulting mixture representation of "cured" and "non-cured" individuals becomes degenerate in the cured component. That is, the survival distribution becomes degenerate at 1 for cured individuals, with infinite failure times. The mixture model originally proposed by Berkson and Gage (1952) models the population survival function as a mixture of a standard survival distribution and a degenerate survival function with point mass at 1. Thus, the survivor function for the entire population is

$$S_{pop}(t) = p\,S(t) + 1 - p \qquad\qquad (1.1)$$

where $S(t)$ is the survivor function for individuals who will experience the event, and p represents the probability of eventually experiencing the event, given enough time. A nice feature of model (1.1) is that because $S(\infty) = 0$, as for an ordinary survival distribution, $S_{pop}(\infty) = 1 - p$, so that $1 - p$ denotes the "cure rate" of the population.

Examples of (1.1) as an LFP mixture model are found in Berkson and Gage (1952), Farewell (1982), and Taylor (1995). Farewell (1982) describes a logistic-Weibull mixture applied to fish toxicology data, where the Weibull model is used for the survival distribution and a logistic model is used for the cure probability. Taylor (1995) and Kuk and Chen (1992) describe semi-parametric versions of this model using, respectively, a Kaplan-Meier estimator for the survival distribution and a proportional hazards structure with an unspecified baseline hazard function.

Although the form of model (1.1) is simple, it can be attractive in certain circumstances, such as clustering of observations and the presence of interval censoring. Because of its simplicity, the use of model (1.1) with a parametric survival distribution can make both estimation and assessment of goodness-of-fit and predictive validation easier. Model (1.1) is also easy to interpret, and its mixture structure may already be familiar to researchers. Thus, it serves as a nice start from which a more flexible model can be constructed. The aim of our paper is to apply this relatively simple model to a dataset with interval- and right-censored measurements, with the addition of random effects to handle clustering of observations.

We fit an LFP model for predicting the time to onset of Grade IV venous gas bubbles. To achieve this, we use a NASA databank 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 of the individuals tested performed more than one test, thereby providing multiple records in the dataset. Measurements on certain explanatory variables known to be associated with DCS were also recorded for each subject.

Since we assume a priori that individuals who are immune to Grade IV bubbles are present in the population under study and also in the dataset at hand, there will be no formal test for the presence of immunes. We base our decision on examination of the physiological circumstances surrounding DCS, and on the fact that our dataset may not have sufficient follow-up to be able to test formally for the presence of immunes. Maller and Zhou (1996) discuss formal testing and its limitations.

The remainder of this paper proceeds as follows. Section 2 gives a description of the data, testing procedure, and explanatory variables measured. Section 3 explores nonparametric survival distributions. Sections 4 and 5 describe the model fitting. Section 6 provides an assessment of goodness of fit. Section 7 discusses predictive validation. Section 8 addresses predictions for the fitted models. And, Sections 9 and 10 include discussion and extensions.

2. Description of the Data and Testing Procedure

NASA's HDSD (Conkin et al., 1992) contains monitoring test results from human volunteer subjects undergoing denitrogenation test procedures prior to exposure to low pressure. The exposure records are from 453 males and 96 females who participated in a total of 28 different test procedures from 1983 to 1998. However, because some subjects participated in more than one test procedure, the number of *individuals* tested was 238 (of which 177 were male). Each test involved one decompression, and is described generally in the next subsection. The highest number of test results contributed by a single individual was 13. The median number of test results was two. The recorded data did not provide information on the order in which the tests were taken.

2.1 Testing Sessions and Recorded Data

Each testing session was scheduled to last anywhere between two and six hours; the median was three hours. Subjects were tested in groups of one or more individuals. A subject's group designation was not included in the recorded data, although multiple records by a single subject were indicated. There were two major phases during a typical testing session. In the first phase, the subject prebreathed 100% oxygen at site pressure while sitting. In the second phase, the subject was brought to "altitude" where he/she was monitored for Doppler-detectable bubbles. During this second phase, the subject performed a variety of repetitive exercises, mostly while standing. He or she also walked to as many as three exercise stations and to a bubble monitoring station. Bubble monitoring was scheduled to begin at approximately every 16 minutes while a group of subjects was at altitude. However, according to the data, monitoring sometimes

began after intervals of over a half hour. Each monitoring session was scheduled to last for four minutes, where the subject flexed the limbs in sequence to improve bubble detection.

If Grade IV bubbles were detected within a monitoring interval, the recorded datum for the subject was the interval between the end of the last monitoring period and the beginning of the interval in which Grade IV bubbles were detected, no matter the point in the monitoring interval in which Grade IV bubbles were detected. Thus, these cases were interval-censored. If Grade IV bubbles were not detected during the entire test session, the recorded onset time was right-censored at the end of the session. These right-censored observations were considered to be Type I right-censored (Lawless, 1982). If the test was stopped for any reason prior to the end of the session, the observation was right-censored at that time. Observations, which corresponded to tests stopped prior to the prescheduled time at altitude, were considered *randomly* right-censored. A test was never stopped during a monitoring interval.

Of the 549 records, 124 were interval-censored (i.e., Grade IV bubbles were detected), leaving over 75% of the cases Type I right-censored and 2% random right-censored. Due to equipment failure, the interval for one observation lasted about 104 minutes. This observation was discarded, leaving 548 records from the 238 individuals tested.

2.2 Explanatory Variables

Explanatory variables included experimental variables and physical characteristics of the subjects. The variables and their summary statistics are given in Table 1. The importance of these variables in DCS is well-documented (Carturan et al., 1999; Conkin and Powell, 2001; Sulaiman et al., 1997; Webb et al., 1999). The first variable (TR360), a measure of decompression stress, is the ratio of the partial pressure of nitrogen at altitude to ambient pressure prior to ascent. A theoretical compartment with a half-time of 360 minutes was used to model nitrogen elimination and obtain the value of the explanatory variable, TR360 (see Conkin et al., 1996, for information on the development of TR360 and its recording in the HDSD). The greater this ratio is above 1.0, the more quickly we would expect to detect high bubble grades. The variable NOADYN indicated whether the test subject was ambulatory (NOADYN = 1) or lower body adynamic (NOADYN = 0) during the session. The variable SEX was coded male = 1 and female = 0. The mean of SEX shown in Table 1 is slightly misleading because although 83% of the 548 test records were contributed by males, only 74.4% of the 238 *individuals* were male.

Table 1: Explanatory Variables Measured on Each Case

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

In what follows, we will refer to the entire dataset of interval- and right-censored observations, along with measured explanatory variables as the Grade IV VGE data.

2.3 Data Characteristics

To provide initial insight into the characteristics of the data and to facilitate further discussion, we constructed several cross-tabulations of explanatory variables by proportion of Grade IV VGE

4

occurrence. TR360 was categorized into quintiles, and AGE was categorized into groups: 19 – +30, 30 – +40, 40 – +60. The categories for AGE were chosen to divide the age group into younger, middle, and older age categories. Three categories were chosen to have enough data fall into each category. Table 2a shows the proportions for TR360 and NOADYN status, Table 2b shows the proportions for SEX and NOADYN status, and Tables 2c and 2d show the proportions for SEX by AGE, and NOADYN status by AGE, respectively. No formal hypotheses were tested using the tabulations.

Not surprisingly, the proportion of Grade IV cases in the dataset increases with higher categories of TR360. According to Table 2a, however, when the sample is categorized by NOADYN status, there is an increasing incidence of Grade IV VGE by TR360 for ambulatory subjects, but there does not appear to be a similar trend for the adynamic subjects. Notice, too, that the proportion of Grade IV cases for Adynamic subjects is higher than the proportion for Ambulatory subjects in the TR360 range [0.94, 1.35). For all other TR360 categories, the proportion for Ambulatory subjects is higher than for Adynamic subjects. This may be indicative of a slight interaction between TR360 and NOADYN in their influence on Grade IV occurrence. However, some of the sample sizes in the Adynamic categories may be too small to provide reliable proportions of Grade IV VGE.

Table 2a: Cross-Tabulation of TR360, NOADYN, and Grade IV VGE

	TR360[*]			
	0.94+ – 1.35	1.35+ – 1.68	1.68+ – 1.77	1.77+ – 1.89
	NOADYN=0 (Adynamic)			
No GIV VGE	16/19 = 0.84	14/14 = 1.0	22/23 = 0.96	24/28 = 0.86
GIV VGE	3/19 = 0.16	0/14 = 0.0	1/23 = 0.04	4/28 = 0.14
	NOADYN=1 (Ambulatory)			
No GIV VGE	107/113 = 0.95	91/124 = 0.73	126/191 = 0.66	13/24 = 0.54
GIV VGE	6/113 = 0.05	33/124 = 0.27	65/191 = 0.34	11/24 = 0.46

[*]The column TR360 ≤ 0.94 was not included because only 12 observations had TR360 in this range, with no cases of Grade IV (GIV) VGE.

Table 2b shows that the incidence of Grade IV VGE is higher among Ambulatory than Adynamic subjects, and this direction does not change with sex. Males have a higher percentage of Grade IV VGE cases than females have.

Table 2b: Cross-Tabulation of NOADYN, SEX, and Grade IV VGE

	NOADYN	
	Adynamic	Ambulatory
	Female	
No GIV VGE	25/27 = 0.93	61/69 = 0.88
GIV VGE	2/27 = 0.07	8/69 = 0.12
	Male	
No GIV VGE	51/57 = 0.90	288/395 = 0.73
GIV VGE	6/57 = 0.10	107/395 = 0.27

Table 2c shows that the proportion of Grade IV VGE cases increases with AGE group regardless of sex – with the exception of the incidence for females over age 40. This exception might also be considered as evidence of a slight interaction between SEX and AGE, although the sample size in this category may be too small for the interaction to be reliable.

Table 2c: Cross-Tabulation of SEX, AGE, and Grade IV VGE

	AGE		
	19+ – 30	30+ – 40	40+ – 60
	SEX=0 (Female)		
No GIV VGE	34/35 = 0.97	39/47 = 0.83	13/14 = 0.93
GIV VGE	1/35 = 0.03	8/47 = 0.17	1/14 = 0.07
	SEX=1 (Male)		
No GIV VGE	206/260 = 0.79	89/124 = 0.72	44/68 = 0.65
GIV VGE	54/260 = 0.21	35/124 = 0.28	24/68 = 0.35

Finally, Table 2d tabulates Grade IV incidence by AGE and NOADYN status. As before, we see an increasing incidence of Grade IV VGE with AGE and with ambulatory individuals. However, this table also shows that the difference in Grade IV incidence across NOADYN status (Adynamic vs. Ambulatory) remains roughly constant across AGE.

Table 2d: Cross-Tabulation of NOADYN, AGE, and Grade IV VGE

	AGE		
	19+ – 30	30+ – 40	40+ – 60
	NOADYN=0 (Adynamic)		
No GIV VGE	33/34= 0.97	27/31 = 0.87	16/19 = 0.84
GIV VGE	1/34= 0.03	4/31 = 0.13	3/19 = 0.16
	NOADYN=1 (Ambulatory)		
No GIV VGE	207/261 = 0.79	101/140 = 0.72	41/63 = 0.65
GIV VGE	54/261 = 0.21	39/140 = 0.28	22/63 = 0.35

Overall, the cross-tabulations show general trends of the importance of TR360, SEX, NOADYN status, and AGE on the incidence of Grade IV VGE. Also, there appears to be some evidence of possible interactions among variables as the variables are categorized in the tables.

In the following sections, we will explore the influence of the explanatory variables on the *time to onset* of Grade IV VGE. For exploratory purposes, we will next describe a nonparametric estimator of the survival curve.

3. Turnball Estimates of Survival Curves

A nonparametric method for estimating $S(t)$, the probability of survival beyond a given time, t, is due to Turnball (1976). Turnball's method generalizes the Kaplan-Meier estimate of survival probabilities to interval-censored data and provides a nonparametric maximum likelihood estimate (MLE) of $S(t)$ computed using an Expectation-Maximization algorithm, as we briefly describe next.

Consider each interval endpoint to be a recorded time, say τ_j, and form new intervals based on the ordered recorded times, $0 = \tau_0 < \tau_1 < \ldots < \tau_m$. Consider the ith individual who experiences an event within the interval $(L_i, R_i]$. For every interval of recorded time, $(\tau_{j-1}, \tau_j]$, that falls within the censored interval $(L_i, R_i]$, set $a_{ij} = 1$. Given an initial estimate of $S(\tau_j)$, $j = 1,\ldots, m$, the algorithm iterates between estimating $p_j = S(\tau_{j-1}) - S(\tau_j)$, the probability of the event occurring within the interval, $(\tau_{j-1}, \tau_j]$, and estimating $S(\tau_j)$, using a Kaplan-Meier estimate involving the pseudo-number of events, d_j, occurring at time τ_j, where

$$d_j = \sum_{i=1}^{n} \left(a_{ij} p_j \Big/ \sum_k a_{ik} p_k \right).$$ Convergence of the algorithm to stable estimates $\hat{p}_j$ then yields the nonparametric MLE of $S(t)$.

However, the nonparametric MLE found by the algorithm is not necessarily unique (Turnball, 1976). Gentleman and Geyer (1994) point out that the maximization is a concave programming problem with linear constraints on the sum of the nonnegative p_j. A sufficient condition for the uniqueness of the MLE computed using this algorithm is the negative definiteness of the Hessian matrix of the log likelihood as a function of the p_j. The Hessian matrix is negative definite if the $n \times m$ matrix, $\mathbf{A}$, of the elements a_{ij} is full column rank. We refer you to Gentleman and Geyer (1994) for further details.

The Turnball estimate of $F(t) = 1 - S(t)$, the probability of experiencing the event by time, t, was computed for the Grade IV VGE data togethre with 95% simultaneous confidence bands. The MLE of F was determined to be unique if we use the method of Gentleman and Geyer. The confidence bands were computed on the logit-transformed $F(t)$, then back-transformed to get an interval on $F(t)$ itself (Meeker and Escobar, 1998). The formulas used for the lower and upper confidence limits were programmed in S-PLUS 2000 and Mathematica 4.0. Figure 1 shows the resulting estimate with 95% simultaneous confidence bands.

The presence of a cured fraction of individuals in a population may be manifested in the data through a plateau in an estimate of the survival curve, such as in a Kaplan-Meier or Turnball survival estimate. Figure 1 shows evidence of a plateau of approximately 0.25, beginning at about five hours. Thus, if the cured fraction in the dataset were actually 0.75, the Turnball estimate as shown in Figure 1 would plateau at around 0.25. However, a plateau can also occur if there is insufficient follow-up in the study, resulting in many right-censored observations. Maller and Zhou (1996) give further details and references.
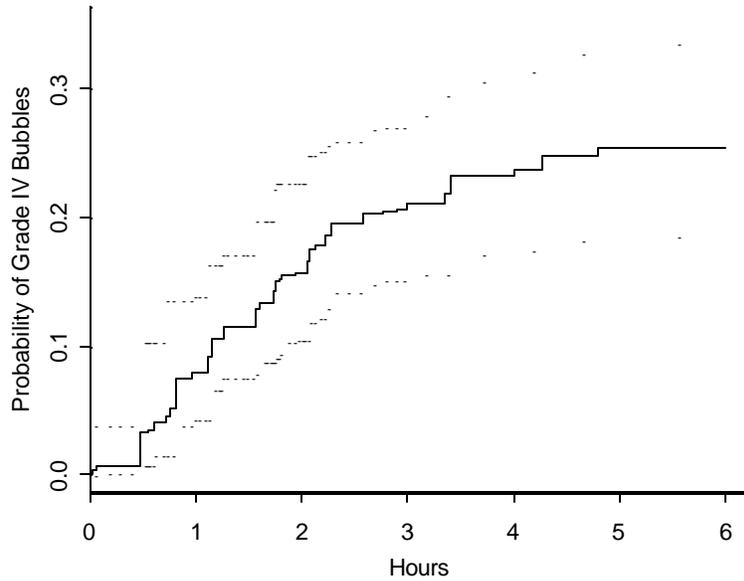
Figure 1. Turnball estimate of probability of Grade IV VGE by time (hours), along with 95% simultaneous confidence bands.

We also stratified the data by sex as well as by NOADYN status (LBA = lower body adynamia) before we computed the Turnball estimate of the probability of Grade IV bubbles. Figure 2 shows the estimates of probability, excluding 95% confidence intervals for clarity. It is important to note that although the strata have different sample sizes, the graphical estimates still give some idea of what to expect from predictions of a parametric model. Figure 2a shows that males may have a higher probability of Grade IV bubbles over females, although the curve for females is based on only 96 out of 548 records. Figure 2b shows that individuals with movement in the lower body (i.e., ambulatory) have a higher probability of Grade IV bubbles at most times; again, the curve for LBA is based on only 84 records.
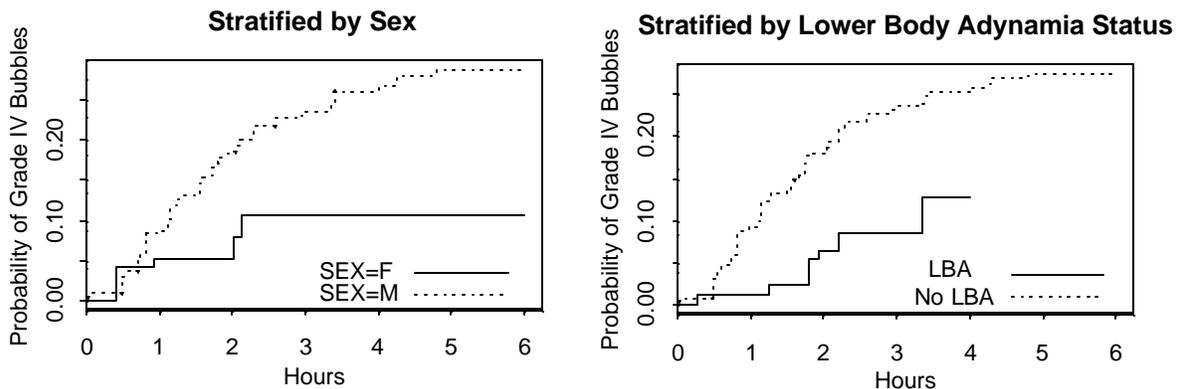
Figure 2. Turnball estimates of probability of Grade IV VGE, stratified by (a) sex and (b) LBA status. No LBA implies that the individual is ambulatory. Sample sizes can be inferred from Table 1.

Figures 3a and 3b stratify the data by quintiles of TR360 and the AGE categories used in Table 2d, and show the Turnball estimate for each stratum. The lowest category was not included for TR360 because there were no instances of Grade IV VGE. Both figures show that onset time decreases with age and with TR360.

Stratified by TR360
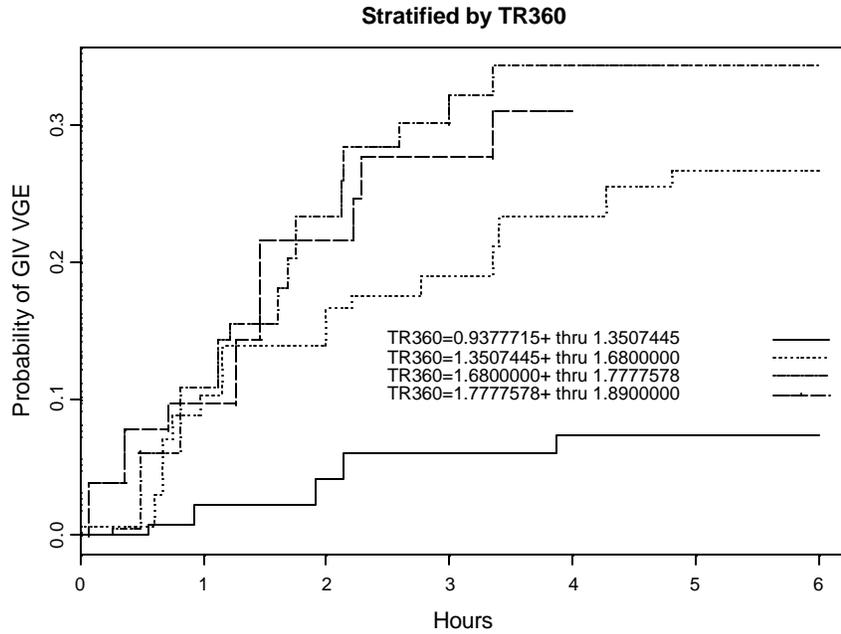
Figure 3a. Turnball estimates of probability of Grade IV VGE, stratified by TR360.
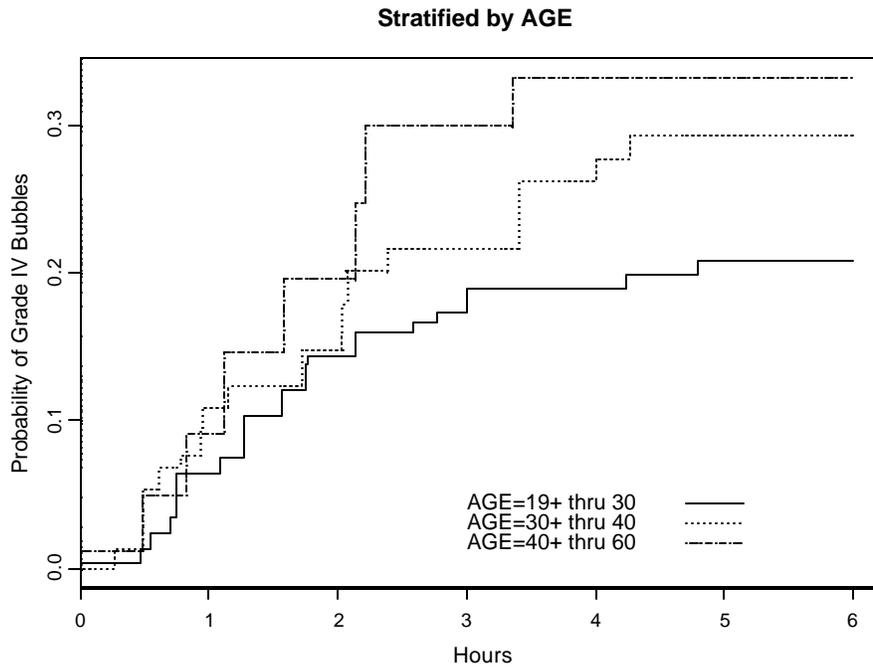
Stratified by AGE

Figure 3b. Turnball estimates of probability of Grade IV VGE, stratified by AGE.

9

4. Model Fitting and Selection

4.1 Lognormal Model for Non-cured Population

A common way to deal with interval-censored observations when fitting parametric models is to use maximum likelihood estimation, where the likelihood contribution for an observation known to fall within the interval $(t_0, t_1]$ is $\int_{t_0}^{t_1} f(u; \boldsymbol{q}) \, du$, where f is the density corresponding to the parametric distribution function F, which depends on $\boldsymbol{q}$. For Grade IV VGE data, the chosen parametric form for the survival distribution for the "non-cured" individuals was lognormal because the form of its hazard function is consistent with the hypothesized potential for experiencing Grade IV bubbles. That is, the lognormal hazard function is zero at time zero, increases to a maximum, then decreases to zero as time goes to infinity. The potential for Grade IV VGE for individuals in hypobaric conditions is purported to follow this trend. 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

$$S\left(t \mid \boldsymbol{b}, \boldsymbol{s}, \mathbf{x}\right) = 1 - \Phi\left(\frac{\log(t) - \boldsymbol{m}(\mathbf{x})}{\boldsymbol{s}}\right) \tag{4.1}$$

where

Φ is the cumulative distribution function for the standard normal distribution

$\boldsymbol{m}(\mathbf{x}) = \boldsymbol{b}_0 + \sum_{k=1}^{p} x_k \boldsymbol{b}_k$

$\boldsymbol{s}$ is the standard deviation of $\log(t)$, and

$x_k \quad k = 1, \ldots, p$ are values of p explanatory variables

An initial model included all the variables listed in Table 1 as covariates. For computational purposes, the variables AGE and TR360 were standardized. The standardized versions are subsequently denoted as S.AGE and S.TR360, respectively, where S.AGE = (AGE − 31.85)/7.17, and S.TR360 = (TR360 − 1.57)/.263 (Table 1 contains the means and standard deviations used for standardization.) The covariates were all included in linear form. Some investigation into nonlinear forms yielded no reason to suspect nonlinear effects of the covariates on the response variable. In addition, pairwise scatterplots between pairs of covariates did not reveal any noticeable relationships.

4.2 Modeling Multiple Observations Per Subject

Some observations originated from the same individual undergoing different tests. That is, some groups of observations represent repeated observations within the same subject, and these observations are correlated. Although analyzing the data as though the observations were independent can be misleading with respect to the precision of the parameter estimates. There are several ways to deal with dependent observations. Statistical methods for handling dependent observations might broadly be categorized in one of two ways: (1) either dependency is accounted for in the statistical model, or (2) an adjustment is made to the variability estimates obtained from a model that assumes independent observations. We chose a relatively simple, but typical, method of the first way for modeling the correlation between pairs of observations from the same individual.

Random Effects Model for Dependent Observations

A random effects model handles repeated measurements by incorporating random subject effects into the linear form representing the conditional mean of the log time to the event. The conditional mean for the jth log event time for the ith subject is modeled as

$$E(\log t_{ij} \mid b_i, \mathbf{x}) = \boldsymbol{m}_{ij}(\mathbf{x}) = \boldsymbol{b}_0 + \sum_{k=1}^{p} x_{kij} \boldsymbol{b}_k + b_i \qquad (4.2)$$

In this representation, b_i is the random effect for subject i, distributed normally with mean zero and constant variance $\boldsymbol{s}_b^2$. Note that the observations made on the same subject share the same random effect. Also, the b_i are assumed to be independent of one another. Using (4.2), the model representation for log t_{ij} is

$$\log t_{ij} = \boldsymbol{m}_{ij}(\mathbf{x}) + \boldsymbol{e}_{ij} \qquad (4.3)$$

In (4.3), $\boldsymbol{e}_{ij}$ represents a mean-zero, normally distributed error term with variance $\boldsymbol{s}^2$, and independent of b_i. The unconditional variance of $\log t_{ij}$ is thus $\boldsymbol{s}^2 + \boldsymbol{s}_b^2$, so the correlation between any pair of log times for a given subject is $corr(\log t_{ij}, \log t_{ij'}) = \boldsymbol{s}_b^2 / (\boldsymbol{s}^2 + \boldsymbol{s}_b^2) = \boldsymbol{r}$, independent of i and j.

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, $\boldsymbol{b}_0 + \sum_{k=1}^{p} x_{kij} \boldsymbol{b}_k$. A large value of $\boldsymbol{s}_b^2$, the variance of random subject effects, relative to $\boldsymbol{s}^2$, the variance due to all other factors besides subject effects, may indicate relatively high variability among observations made between individuals or could indicate a proportionately large amount of unmodeled covariates.

4.3 Estimation of the Parameters from a Random Effects Model

Model (4.3) has $p + 3$ parameters: $(\boldsymbol{b} = (\beta_0, \ldots, \beta_p), \boldsymbol{s}_b, \boldsymbol{s})$. In a classical statistics context, the number of parameters for this model will always be $p + 3$ irrespective of the number of subjects tested. Although the random effects, b_i, may behave like parameters, technically these random effects are just another level of random variation in the model. So, we do not estimate them as we would the other $p + 3$ parameters. However, their distribution must be considered in the likelihood function describing the random process that is assumed to have generated the data. If we denote the likelihood conditional on $\mathbf{b} = (b_1, \ldots, b_n)$ by $L(\boldsymbol{b}, \boldsymbol{s}, \boldsymbol{s}_b, \mathbf{b} = (b_1, \ldots, b_n))$, then MLEs of $(\boldsymbol{b}, \boldsymbol{s}_b, \boldsymbol{s})$ are obtained by maximizing, with respect to $(\boldsymbol{b}, \boldsymbol{s}_b, \boldsymbol{s})$, the likelihood integrated over the distribution of the random effects, $\mathbf{b}$

$$L(\boldsymbol{b}, \boldsymbol{s}, \boldsymbol{s}_b) = \int L(\boldsymbol{b}, \boldsymbol{s}, \mathbf{b}) N(\mathbf{b} \mid \boldsymbol{s}_b) \, d\mathbf{b} = \int \prod_i L(\boldsymbol{b}, \boldsymbol{s}, b_i) N(b_i \mid \boldsymbol{s}_b) \, d\mathbf{b} \qquad (4.4)$$

where $N(b_i \mid \boldsymbol{s}_b)$ is the zero-mean Gaussian probability density function with standard deviation, $\boldsymbol{s}_b$, evaluated at b_i. However, the integration in (4.4) is not always tractable. In this case, other options are available for getting MLEs. Computational methods for standard errors of the estimates depend on the

option chosen. In our case, numerical integration is a viable option because we integrate with respect to univariate variables. That is, the right-most side of (4.4) can be written as

$$L(\boldsymbol{b}, \boldsymbol{s}, \boldsymbol{s}_b) = \prod_i \int L(\boldsymbol{b}, \boldsymbol{s}, b_i) N(b_i | \boldsymbol{s}_b) \, db_i \qquad (4.5)$$

Equation (4.5) is then maximized with respect to the parameters of interest. Gauss-Hermite quadrature can be used to do the integrations (Naylor and Smith, 1982).

Another alternative for obtaining *generalized* MLEs is via Bayesian estimation with diffuse priors on the unknown parameters. With the complicated likelihood functions, Markov Chain Monte Carlo (MCMC) sampling makes estimation easier than other methods. In this type of sampling, repeated draws from the joint posterior distribution follow a first-order Markov chain. The joint posterior distribution of the parameters is proportional to the likelihood function 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 samples from the chain. For symmetric, unimodal distributions, first-order expectations (if they exist) coincide with strict MLEs. So, when histograms of the samples from the chain look unimodal and symmetric, the sample average is fairly close to what would be the MLE had we used traditional maximum likelihood estimation procedures (for example, Gilks et al., 1996).

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 over. This avoids direct integration required to get the integrated likelihood.

We combined the two approaches to estimate parameters. First, we set independent diffuse priors on the parameters in ($\boldsymbol{b}$, $\boldsymbol{s}_b$, $\boldsymbol{s}$), and then used MCMC sampling to obtain approximate generalized MLEs by taking the sample averages from the MCMC output chains. All univariate histograms of the MCMC samples of single parameters were roughly symmetrical. We then used approximate generalized MLEs as starting values in a quasi-Newton Rapheson algorithm to maximize integrated likelihood. The integration was performed using Gauss-Hermite numerical integration. The reason for combining the two methods instead of using the MCMC sampling alone was that the MCMC sampling (as described) included the specification of a joint prior distribution. Although the joint prior was specified to be diffuse and, thus arguably, close to noninformative with respect to the likelihood, we note that this analysis is not strictly a Bayesian analysis. A prior distribution has no meaning in our estimation, and is used together with MCMC sampling to get starting values only. There are certainly methods for using MCMC sampling in frequentist contexts (e.g., Gilks et al., 1996), including the use of a truly noninformative joint prior. However, none of these procedures (that we tried) was as practical to implement computationally as the above combined approaches. Furthermore, all of the methods we did try "converged" to similar estimates.

More details of the estimation are presented later. We will now describe an extension of the initial lognormal model to handle a cured fraction.

5. An LFP Model for Time to Onset of Grade IV VGE

An LFP Model is a mixture model for survival data consisting of two groups: (1) individuals who will eventually experience the event, and (2) individuals who will never experience the event, sometimes called "immune" individuals. The probability that a randomly selected individual will never experience the event in question is sometimes referred to in the literature as the "cure rate". The cure rate may or may not depend on explanatory variables. There are numerous references of the application of these

models in survival and reliability analysis in the literature (e.g., Maller and Zhou, 1996; Meeker and Escobar, 1998).

For our population, the cure rate is the probability of never experiencing Grade IV bubbles. We assume that the cure rate depends on explanatory variables measured on each individual. This means that for a given individual, the probability of experiencing Grade IV bubbles is a function of physical characteristics of the individual measured prior to the test session. 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. In the next subsection, we will define the cure rate as a function of explanatory variables, and give the resulting lognormal likelihood for interval- and right-censored data. Later, we will compare the LFP model to a non-mixture lognormal model, where both models are fitted using random subjects effects.

5.1 Construction of the Limited Failure Population Model

For the jth observation from the ith individual, define the indicator variable as

$$z_{ij} = \begin{cases} 1 & \text{if the } i\text{th subject will eventually experience Grade IV VGE on their} j\text{th test} \\ 0 & \text{otherwise} \end{cases}$$

Although immunity to the event, Grade IV VGE, may be perceived as applying to individuals, the indicator variable is defined for each *observation* within each individual. An individual will not necessarily have a constant propensity for cure across repeated observations, as some covariate values may change across measurements.

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 subsection. This assumption is valid provided 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.

If we assume that censoring times are independent of event times, and that observations from different *individuals* are independent of one another, we can construct the likelihood as follows:
For an interval-censored observation with covariates, $\mathbf{x}_{ij}$, the contribution to the likelihood is
$P(t_{0_{ij}} < T_{ij} \leq t_{1_{ij}}, z_{ij} = 1 | \mathbf{x}_{ij}) = P(z_{ij} = 1 | \mathbf{x}_{ij})(S(t_{0_{ij}} | \mathbf{x}_{ij}) - S(t_{1_{ij}} | \mathbf{x}_{ij}))$, for known left and right interval endpoints $t_{0_{ij}}$ and $t_{1_{ij}}$. Similarly, for a Type I right-censored observation, the contribution to the
likelihood is $P(t_{0_{ij}} < T_{ij}, z_{ij} = 1 | \mathbf{x}_{ij}) + P(t_{0_{ij}} < T_{ij}, z_{ij} = 0 | \mathbf{x}_{ij}) = P(z_{ij} = 1 | \mathbf{x}_{ij}) S(t_{0_{ij}} | \mathbf{x}_{ij}) + P(z_{ij} = 0 | \mathbf{x}_{ij})$.
As noted for model (1.1), the survival distribution S is only defined for the non-cured population. The survival distribution is always 1 for the cured population.

If $G(t_c | \mathbf{x})$ represents the survival function for the randomly right-censored time, t_c, then a randomly right-censored observation's contribution to the likelihood is
$P(t_{0_{ij}} < T_{ij}, z_{ij} = 1 | \mathbf{x}_{ij}) + P(t_{0_{ij}} < T_{ij}, z_{ij} = 0 | \mathbf{x}_{ij}) = P(z_{ij} = 1 | \mathbf{x}_{ij}) G(t_{0_{ij}} | \mathbf{x}_{ij}) + P(z_{ij} = 0 | \mathbf{x}_{ij})$. If the random censoring distribution G does not involve parameters of interest related to the covariates, then $G(t_c | \mathbf{x}) = G(t_c)$. In this case, the contribution to the likelihood from a randomly censored observation becomes $P(z_{ij} = 1 | \mathbf{x}_{ij}) G(t_{0_{ij}}) + P(z_{ij} = 0 | \mathbf{x}_{ij})$. If the randomly right-censored observations can be treated as Type I censored observations, these observations contribute the same term to the likelihood as given

for the Type I censored cases. This situation is desirable because it eliminates the need to specify the form for G, with or without dependence on covariates.

Of the 425 right-censored observations, 11 cases were randomly right-censored. Because there were so few randomly censored observations, we did not feel it was statistically beneficial to include an additional term in the likelihood for only these 11 cases. Moreover, our exclusion of the randomly right-censored cases from parameter estimation procedures made little changes to the resulting estimates. So, we treat the 11 randomly right-censored cases as though they were Type I right-censored.

The contribution by right-censored observations reflects the mixture aspect of the model. A right-censored observation would have a probability $P(z_{ij}=1|\mathbf{x}_{ij})$ of ever experiencing the event, and a probability $P(z_{ij}=0|\mathbf{x}_{ij})$ of being immune to the event. The probability $P(z_{ij}=0|\mathbf{x}_{ij})$ is the cure rate mentioned above.

Define the censoring indicator $\boldsymbol{d}_{ij}$ to be 1 if the ijth observation is interval-censored and zero otherwise. Denote the observed data by $D=\{\mathbf{x}_{ij},t_{0_{ij}},t_{1_{ij}},\boldsymbol{d}_{ij};\ j=1,...,m_i\ ;i=1,...,n\}$, and denote a single data record by D_{ij}. The likelihood is then

$$L(\boldsymbol{b},\boldsymbol{s},\boldsymbol{s}_b\,|D)=\prod_{i=1}^{n}\int\prod_{j=1}^{m_i}L_{ij}(\boldsymbol{b},\boldsymbol{s},\boldsymbol{s}_b\,|D_{ij},b_i)\,N(b_i\,|0,\boldsymbol{s}_b)\,db_i \tag{5.1}$$

where

$$L_{ij}(\boldsymbol{b},\boldsymbol{s},\boldsymbol{s}_b|D_{ij},b_i)=$$

$$\left[P(z_{ij}=1|\mathbf{x}_{ij})(S(t_{0_{ij}}|\mathbf{x}_{ij},b_i)-S(t_{1_{ij}}|\mathbf{x}_{ij},b_i))\right]^{\boldsymbol{d}_{ij}}\left[P(z_{ij}=1|\mathbf{x}_{ij})\,S(t_{0_{ij}}|\mathbf{x}_{ij},b_i)+P(z_{ij}=0|\mathbf{x}_{ij})\right]^{1-\boldsymbol{d}_{ij}}$$

and, $S(t\,|\,\mathbf{x}=(x_1,...,x_p),b)=\mathbf{F}\left(\dfrac{\log t-\boldsymbol{b}_0-\sum_{k=1}^{p}\boldsymbol{b}_k x_k-b}{\boldsymbol{s}}\right).$

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

$$P(z_{ij}=1|\mathbf{x}_{ij})=\exp(\boldsymbol{a}_0+\sum_{k=1}^{p}\boldsymbol{a}_k x_{ijk})\Big/\left(1+\exp(\boldsymbol{a}_0+\sum_{k=1}^{p}\boldsymbol{a}_k x_{ijk})\right) \tag{5.2}$$

where the $\boldsymbol{a}_k$ are parameters relating the covariates to the cure rate.

With these forms for the cure rate and survival distribution, $L_{ij}(\boldsymbol{b},\boldsymbol{s},\boldsymbol{s}_b|D_{ij},b_i)$ in equation (5.1) becomes

$$L_{ij}(\boldsymbol{b},\boldsymbol{s},\boldsymbol{s}_b|D_{ij},b_i) =$$

$$\left(\frac{\exp(\boldsymbol{a}_0+\sum_{k=1}^p \boldsymbol{a}_k x_{ijk})}{1+\exp(\boldsymbol{a}_0+\sum_{k=1}^p \boldsymbol{a}_k x_{ijk})}\right)^{d_{ij}}\left(\Phi\left(\frac{\log t_{1_{ij}}-\boldsymbol{b}_0-\sum_{k=1}^p \boldsymbol{b}_k x_{ijk}-b_i}{\boldsymbol{s}}\right)-\Phi\left(\frac{\log t_{0_{ij}}-\boldsymbol{b}_0-\sum_{k=1}^p \boldsymbol{b}_k x_{ijk}-b_i}{\boldsymbol{s}}\right)\right)^{d_{ij}}$$

$$\times\left[\frac{\exp(\boldsymbol{a}_0+\sum_{k=1}^p \boldsymbol{a}_k x_{ijk})}{1+\exp(\boldsymbol{a}_0+\sum_{k=1}^p \boldsymbol{a}_k x_{ijk})}\left(1-\Phi\left(\frac{\log t_{0_{ij}}-\boldsymbol{b}_0-\sum_{k=1}^p \boldsymbol{b}_k x_{ijk}-b_i}{\boldsymbol{s}}\right)\right)+\frac{1}{1+\exp(\boldsymbol{a}_0+\sum_{k=1}^p \boldsymbol{a}_k x_{ijk})}\right]^{(1-d_{ij})}$$

$$(5.3)$$

We refer to the parameters, $\boldsymbol{b}=(\boldsymbol{b}_0,\boldsymbol{b}_1,....,\boldsymbol{b}_p)$ as *location parameters* and to the parameters $\boldsymbol{a}=(\boldsymbol{a}_0,\boldsymbol{a}_1,...,\boldsymbol{a}_p)$, in the cure rate, as *mixture parameters*. If the location parameters in (5.3) are all distinct from the mixture parameters, the maximum number of distinct parameters in (5.3) is $2(p+1)+1$. In many modeling situations, some of the mixture or location parameters are fixed at zero.

5.2 Estimation of the Parameters from the Random Effects Models

We now fit two LFP random effects model to the Grade IV VGE data and contrast their fit with that of a non-mixture model, which is actually a special case of the LFP model, but with $P(z_{ij}=1)=1$ for all i and j. The models were fit using the structures given above, with the addition of a random effect added to the conditional mean log time to onset for each individual. Estimation was done in two steps. First, we used MCMC sampling to sample realizations of $(\boldsymbol{b},\boldsymbol{a},\boldsymbol{s},\boldsymbol{s}_b,\mathbf{b})$ from their joint posterior distribution. The WinBUGS software version 1.3 (Spiegelhalter et al., 2000) was used for the computations. Independent normal priors with mean 0 and variance 1,000 were chosen for the elements in $(\boldsymbol{b},\boldsymbol{a},\log\boldsymbol{s},\log\boldsymbol{s}_b)$. Results did not appear to be sensitive to the parametric form of the diffuse prior specification. For example, using gamma distributions for the standard deviations instead of normal distributions for the log standard deviations did not change the results. Three chains of 2,000 samples each were run per model, each starting from different initial values. One set of initial values for the MCMC procedure came from direct maximization using Newton-Raphson's algorithm, where observations were treated as independent. (That is, random effects were ignored.) Basic convergence diagnostics, such as trace plots, and Brooks-Gelman-Rubin Statistics (Brooks and Gelman, 1998) showed no evidence of lack of convergence of the Markov chains to a target distribution. Univariate histograms of samples from the chains were all roughly symmetrical, even for the standard deviation parameters.

Sample means from the combined three chains of output were used as starting values for a quasi-Newton-Rapheson maximization of the integrated likelihood in (5.1). The integration was performed using Gauss-Hermite quadrature, and was programmed in Mathematica 4.0. Thirty quadrature points were deemed more than sufficient for the integration. Table 3 gives the approximate MLEs of the coefficients for three models. The first is a non-mixture model, and the others are LFP mixture models. 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 (Tanner, 1996, p. 74). No evidence of the nonidentifiability of parameters was observed for any of the models in Table 3. Asymptotic correlation matrices for Table 3 are in Appendix A.

Table 3. Approximate Maximum Likelihood Estimates for Fitted Models

	Non-Mixture Model	Full LFP Mixture Model	Reduced LFP Mixture Model
−LogLH	552.237	549.683	550.635
AIC	559.237	558.683	557.635
Location Parameter Estimates			
β_0	4.622 (0.491)	3.586 (0.502)	3.376 (0.403)
β_1 (SEX)	−0.885 (0.347)	−0.491 (0.364)	
β_2 (S.AGE)	−0.242 (0.122)	0.047 (0.146)	
β_3 (S.TR360)	−0.892 (0.149)	−0.840 (0.143)	−0.864 (0.144)
β_4 (NOADYN)	−1.354 (0.350)	−1.271 (0.336)	−1.359 (0.332)
σ (scale)	1.299 (0.120)	1.064 (0.143)	1.121 (0.132)
Mixture Parameter Estimates			
α_1 (SEX)		1.448 (0.839)	1.906 (0.917)
α_2 (S.AGE)		1.306 (0.593)	1.416 (0.717)
Random Effects Parameter Estimates			
σ_b (sd of random effects)	1.081 (0.161)	1.017 (0.150)	1.032 (0.150)
r corr($\log t_{ij}$, $\log t_{ij'}$)	0.409 (0.151)	0.477 (0.159)	0.459 (0.157)

Certain physical characteristics may make a person immune to experiencing Grade IV VGE. The variables SEX and S.AGE are suitable for modeling the probability of immunity. The explanatory variables S.TR360 and NOADYN are likely only to influence the onset time of Grade IV VGE. Thus, although we only consider SEX and S.AGE for the mixture portion of any LFP models, we consider all four variables for inclusion in the location portion of an LFP model. Despite slight evidence in Tables 2a-d of possible interactions among covariates, we did not consider interactions among covariates in this report.

The first LFP model represented in Table 3 (called the full LFP mixture model) describes log survival time as depending on all four of the explanatory variables. It also models the cure rate as depending on two explanatory variables: SEX and S.AGE. The second LFP model (the reduced LFP mixture model) describes log survival time as depending on two of the explanatory variables: S.TR360 and NOADYN. It further models the cure rate as depending on two explanatory variables: SEX and S.AGE. Other configurations were tried, but none was uniformly better in terms of fit and accuracy of predictions. Judging from the values of Aikaike's Information Criterion (AIC) in Table 3, the reduced LFP model appears to be a better fit than either the non-mixture or the full model. In addition, two of the location parameter estimates in the full model (SEX and S.AGE) are much less than twice their standard errors, indicating that these two variables are insignificant location parameters in the model.

Before we discuss additional assessments of model fit and predictive accuracy, we will present a brief interpretation of the meaning of the parameter estimates within each model.

Interpretation of the parameter estimates

Negative location parameter estimates for a given explanatory variable imply that for two observations differing by one unit in the variable, the expected log time to onset of Grade IV bubbles decreases by the estimate given. According to the non-mixture model, for a male versus a female, the expected onset time is 0.885 log minutes sooner. Similarly, for two individuals who differ by one standard deviation in AGE (approximately 7 years), the expected time to onset is 0.242 log minutes sooner for the older individual. The remaining parameter estimates are interpreted similarly. For the reduced LFP model in Table 3, the location parameter estimates share the same direction as those for the non-mixture model. In addition, for a male versus a female, the expected log odds of the probability of eventually experiencing Grade IV VGE is 1.906 higher. For two individuals who differ by one standard deviation in AGE, the expected log odds are increased by 1.416 for the older individual. Both of these imply that younger females are most likely to be immune to Grade IV VGE, and older males are least likely to be immune to Grade IV VGE, at least within the age range we modeled.

Two sources of variability are estimated for each model given in Table 3. The parameter s_b describes variability between individuals, and the parameter s describes variability from a measurement error not associated with between-individual variability. A relatively high estimate for s_b as compared to its standard error may imply strong heterogeneity among individuals. An acceptable rule of thumb to apply in this case is that if the estimate of s_b is greater than twice its standard error, there is evidence of significant heterogeneity among individuals. Table 3 shows that the estimates for s_b are much greater than twice their respective standard errors. Heterogeneity may result from idiosyncratic differences among individuals or from differences associated with one or more covariates that were not modeled. The relatively high estimate for s_b results in moderate estimates of correlation between two log responses on the same subject.

Also, the estimates of s_b for both models are comparable to those of s. One reason for this may be due to the small number of tests contributed by most subjects. In fact, 42% of the 238 subjects contributed only one record, 30% contributed two records, and less than 5% contributed more than six records. Thus, much of the variability in log event times is due to variability among individuals.

Next we will discuss goodness of fit and predictive accuracy of the two models represented in Table 3.

6. Goodness of Fit of Two Models for Onset Time of Grade IV VGE

Goodness of fit of parametric models with interval censoring is not easy to check, particularly when the recorded intervals are not fixed across subjects. Therefore, standard test procedures that use the prediction of the percentages of events within certain time intervals are not useful. In addition, if the covariates are continuous, we may not be able to assess goodness of fit well by dividing the data into groups according to the values on the covariates, and then comparing the nonparametric survival estimates to the predicted survival curves. Also, with a high percentage of right-censored observations, it is difficult to get enough interval-censored data within some of the groups.

In addition, residual plots are complicated due to the difficulty in, first, defining a residual for interval-censored data, and, second, determining its sampling distribution under the hypothesis that a model is correct. So, instead we describe a graphical goodness-of-fit approach that is useful in evaluating the model fits.

6.1 Goodness of Fit Using a Graphical Method

The graphical approach we used involved a parametric bootstrap procedure. We bootstrapped interval- and right-censored observations from each of the fitted models with point estimates in Table 3 and fixed covariate patterns. Then, we calculated the Turnball estimate of the probability of Grade IV bubbles for each bootstrapped sample and compared these Turnball curves to the data-based Turnball estimate and 95% simultaneous confidence bands (as we discussed in Section 3). The confidence bands describe the uncertainty that is associated with the data-based Turnball estimate as an estimate of the actual or "true" population-based curve. The population here might be individuals in simulated hypobaric conditions. So, to the extent that the confidence interval endpoints contain curves from the parametric bootstrap from a particular model, that model may be said to describe the population well. Alternatively, if there is the possibility (under 95% confidence) that the "true" population Turnball curve is located in areas where there are no generated curves from a particular model, there is an unacceptably high chance that the data did not originate from the same model. This reasoning assumes that the generated Turnball curves fill the entire range of possible curves generated from the model in question.

To reflect the dependency in the data, the parametric bootstrap procedure generated a random effect for each individual from a normal distribution with mean zero and variance equal to the point estimate of s_b^2. Repeated tests from a single subject all had the same random effect.

One hundred bootstrapped samples were drawn from each of the competing random effects models. Details of the bootstrapping procedure are in Appendix B. After around 50 samples were analyzed, the concentration of the resulting plotted Turnball estimates hardly changed. So, 100 samples were deemed sufficient for graphical purposes. Figure 4 contains the resulting Turnball estimates of "failure" or occurrence probabilities by time for the non-mixture model and the LFP models described in Table 3. Each panel in the figure shows the 100 generated estimates for one of the models (the grayed lines), superimposed with the original data-based curve based on the original data (the dark solid line) and the simultaneous confidence bands (dotted dark lines).

According to Figure 4, both LFP mixture models can generate Turnball nonparametric maximum likelihood estimates that fall within the 95% confidence bands associated with the estimate computed from the original data. If the actual population Turnball curve falls within the dark dotted line (with 95% confidence), that curve is consistent with either LFP model. The patterns for the two LFP models are very similar, except the full model displays more variability – probably due to its two additional coefficients. Also, the distributions of Turnball points at each hour for the reduced LFP model appear to be negatively skewed. Thus, fewer generated datasets had many observations with infinite event times for the reduced LFP model than for the full LFP model.

The plot containing the samples drawn from the non-mixture model shows that almost all the generated curves are close to the data-based curve and also fall within the 95% simultaneous confidence bands on the data-based Turnball curve. However, within the 95% confidence bands are areas that are outside of the range of generated curves. This may be indicative of a poor model fit because the actual curve may correspond to areas outside of the range of generated curves. Adding a mixture part to the model (as was done in the full and reduced LFP models) ensures that this situation will not happen and that the actual population curve (with 95% confidence) is within the range displayed by the generated curves from the model. As we mentioned in a previous section, this conclusion assumes that the bootstrap has generated curves to cover the entire range that the model in question would predict. The assumption seems to hold true, at least approximately, even though there are only 100 generated curves because the difference in the range between 50 and 100 generated curves was minimal.
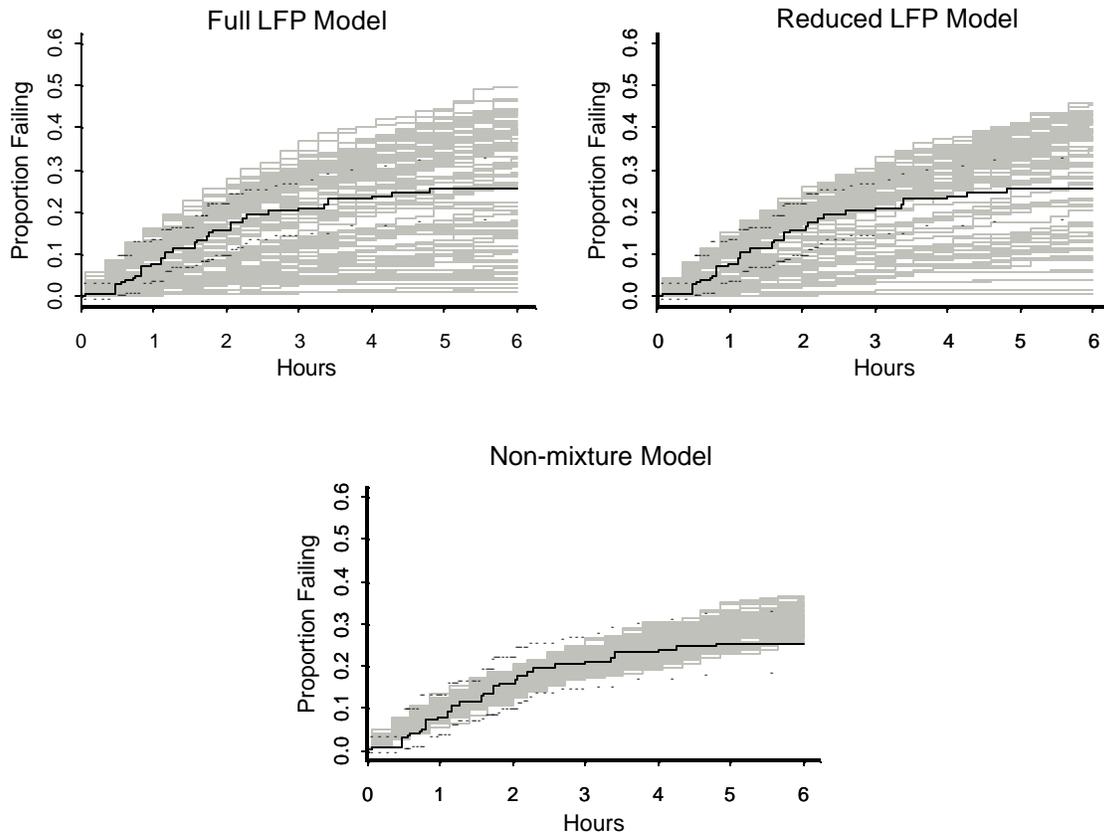
Figure 4. Goodness-of-fit graphs for three models: one non-mixture lognormal model and two LFP mixture lognormal models. The dark solid line represents the Turnball nonparametric probability estimate based on the observed data. The dark dotted lines are 95% simultaneous confidence bands. The 100 grayed lines represent the Turnball estimates based on 100 samples from the respective fitted model.

A generated dataset with many immune observations has a nearly flat Turnball curve. The presence of S.AGE in the mixture portion of the LFP models may cause some generated datasets to contain many immune observations. The mixture coefficient on S.AGE is positive for both models, which implies that with increasing age, the probability of experiencing Grade IV VGE eventually increases as well. The majority of the S.AGE values were negative (slightly less than 60%), however, so age was below the mean age of 31.85 years. A negative S.AGE decreases the estimated probability of eventually experiencing Grade IV VGE. Generated datasets of onset times that have many negative S.AGE values may have nearly flat Turnball estimates.

Based on the plots in Figure 4, it seems that our data more likely originated from a model similar to an LFP mixture model than from the non-mixture model. Furthermore, as given in the next section, general predictive accuracy based on the non-mixture model is not as good as prediction based on the LFP mixture model.

7. Predictive Accuracy of Models

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 are 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*. For details of the adjustment, see Davison and Hinkley (1997). We detail the procedure for our application in Appendix C.

Predictions, conditional on the covariate values and censoring pattern, were obtained on log times to Grade IV bubbles. 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 made on the same individual. For computation of the predicted log time, see Appendix C.

The top half of Table 4 contains 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, $K = \min(10, \sqrt{n})$, where n is the size of the data set (Davison and Hinkley, 1997). 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 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: namely that the LFP model predicts better than the non-mixture model on a new dataset from the same population. Other choices of K led to very similar results as those seen in the table.

Table 4: Measure of Predictive Accuracy of Models
(K-fold Cross-Validated Prediction Error)

	Non-mixture Model	Full LFP Mixture Model	Reduced LFP Mixture Model
Number of Groups (Size of Left-out group)			
K = 10 (*m* = 55)	0.961	0.532	0.596
K = 548 (*m* = 1)	0.961	0.532	0.594
Sample size used to fit model:			
n = 25	1.005	0.751	0.636
n = 35	0.991	0.698	0.636
n = 50	0.920	0.636	0.593

The bottom half of Table 4 contains results to get an idea of the sample size for which overfitting begins to display itself. We randomly selected a small number of observations (n = 25, 35, or 50) from the dataset 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 and took the median of the resulting prediction errors. (The median was used instead of the mean because of some very large prediction

errors). The median is displayed in the table. As we expected, prediction generally gets better for all models as the sample size increases. But notably, the non-mixture model predicts poorly compared to the two LFP models for all sample sizes. In addition, the reduced LFP model appears to perform slightly better than the full LFP model.

We note that although the parameter estimates can be expected to change slightly when all of the data points are not used to fit the model, it is hoped that the parameter estimates do not change grossly for each fitting, or else there is a problem with highly influential points. That did not appear to be the case here, as regards the first two rows of Table 4, where a substantial part of the data was used in the fitting. However, that is not the case when using only 25 or 35 observations. Here, the parameter estimates sometimes changed wildly due to different percentages of right-censoring. So, the last three rows of Table 4 really assess the stability of the structure of the model, not necessarily of the parameter estimates. Even with 25 observations, overfitting does not appear to be a big problem for any of the models, but predictive accuracy is again much worse for the non-mixture model versus the two mixture models.

The two mixture models are very close in predictive accuracy. Either model could be used for prediction, giving similar results. However, the reduced model has a slightly better predictive accuracy in the case of the small sample size. It also has fewer parameters and, thus, a smaller AIC value. All of the coefficient estimates of the reduced model exceed or come close to exceeding twice their standard errors. It is clear that coefficients for SEX and S.AGE do not belong in the location portion of the full model because of their magnitude relative to their standard errors. We will therefore use the reduced LFP for specific predictions in the next section.

8. Further Evaluation of the LFP Model

Based on estimates from the reduced LFP Model in Table 3, the estimated probability of remaining free from Grade IV bubbles by a given time, t, is

$$\hat{P}(T > t) = \hat{P}(z = 1)\hat{S}(t) + \hat{P}(z = 0)(1) = \hat{P}(z = 1|\mathbf{x})\left(1 - \Phi\left(\frac{\log t - \hat{m}(\mathbf{x})}{1.121}\right)\right) + \hat{P}(z = 0|\mathbf{x}) \qquad (8.1)$$

where

$$\hat{m}(\mathbf{x}) = 3.376 - 0.864\ STR360 - 1.359\ NOADYN,$$

$$\hat{P}(z = 1|\mathbf{x}) = \frac{\exp(1.906\ SEX + 1.416\ S.AGE)}{1 + \exp(1.906\ SEX + 1.416\ S.AGE)},$$

$$S.AGE = (AGE - 31.85)/7.17, \quad \text{and} \quad STR360 = (TR360 - 1.57)/0.263.$$

The complement of (8.1), the probability of experiencing Grade IV bubbles by a given time, t, is

$$\hat{P}(T \leq t) = \hat{P}(z = 1)(1 - \hat{S}(t)) = \hat{P}(z = 1|\mathbf{x})\Phi\left(\frac{\log t - \hat{m}(\mathbf{x})}{1.121}\right). \qquad (8.2)$$

For calculating predictions, the random effect term, b_i, that appears in (4.2) is not used. Equation (8.2) is called the cumulative distribution function (CDF). Plots of (8.2) are shown in Figure 5 for several values

of the explanatory variables, along with 95% point-wise confidence intervals that were 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 they are always contained within 0 and 1 (Meeker and Escobar, 1998). The upper points of the intervals are represented by dashed lines, and the lower points are represented by dotted lines. For all four plots, the higher dashed lines and the higher dotted lines correspond to the higher solid lines, representing the point estimates.

Figures 5a and 5b compare probabilities by time at altitude for males and females ages 27 and 40, respectively (to contrast younger and older ages), with a tissue ratio of 1.5, and who are lower body adynamic. The ages 27 and 40 years were chosen to represent "younger" and "older" because of the numbers of males and females at those ages in the dataset. There were nine females and 39 males at age 27, and five females and six males at age 40. (The number of females at any age was less than 10.)

According to the predictions, males have higher probabilities of Grade IV bubbles than do females at almost all hours at altitude, but the difference is very slight for an age of 40. The slight difference between the male and female curves at an age of 40 (as compared to the larger difference at an age of 27) is caused by the mixture probability being dominated by the exponential of S.AGE times its coefficient, when S.AGE increases to a large positive value. The standardization of 40 years is about 1.14, whereas the standardization of 27 years is negative (−0.68). The same reason also explains the larger standard errors when S.AGE is positive: that is, the expression for the standard error (at a given time) contains the exponential of S.AGE times its coefficient. Also, there are more observations at age 27 (n = 48) than there are at age 40 (n = 11).

Figures 5c and 5d 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 (i.e., time = 6 hours). For Figure 5c, the individual is considered to be lower body adynamic. For Figure 5d, the individual has a tissue ratio of 1.5. For both plots, the probability estimates increase with age, then flatten out as age increases beyond its mean in the dataset (about 32 years). Also, at any age, the higher tissue ratios have higher estimated probabilities as do ambulatory individuals, with tissue ratio held constant. The negative acceleration pattern that is seen in Figure 5d actually occurs for ambulatory and adynamic individuals. The range of the vertical scale hides the pattern for the adynamic curve. The same general pattern in the predictions occurred for other hours besides six hours.

Together, the panels in Figure 5 seem to predict interactions between age and the other variables in their influence on probability of Grade IV VGE at a given time point. For example, the predicted difference in probability of Grade IV VGE between an adynamic and ambulatory individual at time = 6 hours depends on the age of the individual. It is greater at a higher age. Figure 5 may signal the need for an interaction term in the model, although we do not explore this idea further here.

A comparison of Figure 5 with Figure 6, which has the analogous predictions from the non-mixture model, shows mean predictions that are generally in the same direction, albeit somewhat larger overall. There are several differences, however. First, when age is fixed at 40 years, the difference in the CDF for males versus females for the non-mixture model widens at almost all time points, as compared to when age is fixed at 27 years. This is in contrast to Figure 5, where the analogous graphs for the reduced LFP model show a narrowing of the CDF when age is fixed at 40 years, as compared to when age is fixed at 27 years. The narrowing in Figure 5 is due to not having S.AGE or SEX in the location portion of the model. Predictions from the full LFP model (not shown) show a slight widening as well, but not as much as for the non-mixture model. The full model had S.AGE and SEX in both mixture and location portions. When AGE begins to increase, both LFP models have mixture portions that are dominated by the S.AGE term, and the mixture portion goes toward 1.0 for both sexes as S.AGE increases. For the reduced LFP

model, however, the location portion stays fixed when S.AGE changes, so the CDFs for males and females grow closer together. For the full LFP model, the location part changes when S.AGE and SEX change, and the change compensates for the mixture portions going toward 1.0, causing the CDFs to remain substantially apart. The full LFP model showed similar predictions to Figures 5c and 5d.
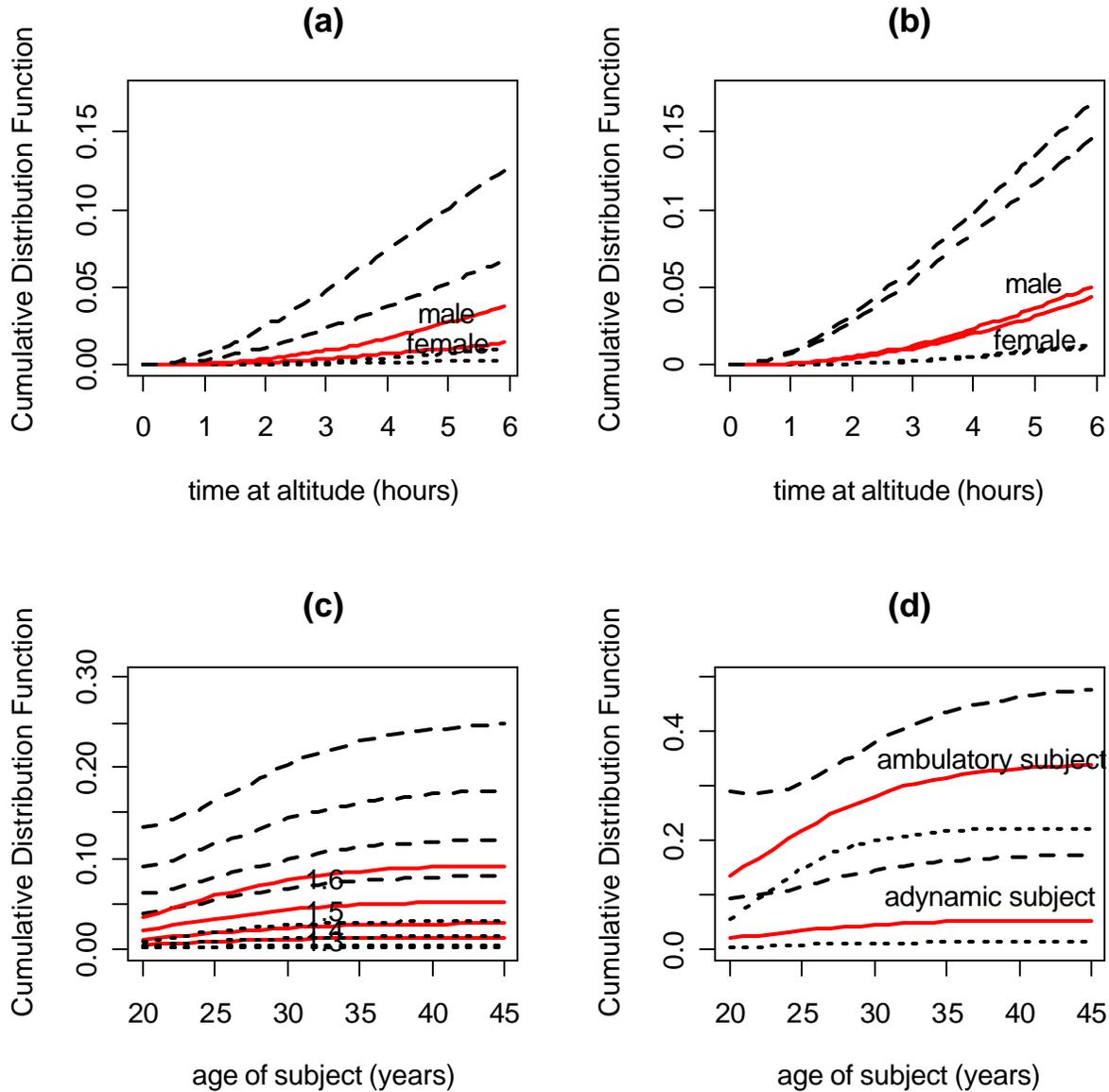
Figure 5. Estimated probabilities of Grade IV VGE, with 95% point-wise confidence intervals, as predicted by the reduced LFP Model. Dashed lines are upper points of the intervals; dotted lines are the lower points. The solid lines are point estimates.
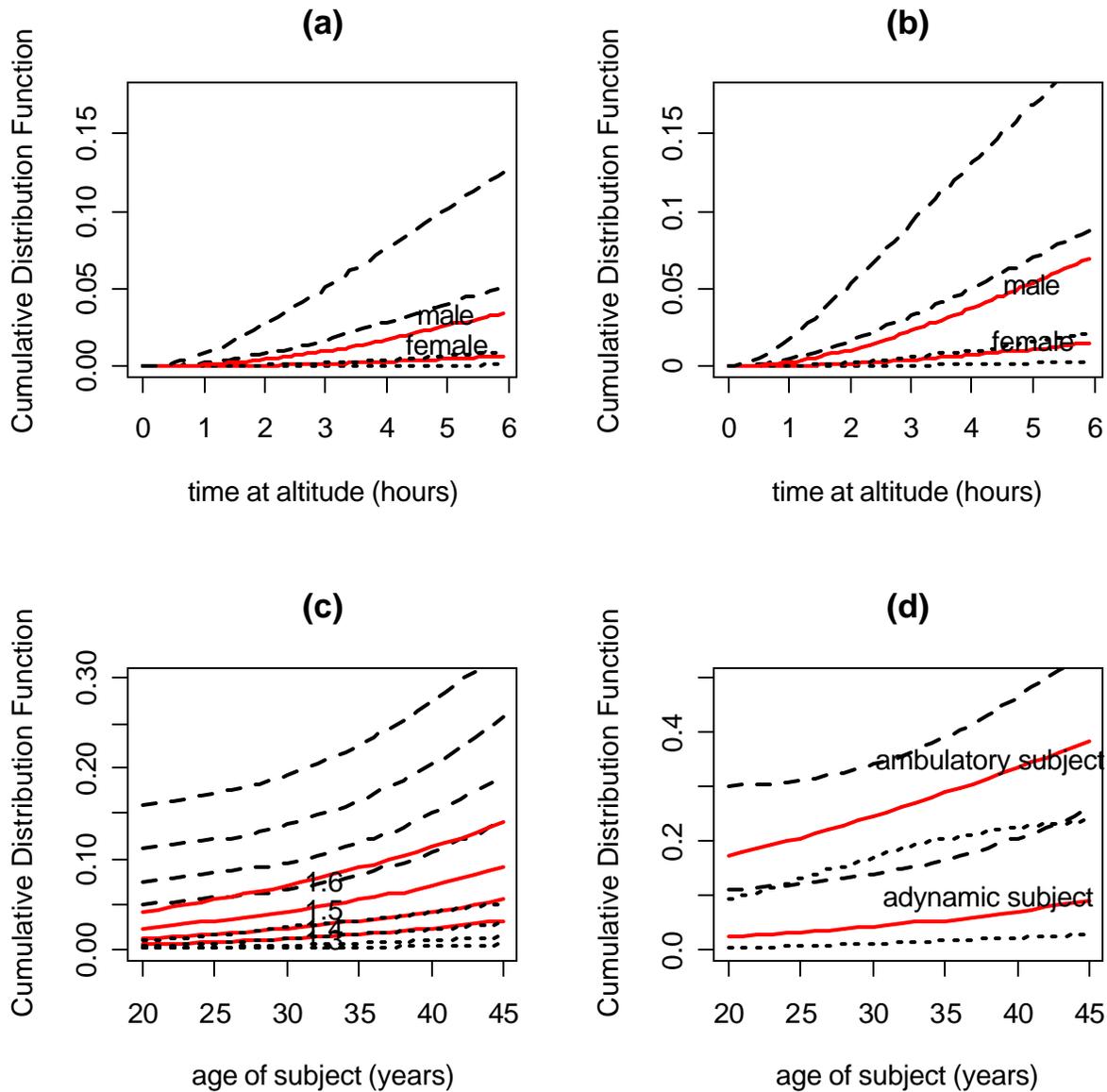
Figure 6. Estimate d probabilities of Grade IV VGE, with 95% point-wise confidence intervals, as predicted by the non-mixture model. Dashed lines are upper points of the intervals; dotted lines are the lower points. The solid lines are means.

Differences in mean predictions across tissue ratio and adynamia status become increasingly larger over age for the non-mixture model. Thus, for the non-mixture model we also see evidence of predicted interaction between age and other explanatory variables. However, in the dataset, the CDFs do not level off after age increases past the mean.

Thus, in comparison with a non-mixture model, the LFP model predicts on average, somewhat smaller probabilities of Grade IV VGE – as judged by comparing the vertical axes across the two figures. The differences in predictions across tissue ratios and adynamia status appear to increase over age, and then stabilize as age increases beyond the mean of about 32 years; for the non-mixture model, the differences continue to increase. The empirical proportions of Grade IV incidence by age categories and adynamia

24

status, as given previously in Table 2d, did not indicate an increasing gap in Grade IV incidence across adynamia status as age increased. Empirical proportions (not shown) of Grade IV incidence by age and TR360 categories also did not indicate increasing differences when comparing tissue ratios across age. However, whether the proportions in Table 2d are consistent with the stabilizing differences seen in Figures 5c and 5d is hard to infer.

Figures 5a and 5b versus Figures 6a and 6b make different predictions regarding the probability of Grade IV VGE over time for males and females at different ages (27 and 40 years). Although the figures show only two ages, generally as age increases, the reduced LFP model predicts that males and females have increasingly the same CDF; the non-mixture model, on the other hand, predicts that the CDFs for males and females remain different as age increases. The reduced LFP model predicts that when a tissue ratio is around 1.5, after a certain age – perhaps 40 years – lower body adynamic males and females are equally susceptible to Grade IV VGE. A comparison of the predicted CDFs to the empirical proportions given in Table 2c appears to corroborate the predictions of the non-mixture model in that the difference in Grade IV incidence for males versus females is higher for the 40 – +60 age category than for the 19 – +30 age category. However, the number of females beyond age 40 is relatively small. Also, the difference is lower for the 30 – +40 category, which would seem to support the reduced LFP model.

9. Discussion

We have fit two different types of models to the Grade IV VGE data: (1) a limited failure population lognormal mixture model and (2) a traditional lognormal non-mixture model. The reduced LFP mixture model appears to be a slightly better fit, as evidenced by its AIC value in Table 3 and the discussion of the goodness-of-fit graphs in Figure 4. Also, the LFP mixture model gives more accurate predictions overall, as seen in Table 4. However, a comparison of specific predictions in Figures 5 and 6 to cross-tabulations in Table 2 (a-d) do not appear to overwhelmingly favor either the non-mixture or the LFP mixture model.

An alternative to the reduced LFP model is to use the full LFP model. This model had good predictive accuracy and a good fit to the data. Moreover, the full model did not predict a narrowing difference in probabilities of Grade IV VGE by time across ages 27 and 40 for males and females. However, according to the asymptotic standard errors that we computed, the two extra parameters the model contained over the reduced model were clearly not significantly different from zero.

In Figure 7, we show boxplots of the predicted probabilities of eventually experiencing Grade IV VGE during an exposure for each of the two LFP models. The probabilities were computed for the covariate combinations in the dataset and the estimates in Table 3 and in equation (5.2). The boxes contain the middle 50% of the probabilities, and the horizontal lines with dots in the center represent the medians. Within the whiskers are 95% of the data. The two boxplots are somewhat similar; indeed, they show predictions almost anywhere between 0.20 and 1.0. However, the distribution of probabilities from the reduced model contains the bulk of its values at a higher probability. So, the reduced model may be more suitable for a population expected to contain few individuals immune to Grade IV VGE, where immunity is determined largely based on age and sex.

As a predictive model, the reduced LFP model of Table 3 is the best of the models presently investigated to predict the onset of Grade IV VGE in the population of volunteer subjects undergoing altitude chamber tests. A third LFP was investigated that contained the same parameters as those of the reduced LFP, but it also contained S.AGE in the location portion. The fitting of this third LFP model was almost identical to that of the reduced LFP model.
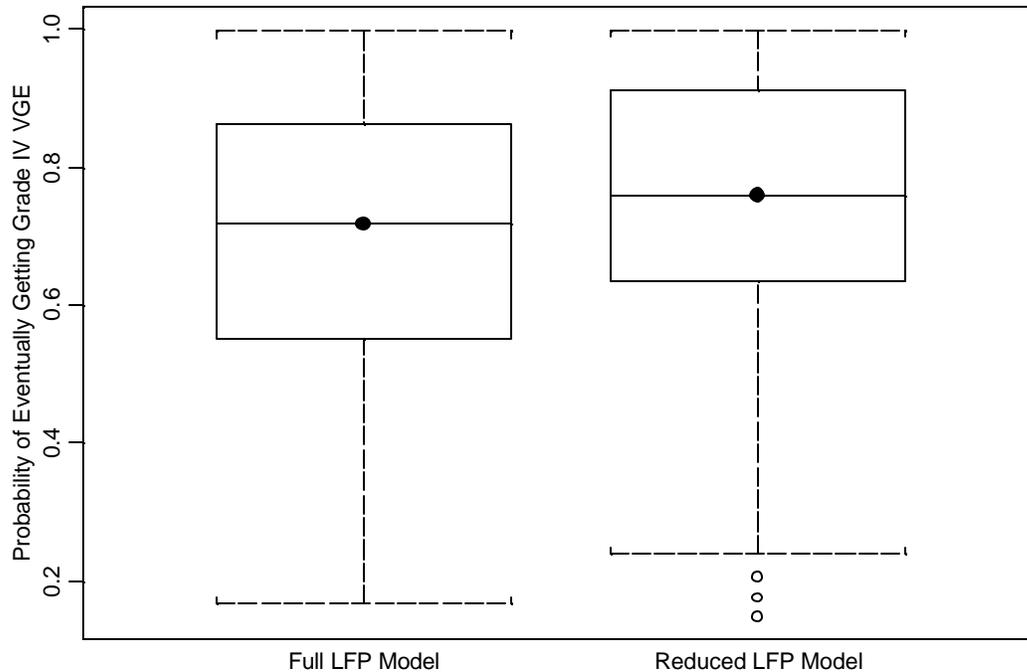
Figure 7. Boxplots of predicted probabilities of eventually experiencing Grade IV VGE for two LFP models.

10. Limitations and Extensions

We have shown that a limited failure population model might suitably describe the onset of Grade IV VGE in volunteer subjects undergoing testing in altitude chambers. The reduced LFP mixture model that is shown in Table 3 seems an appropriate candidate for prediction of onset time. This model is appropriate in a situation where age affects the predisposition for Grade IV VGE, but does not influence its *rate* of occurrence in a situation where Grade IV VGE will occur eventually. The general predictive ability of a non-mixture for future observations similar to the current set of observations does not appear to be as good as that of the LFP model we fit.

It is possible that other physical characteristics beside age and sex influence immunity to Grade IV VGE. One of these variables is body weight or body mass. Body mass was measured for individuals in the current dataset, but a decision was made early by the original researchers not to include this variable in their analysis because it had low predictive ability. Its inclusion in an LFP model may help predictive ability, however. It also may be necessary to look more closely at interactions among certain covariates, such as TR360 and NOADYN, as well as AGE and SEX.

More flexible models might be considered. For example, the generalized-F distribution, which contains the log normal distribution as a special case, can be used (Peng et al., 1996). To apply this distribution to interval-censored data 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 study (Chhikara et al., 2000) 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 we have only a few observed occurrences of Grade IV VGE in the right tail of the survival distribution.

Finally, the dataset we analyzed in this study contained many right-censored observations. Although the likelihood accounts for right-censoring, we did not specifically address the effect, if any, that heavy censoring may have in our goodness-of-fit procedures. Since heavy right-censoring is common in data obtained from altitude chamber tests (Koti et al., 1998; Chhikara et al., 2000; Conkin et al., 1998), this is a potential concern and limitation in our analysis.

References

Berkson, J. and Gage, R.P. (1952) Survival curve for cancer patients following treatment, *Journal of the American Statistical Association*, 47, 501-515.

Bove, A.A. (1998) Risk of decompression sickness with patent foramen ovale. *Undersea and Hyperbaric Medicine*, 25, 175-178

Brooks S. P. and Gelman A. (1998) Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics.* 7 , 434-455.

Carturan, D., Boussuges, A. Burnet, H., Fondaral, J., Vanuxem, P., Gardette, B. (1999). Circulating venous bubbles in recreational diving: relationship with age, weight, maximum oxygen uptake, and body fat percentage. *International Journal of Sports Medicine*, 20, 410-414.

Chhikara, R., Spears, F., and English, T. (2000). Modeling and Analysis for Risk Assessment of Decompression Sickness, ISSO Annual Report for 2000, University of Houston, TX.

Conkin, J., Bedahl, S., and Van Liew, H. (1992). A computerized databank of decompression sickness incidence in altitude chambers. *Aviation, Space & Environmental Medicine*, 63, 819-824.

Conkin, J. and Powell, M.R. (2001). Lower body adynamia as a factor to reduce the risk of hypobaric decompression sickness. *Aviation Space & Environmental Medicine*, 72, 202-214.

Conkin, J., Powell, M.R., Foster, P.P, and Waligora, J.M. (1998) Information about venous gas emboli improves prediction of hypobaric decompression sickness. *Aviation Space and Environmental Medicine*, 69, 8-16.

Conkin, J., Foster, P., Powell, M.R., and Waligora, J.M. (1996). Relationship of the time course of venous gas bubbles to altitude decompression illness, *Undersea and Hyperbaric Medicine*, 23, 141-149.

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

Farewell, V.T. (1982). The use of mixture models for the analysis of survival data with long-term survivors, *Biometrics*, 38, 1041-1046.

Gentleman, R. and Geyer, C. (1994). Maximum likelihood for interval-censored data: Consistency and computation. *Biometrika*, 81, 618-623.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Gordon, N.H. (1990). Application of the theory of finite mixtures for the estimation of "cure" rates of treated cancer patients, *Statistics in Medicine*, 9, 397-407.

Kannan, N. and Raychaudhuri, A. (1998) Survival Models for Predicting Altitude Decompression Sickness. USAF Technical Report AL/CF-TR-1997-0030, Brooks AFB, San Antonio, TX.

Koti, K., Chhikara, R., and Spears, F. (1998) Cox Regression Modeling and Analysis of Decompression Sickness Data. ISSO Annual Report for 2000, University of Houston, TX.

Kuk, A. Y. C. and Chen, C. H. (1992) A mixture model combining logistic regression with proportional hazards regression, *Biometrika*, 79, 531-541.

Kumar, K. and Powell, M. (1994). Survivorship models for estimating the risk of decompression sickness, *Aviation, Space, and Environmental Medicine*, 65, 661-665.

Larson, M.G., and Dinse, G.E. (1985). A mixture model for the regression analysis of competing risks data, *Applied Statistics*, 34, 201-211.

Lawless, J. (1982). *Statistical Models and Methods for Lifetime Data*. Wiley, NY.

Maller, R. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. Wiley, NY.

Meeker, W., and Escobar, L. (1998). *Statistical Methods for Reliability Data*. Wiley: NY.

Naylor, J. C., and Smith, A. F. M. (1982) Applications of a method for the efficient computation of posterior distributions, *Applied Statistics*, 31, 214-225.

Peng, Y., Dear, K., and Denham, J. W. (1996) A generalized –F mixture model for cure rate estimation, *Statistics in Medicine*, 17, 813-830.

Spencer, M.P. (1976). Decompression limits for compressed air determined by ultrasonically detected blood bubbles. *Journal of Applied Physiology*, 40, 229-235.

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

Sulaiman, Z. M., Pilmanis, A. A., and O'Connor, R. B. (1997). Relationship between age and susceptibility to altitude decompression sickness. *Aviation, Space & Environmental Medicine*, 68, 695-698.

Tanner, M. (1996) *Tools for Statistical Inference*. Springer-Verlag, NY

Taylor, J. M. (1995). Semi-parametric estimation in failure time mixture distributions, *Biometrics*, 51, 814-817.

Turnball, B (1976). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association*, 69, 169-173.

Webb, J. T., Pilmanis, A. A., Krause, K. M., Kannan, N. (1999) Gender and altitude-induced decompression sickness susceptibility. 70th Annual Scientific Meeting of the Aerospace Medical Society

Appendix A: Asymptotic Correlation Matrices for Fitted Models

Non-Mixture Model

	β_0	β_1 (SEX)	β_2 (S.AGE)	β_3 (S.TR360)	β_4 (NOADYN)	$\log \sigma$	$\log \sigma_b$
β_0	1.000						
β_1 (SEX)	-0.613	1.000					
β_2 (S.AGE)	-0.207	0.150	1.000				
β_3 (S.TR360)	-0.271	-0.001	0.022	1.000			
β_4 (NOADYN)	-0.671	-0.047	0.111	0.159	1.000		
$\log \sigma$	0.350	-0.098	-0.069	-0.273	-0.130	1.000	
$\log \sigma_b$	0.350	-0.159	-0.112	-0.080	-0.179	-0.075	1.000

Reduced LFP Mixture Model

	β_0	α_1 (SEX)	α_2 (S.AGE)	β_3 (S.TR360)	β_4 (NOADYN)	$\log \sigma$	$\log \sigma_b$
β_0	1.000						
α_1 (SEX)	0.463	1.000					
α_2 (S.AGE)	0.295	0.810	1.000				
β_3 (S.TR360)	-0.317	-0.077	-0.015	1.000			
β_4 (NOADYN)	-0.824	-0.123	-0.126	0.140	1.000		
$\log \sigma$	0.490	0.472	0.242	-0.268	-0.155	1.000	
$\log \sigma_b$	0.310	0.149	0.102	-0.074	-0.163	-0.046	1.000

Full LFP Mixture Model

	β_0	β_1 (SEX)	α_1 (SEX)	β_2 (S.AGE)	α_2 (S.AGE)	β_3 (S.TR360)	β_4 (NOADYN)	$\log \sigma$	$\log \sigma_b$
β_0	1.000								
β_1 (SEX)	-0.426	1.000							
α_1 (SEX)	0.418	0.272	1.000						
β_2 (S.AGE)	-0.365	0.016	-0.335	1.000					
α_2 (S.AGE)	0.194	0.162	0.727	0.085	1.000				
β_3 (S.TR360)	-0.258	-0.075	-0.149	0.077	-0.037	1.000			
β_4 (NOADYN)	-0.653	-0.151	-0.257	0.140	-0.175	0.155	1.000		
$\log \sigma$	0.458	0.164	0.607	-0.309	0.290	-0.300	-0.249	1.000	
$\log \sigma_b$	0.361	-0.036	0.252	-0.169	0.107	-0.094	-0.191	0.056	1.000

Appendix B: Details of the Parametric Bootstrapping procedure

Let $\{\hat{a}_1, ..., \hat{a}_4\}$ and $\{\hat{b}_0, \hat{b}_1, ..., \hat{b}_4\}$ denote point estimates of the mixture and location parameters, respectively, for the models studied. Note that some of these estimates will be zero. In particular, for the non-mixture model all of $\{\hat{a}_1, ..., \hat{a}_4\}$ will be zero. Also, let $\hat{s}$ and $\hat{s}_b$ be point estimates of the scale parameter and standard deviation of random effects (on the log scale), respectively.

For $i = 1$ to 238 subjects, and for $j = 1$ to n_i tests, each with prescheduled time at altitude, ALTTIME_{ij} and recorded interval, $(t_{0_{ij}} \quad t_{1_{ij}}]$:

(1) Draw a random effect, b_i, from a normal distribution with mean 0 and standard deviation, $\hat{s}_b$.

(2) Divide the time axis from 0 to ALTTIME_{ij} hours into alternating four-minute monitoring intervals and 12-minute non-monitoring, or "resting," intervals. This scheme reflects the intervals of assessment designed in the study. (Naturally, not all time axes between 0 and ALTTIME_{ij} divided evenly into four- and 12-minute intervals. In cases where the time axis does not divide evently, we truncated the ending interval to fit the time frame.)

(3) Using the values on the explanatory variables, $\mathbf{x}_{ij}$, with probability in (5.2), generate a random variate, T_{ij}, from a lognormal distribution with location parameter, $m = \hat{b}_0 + \sum_{k=1}^{p} \hat{b}_k x_{ijk}$, and scale parameter, $s = \hat{s}$. Otherwise, let $T_{ij} = \infty$.

That is, use (5.2) to compute the estimated probability that the observation will eventually experience Grade IV VGE. (For the non-mixture model, the probability is one.) With this probability, the generated random variate comes from the indicated lognormal distribution. Otherwise, it is infinite. Note that the set of covariates $\mathbf{x}_{ij}$ used for location and mixture portions may differ.

(4) Simulate Type I right-censoring: If $T_{ij} < \text{ALTTIME}_{ij}$, determine which interval T_{ij} falls in by comparing T_{ij} with the divided time axis in (2). Record an interval for T_{ij}, using the following scheme: (a) If T_{ij} falls within a 12-minute "resting" interval, record the 12-minute interval; (b) if T_{ij} falls within a four-minute monitoring interval (not including the first four-minute interval), record the endpoints of the 12-minute resting interval that comes immediately prior to the four-minute interval; (c) if T_{ij} falls within the first four-minute interval, record the interval from time zero to T_{ij}; and finally, (d) if $T_{ij} > \text{ALTTIME}_{ij}$, or if T_{ij} falls within the final 12-minute interval (if there is a final 12-minute interval), call the observation Type I right-censored at ALTTIME_{ij}.

The interval scheme described in (2) above did not appear to greatly influence goodness-of-fit plots. In fact, we tried many different schemes, including intervals of one fixed width, and the results were virtually identical.

Note that we did not simulate random censoring, because none of our models accounted for the sparse existence of random censoring in the data. These cases were instead treated as Type I censored.

Appendix C: K-fold Adjusted Cross-Validation Procedure for GIV VGE models

We assessed the aggregate prediction error of each model using a K-fold cross-validation algorithm. In general, K-fold cross-validation repeatedly splits the data into K disjoint sets of nearly equal size (say, $C_1,\ldots, C_k$). These K sets define K different splits into training and test sets, with each set C in turn acting as a test set and the remaining sets together acting as a training set. Prediction error is then calculated once for each observation, and the average prediction error is obtained. This average is called the *K-fold cross-validation estimate of prediction error* (e.g., Davison and Hinkley, 1997). A practical rule of thumb is to take $K = \min(n^{1/2}, 10)$.

A further adjustment corrects for a bias on the order of $1/(K-1)$ in the K-fold cross-validation estimate of prediction error. This adjustment adds a term that represents the resubstitution error minus the sum of the averaged prediction errors resulting from using the model fit to each of the K subsets, in the proportion each subset comprises in the data. The following steps may clarify the algorithm (see Davison and Hinkley, 1997, p. 294-295 for full details).

Algorithm for K-fold Adjusted Cross-validation

(1) Fit the model to all cases $\{\ell = 1,\ldots, n = 548\}$. Obtain predicted log times for the ℓ th case using

$$m(\mathbf{x}_\ell, \hat{F}) = \log t_\ell = \begin{cases} \hat{b}_0 + \sum_{j=1}^{k} x_{\ell k} \hat{b}_k & \text{with probability } \left(P(z_\ell = 1 \mid \hat{a}, \mathbf{x}_\ell)\right)^{1-d_\ell} \\ \infty & \text{otherwise} \end{cases}$$

where d_ℓ is defined in (5.1).

(2) Compute the averaged squared error using these predicted values,

$$D(\hat{F}, \hat{F}) = n^{-1} \sum_{\ell=1}^{n} c\left\{\log t_{0_\ell}, \log t_{1_\ell}, m(\mathbf{x}_\ell, \hat{F})\right\}$$

where

$$c\left\{\log t_0, \log t_1, m(\mathbf{x}, \hat{F})\right\} = \begin{cases} 0 & \text{if } \log t_0 \le m(\mathbf{x}, \hat{F}) \le \log t_1 \\ \left(\min\left\{\left|\log t_0 - m(\mathbf{x}, \hat{F})\right|, \left|\log t_1 - m(\mathbf{x}, \hat{F})\right|\right\}\right)^2 & \text{otherwise} \end{cases}$$

(3) Divide the cases into K disjoint groups of sizes (approximate equal) m1, …, mK

For each $k = 1,\ldots,$ K,

(4) Fit the model to all data except cases in the kth group.

(5) Calculate predictions for all observations using this model, and calculate

$$D(\hat{F}, \hat{F}_{-k}) = n^{-1} \sum_{\ell=1}^{n} c\left\{\log t_{0_\ell}, \ \log t_{1_\ell}, \ \mathbf{m}(\mathbf{x}_\ell, \hat{F}_{-k})\right\}$$

(6) Calculate the K-fold cross-validation estimate of prediction error:

$$\hat{\Delta}_{CV,K} = n^{-1} \sum_{k=1}^{K} \sum_{i=1}^{m_k} c\left\{\log t_{0_i}, \ \log t_{1_i}, \ \mathbf{m}_{ik}(\mathbf{x}_{ik}, \hat{F}_{-k})\right\}$$

(7) The adjusted estimate is

$$\hat{\Delta}_{ACV,K} = \hat{\Delta}_{CV,K} + D(\hat{F}, \hat{F}) - \sum_{k=1}^{K} \frac{m_k}{n} D(\hat{F}, \hat{F}_{-k})$$

(8) Repeat steps (3) through (7) M times, and average the prediction errors.

Note that in this algorithm, the predicted values are the log times to onset (which may be infinite), and that the error in prediction is defined by the distance of the predicted value from the nearest endpoint of the recorded interval, if the predicted value falls outside of the interval, and zero otherwise.

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE May 2002	3. REPORT TYPE AND DATES COVERED NASA Technical Paper

4. TITLE AND SUBTITLE Modeling Grade IV Gas Emboli Using a Limited Failure Population Model with Random Effects	5. FUNDING NUMBERS
6. AUTHOR(S) Laura A. Thompson,* Johnny Conkin,** Raj S. Chhikara,* and Michael R. Powell***	

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Lyndon B. Johnson Space Center Houston, Texas 77058	8. PERFORMING ORGANIZATION REPORT NUMBERS S-891

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546-0001	10. SPONSORING/MONITORING AGENCY REPORT NUMBER TP-2002-210781

11. SUPPLEMENTARY NOTES
*UH-CL, School of Natural and Applied Sciences, Houston, TX; **National Space Biomedical Research Institute, Houston, TX; NASA-Lyndon B. Johnson Space Center, Houston, TX

12a. DISTRIBUTION/AVAILABILITY STATEMENT Available from the NASA Center for AeroSpace Information (CASI) 7121 Standard Hanover, MD 21076-1320　　　　Category: 52	12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Venous gas emboli (VGE) (gas bubbles in venous blood) are associated with an increased risk of decompression sickness (DCS) in hypobaric environments. A high grade of VGE can be a precursor to serious DCS. In this paper, we model time to Grade IV VGE considering a subset of individuals assumed to be immune from experiencing VGE. Our data contain monitoring test results from subjects undergoing up to 13 denitrogenation test procedures prior to exposure to a hypobaric environment. The onset time of Grade IV VGE is recorded as contained within certain time intervals. We fit a parametric (lognormal) mixture survival model to the interval- and right-censored data to account for the possibility of a subset of "cured" individuals who are immune to the event. Our 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. We also evaluated some alternative mixture models. Predictions from the best fitted mixture model indicate that the actual process is reasonably approximated by a limited failure population model.

14. SUBJECT TERMS venous gas emobli; bubbles; decompression sickness; bends; extravehicular activity; goodness of fit; accuracy, predictive; random effects; distribution, lognormal		15. NUMBER OF PAGES 46	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unlimited	20. LIMITATION OF ABSTRACT Unlimited