11-27-2007

Longitudinal Data with Follow-up Truncated by Death: Finding a Match Between Analysis Method and Research Aims

Brenda Kurland
Fred Hutchinson Cancer Research Center, bkurland@fhcrc.org

Laura Lee Johnson
National Cancer Institute, lauraleehome@gmail.com

Paula Diehr
University of Washington, pdiehr@u.washington.edu

1. Introduction

Research studies often collect information at multiple timepoints. For example, the Cardiovascular Health Study (CHS), an observational study of 5,888 older adults [1], has conducted semiannual assessments of cardiovascular functioning and other health measures for up to 18 years. With these longitudinal measurements, CHS data may be used to study a disease course [2], health in the years leading up to a diagnosis [3], or the natural history of aging [4, 5].

Missing data are an impediment to longitudinal data analysis. Early analysis methods, such as repeated measures ANOVA, require balanced data (responses at the same timepoints for all individuals), and other conditions unlikely to be met in long-term follow-up [6]. Suppose a cohort of 200 subjects report self-rated health at age 70 years, and only 150 of these subjects are located for follow-up. If the average self-rated health at age 75 is higher than the average at 70, the increase could reflect improvement in individuals' health, attrition of sicker participants, or death of sicker participants.

Dropout and nonresponse in longitudinal data analysis have been studied extensively, and accommodation of selected types of missing data has become routine [7-10]. Comparatively little work in statistical methodology has addressed data missing when *deaths* occur during the period of follow-up [11-18]. In many applications deaths are fundamentally different from data missing due to nonresponse. Data missing due to nonresponse could potentially have been collected, but were not. Data missing due to death are undefined, or have coding (such as a systolic blood pressure of 0) inconsistent with measured responses. We focus here on truncation due to death; however, a brief review of models for longitudinal data with monotone dropout provides a foundation for discussing analysis of longitudinal data truncated by death.

Two common, widely applied analysis methods for longitudinal data are random effects models [19] (related to mixed models [20], multilevel models [21], and latent variable models [22]) and generalized estimating equations (GEE) [23]. By modeling a structure for the correlation between subjects' longitudinal responses, many random effects models will fit unbiased regression models even with unbalanced data. For example, if sicker participants drop out, their trajectory of decline in self-rated health is continued implicitly by a random effects model [24]. If trends for dropouts can be inferred from observed data ("missing at random", such as when scores decline before dropout), the missingness is "ignorable" and the overall rate of change will be measured as if no one has dropped out. If the decline in health that leads to dropout starts after the last recorded measurement, then dropout is "non-ignorable," and random effects models are not an easy solution. Untestable assumptions must be made about non-ignorable dropout processes to model longitudinal trends [24]. GEE can accommodate data missing at random if estimating equations are weighted by the inverse probability of dropout [9]. Giving additional weight to observed data for people who were likely to drop out is similar to implicit or explicit imputation of unobserved data. In fact, under some conditions, weighted GEE and imputation will give the same results [25]. "Missing at random" is often a reasonable assumption, especially when longitudinal observations are closely spaced relative to mechanisms acting on both dropout and response. For example, preclinical cognitive changes could likely be detected by annual assessments before a CHS participant becomes impaired by dementia in a way that would lead to nonresponse. However, analysis of longitudinal data with MAR dropout still requires accurate modeling of the regression model (fixed effects) and either correlation (for random effects models) or dropout (for weighted GEE).

Different approaches have been proposed to distinguish data missing due to death from nonresponse, and to accommodate data missing due to death [11-18]. In this article, we examine

how these models arise from the statistical distribution of longitudinal data and survival information, and give some guidance on appropriate analysis techniques. In Section 2, two hypothetical data examples are used to present five modeling options for longitudinal data truncated by death. Data examples without measurement error illustrate clearly how modeling choices for longitudinal data reflect assumptions about survival. Section 3, an analysis of CHS data, illustrates how standard analysis techniques such as random effects models and GEE may be applied to address different research aims involving longitudinal data truncated by death.

2. Longitudinal data truncated by death

2.1. Data examples

2.1.1. Cardiovascular Health Study (CHS)

The Cardiovascular Health Study (CHS) is a population-based prospective longitudinal study of 5,888 adults aged 65 years and older at baseline [1]. Cognitive functioning was assessed annually for up to 10 years. We examine the 3814 participants aged 70 years and older at baseline, 1356 (36%) of whom died during follow-up. The longitudinal response for this analysis is cognitive functioning, measured by the Modified Mini-Mental State Examination (3MSE, scored from 0 to 100) [26]. Questions of interest include the rate of change and expected cognitive status at specific ages. We will examine how different analysis methods each yield 3MSE trajectories and fitted values, but address different research aims.

Accommodation of deaths in CHS data will be described using simplified hypothetical data. Table 1 shows 3MSE data for 4 hypothetical participants with a baseline age of 70. Participant A, representing normal cognitive functioning, has a 3MSE of 90 points at all assessments. Participant B's linear decline from 84 to 74 points over 5 years reflects a possible trajectory of mild cognitive impairment (which could be interpreted as preclinical Alzheimer's disease). Participants C and D both decline between baseline and age 72, and die before age 73.

2.1.2. Complementary Comfort Care (C3) Study

The Complementary Comfort Care (C3) study (William Lafferty, principal investigator) is a randomized clinical trial testing the effect of complementary and alternative medicine on quality of life (QOL) and symptoms for patients at the end of life. A palliative care intervention of guided meditation and twice-weekly massage is compared to an active control of friendly visits. The longitudinal response variable for the C3 study is a single-item QOL item, "How would you rate your overall quality of life during the past 7 days?" Scores of 0 are defined as "no quality of life and 10 as "perfect quality of life". The aim of the study is to compare QOL in the treatment and control groups.

The study is ongoing, so we present only hypothetical data from a simplified version of the C3 design (Table 2). In both the control group (A) and treatment group (B), the weekly quality of life rating declines by 1 point in each of the three weeks before death [27]. For earlier QOL assessments, the treatment group increases QOL by 0.5 points each week, and the control group QOL is stable at baseline levels.

2.2. Notation

Vector $\mathbf{Y}_i$ represents the longitudinal response (i.e., cognitive functioning or quality of life), measured at multiple timepoints for participant *i*. The dimension (length) of $\mathbf{Y}_i$ may differ for individuals (values of *i*), due to death. For example, in Table 1 (hypothetical CHS data), responses for participant A, $\mathbf{Y}_A$, are the vector (90,90,90,90,90,90), and for participant C, $\mathbf{Y}_C$ is (84,80,76). Scalar variable S_i represents survival time for participant *i*, such as age at death or weeks from baseline until death: in Table 1, S_C is 73 years. The dimension of $\mathbf{Y}_i$ is determined by the value of S_i. The joint distribution $f(\mathbf{Y}_i, S_i)$ describes the probability that $\mathbf{Y}_i$ takes on a vector of specific values, *and* that participant *i* dies at a specific time.

2.3. Statistical models for longitudinal response and survival

Regression models for longitudinal data describe the relationship between predictors and the longitudinal response, $\mathbf{Y}_i$. Because survival S_i determines the length of $\mathbf{Y}_i$ and is not fixed, regression models of longitudinal data truncated by death must explicitly or implicitly model survival as well. A single regression model could be built for the joint distribution $f(\mathbf{Y}_i, S_i)$, or for factorizations based on the definitions of joint and conditional distributions: $f(\mathbf{Y}_i \mid S_i) \cdot f(S_i)$ or $f(S_i \mid \mathbf{Y}_i) \cdot f(\mathbf{Y}_i)$. We will characterize models for $\mathbf{Y}_i$ as unconditional, fully conditional, or partly conditional based on how, or whether, the longitudinal response model conditions on S_i. These models are defined in more detail below, and summarized in Table 3. Each model is applied to both the hypothetical CHS and C3 data. The different research aims of the two studies serve as a useful contrast, and the simplified data illustrate clearly how modeling choices for longitudinal data reflect assumptions about survival.

2.3.1 $f(\mathbf{Y}_i)$ Unconditional

An unconditional model, $f(\mathbf{Y}_i)$, is appropriate if deaths do not occur, are independent of the response process, or do not result in truncation (if the response has a well-defined value following death). If these stipulations are not met, the unconditional distribution $f(\mathbf{Y}_i)$ reflects averaging $f(\mathbf{Y}_i \mid S_i)$ over the survival function $f(S_i)$, as shown in this section. Unconditional regression models cognitive functioning at all timepoints as if nobody died, in an "immortal cohort" [14]. The unconditional average 3MSE at age 75 years in the CHS hypothetical data is:

$$\begin{aligned}\text{average(3MSE at age 75)} &= \text{average(3MSE at 75}\mid\text{alive at 75)} \cdot \text{P(alive at 75)} \\ &\quad + \text{average(3MSE at 75}\mid\text{deceased at 75)} \cdot \text{P(deceased at 75)} \\ &= \frac{90+74}{2} \cdot \frac{2}{4} + \text{X} \cdot \frac{2}{4}.\end{aligned}$$

For analysis methods for which "missing at random" nonresponse mechanisms are ignorable – such as for random effects models fit to unbalanced longitudinal data – the value of X is imputed implicitly because of the structure imposed by correlation between each subject's longitudinal observations. We convey this implicit imputation by projecting the hypothetical 3MSE data of deceased participants based on individual slopes. This extrapolation is more extreme than estimates would be with real data (i.e., Section 3.1.1), since estimation of fixed effects in random effects models will be influenced by regression to the mean and other shrinkage [28]. Extending

the Table 1 trajectories for Participants C and D linearly, the "incomplete" response vectors (84,80,76) and (65,50,35) are imputed to (84,80,76,72,68,64) and (65,50,35,20,5,-15). Participant D's imputed response at age 75 (-15 points) is inappropriate, outside the range of the 3MSE.

An unconditional model uses both observed and imputed data to estimate linear 3MSE slope and 3MSE at age 75 (Table 4, row a). Completing the estimate of 3MSE at age 75 by imputing the value of X:

$$\text{average(3MSE at age 75)} = \frac{90+74}{2}\cdot\frac{2}{4} + \frac{64+(-15)}{2}\cdot\frac{2}{4} = \frac{90+74+64+(-15)}{4} = 53.25.$$

The age 75 fitted 3MSE (53.25 points) and linear 3MSE slope (5.25 point decline per year) both underestimate cognitive functioning among participants alive at age 75, because values are imputed beyond death.

QOL for the hypothetical C3 data may be described by unconditional linear models with quadratic time, fitted separately (via interaction terms) for treatment and control groups:

$$E(QOL_{ij}) = \beta_0 + \beta_1 \cdot week_{ij} + \beta_2 \cdot week_{ij}^2 + \beta_3 \cdot tx_i + \beta_4 \cdot tx_i \cdot week_{ij} + \beta_5 \cdot tx_i \cdot week_{ij}^2$$

for treatments *i*=A,B and weeks *j*=0-5. The fitted models in Figure 1a were generated by using available data and by extending past death (if needed to have data to week 5), using the terminal decline trajectory. The unconditional model reflects that the patients had similar QOL at baseline, but that the treatment group QOL could improve or remain stable on average, while the control group QOL declined on average. Notably, the fitted QOL at week 5 for the control group is -0.3 points, beyond the range of the QOL scale. Extension of QOL trajectories beyond death was explicit when computing this example, but would be implicit if a random effects model were fitted to observed data [14, 15, 24].

In rare cases, implicit imputation beyond death may be reasonable. When local recurrence following ablation of liver tumors is evaluated, some livers may "die" due to transplant. The chance of recurrence ($\mathbf{Y}_i$) and transplant candidacy (S_i) are related. However, since the transplant rate will never be 100%, an unconditional model and implicit imputation beyond transplant may be valid. The research question addressed is, "What would have happened if the person had not died, but the longitudinal response continued along the exact path that led to death?" While this question is relevant to liver transplants, the CHS and C3 hypothetical examples show that unconditional models are generally inappropriate for longitudinal data with considerable imbalance due to death.

2.3.2. $f(\mathbf{Y}_i \mid S_i = s)$ Fully conditional

Sometimes only subjects who survive to the end of the study are included in analysis, or decedents are analyzed separately from non-decedents [29]. An analogy in the missing data literature is pattern-mixture models [10, 30], which stratify by the time of dropout. Pattern-mixture models may be fitted using the same methods as unconditional models (random effects regression, etc.) but are made fully conditional by fitting separate regression models to strata defined by time of death. Generally a categorical variable defined by survival time is used as a main effect (and interaction term) in regression models, so that longitudinal trajectories are fit for groups defined by time of death [13, 31]. An advantage of this approach is accurate representation of individuals' scores over time.

Computing linear slopes separately for decedents and survivors in the hypothetical CHS data (Table 4) demonstrates minimal decline in survivors (row b(1), 1 point per year) and terminal decline for decedents (row b(2), 9.5 points per year). However, since the time of death is not known in advance, these models could not be used to predict an individual's trajectory based on baseline information. Figure 1b shows the average QOL response for weeks 0-4 for the 6 C3 hypothetical patients who die just after week 4, 5, or 6. The shape of the QOL trajectory for each treatment group, and the difference between groups, are similar to the unconditional model. Unlike the unconditional model the fitted QOL values are within the range of values reported by the patients.

In addition to decedent and non-decedent, other stratifications based on time of death have scientific merit. For example, the "dying process" may be examined for decedents only, with years until death as the timescale [5, 32, 33]. This approach is especially appropriate for the C3 data, since hospice patients are expected to die within a relatively short timeframe. Figure 1c changes the x-axis of the C3 longitudinal plot, directly measuring terminal decline [5]. The average QOL values in the weeks before death are shown separately for treatment and control groups. This fully conditional model reflects the data-generating assumption that the treatment group experienced QOL improvement before the onset of terminal decline.

2.3.3. $f(\mathbf{Y}_i \mid S_i > t)$ *Partly conditional*

For partly conditional models, the expected value of Y_{ij} (response of subject i at time t_{ij}) conditions on the subject's being alive at time t_{ij}. This conditioning may seem trivial: after all, data are not collected posthumously. However, as is well-documented for data missing due to dropout [24], analysis methods that model the correlation structure of longitudinal data (such as mixed models) will implicitly impute responses, whether missing due to dropout or death [14, 15]. Partly conditional regression models assume independence among the longitudinal responses, and in that sense are equivalent to linear regression or generalized linear models. However, models are fit using generalized estimating equations (with independence working correlation) in order to estimate sandwich standard errors [23], especially when weights are included to accommodate dropout [9, 15].

Partly conditional "regression conditioning on being alive" [15] (RCA) describes 3MSE scores at different ages among the surviving participants. For the hypothetical CHS data, RCA estimates of average 3MSE at age 75 and linear 3MSE slope are calculated using linear regression (Table 4, row d). Implicit imputation is avoided by treating observations from the same person as independent. RCA accurately shows that the *prevalent* cognitive functioning level is slightly higher at age 75 (82 points is the average 3MSE for survivors A and B, estimated as 80.8 by imposing a single linear slope to all observed data), compared to age 70 (80.8 point average for A-D, estimated as 76.2 by linear regression). The partly conditional slope predicts that average 3MSE increases 0.92 points per year, despite that no individuals have increasing 3MSE scores. Partly conditional regression reflects 3MSE in the dynamic cohort of survivors, not individual subjects' change in cognitive functioning.

A partly conditional model for QOL in hypothetical C3 patients is shown in Figure 1d. The same quadratic time trend is fitted as for the unconditional model, but to unbalanced data (different length of follow-up) without imputation past death. The average QOL for the treatment group survivors is stable over time, reflecting the mixture of treatment group

participants with rising QOL and those with terminal decline. The dynamic cohort of surviving control group participants shows a declining average QOL.

2.3.4. $f(\mathbf{Y}_i, S_i)$ *Unconditional (joint model)*

A joint response encompassing both survival and the longitudinal response also may be of interest. A patient facing a diagnosis may ask not only, "What is the chance that I'll be alive in 5 years?" but also, "What is the chance that I'll be alive *and healthy* in 5 years?" A joint model of the probability of being healthy and alive [12, 34, 35] characterizes the status of the entire cohort with respect to a longitudinal response and death. A related approach rescales response measures to predict the probability of being healthy and alive in a prescribed amount of time, such as one year [29, 36]. Because the joint response (healthy and alive) is defined at all timepoints for all individuals, longitudinal data are balanced. Therefore, analysis methods (random effects, GEE, etc) will not be affected by differential survival. Joint models also may assess treatment effects simultaneously for longitudinal response and survival [37], or integrate morbidity and mortality in utility measures such as quality-adjusted life years.

Defining 3MSE scores ≥ 80 points as healthy, the probability of being "healthy and alive" at age 75 in the hypothetical CHS data is 1/4 (Table 4, row e), which reflects a decline in the cohort, since 3/4 were healthy and alive at baseline (see Table 1). Assuming a linear trend, the decline from 3/4 healthy to 1/4 healthy over 5 years reflects a rate of decline of 1/10 of the cohort losing health or life each year. Note, however, that it is possible to have a transition from an "unhealthy" 3MSE to a "healthy" 3MSE score at a later time.

Figure 1e shows a joint model for the C3 hypothetical data, tracking the percentage in the control and treatment groups that are healthy and alive (where "healthy" is defined as QOL ≥ 4). Like most panels of Figure 1, Figure 1e shows better hypothetical QOL for the treatment group compared to the controls. In addition to comparing the percent of the cohort that is healthy and alive at each timepoint, the treatment effect may be assessed using the area under the curve (AUC), or average weeks of healthy life [38]. The AUC is 1.25 (of 5) weeks of healthy life for the control group, and 3.13 weeks for the treatment group.

2.3.5. Other factorizations and analysis approaches

A factorization of the joint distribution $f(\mathbf{Y}_i, S_i)$ (Section 2.2) not yet discussed is $f(S_i \mid \mathbf{Y}_i) \cdot f(\mathbf{Y}_i)$. This framework is especially applicable to predicting survival (S_i) using information from longitudinal biomarkers ($\mathbf{Y}_i$) [39-41]. Applying this model to hypothetical CHS data, we could conclude that 33% (1/3) of participants with declining 3MSE survive to age 75, while 100% (1/1) with stable 3MSE survive to age 75. This class of models would be categorized as "unconditional" in the framework discussed here, but is not considered in detail because survival, not longitudinal data, is the primary response of interest.

Another approach to modeling longitudinal data truncated by death is to estimate causal models for selected principal strata [16, 18]. In the C3 study, the effect of treatment on QOL would be decoupled from survival rates by estimating a causal effect of treatment versus control at week X in strata of patients expected to live until week X regardless of treatment assignment. The interpretation of these models is similar to interpretation of pattern-mixture models, in that information about future survival status (as well as counterfactual survival status) must be known in order to make use of predicted causal effects within each principal stratum. As for pattern-

mixture models, stratification based on information known at study entry would be preferable [16].

3. Data example

As described in Section 2.1.1, our primary goal in the CHS analysis is to describe the trajectory of cognitive functioning (3MSE) over time, and to estimate 3MSE scores at different ages. We examined 3814 participants age 70 years or older at the beginning of the study. In this cohort, 1356 participants (36%) died during follow-up: 44% (744/1703) of men and 29% (612/2111) of women. Gender effects were explored to add a between-person variable of interest to the longitudinal (within-person changes) model, and to explore the effects of differential survival on different regression approaches for longitudinal and survival data.

Although models may be constructed to accommodate deaths and nonresponse separately [15], we impute data missing due to nonresponse for simplicity of presentation. Data were *not* considered missing if follow-up was truncated by death, or censored by the end of the study period. (A group to boost minority recruitment received 6 annual assessments instead of 10.) Most participants (n=2061, 54%) completed all scheduled 3MSE assessments. About 17% of 3MSE scores (5174 of 31093) were missing due to participant nonresponse. Most nonresponse was intermittent, with one or two scores missing (n=948 participants). Some participants had dropped out of the study, and were missing 7 or more scores (n=134). Nonresponse was accommodated by imputation using the Markov Chain Monte Carlo method of PROC MI in the SAS/STAT software, version 9.1 (SAS Institute, Inc., Cary, NC). Imputation was stratified by time of death (for decedents) and recruitment group (for non-decedents), and was modeled based on observed 3MSE values, baseline age, gender, and recruitment group.

3.1.Statistical models for longitudinal response and survival

CHS analysis results are shown in Figure 2. Using models fit to all participants with baseline age ≥ 70 years (n=3814), each plot shows fitted 3MSE values for a baseline age of 70 years.

3.1.1 $f(\mathbf{Y}_i)$ *Unconditional*

The unconditional model of 3MSE scores by age in the CHS data is a random effects linear regression [19], separating the effects of age and aging [42] in a quadratic model with random intercept and first-order polynomial:

$$\hat{3\text{MSE}}_{ij} = \beta_0 + b_{0i} + \beta_1 \cdot \text{male}_i + \beta_2 \cdot \text{age0}_i + (\beta_3 + b_{1i}) \cdot \text{year}_{ij} + \beta_4 \cdot \text{year}_{ij}^2 + \varepsilon_{ij}$$

where *age0* is the participant-specific baseline age (in years), *year* is the study year (and difference from baseline age), and random intercept, slope, and error (b_{0i}, b_{1i}, and ε_{ij}) are normally distributed with mean 0. Likelihood-based methods such as random effects regression will fit an unconditional model, since they treat any unbalance in the data as "missing at random." Interactions between sex and linear and quadratic terms were explored, but did not contribute to the models. Figure 2a shows the fitted "average" trajectories for males and females (thick lines, $b_{0i}=b_{1i}=\varepsilon_{ij}=0$), and several fitted trajectories for individuals (thin lines, selected b_{0i} and b_{1i}). The random intercepts and slopes allow a wide range of individual fitted trajectories.

However, time trends and other covariate effects are generally interpreted based on the mean model (thick lines), which reflects both observed data and data implicitly imputed beyond death. According to this average trajectory, the expected 3MSE for both males and females is 86 points at age 75, and 77 points at age 79. The unconditional model suggests rather strong age-associated declines in cognitive functioning.

3.1.2. $f(\mathbf{Y}_i \mid \mathrm{S}_i = s)$ *Fully conditional*

The first fully conditional model fitted to the CHS data is a pattern-mixture model, in which quadratic time trends are fitted (as for the unconditional model), stratified by year of death relative to baseline. Figure 2b shows fitted mean 3MSE trajectories for a baseline age of 70 years. The pattern-mixture model demonstrates terminal decline in participants who die (fitted lines that end before age 79), and reasonably stable cognitive functioning in participants who survive. The fitted 3MSE at age 75 ranged from 75 points in males who died by age 76, to 91 points in males and females who enrolled at age 70 and survived at least to age 79. The fitted mean trajectories are closer to trajectories observed for individuals than the unconditional model, but require conditioning on survival time, which is not known at baseline.

The second fully conditional model examines terminal decline. Rather than counting forward in years of age, the time scale for this analysis counts backward from death. The 2458 participants who are alive at the end of follow-up (64%) are excluded, since their age at death is not known. As in earlier models, a random effects model is fitted with quadratic time and random intercept and slope. However, the time scale is changed:

$$\hat{3\mathrm{MSE}}_{ij} = \beta_0 + b_{0i} + \beta_1 \cdot \mathrm{male}_i + \beta_2 \cdot \mathrm{age0}_i + (\beta_3 + b_{1i}) \cdot \mathrm{yr\_fr\_death}_{ij} + \beta_4 \cdot \mathrm{yr\_fr\_death}_{ij}^2 + \varepsilon_{ij},$$

where yr_fr_death ranges from -1 (one year before the year of death) to -9 (9 years before), and other variables are as described in Section 3.1.1. Figure 2c shows fitted average terminal decline trajectories for men and women with baseline age 70. The estimated rate of decline is about 4.7 points per year, plus the effect of a negative quadratic coefficient (-0.3). The fitted 3MSE score 6 years before death is about 87 points, close to the average baseline value (88 points) for the 70-year-olds at baseline. For this cohort of decedents, averaging over yr_fr_death and sex, the expected 3MSE at age 75 is 85 points, and at age 79 is 82 points. The rate of terminal decline reflects the combined influence of the most dramatic declines and the more stable 3MSE patterns observed in the multiple pattern-mixture fitted trajectories for decedents.

3.1.3. $f(\mathbf{Y}_i \mid \mathrm{S}_i > t)$ *Partly conditional*

The partly conditional model avoids implicit imputation of data for deceased subjects, and describes longitudinal 3MSE for the dynamic cohort of survivors [14, 15]. The regression equation is similar to that for the unconditional and pattern-mixture models, but does not include a random intercept or slope:

$$\hat{3\mathrm{MSE}}_{ij} = \beta_0 + \beta_1 \cdot \mathrm{male}_i + \beta_2 \cdot \mathrm{age0}_i + \beta_3 \cdot \mathrm{year}_{ij} + \beta_4 \cdot \mathrm{year}_{ij}^2 + \varepsilon_{ij}.$$

Figure 2d shows the expected 3MSE score for participants who entered the study at age 70, given that they were alive at the time 3MSE was measured. The difference in expected 3MSE at different ages is smaller than for the unconditional or fully conditional models. The expected 3MSE score of participants who entered CHS at age 70 is 91 points for surviving 75-year-olds, and 87 points for surviving 79-year-olds. However, this does not imply an average decline of 1 3MSE point per year, since the average for survivors is not the same as the trajectory from following individuals. The partly conditional model tracks the prevalent average 3MSE score in the survivors at each timepoint.

3.1.4. $f(\mathbf{Y}_i, S_i)$ *Unconditional (joint model)*

Figure 2e shows the proportion of CHS participants who are healthy (3MSE ≥ 80) and alive at ages 70-79. The percent alive and healthy (PAH) shown is a proportion reflecting the status of study participants enrolled at age 70. Modeling of PAH also is possible, as well as constructing confidence intervals around the PAH estimate [35]. The percent alive and healthy is not always decreasing: individuals may regain "health", such as when a low 3MSE score was due to short-term side effects of medication. Like the partly conditional model, the joint model describes the cohort, rather than trends for individuals. Unlike the partly conditional model, the entire cohort is described at each timepoint, not only survivors. (The end of follow-up for the minority recruitment group could affect differences seen between ages 70-76 and 77-79. We avoid implicit imputation beyond 6 years by computing the empirical PAH as a simple proportion.)

For participants aged 70 years at baseline, the probability of being alive and having 3MSE ≥ 80 at age 75 is 0.82 for females and 0.75 for males. At age 79, the PAH is 0.70 for females and 0.54 for males. Summing the area under the curve, the average years of healthy life (of 9 possible) is 7.6 for females, and 6.7 for males. The average gender difference appears to be greater for the joint model than for the other fitted models in Figure 2. This reflects a survival advantage for females, which was not apparent in models that focused on the 3MSE.

4. Discussion

Through analysis of hypothetical and actual data sets, we have shown that choice of analysis has a great influence on interpretation of longitudinal data truncated by death. No single approach is appropriate in all situations, so the analysis should be chosen to address the aims of a research project. Summaries of individual trajectories and descriptions of terminal decline were achieved with fully conditional models, in which analysis of longitudinal response is stratified by time of death. The partly conditional model addressed situations where prevalence, rather than individual trajectories, was of interest. For example, survival information could estimate the number of new Medicare recipients who will be alive in 10 years, and a partly conditional model could then estimate the need for dementia services in those survivors. In our hypothetical palliative care example, treatment effects were reflected by a joint model for longitudinal response and survival. The area under the joint density curve summarized treatment differences in both survival and quality of life response. In our examples, observational data (CHS) were best described using models based on individual trajectories and the dynamic cohort of survivors, while the clinical trial (C3) aims were addressed by joint models. However, due to differences

in study aims and the vagaries of data collection, this heuristic will not apply in all situations.

Once the statistical model is clarified by research aims, choice of analysis method should be apparent (Table 3). An unconditional model is fit by random effects and other multilevel approaches, when the time scale does not depend on survival times. A fully conditional model may be fit by random effects or other analysis methods, with time scale or stratification depending on survival time [5, 13]. Partly conditional models are fit directly by GEE with independence working correlation [15]. Joint models may be fit as a multivariate joint distribution [37], or as a composite response incorporating both survival and longitudinal response [29].

We have focused on research in which estimation of a parameter, such as a treatment effect, slope, or patient trajectory is of primary interest. Other classes of statistical analysis are available to address different research questions. Additionally, models fit using one factorization of the joint distribution of survival and longitudinal response may be transformed to address aspects of another model. For example, a fully conditional model may be marginalized [43] to estimate a partly conditional entity, such as the expected 3MSE among CHS survivors at age 85 [10]; a joint model may estimate fully conditional trajectories [31].

In longitudinal studies in which some subjects die yet another response, such as quality of life, is of primary interest, careful modeling is required to identify an analysis method to address research aims. When deaths occur at many different times along the time frame for which responses are measured (i.e., age or time from baseline), random effects models (which are unconditional with respect to survival) may implicitly impute data beyond death. Implicit imputation is a fundamental strength of random effects models in the missing data context, but limits the suitability of these unconditional models in analyzing longitudinal data with great imbalance due to death. When the time scale describes time from (not until) death, the model becomes fully conditional. For terminal decline models, implicit imputation beyond death will not occur when random effects models are fit. Analysts concerned about the potential impact of implicit imputation may fit a generalized linear model or generalized estimating equations with independence correlation (which fit partly conditional models) and compare fitted parameters to an unconditional model.

Acknowledgments

We are grateful to Dr. William Lafferty for discussion of the C3 study design (CA106204). Research reported in this article was supported by contracts N01-HC-35129, N01-HC-45133, N01-HC-75150, N01-HC-85079 through N01-HC-85086, N01 HC-15103, N01 HC-55222, and U01 HL080295 from the National Heart, Lung, and Blood Institute, with additional contribution from the National Institute of Neurological Disorders and Stroke. A full list of participating CHS investigators and institutions can be found at http://www.chs-nhlbi.org.

References

1. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, Kuller LH, Manolio TA, Mittelmark MB, Newman A, et al. The Cardiovascular Health Study: design and rationale. *Annals of Epidemiology* 1991; 1(3): 263-276.
2. Kaplan RC, Tirschwell DL, Longstreth WT, Jr., Manolio TA, Heckbert SR, Lefkowitz D, El-Saed A, Psaty BM. Vascular events, mortality, and preventive therapy following ischemic stroke in the elderly. *Neurology* 2005; 65(6): 835-842.
3. Diehr P, Williamson J, Patrick DL, Bild DE, Burke GL. Patterns of self-rated health in older adults before and after sentinel health events. *Journal of the American Geriatrics Society* 2001; 49(1): 36-44.
4. Burke GL, Arnold AM, Bild DE, Cushman M, Fried LP, Newman A, Nunn C, Robbins J. Factors associated with healthy aging: the cardiovascular health study. *Journal of the American Geriatrics Society* 2001; 49(3): 254-262.
5. Diehr P, Williamson J, Burke GL, Psaty BM. The aging and dying processes and the health of older adults. *Journal of Clinical Epidemiology* 2002; 55(3): 269-278.
6. McCulloch CE. Repeated Measures ANOVA, R.I.P.? *Chance* 2005; 18: 29-33.
7. Little RJA, Rubin DB. *Statistical analysis with missing data*. Wiley: New York, 1987.
8. Rubin DB. *Multiple imputation for nonresponse in surveys*. John Wiley and Sons: New York, 1987.
9. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 1995; 90: 106-121.
10. Fitzmaurice GM, Laird NM. Generalized linear mixture models for handling nonignorable dropouts in longitudinal studies. *Biostatistics* 2000; 1(2): 141-156.
11. Revicki DA, Gold K, Buckman D, Chan K, Kallich JD, Woolley JM. Imputing physical health status scores missing owing to mortality: results of a simulation comparing multiple techniques. *Medical Care* 2001; 39(1): 61-71.
12. Diehr P, Patrick DL. Trajectories of health for older adults over time: accounting fully for death. *Annals of Internal Medicine* 2003; 139(5 Pt 2): 416-420.
13. Pauler DK, McCoy S, Moinpour C. Pattern mixture models for longitudinal quality of life studies in advanced stage disease. *Statistics in Medicine* 2003; 22(5): 795-809.
14. Dufouil C, Brayne C, Clayton D. Analysis of longitudinal studies with death and drop-out: a case study. *Statistics in Medicine* 2004; 23(14): 2215-2226.
15. Kurland BF, Heagerty PJ. Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics* 2005; 6(2): 241-258.
16. Rubin DB. Causal inference through potential outcomes and principal stratification: Application to studies with "censoring" due to death. *Statistical Science* 2006; 21(3): 299-309.
17. Harel O, Hofer SM, Hoffman L, Pedersen NL, Johansson B. Population inference with mortality and attrition in longitudinal studies on aging: a two-stage multiple imputation method. *Experimental Aging Research* 2007; 33(2): 187-203.
18. Egleston BL, Scharfstein DO, Freeman EE, West SK. Causal inference for non-mortality outcomes in the presence of death. *Biostatistics* 2007; 8(3): 526-545.
19. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; 38(4): 963-974.

20. Littell RC, Milliken GA, Stroup WW, Wolfinger RD. *SAS system for mixed models*. SAS Institute Inc.: Cary, N.C., 1996.
21. Bryk AS, Raudenbush SW. *Hierarchical linear models : applications and data analysis methods*. Sage: Newbury Park, CA, 1992.
22. Bollen KA, Curran PJ. *Latent curve models : a structural equation perspective*. Wiley-Interscience: Hoboken, N.J., 2006.
23. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; 7: 13-22.
24. Laird NM. Missing data in longitudinal studies. *Statistics in Medicine* 1988; 7(1-2): 305-315.
25. Paik MC. The generalized estimating equation approach when data are not missing completely at random. *Journal of the American Statistical Association* 1997; 92: 1320-1329.
26. Teng EL, Chui HC. The Modified Mini-Mental State (3MS) examination. *Journal of Clinical Psychiatry* 1987; 48(8): 314-318.
27. Diehr P, Lafferty WE, Patrick DL, Downey L, Devlin SM, Standish LJ. Quality of life at the end of life. *Health and Quality of Life Outcomes* 2007; 5: 51.
28. Robinson GK. That BLUP is a good thing: The estimation of random effects (with discussion). *Statistical Science* 1991; 6(1): 15-51.
29. Diehr P, Patrick DL, Spertus J, Kiefe CI, McDonell M, Fihn SD. Transforming self-rated health and the SF-36 scales to include death and improve interpretability. *Medical Care* 2001; 39(7): 670-680.
30. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* 1995; 90: 1112-1121.
31. Ribaudo HJ, Thompson SG, Allen-Mersh TG. A joint analysis of quality of life and survival using a random effect selection model. *Statistics in Medicine* 2000; 19(23): 3237-3250.
32. Siegler IC. The terminal drop hypothesis: fact or artifact? *Experimental Aging Research* 1975; 1(1): 169-185.
33. Hwang SS, Chang VT, Fairclough DL, Cogswell J, Kasimis B. Longitudinal quality of life in advanced cancer patients: Pilot study results from a VA medical cancer center. *Journal of Pain and Symptom Management* 2003; 25(3): 225-235.
34. Diehr P, Patrick D, Hedrick S, Rothman M, Grembowski D, Raghunathan TE, Beresford S. Including deaths when measuring health status over time. *Medical Care* 1995; 33(4 Suppl): AS164-172.
35. Johnson LL. Incorporating Death into the Statistical Analysis of Categorical Longitudinal Health Status Data. Seattle, WA: University of Washington; 2002.
36. Diehr P, Johnson LL, Patrick DL, Psaty B. Methods for incorporating death into health-related variables in longitudinal studies. *Journal of Clinical Epidemiology* 2005; 58(11): 1115-1124.
37. Gray SM, Brookmeyer R. Multidimensional longitudinal data: estimating a treatment effect from continuous, discrete, or time-to-event response variables. *Journal of the American Statistical Association* 2000; 95: 396-406.
38. Diehr P, Patrick DL, Bild DE, Burke GL, Williamson JD. Predicting future years of healthy life for older adults. *Journal of Clinical Epidemiology* 1998; 51(4): 343-353.
39. De Gruttola V, Tu XM. Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* 1994; 50(4): 1003-1014.

40. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; 53(1): 330-339.
41. Wang Y, Taylor JMG. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* 2001; 96(455): 895-905.
42. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 1998; 54(2): 638-645.
43. Heagerty PJ, Zeger SL. Marginalized multilevel models and likelihood inference (with discussion). *Statistical Science* 2000; 15: 1-26.

Table 1: Longitudinal 3MSE scores for 4 hypothetical CHS participants (X = deceased)

age	70	71	72	73	74	75
A ("normal")	90	90	90	90	90	90
B ("mild cognitive impairment")	84	82	80	78	76	74
C ("terminal decline")	84	80	76	X	X	X
D ("terminal decline")	65	50	35	X	X	X

Table 2: Longitudinal QOL scores for 8 hypothetical C3 participants (X = deceased)

	week										
	0	1	2	3	4	5	6	7	8	9	10
A1	3	3	3	3	2	1	0	X	X	X	X
B1	3	3.5	4	4.5	3.5	2.5	1.5	X	X	X	X
A2	4	4	3	2	1	X	X	X	X	X	X
B2	4	4.5	3.5	2.5	1.5	X	X	X	X	X	X
A3	1	0	X	X	X	X	X	X	X	X	X
B3	1	1.5	2	2.5	3	3.5	4	3	2	1	X
A4	5	5	5	4	3	2	X	X	X	X	X
B4	5	5.5	6	5	4	3	X	X	X	X	X

A = control group (friendly visits, which might include help with chores)
B = treatment group (guided meditation and twice-weekly massage)

Table 3: Summary of statistical models for longitudinal response and survival (time of death)

Row		Statistical Model	Sample Research Setting	Primary Analysis Method	Comments
a.	Unconditional $f(\mathbf{Y}_i)$	Describe $\mathbf{Y}_i$ (longitudinal response) in setting where survival (S_i) is unrelated to $\mathbf{Y}_i$, or when death does not result in missing data	Rate of local recurrence following ablation of liver tumors*	Mixed effects/random effects/latent variable regression	May implicitly impute data beyond death
b.	Fully conditional: pattern-mixture $f(\mathbf{Y}_i \mid S_i = s)$	Describe $\mathbf{Y}_i$ separately for groups defined by survival time	Longitudinal change in physical functioning following stroke, separately for 6+-month and 5+-year survivors	Mixed effects/random effects regression stratified by survival time	Describes individual trajectories, but uses future survival information to predict earlier responses
c.	Fully conditional: terminal decline $f(\mathbf{Y}_i \mid S_i = s)$	Describe $\mathbf{Y}_i$ counting backward from time of death	Terminal decline studies	Mixed effects/random effects regression	Time scale is retrospective
d.	Partly conditional $f(\mathbf{Y}_i \mid S_i > t)$	Describe $\mathbf{Y}_i$ in the dynamic cohort of survivors at each timepoint	Average physical functioning in survivors at 6 months and 5 years after stroke	Generalized estimating equations (GEE) with independence working correlation	Describes longitudinal trend of dynamic cohort, not individuals
e.	Joint model $f(\mathbf{Y}_i, S_i)$	Describe both $\mathbf{Y}_i$ and S_i – for example, "probability of being healthy and alive"	Percent of stroke patients who are alive and can perform self-care 6 months after stroke	Logistic regression, GEE (binary outcome), or specialized methods for multiple responses	Continuous longitudinal outcomes may need to be categorized for analysis

* Some livers may "die" due to transplant, but the transplant rate will never be 100% due to availability. Therefore, "Would recurrence have occurred if transplant had not?" is a valid question.

Table 4: Hypothetical CHS data (Table 1) estimated age 75 3MSE score and 3MSE slope, using models that accommodate deaths in different ways

Row	Statistical Model	Sample Research Question(s)	3MSE at age 75	linear 3MSE slope (annual change in 3MSE)
a.	Unconditional $f(\mathbf{Y}_i)$	What is the expected 3MSE at age 75, in an immortal cohort?	$\frac{90+74+64+(-15)}{4} = 53.25$ points	$\frac{0+(-2)+(-4)+(-15)}{4} = -5.25$ points/year
b(1).	Fully conditional: survivors $f(\mathbf{Y}_i \mid S_i > 75)$	What is the expected 3MSE at age 75/70 for people who live to be at least 75?	$\frac{90+74}{2} = 82$ points	$\frac{0+(-2)}{2} = -1.0$ points/year
b(2).	Fully conditional: decedents $f(\mathbf{Y}_i \mid S_i \leq 75)$	What is the expected 3MSE at age 70 for people who die at age 71-75?	(both deceased at age 75)	$\frac{-4+(-15)}{2} = -9.5$ points/year
c.	Fully conditional: terminal decline $f(\mathbf{Y}_i \mid S_i = s)$	What is the expected 3MSE two years before death?	(not estimated directly)	-9.5 points/year (same as previous row – changing time scale does not make a difference with only one stratum of decedents)
d.	Partly conditional $f(\mathbf{Y}_i \mid S_i > t)$	What is the expected 3MSE at age 75/70 for people who live to be at least 75/70?	76.2 + 0.92*5 = 80.8 points	0.92 points/year
e.	Joint model $f(\mathbf{Y}_i, S_i)$	What is the probability of being healthy and alive at age 75 for people who were alive at age 70?	1/4 alive and 3MSE ≥ 80 at age 75	on average 1/10 lose health/life each year (since 3/4 healthy at age 70, and 1/4 healthy at age 75)

Figure 1: Hypothetical C3 data (Table 2) fitted quality of life (QOL) trajectories, using different models to summarize longitudinal response (QOL) and survival (S)

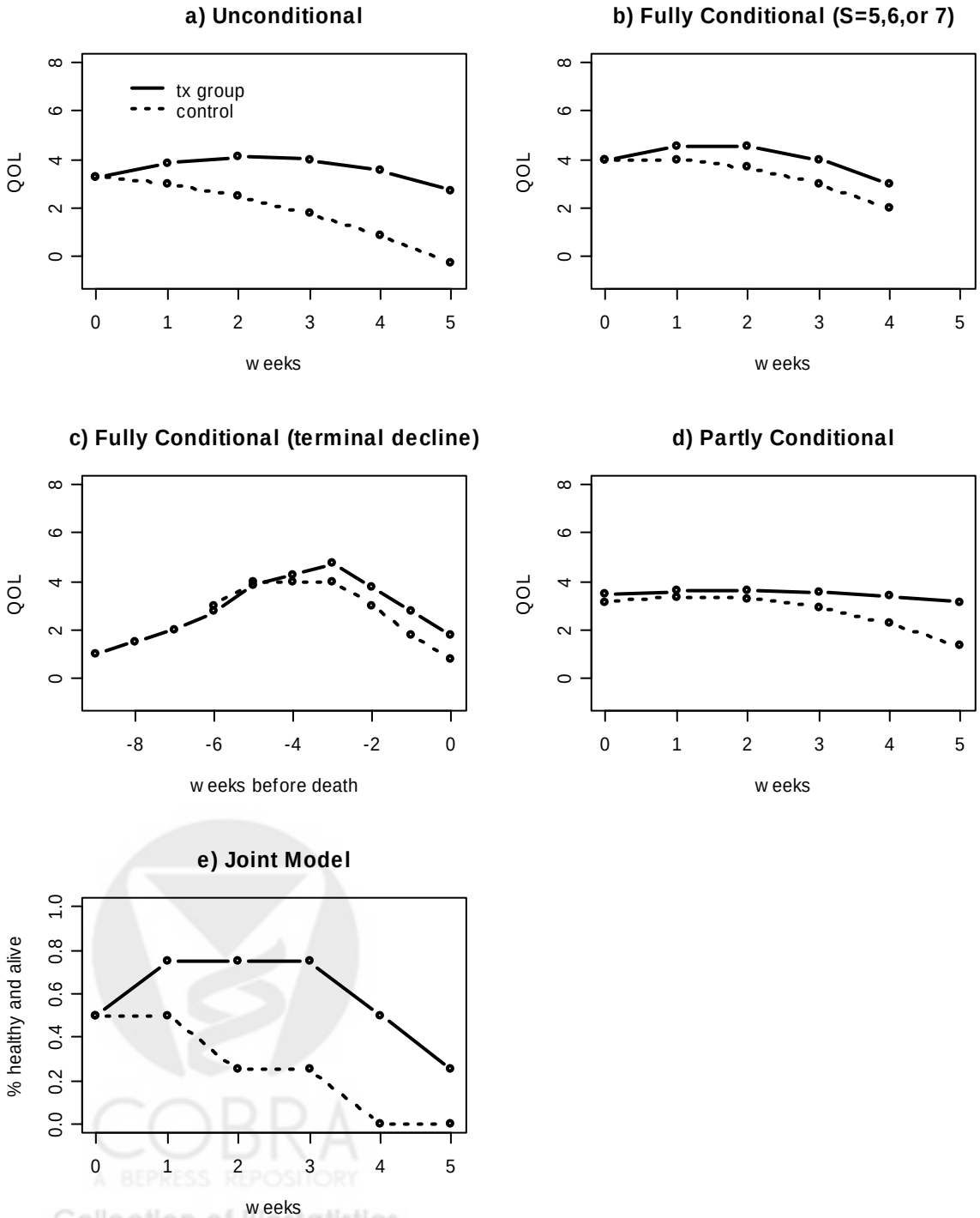

Figure 2: Fitted Modified Mini-Mental State Examination (3MSE) trajectories for CHS participants aged 70 years at baseline, using different models to summarize longitudinal response (3MSE) and survival