Practice of Epidemiology

Inverse Intensity Weighting in Generalized Linear Models as an Option for Analyzing Longitudinal Data with Triggered Observations

Peter H. Van Ness*, Heather G. Allore, Terri R. Fried, and Haiqun Lin

* Correspondence to Dr. Peter H. Van Ness, Program on Aging, Yale University School of Medicine, 300 George Street, Suite 775, New Haven, CT 06511 (e-mail: peter.vanness@yale.edu).

Initially submitted August 20, 2008; accepted for publication September 17, 2009.

Longitudinal epidemiologic studies with irregularly observed categorical outcomes present considerable analytical challenges. Generalized linear models (GLMs) tolerate without bias only values missing completely at random and assume that all observations contribute equally. A triggered sampling study design and an analysis using inverse intensity weights in a GLM offer promise of effectively addressing both shortcomings. A triggered sampling design generates irregularly spaced outcomes because, in addition to regularly scheduled follow-up interviews, it specifies that data be collected after a “trigger” (a decline in health status during follow-up) occurs. It is intended to mitigate bias introduced by study participant loss to follow-up. For each observation, an inverse intensity weight is calculated from an Anderson-Gill recurrent-event regression model whose events of interest are observed interviews; the weights help to equalize observation contributions. Investigators in the Longitudinal Examination of Attitudes and Preferences (LEAP) Study (1999–2002), a Connecticut study of seriously ill older adults at the end of life, used a triggered sampling design. In this paper, the authors analyze data from the LEAP Study to illustrate the methods and benefits of inverse intensity weighting in GLMs. An additional benefit of the analytical approach presented is that it allows for assessment of the utility of triggered sampling in longitudinal studies.

data collection; generalized linear model; longitudinal studies; sampling plan; weighting

Abbreviations: GLM, generalized linear model; LEAP, Longitudinal Examination of Attitudes and Preferences.
The method we propose in this article follows that of Robins et al. (3) in using inverse intensity weighting in GLMs in a way that attains the advantages of linear mixed models for handling missing data. Developing techniques introduced by Lin et al. (4), the model proposed here will also achieve 2 other important goals. First, it will inversely weight participants’ data according to their intensity of occurrence—that is, according to their frequency of occurrence per unit of time. These weights will be obtained by fitting an Andersen-Gill recurrent-event regression model (5) in which the recurring events are participant interviews—sometimes regularly scheduled, sometimes triggered by health declines. Through inclusion of these inverse intensity weights in a GLM fitted to the longitudinal data, each observation of each participant will contribute more equally to the study results. A practical result of weighting the interviews is that observations of participants who leave the study early but whose changes prior to departure are captured by the triggered sampling will not be dominated by the observations of participants who persevered in the study but show potentially less dramatic changes over time. A second result of the weighting is that the outcome of interest can be assessed at the population level.

The proposed model will allow for the inclusion of an indicator variable for a triggered interview. Results for this variable will provide insight as to whether the triggered sampling design achieved its goal of providing additional new information that would not have been obtained without triggering. Study designs characterized by triggered sampling involve increased financial costs for researchers, who must collect the additional data, and an increased burden for participants, who must answer additional questions after a decline in health. Thus, they should not be used unless evidence from comparable studies indicates that intended benefits are likely.

In the Longitudinal Examination of Attitudes and Preferences (LEAP) Study, Fried et al. (6) investigated a cohort of seriously ill Connecticut residents at the end of life with regard to their treatment preferences. The overall objective of the LEAP Study was to characterize changes in treatment preferences among older persons with advanced illness over time and to determine the associations between participants’ sociodemographic, health, and psychosocial status and their preferences, as measured by several different outcomes. The objective of the specific LEAP analysis most relevant to this methodological proposal was to identify associations between health and health-related predictors, represented in a regression model as longitudinal trajectories of observations, and the acceptability of being in moderately severe pain, also represented as a trajectory of observations collected over time. Since study participants were told that finding the level of pain acceptable implied a willingness to undergo additional treatment, results from this specific analysis helped to fulfill the overall goal of the LEAP Study.

Many patients were expected to die or withdraw from the study because of illness during follow-up, and their assessments of quality of life prior to serious health crises were anticipated to be a crucial factor in determining their treatment preferences (7). These analytical circumstances made joint modeling and multiple imputation unfeasible. Ignoring either death or loss to follow-up due to illness might introduce bias into statistical inferences, because it is known from previous studies that serious declines in health are accompanied by changes in people’s treatment preferences. A triggered sampling design was used to meet these challenges; thus, the LEAP Study offers an appropriate context for application of the methods proposed here. Our objective in this article is to provide an alternative way of analyzing longitudinal triggered data that yields insight into both the clinical question of interest—for example, its interpretation at the population level—and the efficacy of the triggered sampling design.

MATERIALS AND METHODS

Participants

A total of 226 community-dwelling older persons with advanced chronic illness participated in the LEAP Study during 1999–2002. The study has been described in detail elsewhere (6). Briefly, sequential charts were screened for persons aged ≥60 years with a primary diagnosis of cancer, congestive heart failure, or chronic obstructive pulmonary disease for the primary eligibility requirement: advanced illness, as defined by Connecticut Hospice criteria (8) or SUPPORT criteria (9). Charts were identified according to the patient’s age and primary diagnosis in subspecialty outpatient practices in the greater New Haven area and in 3 hospitals: a university teaching hospital, a community hospital, and a Veterans Affairs hospital. An additional eligibility criterion was a need for assistance with at least 1 instrumental activity of daily living (10). This second criterion was established during a telephone interview and was included to improve the identification of patients with advanced disease. Because of research interest in the relation between disease diagnosis and preferences, screening and enrollment were implemented in a stratified manner in order to enroll approximately equal numbers of patients with cancer, congestive heart failure, and chronic obstructive pulmonary disease. Each participant provided written informed consent, and human investigations committees from each of the hospitals participating in the study approved the study protocol.

Data collection

Patients were interviewed in their homes, and data on all variables were obtained by self-report. Patients were subsequently interviewed at least every 4 months for up to 2 years or until death. If a patient had a decline in health status, as determined by a monthly telephone call, the next interview was scheduled immediately. Such interviews are here called “triggered interviews.” Interviews subsequent to both scheduled and triggered interviews were conducted every 4 months, unless the patient experienced another decline. Thus, indirectly, monthly telephone interviews informed the regression results reported here; however, the observations recorded by study variables were collected only in face-to-face interviews occurring at least
and psychosocial measures. Ordinal variables were dichotomized so that at least 10% of respondents were in the most severe category. Sociodemographic variables included age and education as continuous variables and gender, ethnicity, sufficiency of monthly income (12), marital status, and living arrangement as binary variables. Health status variables included self-rated health, with response categories of “excellent/very good/good,” and “fair/poor,” and the number of disabilities related to instrumental activities of daily living (range, 0–14). Self-rated life expectancy was assessed by asking patients, “If you had to take a guess, how long do you think you have to live?” Level of pain was measured by asking patients, “How would you describe your worst pain during the last 24 hours?” The 2 response categories were “no pain/mild pain” and “moderate pain/severe pain.” Psychosocial variables included the existence of a living will and quality of life, with response categories for the latter variable consisting of “best possible/good” and “fair/poor/worst possible.” Depression was measured using the 2-item PRIME-MD instrument (13). Data on health and psychosocial variables were obtained at each interview.

The outcome variable for this illustration, also assessed at each interview, was taken from an instrument asking participants to consider whether a number of diminished health states that could result from treatment represented an acceptable or unacceptable quality of life (6). The specific item used in this study asked participants about pain, described as being in moderately severe pain daily as a result of treatment, like having a broken bone or appendicitis. Similar to the concept of “states worse than death” (14), participants were told that rating the health state as acceptable meant that they would want to undergo treatment and that rating the health state as unacceptable meant that they would prefer to die rather than undergo treatment and experience an unacceptable level of pain.

**Statistical analysis**

To describe the population, we used mean values and standard deviations for continuous variables and counts and percentages for categorical variables. Inverse intensity weights were obtained in the following way. First, a recurrent-event model was fitted in which the events of interest were the participant interviews (see Appendices 1 and 2). In this model, the risk sets constructed by the model change over time because of death and other losses to follow-up. The objective in fitting this model for the intensity of participant interviews was to include independent variables that explained as much of the variability in the dependent variable as possible while maintaining goodness of fit. Potential explanatory variables consisted of a variable for identifying triggered interviews and variables identified as clinically important in the original analysis of this outcome. Aside from demographic variables, all of the variables in this model-fitting process were time-varying. For persons in the risk set who had an event (i.e., an interview) at a given time, the values of the explanatory variables for that individual were from the current interview in the Andersen-Gill model data set. Baseline values for most of them are listed in Table 1. Variables for activities of daily living disability and hospitalization were forced into the model because they contributed to the criterion for triggered observations.

Other variables were selected for the multivariable model by means of forward selection, with the variables with the lowest P values in bivariate models being entered first and the variables with P values less than 0.10 being retained in the model. A robust sandwich estimator was used to account for correlation of observations within a subject. Inverse intensity weighting was estimated using the method described by Gill and was implemented by programming a tool for R (15). Using this method, we fitted a model based on the Andersen-Gill model data set. Baseline values for most of them were entered first and the variables with the lowest P values in bivariate models being entered first and the variables with P values less than 0.10 being retained in the model. A robust sandwich estimator was used to account for correlation of observations within a subject. Inverse intensity weighting was estimated using the method described by Gill and was implemented by programming a tool for R (15). Using this method, we fitted a model based on the Andersen-Gill model data set.

### Table 1. Characteristics of the Study Sample (n = 226) at Baseline, Longitudinal Examination of Attitudes and Preferences (LEAP) Study, Connecticut, 1999–2002

<table>
<thead>
<tr>
<th>Variable</th>
<th>No.</th>
<th>%</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>79</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td>81</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>66</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Age, years</td>
<td>73</td>
<td>7</td>
<td>(7)</td>
</tr>
<tr>
<td>Education, years</td>
<td>12</td>
<td>3</td>
<td>(3)</td>
</tr>
<tr>
<td>White race</td>
<td>205</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>Female sex</td>
<td>97</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Currently married</td>
<td>130</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>High income (some money left over monthly)</td>
<td>98</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Has a living will</td>
<td>120</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Self-rated health reported as excellent/very good/good</td>
<td>81</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>At least 1 disability related to activities of daily living</td>
<td>85</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Self-rated quality of life reported as high (best possible/good)</td>
<td>145</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>Depressed</td>
<td>107</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>In moderate/severe pain</td>
<td>62</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>≥2 hospitalizations in past year</td>
<td>106</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Self-rated life expectancy, years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2</td>
<td>33</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>≥2</td>
<td>94</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Uncertain</td>
<td>97</td>
<td>43</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: SD, standard deviation.

a Seven persons had missing values for this variable.
b Two persons had missing values for this variable.
estimator was used to estimate standard errors and determine which explanatory variables met our a priori level of statistical significance for retention in the multivariable model. Its use did not influence the weights obtained from the model. Ties were handled using a discrete method, because most observations occurred at regular intervals.

An advantage of the triggered sampling design is that participants were interviewed when first indications of a decline in health occurred, thereby reducing the number of interviews missing due to seriously poor health. Thus, the Cox model for the intensity of participant interviews is one in which the occurrence, and nonoccurrence, of interviews is at least partially explained by the model’s design and health-related covariates. Thus, missing data are plausibly missing at random.

Next, the linear predictors from this model were exponentiated to reverse their transformation to a natural logarithmic scale, and then inverse values of the exponentiated linear predictors were calculated. These values were normalized in the sense that the mean inverse value was subtracted from each and 1 was added to the result, thereby making 1 the neutral weight. This provided inverse intensity weights whose overall mean was 1. Finally, 2 adjustments were made. Since the last data line for each individual in our recurrent-events data set did not record an event—the stop time represented the end of follow-up when no interview was obtained—weights were advanced 1 time point within the sequence of observations for each participant. As a result of this adjustment, the baseline interview for each individual lacked a weight: A weight of 1 was assigned to each participant’s baseline interview, because all participants had a baseline interview. A second consequence of this realignment was that in the weighted GLM, the weight for a given wave of time-varying covariate and clinical outcome data was obtained from information gathered at the interview conducted during the previous wave of data collection. Box plots were constructed of normalized inverse intensity weights for the observed interviews in the entire cohort and for observed interviews as stratified for cancer, depression, and hospitalization.

In order to determine factors associated with the acceptability outcome, we utilized GLMs by implementing repeated-measures logistic regression using generalized estimating equations (15, 16) (see Appendices 1 and 2). For each model, at each time point the dependent variable was the rating of pain as acceptable or unacceptable as a result of treatment. The objective of the GLM analyses was to identify explanatory variables associated in a statistically significant way with the outcome of interest. Model-fitting was conducted using the data set with triggered observations and without inverse intensity weights; for comparative purposes, the multivariable model identified for this data set was used in the other 2 GLM analyses. The selection of potential confounders took advantage of work previously done for the published analysis (6) of this data by beginning with the multivariable model identified there. Hence, we began with a multivariable model that included potential control variables for age, gender, race, marital status, and months since follow-up began. By a process of backward elimination, we individually removed variables from the model that were not statistically significant at the $P < 0.10$ level.

A first-order autoregressive covariance structure was used in this model because there was some evidence of serial correlation in the longitudinal data. An independent covariance structure was used for the GLM with weights, because the weights effectively represent the information that would be conveyed by the choice of another type of covariance structure. Bootstrapping was used to obtain correct standard errors for the weighted model.

In addition to unweighted and weighted triggered sampling models, we fitted a third model that neither used weights nor included the additional information obtained by the triggered sampling design. Finally, to obtain yet further insight into the impact of the identification of triggered interviews, both multivariable models using the triggered data were refitted including an indicator variable for whether or not an observed interview was triggered. Since this variable indicates a self-reported change in health status (i.e., a new disability in activities of daily living) or in health-service utilization (i.e., a serious hospitalization or introduction of hospice services), a positive response to it is interpretable as evidence of a decline in health status that might prefigure death or withdrawal from the study.

All analyses were carried out with SAS software, version 9.1.3 (SAS Institute Inc., Cary, North Carolina) (17). We used the PHREG procedure to model intensities and the GENMOD procedure to fit GLMs.

### RESULTS

#### Descriptive statistics

Table 1 describes the patient population. A total of 125 participants died before the end of the 2-year follow-up period (1999–2002). Of the cohort members, 68% had at least 3 interviews and 36% had 5 or more. The median number of interviews was 2 for patients with cancer, 4 for patients with congestive heart failure, and 5 for patients with chronic obstructive pulmonary disease. Among nondropouts, ascertainment of outcome data was 90% complete, and among the 10% with missing data, 89% of the missing data was due to the participant’s being too cognitively impaired or too ill to participate in the interview. Of the 1,006 observations, 167 were triggered and added new information to the longitudinal data set. Sixty-five percent of participants had a triggered interview: 72 participants had 1 triggered interview, 43 had 2, 18 had 3, 10 had 4, 2 had 5, 2 had 6, and 1 had 7.

Box plots for the normalized version of the inverse intensity weights are provided in Figure 1. The plot of the weights for the entire cohort simply shows the magnitude and variability of the weights used in the weighted GLM. Others are stratified by each level of the binary health-status variables used in the calculation of the weights. The plots for these other variables—cancer,
depression, and hospitalization status—show how the weights adjust for the intensity of interviews in the weighted GLMs for these variables, which had a strong positive association with the occurrence of an interview. For instance, cancer patients had a high intensity of interviews in the sense that they had more interviews per unit of time. This reflects the fact that cancer patients had a higher percentage of triggered interviews yielding new information than did patients without cancer. Accordingly, cancer patients received relatively low weights—were inversely weighted—to make observations contribute with approximate equality in the weighted analysis. In addition, observations that were triggered had, on average, larger inverse intensity weights than did observations that were not triggered. Since the occurrence of triggered interviews was not very intense, occurring with relatively low frequency per unit of time, their inverse intensity weights were relatively large.

Loss to follow-up was primarily due to death. Withdrawals from the study for other reasons were relatively few, with 11 participants (4.9% of the sample) refusing to provide information after the baseline interview. Although 28 participants declined renewal of consent after the first year of the study, we do not regard these missing values, which occurred after the completion of some follow-up interviews, as posing the same risk of bias as in the case of participants providing only baseline interviews. The weighted model assumes that missing observations are missing at random, thereby making the same assumption that one does when fitting a mixed-effects model for longitudinal data.

Multivariable results

Table 2 summarizes the results from the 3 multivariable models. Although results from the 3 models cannot be strictly compared because they show results based upon different data sets and using different covariance structures, we believe that comparisons nevertheless provide some general analytical insights.

In each model, 3 variables were consistently strong predictors of the acceptability of pain upon treatment: the existence of a living will, the occurrence of moderate-to-severe pain, and higher self-reported quality of life. These findings confirm the published results obtained from a generalized linear mixed model (6). The depression and high income predictors showed more variability with regard to their inferential status. Depression became a stronger predictor when triggered observations were used, both with and without inverse intensity weights. On the other hand, the income variable had generally weaker associations with the outcome in these same models. The standard errors for the model without additional data provided by triggered sampling were mostly larger, as one would expect. The bootstrapped standard errors in Table 2 for the model with inverse intensity weights were generally smaller than those in either of the other 2 models. This result is in accord with the finding of Robins et al. (3) that use of inverse weighting procedures may lead to an increase in the efficiency of parameter estimation.

Inclusion of an indicator variable for interviews identified as triggered was marginally significant in the original model with triggered information but without weights. However, in

Figure 1. Box plots of normalized inverse intensity weights for the observed interviews in the entire cohort and for observed interviews as stratified for cancer, depression, and hospitalization status, Longitudinal Examination of Attitudes and Preferences (LEAP) Study, Connecticut, 1999–2002. T-shaped bars, range of values.
the weighted model with bootstrapped standard errors, this variable became marginally nonsignificant.

**DISCUSSION**

The strategy of using inverse intensity weights in GLMs provides an excellent investigative tool for evaluating the impact of data collected according to a triggered sampling design. While we still believe that the mixed-effects approach provides a satisfactory analysis of this type of data, the weighted GLM analysis provides a sensitivity analysis that supplements a more traditional analysis of this type of longitudinal data, because it provides results that emphasize equal contributions by individual study observations. In most respects, the analyses conducted for this study showed that results using triggered sampling data are similar to results yielded by an unweighted model. This need not be the case, however, and it was not entirely the case in this instance. For example, both models using triggered data suggested that the income variable was not as potent a predictor of the acceptability of a painful posttreatment health state as would have been indicated had data been gathered only at regularly scheduled intervals.

Additionally, the GLM without triggered observations found a somewhat different association for depression than did the approaches that used triggered observations. This change reflects the fact that approximately two-thirds of triggered observations yielding new information occurred among depressed participants as compared with nondepressed participants. This was the highest proportion in this respect among the 5 explanatory variables in the multivariable model. In addition, 22% of observations for depressed participants were triggered; only the pain predictor had a higher proportion in this respect, at 26%. The finding that the depression variable became more strongly associated with the study outcome implies that the new information, especially when weighted—and the weights increased the impact of triggered observations—accentuated the acceptability to depressed participants of the hypothesized pain condition; that is, it made the participants more inclined to undergo treatment.

The evidence for the benefit of triggered sampling is somewhat weaker than was presented by Dubin et al. (2) in their examination of this issue in the same LEAP data set. There it was reported that participants who dropped out after their first triggered interview had a change in their mean outcome value, while those who did not drop out after their first triggered interview had no noticeable such change. In the Dubin et al. study, however, the inferential goal was different, and an ordinal outcome from the LEAP Study was also investigated. Since binary variables are generally less informative than ordinal, count, or continuous variables, it might be that regression models with binary outcomes benefit less from triggered sampling designs than do other regression models. Nevertheless, the results of the current analyses provided some evidence that the triggered data contributed information to the study and thus was worth the time and expense required to collect and analyze it.
Finally, the approach we have recommended and applied offers an alternative for handling missing data when remedies like multiple imputation are not suitable. For this reason and for the other reasons given above, we conclude that inverse intensity weighting in GLMs is a valuable new option for analyzing longitudinal data with triggered observations.

ACKNOWLEDGMENTS

Author affiliations: Department of Internal Medicine, Geriatrics Section, Yale University School of Medicine, New Haven, Connecticut (Peter H. Van Ness, Heather G. Allore); Clinical Epidemiology Unit, Veterans Affairs Connecticut Healthcare System, West Haven, Connecticut (Terri R. Fried); Department of Medicine, Yale University School of Medicine, New Haven, Connecticut (Terri R. Fried); and Department of Epidemiology and Public Health, Biostatistics Section, Yale University School of Medicine, New Haven, Connecticut (Haiqun Lin).

This work was supported in part by the Claude D. Pepper Older Americans Independence Center of the Yale University School of Medicine (grant 2P30AG021342-06).

The authors thank John O’Leary for his assistance with data management.

Conflict of interest: none declared.

REFERENCES


APPENDIX 1

Mathematical Model Specifications

Our analytic approach consists of 2 stages. In the first stage, we fit the following intensity model:

\[ \lambda_i(t) = \lambda_0(t) \exp[\mathbf{z}_i(t) \mathbf{\alpha}], \]

where \( \lambda_i(t) \) is the intensity of the visit time for subject \( i \) at time \( t \), \( \lambda_0(t) \) is the unspecified baseline intensity at time \( t \), \( \mathbf{z}_i(t) \) is the row vector of possibly time-dependent covariates that may affect the frequency and timing of the visits by subject \( i \), and \( \mathbf{\alpha} \) is the vector of regression coefficients associated with the covariates. The regression coefficients in \( \mathbf{\alpha} \) are estimated via the Cox partial likelihood method.

In the second stage, we fit the following generalized linear model:

\[ h[\mu_i(t)] = \mathbf{x}_i(t) \mathbf{\beta}, \]

where \( h \) is the link function for the mean of the longitudinal response at time \( t \) for subject \( i \), \( \mathbf{x}_i(t) \) is the row vector of possibly time-dependent covariates that may affect the longitudinal response, and \( \mathbf{\beta} \) is the vector of regression coefficients associated with the covariates. The regression coefficients in \( \mathbf{\beta} \) are estimated using weighted generalized estimating equations with the weight of the observation of the \( i \)th subject at time \( t \) being a normalized form of the inverse of the intensity, specified as \( 1/\exp[\mathbf{z}_i(t) \mathbf{\alpha}] \).
APPENDIX 2

SAS Code for Key Regression Models

```
proc phreg data=van.leap_iw covs(aggregate);
   title 'Cox Model for the Determination of Weights';
   model (start,stop)*event(0)= adl depressed hosp cancer
      triggered / ties= discrete;
   id id;
   output out=out1 xbeta= xbeta dfbeta=dfbeta;
run;
proc genmod data=van.leap_iw_combined descending;
   title 'Pruned (0.10 level) Weighted GEE Model';
   class id;
   model q18_r = highinc livewill qol pain depressed
      / dist= bin link= logit obstats;
   repeated subject=id / type= ind;
   scwgt lag_norm_inv_intense;
run;
```