Using concurrent cardiovascular information to augment survival time data for evaluating orthostatic tilt test performance

Alan H. Feiveson[1], James Fiedler[2], Stuart M. C. Lee[3],
Matthew D. Koslovsky[3], Michael B. Stenger[1], and Steven H. Platts[1]

[1]National Aeronautics and Space Administration, Johnson Space Center
[2]Universities Space Research Association, Division of Space Life Sciences
[3]KBRwyle, Science & Space

January 11, 2018

Abstract

Head-up tilt (HUT) tests often are used in research to measure orthostatic intolerance (OI) (inability to appropriately control blood pressure while upright) in clinical populations and otherwise healthy individuals after interventions. Post-spaceflight orthostatic intolerance is a well-known phenomenon, and countermeasures to its development has been an active area of research at NASA. In the NASA HUT protocol, subjects lie horizontally on an automatic tilt table for baseline measurements before being raised to 80° head-up tilt for a defined period of time or until signs or symptoms of presyncope ensues (light-headedness, nausea, dizziness, sweating, weakness or fainting). Multiple measures are collected to evaluate the cardiovascular system's ability to respond appropriately to the orthostatic challenge. However if the intended duration of the HUT is short, the ability to detect changes in OI due to an intervention or its prevention by a countermeasure may be limited by a small number of failures to permit comparisons based on survival time alone. Thus, the time-trajectory of the cardiovascular data becomes an important additional source of information. In

particular, we will show how various measures of trajectory variability can effectively augment survival analysis for the assessment of OI in a joint model when high censoring rates are present.

Keywords: orthostatic intolerance, joint models, longitudinal, measurement error, spaceflight, surrogate measures

1 Introduction

Cardiovascular deconditioning is a well-accepted consequence of spaceflight, with manifestations including an increased incidence of orthostatic intolerance (OI) (i.e., an inability to maintain appropriate blood pressure and cerebral blood flow in the upright posture), after both short- (Buckey et al., 1996; Fritsch-Yelle et al., 1996) and long-duration missions (Meck et al., 2001; Lee et al., 2011). For US astronauts, post-spaceflight OI was first observed in the Mercury program (Catterson et al., 1963) and was consequently noted in subsequent space programs, from the early Gemini and Apollo missions to those involving the International Space Station (Lee et al., 2011). Although different methodologies exist, OI has been assessed in the US space program since 1997, primarily using a protocol in which subjects are supine for a period of time and then moved to 80° head-up tilt (HUT). In related research, HUTs have also been used to evaluate or compare prospective countermeasures to OI in ground studies that use either bed rest (Platts et al., 2009) or pharmacologically-induced hypovolemia (Platts et al., 2009) to simulate some of the cardiovascular effects seen in spaceflight.

In practice, the most often used and operationally-relevant HUT outcome for assessing OI has been the amount of time a subject can remain upright without experiencing considerable decrease in blood pressure or the event of presyncope (i.e., lightheadedness, dizziness, nausea). However, subjects' HUT survival times become censored at the intended length of test; therefore censoring rates vary across studies. For example, one group of researchers has run HUTs for more than two hours (Ramsdell et al., 2001), but shorter-duration tests are more common. This is particularly the case for spaceflight studies conducted on the day when the crew return to Earth, since there are a large number of investigators studying the astronauts, and the availability of the crew to participate in testing is limited. In studies of

Space Shuttle astronauts, the upright portion of the HUT was limited to 10 minutes, which produced a high percentage of censored survival times following short-duration flight (e.g., 73% in a recent study (Lee et al., 2011)). In this setting, using survival analysis alone may fail to distinguish between groups of subjects, interventions, or countermeasures. Thus, it would be highly desirable to make better use of the longitudinal cardiovascular information typically obtained during HUTs (e.g., heart rate, blood pressure, stroke volume) to improve inferential efficiency. In some NASA studies or as part of mission operations, summaries of this ancillary data have been reported, but have not been used in a formal statistical analysis to compare or evaluate countermeasures.

In practice, the amount of censoring observed in a study can guide the analytical approach. For example in a study where all subjects fail to complete the test, it could be argued that survival time is the only necessary information regarding OI and the concurrent cardiovascular data is superfluous. Alternatively, if all subjects completed the HUT to the intended test duration, all survival times would be censored and the longitudinal cardiovascular data would be the only outcome that would vary. Often these type of data are modeled with multivariate functional data analysis methods, as described in Ramsay and Silverman (2005).

Here we consider the middle ground, where some tests go to completion and others are cut off at presyncope. In this setting, an analysis based only on survival times does not take full advantage of the information contained in the longitudinal cardiovascular data and would thus be inefficient. We expect that a subject's cardiovascular condition at the time of a HUT would be directly related to the probability of experiencing presyncope that would terminate the test, but we cannot measure a subject's cardiovascular condition directly; instead we approximate it with surrogate measures, derived from the time course of heart rate, blood pressure, etc. during the HUT. Therefore, measurement error is introduced

due to this latency. Furthermore, the very fact that the surrogate measures could be observed at a particular time implies that the subject has not yet failed the test - a defining characteristic of an internal or endogenous covariate.

Joint models for longitudinal and survival data can be used to model survival outcomes that are related to endogenous, time-dependent covariates with measurement error (Rizopoulos, 2014, 2012; Tsiatis and Davidian, 2004). Wu et al. (2011) provide a detailed review of standard joint model formulations and approaches for inference. Typically, linear, mixed effects models are fit to the observed longitudinal data to accommodate measurement error and provide an estimate for their underlying true trajectories. When linear assumptions for the underlying trajectories are violated and the distribution of the longitudinal data is unknown, semi-parametric approaches have been employed (Song et al., 2002; Ye et al., 2008). However, a semiparametric approach may produce over-fitted models when sample sizes are small. In some research settings, it may be of interest to additionally investigate the relation between the true rate of change and time-to-event (Ye et al., 2008). Typically, when fitting joint models, the time-to-event data is modeled with Cox proportional hazards models. However, other methods for handling time-to-event data structures (e.g., fully parametric models, recurrent events, or clustered data) have previously been explored (Ratcliffe et al., 2004; Han et al., 2007).

This work was motivated by data collected during a single-arm 30-day bed rest study in which 27 subjects were enrolled to perform HUTs pre- and post-bed rest; however, 8 subjects dropped out before the bed rest was completed and thus did not perform a second HUT. For each of these 46 HUT sessions, survival times and eight cardiovascular measures were recorded, the latter as 1-minute averages. Testing was terminated after 30 minutes if presyncope did not occur, resulting in a fairly high degree of censoring for HUT pre-bed rest. In this data, we observed that the time histories of the cardiovascular data

were considerably more erratic during post-bed rest HUTs, especially when survival times were short. Since it is also well known that bed rest tends to increase OI (Meck et al., 2009), we conjectured that a summary index of change in multivariate variability of the cardiovascular, longitudinal data within HUTs, especially when combined with survival times, would be an important predictor for comparing levels of OI between experimental conditions (in this case pre- versus post-bed rest). Heart rate and blood pressure normally vary across time but are relatively stable in subjects who are able to compensate for the gravity-induced shift of blood to the lower body while standing. In contrast, subjects who develop presyncope or syncope typically have greater variability in heart rate and blood pressure while upright, with the magnitude and/or frequency of the oscillations increasing as the duration of standing increases and decompensation ensues (Julu et al., 2003; Lipsitz et al., 1997).

Note that the bed rest data arose from a partially repeated-measures design, since 19 subjects participated in HUTs both pre- and post-bed rest. In this sense, there are actually two longitudinal time scales to consider - within and between the two HUTs for these 19 subjects. Therefore, our main objective is to develop a strategy for characterizing the variability in the within-HUT longitudinal cadiovascular data that can be jointly modeled with HUT time-to-event data. For this and similar data, joint models must also take into account the larger-scale longitudinal character of the data imposed by the partially repeated-measures design. For the joint model, we take a two-stage approach. First, we model the trajectory of each individual's variability summary index over time with median regression pre- and post-bed rest. Next, instead of directly using fitted values from stage one as covariates in a survival model, we jointly model the estimated rates of change obtained from the median regression along with survival times to accommodate trend model uncertainty, subject-specific random effects, and measurement error. We validate our

proposed method through simulation to assess its performance in evaluating a hypothetical countermeasure against OI.

Our secondary objective is to compare different methods of summarizing the behavior of cardiovascular variables collected during HUT sessions and identify which sets of cardiovascular measures (e.g., heart rate, systolic, diastolic blood pressures) collected during a HUT should be used to form variability indices that best relate to impending presyncope. Finally, we apply our method to actual HUT data to make inference on the effect of bed rest in the presence of a high degree of censoring using the optimal variability index. This analysis may help NASA researchers effectively design HUT studies with limitations on time and measurement devices available to identify effective countermeasures for bed rest and ultimately help mitigate the effects of microgravity on astronauts returning to Earth.

The remaining sections of this paper are outlined as follows. In Section 2, we propose our joint model and define a set of potential variability summary indices. In Section 3, we conduct a simulation study to evaluate the performance of our method. In Section 4, we apply our method to data collected during HUTs to identify variability indices that best characterize the relation between a set of cardiovascular measures and presyncope before and after bed rest. Finally in Section 5, we provide a brief discussion and concluding remarks.

2 Methods

2.1 Subjects

Healthy men and women (normotensive, non-smokers of normal body weight) volunteered to participate in 6° head-down bed rest studies at the General Clinical Research Center

(GCRC) Satellite Flight Analogs Research Unit (FARU) at the University of Texas Medical Branch (UTMB) in Galveston, TX. Subjects were screened using a modified Air Force Class III physical as well as psychological tests to determine suitability for participation in the bed rest study (Meck et al., 2009). Exclusion criteria included hypertension; electrocardiogram (ECG) abnormalities; chronic medication usage that might complicate interpretation of the results; recent sub-standard nutritional status; history of thyroid dysfunction or renal stones; mental illness; gastroesophageal reflux; cardiovascular disease; musculoskeletal or sensorimotor dysfunction; history of thrombosis; a body mass index outside of 20-30; and abnormal blood or urine clinical values (Meck et al., 2009). Bed rest and test protocols were reviewed and approved by the Institutional Review Boards of NASA Johnson Space Center and the University of Texas Medical Branch (UTMB) as well as the UTMB GCRC Science Advisory Committee. Subjects received verbal and written explanations of the bed rest and test protocols prior to providing written informed consent.

2.2 Bed rest conditions

Pre-, in-, and post-bed rest conditions for subjects were maintained in a manner consistent with the NASA standard protocols, as described previously (Meck et al., 2009). Subjects were admitted to the FARU and remained ambulatory before the bed rest portion of the study commenced. During this pre-bed rest period, subjects were acclimated to the diet and wake-sleep times, and were familiarized with the in-bed rest procedures (e.g., movement restrictions, eating, showering, bed pan use for urination and defecation). Within the bed rest portion of the study, subjects assumed a posture of lying 6° head-down for 24 hours/day, while similar conditions of diet, sleep/wake times, etc. were maintained. During the post-bed rest phase of the study, subjects were again ambulatory and remained in the FARU, during which time post-bed rest data were collected. Subjects participated in a

reconditioning program before they were discharged from the unit.

In this study, 27 healthy volunteers (17M, 10F) participated in a 30-day bed rest without any countermeasures. These subjects performed HUTs five days before the start of the bed rest period. Nineteen subjects also performed the post-bed rest HUT, which actually took place on the last day of bed rest and was the first time that the subjects assumed an upright posture since bed rest began. Eight subjects did not complete the 30-day bed rest and were part of a longer study.

2.3 Tilt test

Upon arrival at the testing station, subjects were directed to lie supine upon an automated tilt table. Prior to data collection, 2-dimensional echocardiography was used to obtain the aortic annulus diameter from the parasternal long axis during supine rest. Subjects were instrumented to measure ECG (Escort II, MDE, Arleta, CA), beat-to-beat blood pressure in the finger (Finapres 2300 blood pressure monitor, Ohmeda, Englewood, CO), and blood pressure each minute in the brachial artery (Dinamap XL Vital Signs Monitor, GE Medical Systems Information Technologies, Milwaukee, WI). The hand utilized for the finger blood pressure measurement was supported to the side using an arm board; the height of the arm support was adjusted so that the finger remained approximately at heart level during upright tilt. Six minutes of supine data were collected prior to tilting the subjects to 80° head-up tilt at a rate of approximately 10° per second. In this study, subjects remained in the tilted position for 30 minutes or until symptoms of orthostatic hypotension (systolic blood pressure below 70 mmHg, drop in heart rate $>$ 20 beats/min) and/or presyncope occurred (lightheadedness, dizziness, or nausea).

During both the supine and tilt periods, in addition to heart rate (HR), systolic blood pressure (SBP), and diastolic blood pressure (DBP), ascending aortic blood velocity time

integral was measured each beat using pulse wave Doppler measurements made at the suprasternal notch using a 2 MHz probe (Biosound MyLab 30, Esoate, Indianapolis, IN). Images from three cardiac cycles collected at a time corresponding to acquisition of heart rate and blood pressure data were digitally recorded and stored for offline analysis. Ascending aortic blood velocity time integral images were independently analyzed by two sonographers (Meck et al., 2001). Stroke volume (SV) was calculated as annulus diameter $\times$ velocity time integral. Other calculated variables, all thought to reflect important aspects of cardiovascular function during orthostasis were cardiac output (CO; $CO = SV \times HR$), mean arterial pressure (MAP; $MAP = (SBP + 2 * DBP)/3$), total peripheral resistance (TPR; $TPR = MAP/CO$), and pulse pressure (PP; $PP = SBP - DBP$). For each HUT, the data we analyzed consisted of time to presyncope (right-censored at 30 minutes) and values of the above eight cardiovascular responses averaged over the last minute of the supine period as well as over each minute of tilt.

2.4 Development of analysis methods

2.4.1 Motivation

Our approach is predicated on the assumption that there are aspects of the joint behavior of the multivariate, cardiovascular data collecting during the observation window, which provide more information about a subject's state of OI when combined with survival time than the survival time itself. This assumption was motivated by data observed in the HUT study. For example, Figure 1 shows the time trajectories of HR, DBP, and MAP for Subject 1 pre- and post-bed rest. In both cases, this subject completed the 30-minute test (i.e., time-to-events were censored at 30 minutes), but note the increased variability in DBP and MAP after bed rest as well as the increased jump in heart rate from the supine

measurement ($t = 0$). Also, this subject's heart rate is much higher after bed rest overall. Presumably, all these effects are manifestations of increased OI after bed rest, which for this subject would not be detectable in terms of survival time alone with a 30-minute test. In such situations, we assume knowledge of the behavior of the cardiovascular data would enable us to better distinguish between these two results.

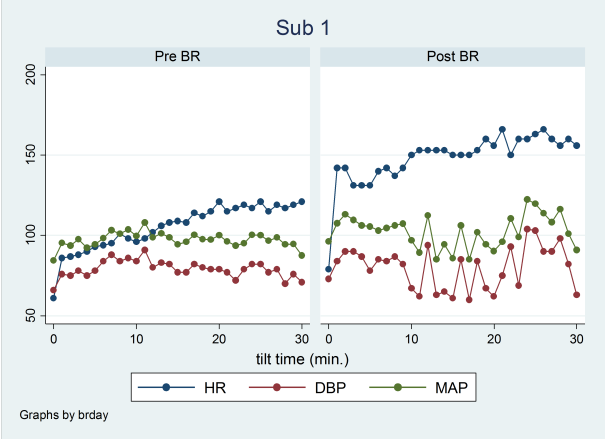

Figure 1: Time traces of heart rate (HR), diastolic blood pressure (DBP) and mean arterial pressure (MAP) pre- and post-bed rest (BR) for Subject (Sub) 1, who completed both pre- and post-bed rest HUT sessions.

Figure 2 shows a similar plot for Subject 2, who completed the 30 minutes of HUT pre-bed rest test, but failed to complete the post-bed rest test (survival time 9 minutes). Again, there is considerably more variability in these cardiovascular measures for the post-bed rest session.

Figures 1 and 2 and others like them (not shown) suggest that increased OI results in not only shorter survival times, but also in more erratic behavior of the cardiovascular trajectories. This led us to seek measures of multivariate variability of the cardiovascular data that could be combined with survival time data in such a way as to allow valid inference

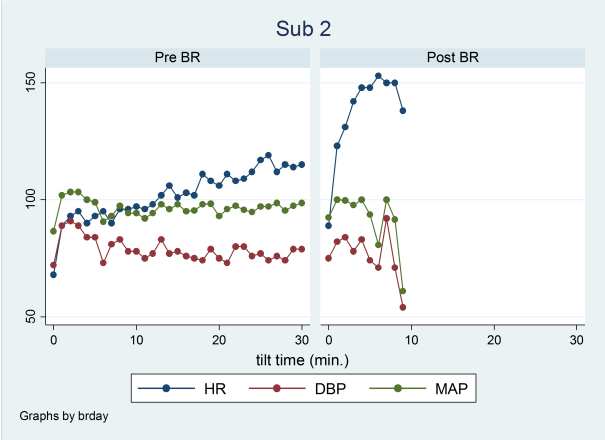

Figure 2: Time traces of heart rate (HR), diastolic blood pressure (DBP) and mean arterial pressure (MAP) pre- and post-bed rest (BR) for Subject (Sub) 2, who completed the pre-bed rest session but failed the post-bed rest session in the 10^{th} minute.

on the effect of bed rest on OI, even with a high percentage of censored outcomes.

2.4.2 Variability summary index

In this section, we propose multiple approaches for summarizing the joint variability of the cardiovascular data in terms of their trajectory in an M-dimensional Euclidean coordinate space, $\mathbb{R}^M$ traced out by the vector-valued function $\boldsymbol{Y}(t) = (Y_1(t), .., Y_M(t))'$ over time. Note that M is the number of cardiovascular measures taken at a particular time. In general, the more "space" in $\mathbb{R}^M$ this trajectory covers, the more variability. As an example, Figure 3 shows the trajectories of HR and DBP for Subject 1 ($M = 2$). Note how the trajectory in $\mathbb{R}^2$ takes up considerably more space post-bed rest than for pre-bed rest.

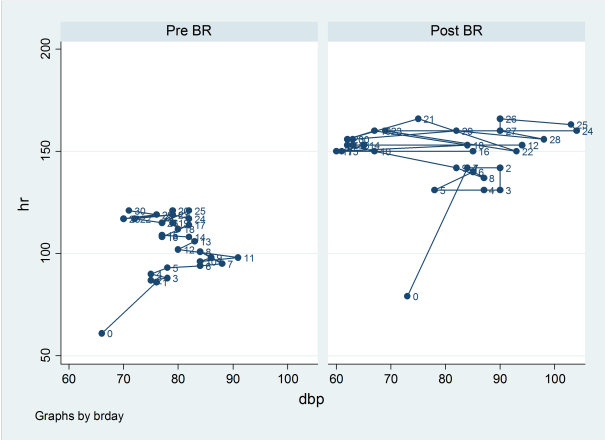

Figure 3: Trace of heart rate (HR) vs diastolic blood pressure (DBP) during HUT pre- and post-bed rest (BR) for Subject 1 who completed both tests. Plotted numbers are times (t_k) of tilt in minutes. The supine time point is denoted by $t = 0$.

Motivated by the patterns in Figure 3, we propose three approaches to summarizing the cumulative multivariate variability of $\boldsymbol{Y}(t)$ up to time t_k in terms of an index V_k:

Approach 1: M^{th} Root of Convex Hull Content

$$V_k = H_k^{1/M}, \tag{1}$$

where H_k is the content of the convex hull formed by the points $\boldsymbol{y}(t_1), \boldsymbol{y}(t_2), \ldots, \boldsymbol{y}(t_k)$ in $\mathbb{R}^M$. Here, $\boldsymbol{y}(t_1), \boldsymbol{y}(t_2), \ldots, \boldsymbol{y}(t_k)$ are M-dimensional vectors consisting of $\boldsymbol{Y}(t)$ at discrete, common time points $\{t_1, t_2 \ldots, t_K\}$ (see Supplementary Appendix I for details)

Approach 2: M^{th} Root of Determinant of Covariance Matrix

$$V_k = |\boldsymbol{S}_k|^{1/M}, \tag{2}$$

where the sample covariance matrix $\boldsymbol{S}_k = 1/(k-1)\sum_{r=1}^{k}(\boldsymbol{y}(t_r) - \bar{\boldsymbol{y}}_k)(\boldsymbol{y}(t_r) - \bar{\boldsymbol{y}}_k)'$ and $\bar{\boldsymbol{y}}_k = 1/k\sum_{r=1}^{k}\boldsymbol{y}(t_r)$.

Approach 3: Path Length

$$V_k = \sum_{r=1}^{k} |\boldsymbol{y}^*(t_r) - \boldsymbol{y}^*(t_{r-1})|, \tag{3}$$

which is equal to the sum of the absolute differences between the observed M-dimensional vectors of cardiovascular data at successive time points. In Equation (Eq.) 3, $\boldsymbol{y}^*$ denotes a vector of normalized components. Normalization is necessary to prevent change in one or a few components from dominating the change in path length. We compute the normalized value of a component as the ratio of its original value to the pre-bed rest average of that component over all subjects in the supine position prior to tilt. Note, normalization in Eq. 1 and 2 is unnecessary since scaling any one of the components of $\boldsymbol{y}$ also scales the root determinant or convex hull accordingly.

2.4.3 Model formulation

Suppose subjects perform HUTs for a scheduled maximum of T^* minutes, where we assume T^* is the same for all experimental sessions indexed by j. In our application, $j = 1$ refers to pre-bed rest and $j = 2$ refers to post-bed rest. Let v_{ijk} denote one of the summary index measures, V_k, given by Eqs. 1, 2, or 3, evaluated for all observations of $\boldsymbol{y}(t_r)$ up to time point t_k $(k = 1, \ldots, K_{ij})$ for the i^{th} subject during the j^{th} session. Here, $t_{K_{ij}}$ is the last

time at which $\boldsymbol{Y}(t)$ is observed. Then, the survival time $T_{ij} = t_{K_{ij}}$ if $t_{K_{ij}} < T^*$; otherwise T_{ij} is censored at T^*.

2.4.4 Survival model

We assumed that given the i^{th} subject, the survival time distribution for the j^{th} session would follow a Weibull distribution with survival function

$$S_{ij}(t) = P(T_{ij} > t) = \exp\left\{-\left(\frac{t}{\theta_{ij}}\right)^{p_j}\right\}. \tag{4}$$

We chose to use a fully parametric survival model to support inference on the session effect (in this case, bed rest), as well as for clinical assessment of OI, where it is desirable to predict the median survival time from the cardiovascular history in censored cases. In the above parameterization of the Weibull distribution (Mudholkar and Srivastava, 1993), θ_{ij} is the scale parameter and p_j is known as the shape parameter. Here, we allow p_j to depend on the session, j, as well as other time-independent covariates $\boldsymbol{W}_j$ in a loglinear submodel $\log p_j = \boldsymbol{\alpha}'\boldsymbol{W}_j$.

2.4.5 Longitudinal model

Recall, our primary interest was to model the relation between survival and the rate of change over time in the summary index, as opposed to modeling the summary index directly. Let us first assume that v_{ijk} is an estimate of a continuous process $V^*_{ij}(t)$ $(t_1 \leq t \leq T_{ij})$ measured at the time points $t = t_k$:

$$v_{ijk} = \boldsymbol{V}^*_{ij}(t_k) + e_{ijk}, \tag{5}$$

where e_{ijk} represents measurement error and v^*_{ijk} is equivalent to $V^*_{ij}(t)$ evaluated at time, t_k. In the study data, we observed a generally linear, positive relation between v_{ijk} and time. At this stage, we could have modeled v_{ijk} directly with Eq. 5 in the joint model, assuming normally distributed errors and random slopes, and set the random slopes as the shared parameter between the within-HUT longitudinal and survival data following Rizopoulos (2014). However, we conjectured these assumptions were too strong for the study data. Thus, we assume $V^*_{ij}(t)$ is approximately linear such that B_{ij}, the average slope of $V^*_{ij}(t)$ over the interval $[t_1, T_{ij}]$, is still a good single indicator of how fast $V^*_{ij}(t)$ (and hence cardiovascular instability) increases.

Rather than applying typical mixed model assumptions to Eq. 5 (e.g., normally distributed errors and random slopes), we used median regression for each combination of i and j to obtain a slope b_{ij} as an estimate of B_{ij}. Use of median regression avoids having to specify a distributional model for e_{ijk} in Eq. 5 and also is robust to moderate departures from linearity. To emphasize the higher relative importance of the variability trend in early in a HUT, the median regression was run using weights proportional to $1/t_k$.

Now, we treat the slopes of these individual models, b_{ij}, as our observed measures for each subject i at session j, and accommodate error introduced into these measures from potential misspecification of the mean trajectory, as well as the distribution for the summary index around the mean. Since two of the candidate dispersion measures (path length and convex hull content) are non-decreasing, while the other measure (determinant) should generally increase with only occasional slight decreases, we expect B_{ij} to be positive. Thus, given some session-dependent covariates $\boldsymbol{X}_j$, this suggests the use of a log-linear mixed model:

$$\log b_{ij} = \boldsymbol{\beta}' \boldsymbol{X}_j + a_i + e_{ij} \tag{6}$$

where $\boldsymbol{\beta}$ is a coefficient vector, $a_i \sim N(0, \sigma_a^2)$, and $e_{ij} \sim N(0, \sigma^2)$. It follows that conditional

on the subject and session,

$$B_{ij} \equiv E(b_{ij}) = \exp(\boldsymbol{\beta}'\boldsymbol{X}_j + a_i + \sigma^2/2).$$

Note, Eq. 6 formalizes the between-HUT longitudinal component of the joint model by accounting for repeated measures of subjects' variability indices' slopes pre- and post-bed rest with a subject specific random intercept, a_i.

2.4.6 Joint model

Given the formulation of the longitudinal and survival models, we now construct our joint model under the assumption that cardiovascular response instability during a HUT would increase at a faster rate for subjects with greater OI. Accordingly, we would expect to see shorter survival times when B_{ij} is large (see Supplementary Appendix II for an illustrative example). We thus model θ_{ij} inversely proportional to B_{ij}:

$$\theta_{ij} = \frac{K_o}{B_{ij}} = K/\exp(\boldsymbol{\beta}'\boldsymbol{X}_j + a_i), \tag{7}$$

where $K = K_o \exp(-\sigma^2/2)$.

Note that Eq. 7 provides the link between the longitudinal and survival models through the shared parameters $\boldsymbol{\beta}$ and σ_a^2. Under the assumption that the two within-subject observations ($j = 1, 2$) are conditionally independent given a_i, the contribution to the joint likelihood for the i^{th} subject has the form

$$L_i = \int \left(\prod_{j=1}^{2} p_W(T_{ij}, \delta_{ij} | K, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{X}_j, \boldsymbol{W}_j, a_i) f_z(z_{ij} | \boldsymbol{\beta}, \boldsymbol{X}_j, a_i) \right) f_a(a_i) da_i \tag{8}$$

In this formulation, the parameters of interest are $K, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_a^2$ and σ^2, $p_W(\cdot)$ is the Weibull density (when $\delta_{ij} = 0$) or survival function (when $\delta_{ij} = 1$), $f_z(\cdot)$ and $f_a(\cdot)$ are

normal densities, and $z_{ij} = log(b_{ij})$. For estimation, the integral in Eq. 8 is approximated using 20-point (pt.) Hermitian Gaussian integration and maximization of the likelihood is performed using the *ml* command in Stata 14 software (StataCorp, 2015). See the Supplementary Material for a listing of Stata *.ado* files needed to implement the calculation and maximization of the likelihood along with some example data.

3 Simulation

To test the performance of our joint model, we simulated multiple sets of data from a hypothetical HUT experiment with two treatment arms ("treatment" and "control"), where each subject performs a HUT pre- and post-bed rest. Details regarding data simulation and parameter specification are found in Supplementary Appendix III. Briefly in six scenarios (Table A-1), we generated 500 random sets of HUT data for 40 control and 40 treatment subjects. The scenarios differed in terms of the amount of censoring (high and low) and treatment efficacy (high, moderate, and none). Values of true effects that define the parameters K, $\boldsymbol{\alpha} = (\alpha_0)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$, σ_a^2, and σ^2 were motivated by characteristics of the bed rest study data. Here, we included an effect for an intercept (β_0), treatment arm assignment (β_1), evaluation time (β_2), and their interaction (β_3) (see Eq. 6). The simulation study was evaluated in terms of bias, average standard error, Monte Carlo error (Koehler et al., 2009), and 95% CI coverage probability for the estimated model parameters. Additionally, we assessed the power of test $H_0 : \beta_3 = 0$ (no treatment $\times$ bed rest interaction) for our joint model versus power for a random-effects analysis of survival times only. All estimation runs for the joint model were implemented with 20-pt. Hermitian Gaussian integration.

4 Data analysis

4.1 Cardiovascular subset selection

In this section, we compared different methods of summarizing the variability of cardiovascular variables collected as part of the study data and identify which subsets of cardiovascular measures should be used to form the variability indices that best relate to impending presyncope. Recall candidate cardiovascular variables included heart rate (HR), diastolic blood pressure (DBP), systolic blood pressure (SBP), mean arterial blood pressure (MAP), total peripheral resistance (TPR), stroke volume (SV), cardiac output (CO) and pulse pressure (PP). In constructing each of the three variability indices, we always included HR because a rapid increase in HR is by far the most important single indicator of impending presyncope, at which time a HUT is terminated. Furthermore, due to the relatively small sample size ($N = 27$), we evaluated the three variability indices (Eqs. 1-3) formed from subsets of three or less dimensions. We assumed that with this data, calculation of variability indices with more than three dimensions would likely result in over-fitting and produce unreliable results.

For each combination of variability index and subset of cardiovascular variables, we fit our joint model (a) with the full data (30-minute HUT) and (b) using only the first 10 minutes of the data to simulate the hypothetical scenario where censoring is heavy. For each censoring time, we aimed to compare combinations of variability index and cardiovascular subset on their ability to detect the effect of an experimental condition (in this case, bed rest) on OI. Accordingly we chose the p-value based on the test statistic for the overall effect of bed rest as the comparison criterion. P-values were obtained using permutation tests (see Supplementary Appendix IV for details).

4.2 Survival prediction

Recall that the premise of our joint model is to take advantage of information contained in the variability of longitudinal data collected during a HUT to provide more insight into the effect of an experimental treatment or condition than survival times alone. In particular, researchers may be interested in predicting the conditional distribution of HUT survival times, when high amounts of censoring occurs within a study. Using the estimated parameters from our joint model, predicted survival plots can be generated as follows:

Given that subject i survived the HUT for t_0 minutes in the j^{th} test session, the model-based conditional probability that this subject would still be surviving at time $t > t_0$ is

$$S(t|t_0) = \exp\{(t_0/\theta_{ij})^p - (t/\theta_{ij})^p\}, \tag{9}$$

where θ_{ij} is given in Eq. 7. After fitting the model, we can predict this probability with $\hat{S}(t|t_0)$, obtained by evaluating Eq. 9 with predicted values $\hat{\theta}_{ij}$ and $\hat{p}$ substituted for θ_{ij} and p. Values of $\hat{\theta}_{ij}$ are calculated using Eq. 7, where the predicted values of $\boldsymbol{\beta}$ and the best linear unbiased predictors of the a_i are obtained from the mixed-effects submodel $\log b_{ij} = \boldsymbol{\beta}'\boldsymbol{X}_j + a_i + \epsilon_{ij}$ (Bates and Pinheiro, 1998). Values of $\hat{p}$ can be obtained directly from the model fit.

4.3 Goodness of fit

The values $\hat{\theta}_{ij}$ and $\hat{p}$ can also be used to obtain a measure of goodness of fit for the joint model by comparing the proportion of surviving subjects, f_j, in session j at time t_0 to the average value of the unconditional predicted survival probability for the i^{th} subject in the j^{th} HUT session at time t_0, $\hat{S}_{ij}$, given by

$$\hat{S}_{ij} = \exp\{(t_0/\hat{\theta}_{ij})^{\hat{p}}\} \tag{10}$$

Note that in Eq. 10, $\hat{\theta}_{ij}$ can be calculated after estimating the model, regardless of whether the i^{th} subject actually had survived up to time t_0.

5 Results

5.1 Descriptive results

In the study data with HUTs set at 30 minutes, median failure times were 17 and 9 minutes for pre- and post-bed rest sessions, respectively. Table 1 shows the number of subjects who performed HUTs pre- and post-bed rest as well the number who survived or failed the test. Additionally, Supplementary Table A-2 reports the times to event for each of the 27 subjects pre- and post-bed rest.

Table 1: Number of subjects who performed HUTs pre- and post-bed rest (BR) as well the number who survived or failed the test

Fail	Pre-BR	Post-BR	Total
No	11	1	12
Yes	16	18	34
Total	27	19	46

Table 2 shows descriptive statistics for the last measurement collected during pre- (stratified by failure) and post-bed rest HUTs for all eight cardiovascular variables. Note that post-bed rest measures for all subjects were combined, as only one subject was able to complete the HUT is this session.

Table 2: Descriptive statistics for study subjects' last measurements during the HUT; SD represents standard deviation.

	HR	DBP	SBP	MAP	TPR	SV	CO	PP
Pre-Bed Rest, No Fail (N=11)								
Mean	115.6	72.1	117.6	87.3	30.7	26.4	3.1	45.5
SD	13.0	8.7	13.3	9.8	10.3	8.9	1.1	7.7
Pre-Bed Rest, Fail (N=16)								
Mean	106.6	55.1	84.0	64.8	26.4	26.0	2.7	28.9
SD	21.7	15.4	22.1	17.2	11.1	5.6	0.6	10.9
Post-Bed Rest (N=19)								
Mean	127.2	54.8	83.8	64.5	26.9	19.1	2.5	29.0
SD	31.4	13.2	23.1	15.5	7.6	4.8	0.7	15.6

5.2 Comparison of subsets of cardiovascular variables and variability indices

For 10- and 30-minute HUTs, the top 10 combinations of cardiovascular measure subsets and variability index are shown in Table 3, ranked by p-values for testing the effect of bed rest. We reiterate that only subsets including HR were evaluated, as others were clearly more poorly performing. The most salient feature of Table 3 is that only one path length variability index appeared among the top 10 in either length of HUT. Furthermore, the variability index (V_k) based on convex hull content appeared in 9 of the 10 entries for 30-minute HUTs, and in 5 of 10 for 10-minute HUTs. Four determinant-based indices also appeared in the top 10, 10-minute results. It is also evident that the cardiovascular measurements describing aspects of blood pressure (DBP, SBP, MAP, and PP), were by far the most commonly appearing in these two lists. Due to collinearity of these measures, all six 3-dimensional subsets comprised of HR and any two of the blood pressure measures give the same results for convex hull and determinant variability indices so that only the one with HR, DBP, and SBP is actually shown in Table 3. Given the relatively small HUT maximum time of 10 minutes with observations recorded only once per minute, it is not surprising that 2-dimensional versions of V_k dominated the 10-minute list (8 of 10 cases); whereas 2- and 3-dimensional versions of V_k appeared five times each in the 30-minute list.

5.3 Survival prediction

Figure 4 shows the predicted conditional survival functions (Eq. 10) 10 minutes into HUTs (pre- and post-bed rest) for Subject 9. Both survival times for this subject would have been censored had the test stopped at 10 minutes. For this illustration, we fit the joint model using the variability index V^*, the convex hull area encompassed by the joint trajectories

Table 3: Results of the comparison of subsets of cardiovascular variables and variability indices. N represents the number of runs that converged out of 500, Test Limit is the length of the HUT in minutes, p_{KS} is the p-value for the KS-test comparing the empirical null distribution of the test statistic for $H_0 : \beta_3 = 0$ to the generalized gamma approximation, and p is the p-value from the generalized gamma distribution derived from the permutation test (See Supplementary Appendix IV for details)

Summary Index	Measures	Test Limit	N	p_{KS}	p
Convex Hull	HR, DBP	10	398	0.892	1.9E-6
Determinant	HR, DBP	10	466	0.400	2.8E-6
Convex Hull	HR, DBP, SBP	10	470	0.866	8.9E-6
Convex Hull	HR, PP	10	47	0.407	2.9E-5
Determinant	HR, PP	10	242	0.512	3.4E-5
Determinant	HR, SBP	10	162	0.362	5.8E-5
Convex Hull	HR, SV, PP	10	487	0.954	6.4E-5
Path Length	HR, MAP	10	23	0.231	8.6E-5
Determinant	HR, MAP	10	456	0.335	1.2E-4
Convex Hull	HR, MAP	10	150	0.543	1.3E-4
Convex Hull	HR, PP	30	483	0.894	8.3E-7
Convex Hull	HR, SBP	30	481	0.763	2.4E-6
Convex Hull	HR, TPR, PP	30	498	0.820	5.9E-6
Convex Hull	HR, SV, PP	30	499	0.808	6.4E-6
Convex Hull	HR, DBP	30	478	0.895	8.8E-6
Convex Hull	HR, CO, PP	30	464	0.988	1.9E-5
Convex Hull	HR, MAP	30	498	0.998	1.9E-5
Convex Hull	HR, SV	30	496	0.945	4.2E-5
Convex Hull	HR, SBP, SBP	30	499	0.735	5.2E-5
Path Length	HR, DBP, SBP	30	49	0.157	7.8E-5

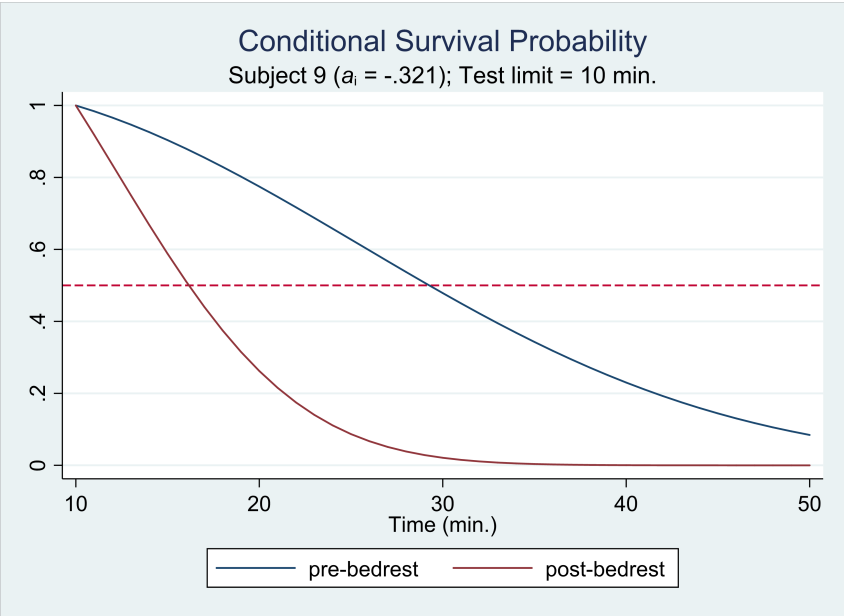

Figure 4: Conditional survival probability for Subject 1 after a 30-minute HUT.

of HR and DBP, which was identified as the best-performing variability index in Table 3 for 10-minute HUTs. For the sake of simplicity of this demonstration, we assumed the model has estimated a common value of p for both HUT sessions. Clearly after 10 minutes, the two censored survival times by themselves tell us nothing about how bed rest affected this subject. However, after fitting the joint model and using Eq. 10, we can see that the predicted pre-and post- bed rest survival functions are actually quite different. In particular from these curves, it can be seen that the estimated median conditional survival times were approximately 29 minutes and 16 minutes for pre- and post-bed rest, respectively. In other words, had the test been stopped at 10 minutes with this subject still tolerating the tilt both pre- and post-bed rest, one could still find evidence that bed rest substantially increased this subject's state of OI.

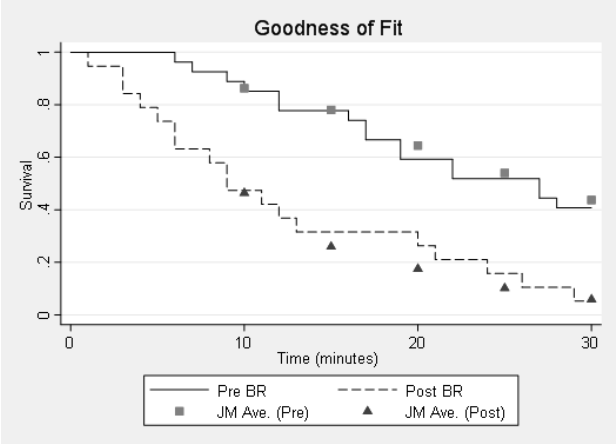

Figure 5: Goodness of fit plot. Average (Ave.) value of $\hat{S}_{ij}$ (Eq. 10) estimated by the joint model (JM) and the Kaplan-Meier estimate of the survival functions pre- and post-bed rest (BR).

5.4 Goodness of fit

Using the study data, we repeatedly fitted the joint model with V^* as the variability index and data at different censoring times $t_0 = 10, 15, 20, 25$, and 30 minutes and then calculated the $\hat{S}_{ij}$ after each model fit (Eq. 10). Figure 5 shows the average value of $\hat{S}_{ij}$ and the Kaplan-Meier estimate of the survival functions pre- and post-bed rest. Note that in this plot, the discrepancies between the two estimates are fairly small compared to the range of values they span, suggesting negligible systematic estimation bias.

5.5 Simulation

Results of the six simulation scenarios are presented in Table 4. Convergence rates for maximum-likelihood estimation of the joint model ranged from 475/500 to 486/500. Estimation of σ^2, the within-subject variance of $\log(b_{ij})$ were consistent with average esti-

mates of $\log \sigma = -0.71$ ranging from -0.64 to -0.72 for the six scenarios. On the other hand, estimates of σ_a^2, the between-subject variance were somewhat biased downwards, with $\log \sigma_a = -0.36$ and average estimates thereof ranging from -0.49 to -0.41. Of the six fixed model parameters, the one of most interest is β_3, which quantifies the interaction effect between treatment and bed rest. For the four non-null scenarios, estimates of β_3 were somewhat biased towards zero, with biases ranging from 9% (30-minute HUTs) to 25% (10-minute HUTs). However for all six scenarios, coverage of the 95% two-sided confidence interval for β_3 was close to nominal (i.e. 95%), ranging from 90 to 96%. Comparisons of power of the test of $H_0 : \beta_3 = 0$ versus $H_1 : \beta_3 < 0$ between our joint model using the cardiovascular information and an analysis of only survival time data with a Weibull lognormal frailty model resulted in substantially higher power with the joint model for the four scenarios where H_0 did not hold. Both methods provided power of close to the nominal alpha level (0.05) for the last two scenarios under which H_0 was true (Table 4). All simulation results are shown in more detail in Supplementary Appendix III.

6 Discussion

We developed a joint model for survival and two time scales of longitudinal data to make inference about treatment effects in the setting of HUTs during which concurrent, multivariate cardiovascular data are collected. The main feature of our model is that it exploited the rate of change of the cardiovascular data's variability within HUTs, manifested in three proposed variability indices, which represent the within-HUT longitudinal data in a way that can be directly related to survival time distributional parameters. Our method accommodates repeated-measures survival data, does not make any strict assumptions regarding the distribution of the rate of change of the variability index, and accounts for trend model

Table 4: Results of power comparison. N represents the number of runs that converged out of 500, Power_{JM} represents power of test of $\beta_3 = 0$ using the joint model and, and Power_S represents power using the survival model.

Set	N	Power_{JM}	Power_S
1	484	0.508	0.298
2	486	0.887	0.574
3	477	0.233	0.155
4	487	0.869	0.466
5	486	0.043	0.064
6	475	0.040	0.051

uncertainty and measurement error. Through simulation, we demonstrate the effectiveness of our model compared to traditional survival methods aimed at evaluating a hypothetical countermeasure against OI. In our application, we propose the most informative combination of cardiovascular data and variability index to assess the effect of bed rest on a HUT. Additionally, we provide a strategy for constructing dynamic prediction plots that can be used to estimate survival in a HUT in the presence of highly censored data.

Previous observations from HUT studies have suggested that large oscillations in heart rate and blood pressure while standing are indicative of an inability of the cardiovascular system to fully compensate for the stress associated with standing and that syncope will ensue if the upright posture is maintained (Lipsitz et al., 1997; Julu et al., 2003). In our study, we have additionally observed that typically, large jumps in the variability index early in a HUT were more predictive of presyncope. Thus, we weighted the median regression models inversely with time when estimating the slopes of the variability indices.

We assume that a reduced ability to tolerate the HUT due to greater OI in a subject would be manifested in increases of all three versions of the variability index. However, there are differences in the way these versions change with time as a HUT proceeds. The cumulative path length is strictly increasing and ignores direction. Thus, long excursions in $\mathbb{R}^M$ are penalized whether they exhibit a trend in some direction or whether they are oscillatory around some common point. The M^{th} root determinant is generally increasing, but can decrease if a new point is near the mean of the previous points. In this case, the new point is "rewarded" in the sense that the updated value of the variability index is lower. The M^{th} root convex hull content is dependent on direction and is non-decreasing in time. However, this index is unchanged for new points in the interior of the previous convex hull and only increases if a new point is outside the previous hull. Thus, in a sense, excursions within the scope of previous values are not penalized.

Our study is limited in that we were unable to explore larger dimensions for the variability indices due to the limited number of subjects and dependency among covariates. For example, the blood pressure measures (DBP, SBP, MAP, PP) form a linearly dependent group of rank two. Therefore, all pairwise combinations of these measures would produce identical values for the convex hull and determinant variability indices, and using three or more of these measures would result in indices equal to zero. However since the calculation of the variability index is performed prior to model estimation, exploring larger dimension spaces would not affect computation time for the joint model estimation. While our approach seemingly takes the form of a dimension reduction technique for the multivariate, within-HUT longitudinal data, we emphasize that including superfluous, but erratic covariates in the variability index may overshadow the underlying relations. Thus when determining the best combination of covariates within a variability index, we recommend using the convex hull or determinant, as the path length is sensitive to subjective

or arbitrary normalization factors which may exacerbate the effect of uninfluential covariates' inclusion. Furthermore in our simulation study, the joint model's estimation was most sensitive to initialization with path length as the longitudinal outcome, often failing to converge. Another limitation is that we estimate the slope of the variability index separately for each individual outside of our joint model's estimation. While this has its advantages computationally, the two-stage approach may potentially bias model estimates (Lawrence Gould et al., 2015). As a result, our joint model contains two sources of error: (1) approximating a subject's "cardiovascular condition" with surrogate measures and (2) possibly mis-specifying not only the mean trajectory, but also the distribution for the summary index around the mean.

Our method is applicable to settings in which data are collected more frequently, however in our application, we were limited to cardiovascular measures measured only once per minute. In studies where the cardiovascular data is measured at higher frequency within HUTs, one could incorporate higher-dimensional data into variability indices and then model their trajectories parametrically or semi-parametrically.

While the focus of this work was to capture the essential information contained in the multivariate trajectories of cardiovascular measures into summary variability indices that could be used to evaluate the effect of experimental conditions on OI, future studies also could take advantage of this methodology to investigate clinical populations, such as postural orthostatic tachardia syndrome and acute autonomic failure patients.

SUPPLEMENTARY MATERIAL

Supplementary Appendix I Description of convex hull. Contained in Supplementary Material pdf.

Supplementary Appendix II Heuristic example of relation between slope of variability indices and survival time. Contained in Supplementary Material pdf.

Supplementary Appendix III Full description of how data were simulated. Contained in Supplementary Material pdf.

Supplementary Appendix IV Description of how p-values were obtained for bed rest study application. Contained in Supplementary Material pdf.

Tutorial This tutorial uses an example data set to illustrate the estimation of our joint model using Stata software. Contained in HUT Tutorial Word document.

Code Code to run joint model. Stata files - ott.ado, ottlik.ado, gausshermite5.dta, gausshermite10.dta, gausshermite15.dta, and gausshermite20.dta

Example Data File contained example data for tutorial. Stata file - ott_testdata.dta

7 BibTeX

References

Bates, D. M. and J. C. Pinheiro (1998). Computational methods for multilevel modelling. *University of Wisconsin, Madison, WI*, 1–29.

Buckey, J., L. D. Lane, B. D. Levine, D. E. Watenpaugh, S. Wright, W. Moore, F. Gaffney, and C. Blomqvist (1996). Orthostatic intolerance after spaceflight. *Journal of Applied Physiology 81*(1), 7–18.

Catterson, A. D., H. A. McCutcheon, M. H. A, and R. A. Pollard (1963). Aeromedical observations. *Mercury Project Summary*. `http://www.hq.nasa.gov/office/pao/History/SP-45/ch18.htm`. Accessed: 2016-08-01.

Fritsch-Yelle, J. M., P. A. Whitson, R. L. Bondar, and T. E. Brown (1996). Subnormal norepinephrine release relates to presyncope in astronauts after spaceflight. *Journal of Applied Physiology 81*(5), 2134–2141.

Han, J., E. H. Slate, and E. A. Peña (2007). Parametric latent class joint model for a longitudinal biomarker and recurrent events. *Statistics in Medicine 26*(29), 5285–5302.

Julu, P., V. Cooper, S. Hansen, and R. Hainsworth (2003). Cardiovascular regulation in the period preceding vasovagal syncope in conscious humans. *The Journal of Physiology 549*(1), 299–311.

Koehler, E., E. Brown, and S. J.-P. Haneuse (2009). On the assessment of Monte Carlo error in simulation-based statistical analyses. *The American Statistician 63*(2), 155–162.

Lawrence Gould, A., M. E. Boye, M. J. Crowther, J. G. Ibrahim, G. Quartey, S. Micallef, and F. Y. Bois (2015). Joint modeling of survival and longitudinal non-survival data: Current methods and issues. Report of the DIA Bayesian joint modeling working group. *Statistics in Medicine 34*(14), 2181–2195.

Lee, S., A. H. Feiveson, M. B. Stenger, S. P. Stein, and S. H. Platts (2011). Orthostatic hypotension after long-duration space flight: NASA's experiences from the International Space Station. *Medical Science Sports Exercise*.

Lipsitz, L. A., R. Morin, M. Gagnon, D. Kiely, and A. Medina (1997). Vasomotor insta-

bility preceding tilt-induced syncope: Does respiration play a role? *Journal of Applied Physiology 83*(2), 383–390.

Meck, J. V., S. A. Dreyer, and L. E. Warren (2009). Long-duration head-down bed rest: Project overview, vital signs, and fluid balance. *Aviation, Space, and Environmental Medicine 80*(Supplement 1), A1–A8.

Meck, J. V., C. J. Reyes, S. A. Perez, A. L. Goldberger, and M. G. Ziegler (2001). Marked exacerbation of orthostatic intolerance after long-vs. short-duration spaceflight in veteran astronauts. *Psychosomatic Medicine 63*(6), 865–873.

Mudholkar, G. S. and D. K. Srivastava (1993). Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability 42*(2), 299–302.

Platts, S. H., D. S. Martin, M. B. Stenger, S. A. Perez, L. C. Ribeiro, R. Summers, and J. V. Meck (2009). Cardiovascular adaptations to long-duration head-down bed rest. *Aviation, Space, and Environmental Medicine 80*(Supplement 1), A29–A36.

Platts, S. H., J. A. Tuxhorn, L. C. Ribeiro, M. B. Stenger, S. Lee, and J. V. Meck (2009). Compression garments as countermeasures to orthostatic intolerance. *Aviation, space, and environmental medicine 80*(5), 437–442.

Ramsay, J. O. and B. W. Silverman (2005). *Functional data analysis*. New York: Springer.

Ramsdell, C. D., T. J. Mullen, G. H. Sundby, S. Rostoft, N. Sheynberg, N. Aljuri, M. Maa, R. Mukkamala, D. Sherman, K. Toska, et al. (2001). Midodrine prevents orthostatic intolerance associated with simulated spaceflight. *Journal of Applied Physiology 90*(6), 2245–2248.

Ratcliffe, S. J., W. Guo, and T. R. Ten Have (2004). Joint modeling of longitudinal and survival data via a common frailty. *Biometrics 60*(4), 892–899.

Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. Boca Raton: CRC Press.

Rizopoulos, D. (2014). The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *arXiv preprint arXiv:1404.7625*.

Song, X., M. Davidian, and A. A. Tsiatis (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics 58*(4), 742–753.

StataCorp (2015, College Station, TX: StataCorp LP). Stata statistical software: Release 14.

Tsiatis, A. A. and M. Davidian (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*, 809–834.

Wu, L., W. Liu, G. Y. Yi, and Y. Huang (2011). Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Journal of Probability and Statistics 2012*.

Ye, W., X. Lin, and J. M. Taylor (2008). Semiparametric modeling of longitudinal measurements and time-to-event data–a two-stage regression calibration approach. *Biometrics 64*(4), 1238–1246.