PREDICTING READMISSION FOR SUBSTANCE ABUSE INPATIENTS

BY

JOHN D. PIETTE, SAEID NAZARI, AND RICHARD A. OLSHEN

TECHNICAL REPORT NO. 198

JULY 1998

PREPARED UNDER THE AUSPICES

OF

PUBLIC HEALTH SERVICE GRANT 2 R01 CA59039-21

DIVISION OF BIOSTATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
Predicting Readmission for Substance Abuse Inpatients

By

John D. Piette, Saeid Nazari, and Richard A. Olshen

Technical Report No. 198

July 1998

Prepared Under the Auspices

Of

Public Health Service Grant

2 R01 CA59039-21

Department of Biostatistics

Stanford University

Stanford, California
Predicting Readmission for Substance Abuse Inpatients


Saeid Nazari, Center for Health Care Evaluation, VA Palo Alto Health Care System, Palo Alto, CA 94025.

Richard A. Olshen, Department of Health Research and Policy, Stanford University, Stanford, CA 94305-5405.

John D. Piette is a Senior Research Associate at the Center for Health Care Evaluation, VA Palo Alto Health Care System, 795 Willow Road (MPD-152), Menlo Park, CA 94025, (piette@odd.stanford.edu). Dr. Piette also is a Consulting Assistant Professor in the Department of Health Research and Policy at Stanford University. When this research began Saeid Nazari was a Research Analyst in the Center for Health Care Evaluation. Now he is Actuarial Analyst, Actuarial Department, California State Automobile Association, 100 Van Ness Avenue, 21st floor, San Francisco, CA 94102 (saeid_nazari@csaa.com). Richard A. Olshen is a Professor of Biostatistics in the Department of Health Research and Policy (HRP Redwood Bldg.), Stanford University, Stanford, CA 94305–5405, (olshen@stat.stanford.edu). He also has appointments (by courtesy) in the Departments of Electrical Engineering and Statistics. Support for this project was provided by the Mental Health and Behavioral Science Service and Health Services Research and Development Service Funds, Department of Veterans Affairs, and by Public Health Service Grant 2 RO1 CA59039 to Stanford University. We gratefully acknowledge Ciaran S. Phibbs, of the Center for Health Care Evaluation, for his advice and assistance with data transfer and statistical computing. We also thank Leo Breiman, of the Department of Statistics, University of California, Berkeley, for comments that led us to consider more carefully some of the methods we used.
ABSTRACT

This paper is a case study of readmission patterns among patients treated for substance abuse disorders in US Department of Veterans Affairs hospitals. We developed our prediction models using routinely collected administrative data for 42,648 discharges occurring between October 1, 1990 and September 30, 1991. The models are of two-parameter Weibull form, with covariates. Our focus here is unusual in that it is almost entirely upon prediction as opposed to testing hypotheses. We argue on the basis of empirical testing for the appropriateness of the Weibull models and devote considerable attention to how well they perform. With data as plentiful as in our study, tests of significance provide limited insight into goodness-of-fit. Thus, we used simulations to choose among competing models and interpret the final model’s fit. As part of this process, we developed two related criteria by which to quantify the explanatory power of our models.

Analyses suggest that VA substance abuse inpatients are at extremely high risk of readmission, particularly if they are more than 65 years of age and have chronic medical problems. Risk of readmission is highest immediately following discharge from the hospital and declines subsequently. Our statistical diagnostics suggest that the chosen model is most predictive for patients with intermediate readmission times and most inaccurate for patients with relatively short or long readmission times. However, our comparisons of the model’s prediction errors with those from a model using simulated Weibull data suggest that intrinsic difficulties in the prediction problem, rather than any limitations of the model, account for its variable accuracy. We conclude that information other than that available from discharge files may dominate patients’ readmission risk soon after and long
after, they leave the hospital. Depending on the criterion of prediction error used, the model we developed explains between 16% and 23% of the variability in patients' risk of readmission. While this is not much by criteria of some physical experimentation, it could have important implications for health policy and expenditures.

KEY WORDS: Weibull models for survival, prediction, substance abuse, goodness-of-fit, hospital readmission, health services research
1. INTRODUCTION

Alcohol and drug abuse are among the most significant causes of mortality and morbidity in the United States (McGinnis and Foege 1993). As many as 8% of all Americans have drinking problems during some period of their lifetimes, and 5% abuse prescription or illicit drugs (Grant 1996). Aside from the direct effects of intoxication including car accidents, suicides, homicides, and overdose, chronic substance abuse leads to a variety of long-term medical complications affecting multiple body systems such as the heart, liver, and nervous system (Piette, Barnett, and Moos 1998). Consequently, substance abuse contributes to more than one quarter of all hospitalizations in the United States and is implicated in a quarter billion dollars in Medicare expenditures each year (NIAAA 1993).

The Department of Veterans Affairs (VA) health care system includes the largest integrated system of care for substance abuse patients in the world. In 1997, 97,800 VA inpatients, 24% of all VA inpatients seen that year, had a substance abuse diagnosis in their discharge summaries. These patients accounted for over 2.3 million days of VA hospital care that year, or 28% of all days (Piette, Baisden and Moos 1998). As a result, VA policy makers like those in other health care systems are interested in predicting the patterns of health service use for patients with substance abuse disorders. Not only could this information provide insight into the clinical processes that determine the natural history of these patients' illnesses, but more pragmatically, information on these patients' patterns of hospitalization is needed to plan effectively for fluctuations in service demand across facilities and over time.
Efforts from within and outside of VA have yielded limited insight into the predictors of service utilization for patients with substance abuse disorders. Some studies have concluded that sociodemographic information has little relationship to substance abuse in-patients' risk of relapse or readmission (Holland and Evenson 1984; Ornstein and Cherepon 1985). For example, in the study by Ornstein and Cherepon (1985) discriminant function analyses of 18 demographic factors explained only 5% of the variance in alcoholism treatment outcome. Readmission models using a more varied set of determinants and techniques such as logistic regression have yielded reductions in deviance (i.e., reductions in the log-likelihood chi-square attributable to the chosen covariates; Hosmer and Lemeshow 1989) in the range of 9% (Moos, Brennan and Mertens 1994a) to 12% (Moos, Mertens and Brennan 1994b). Other studies have failed to report any statistics that describe the predictive power of the models that are presented.

Most studies of health service use among substance abuse patients have been motivated by an interest either in the natural history of the disorder or the effectiveness of treatment. These efforts have focused mainly on outcomes directly attributable to inebriation rather than on the medical problems that these patients develop. As a result, an overall picture of the pattern of readmission among this population has not emerged.

Furthermore, prior studies of substance abuse patients have failed to provide a comprehensive picture of the probability of recurrent events due to the limitations of the statistical tools available to analysts. Some studies have used logistic regression models that allow one to estimate the predicted probability of readmission at a single time point given a fixed array of covariate values. Other studies have used Cox proportional hazards models
(Kalbfleisch and Prentice 1980) in which the population baseline hazard determining the recurrence process is not estimated, thereby making it impossible for readers to estimate directly either a patient's expected time to readmission or the probability of readmission given a fixed time.

We analyzed readmission patterns using a nationwide sample of more than 42,000 VA discharges for substance abuse patients and developed a model to predict these patients' likelihood of readmission to VA hospitals over the subsequent three years. In our effort to identify predictively meaningful models, we developed relatively novel analytic techniques that might be applicable in other contexts including: (a) the use of multivariate Weibull models to predict readmission time; (b) measures that assess the appropriateness of the Weibull model for a given application; and (c) measures that describe the relative predictive power of competing models. Based on the model we selected, we present the predicted probability of readmission for five prototypical groups of VA substance abuse patients that have policy relevance because they are costly to treat and prevalent among VA system users.

2. DATA AND VARIABLES

Data for this study were drawn from a nationwide repository in which information is maintained for all VA health care contacts. Specifically, we analyzed data from the VA Patient Treatment Files (PTF) that contain information on all VA inpatient discharges each fiscal year (i.e., from October 1 through September 30). PTF records include information about the location of the hospital, patient (e.g., race, age, marital status, and zip code of
residence), and the hospitalization event (e.g., dates of admission and discharge and up to 10 discharge diagnoses). All PTF records are linkable via unique patient identifiers.

The data in the PTF are at least as reliable as Medicare or other similar large health service databases and have been used successfully in several past studies of VA patients' risk of readmission. For example, one reliability study of PTF records (Lloyd and Rissing 1985) found that only 4% of PTF records had inappropriate primary diagnoses, and that most of these errors related to a level of detail that was not used in the current study (e.g., differentiating between alcoholic cardiomyopathies and viral cardiomyopathies). Another study also concluded that the diagnostic data in the PTF were reliable, although some missing secondary diagnoses were noted (Hunter-Young, Hamman and Cagen 1994).

2.1 Identification of the Study Population

In order to be eligible to be included in the analytic sample, patients had to have a discharge in fiscal year 1990 from a VA inpatient substance abuse treatment unit or from a medical unit with information indicating that the purpose of the hospital stay was detoxification. We chose to identify our sample from 1990 data so that all discharged patients could be followed for three years and so that we could identify a second sample from 1993 data with which to assess the temporal stability of our models. For each patient in the analytic sample, we identified all discharges occurring during that fiscal year regardless of whether they were for treatment of a substance abuse, psychiatric, or medical problem. Thus, each patient may be represented in the data set more than one time. Data were transferred for analysis to our local Sun Sparcstation. Because the data set was so large,
we were unable to transfer it in entirety over the VA's wide area network. Consequently, we selected sixteen large, geographically-representative hospitals that were used in another VA study of substance abuse treatment (Ouimette, Finney, and Moos 1997) from which to select our patient sample: Palo Alto, CA; Los Angeles, CA; Bay Pines, FL; Danville, IL; Decatur, IL; Hines, IL; Cleveland, IN; New Orleans, LA; Minneapolis, MN; Kansas City, MO; Brooklyn, NY; Philadelphia, PA; Memphis, TN; Salt Lake City, UT; Richmond, VA; and Seattle, WA. Each of these facilities contributed between 3% and 9% of the data.

2.2 Follow-up and Follow-back

For each index discharge, we identified the date of the first VA rehospitalization within the subsequent three years. Thus, each observation includes a failure time of three years if censored (i.e., if the patient was not readmitted within that period) or a failure time of less than three years if the patient was readmitted. In addition, we determined whether each patient had a VA discharge during the twelve months predating his/her index hospital stay.

We identified patients who died before the end of the observation period by matching their social security numbers against a database maintained by the VA as part of the process of identifying and compensating beneficiaries of veterans' death benefits (the Beneficiary Identification Record Locator System). Because the VA death benefit is known to virtually all mortuary administrators and family members of eligible patients, reporting to this file is estimated to be at least 95% complete (NAS 1985; page 1992) and at least as reliable as the National Death Index for this population (Fisher, Weber, Goldberg and Davis 1995).
A total of 4% of all observations had failure times censored by death. These observations were excluded from the analyses because, as a censoring mechanism, death violates the standard assumption that censoring should be independent of the event of interest (in this case readmission), or at least conditionally independent given applicable covariates.

2.3 Creation of Analytic Variables

We used the Andersen-Newman model (Andersen and Newman 1973) of predisposing, enabling, and need factors to focus our selection of variables. Sociodemographic characteristics (specifically, age, race, and marital status) were the predisposing characteristics that we used because such factors independently contribute to patients' patterns of inpatient substance abuse service use (Bannenberg, Raat and Plomp 1992).

Enabling factors allow patients to avail themselves of health services. We considered the presence of a “service-connected disability” as an enabling factor because substance abuse patients with such disabilities are eligible for a wider range of VA services than patients who are eligible for VA care solely on the basis of their incomes. Studies have shown that patients with such disabilities are more likely to use VA substance abuse services (Peterson, Swindle, Phibbs, Recine and Moos 1994) and other services (Holloway, Medendorp and Bromberg 1990). We also examined patients' distance from their source of VA care as an enabling factor because distance influences the amount of VA care patients use as well as whether they choose to use the VA at all (Burgess and DeFiore 1994; Piette and Moos 1996). The distance between each patient’s zip code of residence and the facility from which he or she was discharged was calculated using latitudes and longitudes
obtained from a zip code locator file maintained by the VA. Distances were calculated using a trigonometric formula described by Garnick, Luft, and Robinson (1987).

Need for inpatient care was measured by patients’ discharge diagnoses coded using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM; CPHA 1986). We treated illness as a multi-dimensional construct encompassing patients’ alcohol and drug abuse disorders, psychiatric disorders, and medical disorders. Alcohol and drug abuse were defined as in past studies (e.g., Moos et al. 1994a). Although the distinction between alcohol “abuse” and “dependence” is clinically important, we considered these conditions as a single group because the distinction is not made reliably in the PTF files and because preliminary analyses indicated that alcohol dependence syndrome was by far the largest category of alcohol abuse diagnoses in the files. For similar reasons, we combined “drug abusing” and “drug dependent” patients into a single group. Aside from the reliability problems, the distinction between abuse and dependence among patients with drug abuse disorders has been shown to be unrelated to relapse (Babor, Conney and Lauerman 1987). Medical disorders were defined using the Charlson comorbidity index (Charlson, Pompei, Ales and Mackenzie 1987) as modified by Deyo for use with ICD-9-CM codes (Deyo, Cherkin and Ciol 1992).

3. METHODOLOGY

We turn now to a description of our statistical approach. Thus, write \( T > 0 \) as the nonnegative random variable representing the failure time of an individual member of a heterogeneous population (in this case, discharges). We quantify the heterogeneity by
assigning a vector of covariates \( \mathbf{x} = (x_1 \ldots x_p) \) to each member of the population. We further assume that we have observed a randomly selected sample of the population for an a priori fixed length of time \( l \) and recorded the following information in a data set \( \mathcal{L} \), given in the following form \( \mathcal{L} = \{ (x_1, c_1, t_1), (x_2, c_2, t_2), \ldots (x_n, c_n, t_n) \} \) where \( c_j = 1 \) if the \( j \)th member of the population "failed" (i.e., experienced the event of interest) during the observation period, and \( c_j = 0 \) if it did not fail. If it did fail, \( t_j \) is the observed failure time, and if it did not fail, the observation is considered "censored," and \( t_j = l \). Vectors of real numbers \( x_1, \ldots x_n \) are taken, as are all unprimed vectors, to be column vectors. Vectors are in bold face. A superscripted prime to the right of a vector denotes "transpose." We analyze the following prediction problem: if we randomly select a member of the population and we know values of its covariate vector as well as the information contained in the data set \( \mathcal{L} \), then what will be the value of \( T \)?

3.1 The Accelerated Failure Time Model

Because it is a relatively simple but powerful approach, we assume that the probability distribution of \( T \) can be characterized by the accelerated failure time model with a baseline Weibull distribution (Kalbfleisch and Prentice 1980). According to this model, we assume that:

\[
P(T > t \mid \mathbf{x}) = P(T_0 > \exp(-\mathbf{\beta}'\mathbf{x})t) \quad \text{where} \quad P(T_0 > t) = \exp(-\alpha t^\gamma) = S_{T_0}(t) \quad (1)
\]

The vector \( \mathbf{\beta} \), and positive scalars \( \alpha \), and \( \gamma \) are unknown but are inferred from data. Thus, we have a baseline survival distribution function, \( S_{T_0}(t) \), which has a Weibull distribution that is functionally independent of covariates; and for a member of the population with
the covariate vector $\mathbf{x}$, the survival distribution function, denoted by $S_T(t)$, is given by $S_T(t) = S_{T_0}(\exp(-\mathbf{\beta}'\mathbf{x})t)$. A simple computation shows that the hazard functions for individual patients are proportional if the model applies.

For any member of the population with covariate vector $\mathbf{x}$, we have a parametric model for the probability distribution of failure times, and we use its expectation as our prediction of the member’s failure time. The predicted failure time is

$$E(T) = \exp (\mathbf{\beta}'\mathbf{x})\alpha^{-\frac{1}{\gamma}}\Gamma(1 + \frac{1}{\gamma})$$ (2)

where $\Gamma(a)$ denotes the gamma function.

3.2 Estimating Parameters

To use (2) for prediction, we need to estimate the unknown values of the $\mathbf{\beta}, \alpha$ and $\gamma$. To do so, we use the information contained in the sample $L$,

$L = \{(x_1, c_1, t_1), (x_2, c_2, t_2), \ldots (x_n, c_n, t_n)\}$. We use as our estimates those values of $\mathbf{\beta}, \alpha,$ and $\gamma$ that maximize $\log(L)$ where $L = L(\beta_1, \ldots \beta_p, \alpha, \gamma)$ is the likelihood, according to the model, of observing $L$. Specifically, we maximize the following function

$$\log(L) = \sum_{c_j=1} \left( \log(\alpha t_j^{\gamma-1}) - (\alpha t_j^{\gamma} + \gamma(\mathbf{\beta}'\mathbf{x})) \right) + \sum_{c_j=0} -\alpha(\exp(-\mathbf{\beta}'\mathbf{x})t_j)^{\gamma}$$ (3)

Equation (3) for $\log(L)$ was constructed under the premise that for each given observation in $L$, given by $(x_j, c_j, t_j)$, and if $c_j = 1$, then the contribution of this observation is given by evaluating its modeled probability density function at $t_j$. Alternatively, if $c_j = 0$, then the contribution of this observation is given by evaluating its modeled survival distribution function at $t_j = l$, the length of the time under observation. Indeed, the information
provided by a censored observation is that the time to failure is an unknown value greater than the length of the period of observation.

Maximizing $\log(L)$ as a function of $\beta, \alpha, \text{ and } \gamma$ leads to a system of $p + 2$ nonlinear equations which we solved using a Newton-type method. The computer programs for performing the calculations were taken from routines available in SAS (SAS 1997). Inserting the estimated values of $\beta, \alpha, \text{ and } \gamma$ into (2) allowed us to predict the failure time for each member of the population.

3.3 Verifying the Model's Assumptions

In choosing an accelerated failure time model with a baseline Weibull distribution, we assume that the probability distribution for the failure time of each member of the population is Weibull but with parameters that are dependent on the particular member's vector of covariates. More precisely, we assume that the failure time distribution for a member with the covariate vector $\mathbf{x}$ is Weibull with parameters $\alpha \exp(-\gamma(\beta' \mathbf{x}))$ and $\gamma$, where $\alpha, \text{ and } \gamma$ are the parameters of the baseline Weibull distribution.

We examined the validity of this assumption using the following three data sets:

- **real time to failure data**, $\mathcal{L} = \{(x_1, c_1, t_1), ..., (x_n, c_n, t_n)\}$;
- **predicted time to failure data**, $\mathcal{L}^P = \{(x_1, c_1^P, t_1^P), ..., (x_n, c_n^P, t_n^P)\}$;
- **simulated time to failure data**, $\mathcal{L}^S = \{(x_1, c_1^S, t_1^S), ..., (x_n, c_n^S, t_n^S)\}$.

$\mathcal{L}$ is the previously discussed sample of the population containing observed or censored failure times. $\mathcal{L}^P$ is generated by using the covariate vectors of members of $\mathcal{L}$ and using (2) to predict failure times. If the predicted failure time for the $j$th member is less than
the observation time $l$ then $c_j^P = 1$, but if the predicted time is greater than $l$ then $c_j^S = 0$ and $t_j^S = l$. $\mathcal{L}^S$ is generated by using the covariate vectors of members of $\mathcal{L}$ and using the probability distribution of our model (as given by (1)) to simulate failure times. If the simulated failure time for the jth member is less than the observation time $l$ then $c_j^S = 1$ but if the simulated time is greater than $l$ then $c_j^S = 0$ and $t_j^S = l$.

If each of the assumptions underlying the model is correct, then $\mathcal{L}$ and $\mathcal{L}^S$ are two realizations of the same probability distribution. Hence, a graphical test of the validity of our model’s assumptions is to plot the quantiles of the distribution of patients readmitted as predicted by the model against the corresponding quantiles predicted by the simulation. If the model’s assumptions are correct, the resulting (Q-Q) plot will be a straight line. In some cases with our work the graphs are convex. This indicates that the ratio of simulated to fitted densities is decreasing, as can be inferred from (2) of Lemma 4.1.3 of (van Zwet 1970).

A second measure of the reasonableness of the model’s assumptions is to compare the prediction error observed with what would be expected if the assumptions were perfectly correct. If no failure times were censored, then a model’s prediction error could be quantified using the following familiar measure of its Total Mean Squared Error (TMSE):

$$TMSE = \sum_{j=1}^{n}(t_j - t_j^P)^2$$

(4)

Even if our model’s assumptions were correct, TMSE would be greater than zero due to intrinsic variability. We call the inevitable amount of prediction error, as measured by
TMSE, the Ideal Total Mean Squared Error (ITMSE) and, it is given by:

\[ ITMSE = \sum_{j=1}^{n} (t_j^S - t_j^F)^2 \] (5)

A natural measure of the validity of the model's assumptions is to compute the percentage deviation of TMSE from ITMSE.

3.4 Quantifying the Impact of the Covariates

We evaluated the impact of covariates by computing a measure of the model's prediction error (e.g., TMSE) under the assumption of a homogenous population and then calculating the percentage reduction in prediction error under the assumption of a heterogeneous population (i.e., the prediction error observed given the covariates). Our predicted failure time assuming a homogenous population (for which each member has exactly the same Weibull distribution for its failure time) is given by

\[ \text{predicted time to failure} = \alpha^{-\frac{1}{\gamma}} \Gamma(1 + \frac{1}{\gamma}) \] (6)

where \( \alpha \), and \( \gamma \) are estimated by maximizing the log-likelihood of observing the data set \( \mathcal{L} \) as was outlined in Section 3.1. Note that this current maximization problem leads to a system of two nonlinear equations instead of the \( p + 2 \) nonlinear equations as was the case previously. Thus, we compute the model's prediction error using (6), and calculate the percentage reduction in prediction error gained when we use (2).

3.5 Two Methods for Quantifying Prediction Error in the Context of Censored Failure Times
We developed two measures of prediction error for multivariate Weibull models such as these. The first method measures only average deviations between predicted and observed failure times; the second also incorporates a measure of the dispersion in errors across observations.

Computing TMSE in the Context of Censored Failure Times. As alluded to in Section 3.3, prediction error cannot be quantified as described in (4) if some of the observed failure times are censored. To modify (4) in a manner that allows for censored failure times, we define a new random variable $T_i$ as $T_i = \min\{\text{time to readmission}, l\}$ where $l$, as usual, denotes the length of the observation time. Thus, we consider our data set $\mathcal{L}$ as $\mathcal{L} = \{(x_1, \tau_1), (x_2, \tau_2), ..., (x_n, \tau_n)\}$ where $\tau_j$ is the observed value of $T_i$ for a patient with $x_j$ as its vector of covariates. Since there is no censoring in $T_i$ we have that

$$TMSE = \sum_{j=1}^{n} (\tau_j - t_j^P)^2$$  \hspace{1cm} (7)

with $t_j^P$ denoting our prediction for $T_i$ for a patient with $x_j$ as its vector of covariates.

To calculate $t_j^P$, we first note that $E(T_i) = \int_{-\infty}^{\infty} tf_{T_i}(t)\,dt = \int_{0}^{l} tf_{T}(t)\,dt + l \int_{l}^{\infty} f_{\gamma}(t)\,dt$. By using our model for $f_{\gamma}$ in the above equation for $E(T_i)$ we obtain $t_j^P$. So

$$t_j^P = \int_{0}^{l} tae^{-(\beta'x)\gamma t^\gamma - 1} \exp(-ae^{-(\beta'x)\gamma t^\gamma})\,dt$$

$$+ l \int_{l}^{\infty} ae^{-(\beta'x)\gamma t^\gamma - 1} \exp(-ae^{-(\beta'x)\gamma t^\gamma})\,dt$$ \hspace{1cm} (8)

It follows that

$$t_j^P = \alpha^{-\frac{1}{\gamma}} e^{-(\beta'x)} \Gamma(1 + \frac{1}{\gamma}, \alpha l \gamma e^{-(\beta'x)} \gamma + l \exp(-\alpha l \gamma e^{-(\beta'x)} \gamma),$$

where $\Gamma(a, b)$ denotes the incomplete gamma function: $\Gamma(a, b) = \int_{0}^{b} x^{a-1} e^{-x}\,dx$.  

16
Computation of the Total "Hilden" Error (THE). In this section, we derive the formula used to compute a measure of prediction error motivated by the work of Habbema and Hilden (1981) and discussed with us by Professor Hilden. Thus, we refer to this measure as Total Hilden Error (THE). Using the line of reasoning, definitions, and notations just presented for TMSE, the THE is given by

\[
THE = \sum_{j=1}^{n} \left[ \sigma(T_m) + \frac{1}{\sigma(T_m)} (\tau_j^P - t_j)^2 \right]
\]  

(9)

where \( \sigma(T_m) \) denotes the standard deviation of the distribution of \( T_m = T_m(x_j) \) that models \( T_l \) for a patient with applicable covariates \( x_j \). The motivation for (9) is illustrated by a simple example. A true survival distribution with "death for sure at one year" would be modeled perfectly in terms of (4) by a predicted survival distribution with probability of .99 of "instant death" and .01 of "death in 100 years." Thus, in some contexts, both the variability as well as the expectation of the predictive distribution should be considered. Obviously, our example fares badly, as intuitively it should, according to (9).

The above intuitive remarks can be illustrated quantitatively by the following computation. We denote the difference of expectations of TMSE under the true distribution, with an arbitrary \( (T_j) \) and the true \( (T_{m,j}) \) distribution for \( T_m \), by \( \sum_j [E(T_j) - E(T_{m,j})]^2 \). Note this depends only upon expectations. Computing the same difference for THE, we obtain after some algebra \( \sum_j \frac{1}{\sigma(T_{m,j})} [(E(T_j)) - E(T_{m,j}))^2 + (\sigma(T_j) - \sigma(T_{m,j}))^2] \), a function of expectations and standard deviations.

We saw previously that

\[
\tau_j^P = E(T_m) = \alpha^{-\frac{1}{\gamma}} e^{-(\beta'x)} \Gamma(1 + \frac{1}{\gamma}, \alpha \gamma e^{-(\beta'x)} \gamma) + l \exp(-\alpha \gamma e^{-(\beta'x)} \gamma).
\]

Because

\[
\sigma(T_m) = \sqrt{E(T_m^2) - (E(T_m))^2},
\]

to use (9) we need only to know further \( E(T_m^2) \), which one
computes thus

\[ E(T_m^2) = (\alpha^{-\frac{1}{2}} e^{\beta' x})^2 \Gamma \left( 1 + \frac{2}{\gamma}, \alpha t e^{-(\beta' x) \gamma} \right) + t^2 \exp(-\alpha t e^{-(\beta' x) \gamma}). \]

### 3.6 Variable Selection

In selecting our final model, we began with the largest possible model given the covariates available to us, and then examined successively smaller models comparing them based on their predictive accuracy as measured by reductions in TMSE and THE.

After selecting the set of candidate main effects, we explored the data for interactions between covariates. It is well known that sometimes in observational studies such as this, “main effects” are “interactions,” or synergistic effects between predictors and the particulars of the haphazard environment in which data were collected. CART and related binary tree-structured technologies are suited by design to look for synergistic effects of predictors upon outcome (Brieman, Friedman, Olshen and Stone 1993). We used CART\textsuperscript{TM} to look for interactions among candidate covariates described in Section 2. Our main finding was that the predictive power for readmission within these covariates is almost entirely characterized by their value as “main effects.”

### 3.7 Using Cox Models to Validate the Final Covariate Set

It may be that a given covariate could be included or excluded from our model owing to the idiosyncrasies of the particular statistical technique we used. However, any truly predictive variable in the Weibull model also should predict readmission for a somewhat different model, provided that that model also is predictive. We validated our choice of the
final model by examining the covariates' performance when used in a related but different approach to survival analysis: Cox proportional hazards modeling. Especially given the results of the analyses with CART, we focused in the Cox modeling on covariates’ main effects on failure time.

4. RESULTS

4.1 Identifying the Weibull Model

A total of 42,648 hospital discharges for 24,051 patients were included in the analytic sample. Of these, 9,718 (23%) were discharges from substance abuse treatment units, 12,628 (30%) were discharges from psychiatric units, and the remaining 20,302 (48%) were discharges from medical units. As is shown in Table 1, patients in the three types of units differed in terms of their substance abuse, psychiatric, medical, and sociodemographic characteristics. Not surprisingly, when discharged from substance abuse treatment units, patients were most likely to have an alcohol or drug abuse diagnosis. Patients discharged from psychiatric units were most likely to have diagnoses of schizophrenia, post-traumatic stress disorder, and depression. In contrast, patients discharged from medical treatment units were most likely to have diagnoses of chronic medical illnesses such as congestive heart failure (6%), malignancies (9%), congestive obstructive pulmonary disease (21%), or cirrhosis of the liver (7%). Medical patients also were older than patients treated in the other two unit types and more likely to be married (36%), and to have service-connected disabilities (48%).

| Table 1 about here | 19 |
The bulk of all readmissions had failure times of six months or less (Figure 1); only 26% of all failure times were two years or greater. Readmission rates were high for patients discharged from each of the three types of treatment units (Table 2 and Figure 2). Overall, 25% of all patients were readmitted within 33 days of discharge and half were readmitted within 160 days. Among patients treated in substance abuse treatment units, 25% were readmitted to a VA hospital within 147 days of their discharge and half were readmitted within 604 days. No readmissions were observed within the three year follow-up for 40% of these patients. In contrast, one in four patients discharged from either a psychiatric or a medical treatment unit was readmitted within the first month after discharge and, more than 80% of the patients from both types of units were readmitted at least once within three years.

Figure 1, Figure 2, and Table 2 about here

Figure 3 illustrates the observed hazard function for patients discharged from each the three unit types. Patients had their highest risk of readmission immediately following discharge, and the probability of readmission dropped precipitously during the first six months. Subsequently, the hazard continued to decline for patients treated in each of the three treatment unit types although the rate of decline lessened.

Figure 3 about here

Choosing from the covariates described in Section 2, a final prediction model was selected (Table 3a). Predictive value was identified across a wide range of variables, including
patients' treatment unit type, mental health and medical diagnoses, and sociodemographic characteristics. Factors with the greatest marginal contribution to the model (as defined by their log-likelihood ratio chi-square statistics) included unit type, substance abuse diagnoses, whether there was a prior admission in the past year, and whether the patient had a service-connected disability. Significant quadratic terms were observed for age and distance from the discharging facility. Patients without a prior admission who also had no medical diagnoses upon discharge were less likely to be readmitted than was predicted by the two main effects. As was described in Section 3.7, we fit a Cox proportional hazards model to the same data using the same regressors that were selected for the Weibull model. Table 3b makes clear that the Weibull parameter estimates are about the same in the Cox replication in the sense that their signs agree, and the chi-square values are nearly identical. The magnitude of the parameters in the Cox model are just over half the corresponding Weibull fitted values. Explanation is easy if one focuses on the exponential parts of the hazards in the two models. Algebra shows that \( \exp\{-\beta'x_t\} \) in the Cox model corresponds to \( \exp\{-\gamma(\beta'x)t\} \) in the Weibull. Note that our overall maximum likelihood estimate of \( \gamma \) is .5913. Thus, the ratio of Cox to Weibull parameters should be exactly that if the exponential parts of the hazards are to be the same, that is, if the Weibull is the perfect fit for the more general Cox situation. To check, we took the ratio of corresponding values for the 23 parameters for which we had more than one significant figure for each regression coefficient. The sample mean ratio is .560 and sample standard deviation .024. All but one of the 23 are actually less than .5913 (Hospital #6). Therefore, in Cox terms, our estimate of hazard is slightly too steep when judged by its exponential
part. There can be compensation in the Cox fit that comes from its baseline component. In any case, the Cox and Weibull fits of actual survival would be very close for all practical purposes for any particular subject. Our model is simpler, being finite-dimensional in its parameterization.

Table 3a and Table 3b about here

4.2 Assessing the Appropriateness of the Weibull Model

Using the method described in Section 3.3, we assessed the extent to which the hazard function was Weibull conditional on the chosen covariates. The Q-Q plot for the chosen model was almost, although not perfectly, straight indicating a reasonable fit (Figure 4). In comparison, the curve based on a Weibull model without covariates was more convex. The “ideal” TMSE that is expected for a Weibull model without covariates was $7.07 \times 10^9$ or 12% less than the TMSE that we observed (i.e., $7.94 \times 10^9$). In Table 4a-d, we refer to this comparison between simulated and predicted error as the “Horizontal Deviation” or HD. When the “ideal” and observed TMSE are compared for the chosen model with covariates (i.e., the HD for the model with covariates), the difference decreases to 6% suggesting that the hazards are, in fact, nearly Weibull, conditional on the chosen covariates. The HD calculated using the Hilden Error Criterion was 8% for a model with no covariates and 4% when the chosen covariates were included (Table 4b).
To better understand the behavior of these new measures of fit, we estimated patients’ failure times using the observed data set and a model of the underlying hazards that we knew was erroneous. Specifically, we used a hazard function with a fixed and positive slope (i.e., corresponding to a constant rate of increasing readmission risk). The Q-Q plot associated with this model is more markedly convex than is the approximately linear curve for the Weibull model with covariates (see Figure 4). As shown in Tables 4c and 4d, the HD statistics for this model were markedly greater than those for the Weibull model demonstrating that the distributional assumption is untenable.

4.3 Assessing the Impact of the Covariates

As described in Section 3.4, we assessed the impact of the covariates by comparing the prediction error for the chosen model against the error for a model with no predictors. In Tables 4a-d, we refer to this comparison as the Vertical Deviation or VR. The results indicate that the VR calculated using the MSE criterion is 25%. The VR using the Hilden criterion is 16% (see Tables 4a and 4b). For each of these two measures, the VR is somewhat higher than would be expected if the hazards were perfectly Weibull-distributed (i.e., the IVR) indicating that some of the reduction is associated with deviations from that distribution. Because the VR is less when the Hilden criterion is used, it appears that there is heterogeneity across observations in the size of the prediction errors. In other words, our
model better predicts failure times for some patients than for others. This characteristic of the model is illustrated graphically in Figure 5. Five percent of all observations are responsible for 25% of the TMSE and 14% are responsible for half of the TMSE. When THE is used as the benchmark, 9% of all observations contribute 25% of the total error and 26% contribute half.

Figure 5 about here

Further exploration of the distribution of prediction errors led to two additional observations. First, the model more accurately predicts readmissions for patients with intermediate readmission times (roughly, patients with actual readmission times of between 200 days and two years). The average MSE decreases for patients with actual readmission times between one and 420 days and then increases substantially for patients with readmission times greater than 420 days (Figure 6a). We observed a similar pattern using the Hilden criterion (Figure 6b). When the underlying distribution is forced to be Weibull, as in the simulation data set, a very similar pattern of errors is observed. We conclude that the increased error in prediction for patients with relatively short and long observed failure times results from the increased difficulty of predicting readmission for these patients given the covariates used in the model rather than to any greater intrinsic propensity of the model to make errors in these regions.

Figure 6a and Figure 6b about here
4.4 Subgroup Analyses To Improve Model Fit and Prediction

Treatment goals often vary among substance abuse units, psychiatric units, and medical units. To assess whether our prediction models were better suited to discharges from some unit types than others, we analyzed separately the readmission times for patients treated in each unit type. As a first step, we used the covariate set from our overall model and fit models separately for discharges from each of the three unit types. The $\alpha$ and $\gamma$ for the model fit to psychiatric unit data (0.0054 and 0.57, respectively) were essentially the same as those for the model fit to medical unit data. However, the $\alpha$ and $\gamma$ for the model fit to substance abuse unit data were somewhat different (0.0096 and 0.65). While the measures of fit (i.e., the Q-Q plots and HD calculations) for both the psychiatric unit model and the medical unit models were essentially the same as those for the overall model, the fit statistics for the substance abuse model were better. Specifically, the Q-Q plot for the substance abuse model was more nearly linear, and the HD calculation showed that the data were nearly perfectly modeled as a Weibull process. Because of their similarities in fit statistics, we grouped the data from psychiatric units and medical units for subsequent analyses.

In an effort to improve the fit of the combined psychiatric/medical unit model, we examined "enhanced models" that included additional two-way interaction terms based on the binary covariates we had available. Each pair of covariates yielded four possible interactions for approximately 100 candidate interaction terms overall. From this set, we excluded interaction terms for which less than 33% of the observations had the characteristic of interest because terms with little variance in the data could not possibly improve
prediction. Of the remaining 25 candidate terms, only nine had any discernable effect on psychiatric/medical unit patients' readmission risk. Including all of these nine terms in the psychiatric/medical unit model led to no improvement in the Q-Q plot. Furthermore, compared to a psychiatric/medical unit model without these additional terms, the HD for this enhanced model decreased only 1%. Thus, fitting a model separately for combined psychiatric and medical unit discharges led to no improvements in the appropriateness of the Weibull characterization of the problem.

We further considered the impact of stratification by unit type on the prediction of readmissions from substance abuse units. As stated previously, when only these discharges were considered, the Q-Q plot was more nearly linear, and the HD was nearly 0%. To determine whether these discharges were truly "better served" in a separate model than in the overall model, we calculated the total MSE associated with the separate model and compared it to the total MSE for the same observations in the overall model. The results suggested that fitting a model separately for these admissions improved the prediction error for their readmission times only by 2%.

Finally, we conducted analyses to further explore the meaning of the differences in measures of model fit (Q-Q plot and HD) across unit types in the context of the substantially differing censoring rates for substance abuse discharges compared to psychiatric and medical unit discharges. Because the outcome we are predicting is the minimum of patients' readmission and censoring times, shorter observation time, associated with greater censoring rates, increase the likelihood that our model will predict accurately. For the psychiatric/medical unit model, we found that incrementally decreasing the length of the
observation time from three years to 180 days decreased the HD from 10% to 5%. Further decreases in the time interval were associated with little additional decreases in HD. Thus, we conclude that at least half of the difference in HD between the psychiatric/medical unit model and the substance abuse model was a function of the difference in the censoring rates.

In summary, we found that fit statistics for a model based only on substance abuse unit discharges were better than those for a model based only on psychiatric and medical unit discharges. However, for psychiatric and medical unit discharges, consideration of a large number of alternative covariate sets failed to disclose any better models, and the prediction error for the separate substance abuse unit model was only slightly better than the error for those data in the overall model. Differences in fit statistics for the two separate models are largely a function of differences in the observed censoring rates for the two discharge types. Hence, we maintained as our solution to this problem the overall model including all substance abuse unit, psychiatric unit, and medical unit discharges. Aside from the rationale just presented, the one comprehensive model is both a more parsimonious result and, not less important, provides a clearer solution to this problem for policymakers.

4.5 Predicted Readmission Time for Important Subgroups of Patients

Using our final model, Figure 7 illustrates the predicted readmission probability over time for five prototypical groups of patients with varying characteristics. As the figure shows, the model predicts that patients discharged from substance abuse treatment units without a prior admission will have a lower readmission risk over time than any of the other
four patient groups. In contrast, patients with the greatest risk of readmission are elderly (this is more than 65 years of age) and discharged from a medical unit with a diagnosis of chronic obstructive pulmonary disease. Patients discharged from psychiatric treatment units have intermediate predicted readmission risks.

Figure 7 about here

5. CONCLUSIONS

The purpose of this study was to use readily available data maintained by the VA health care system to develop prediction models that inform policy on the demand for inpatient care among substance abuse patients. Through the process of developing these models, we identified relatively novel approaches to survival analysis that could be useful to other researchers addressing similar questions with similar data sets.

In concert with prior studies, we found that many VA hospital patients with substance abuse disorders are at extremely high risk for readmission within the three years following discharge. Overall, our model predicts that half of all patients will be readmitted within 500 days and 60% will be readmitted within the first two years. In addition, we found that information obtainable at the time of a patient's discharge is important in predicting his or her subsequent readmission risk. In particular, patients' sociodemographic characteristics, psychiatric characteristics, and access to VA care are important predictors of their course. Independent of these factors, treatment factors such as the hospital unit in which the patient received care and whether the admission was the patient's first during the period of observation were important determinants of their course. We also learned from this study
that routinely-collected administrative data on substance abuse inpatients may explain more of the variation in readmission risk than was known previously. Prior studies using logistic regression typically resulted in improvements in deviance of less than 12%. In this study, we were able to achieve a reduction in mean squared error of 25% using the criterion focusing on the average deviation between predicted and observed readmission times and 16% using the Hilden criterion. This finding should be encouraging to researchers and policy makers interested in using analyses such as these to inform program planning.

Further, we learned that the hazard function determining the readmission time for these patients is conditionally Weibull and that it decreases over time. This suggests that the longer patients remain out of the hospital, the lower their risk of readmission will be. A corollary to this finding is that patients are at greatest risk for readmission during the first months following discharge. Preventive services for these patients such as intensive outpatient care, reassessment, and telephone follow-up may be most important and cost-effective during this early, critical phase. The finding that a patient's risk of readmission decreases over time corroborates past studies from the addictions treatment literature demonstrating that the probability of relapse after initial abstinence decreases over time (Vaillant 1995).

A potential use of models such as this is to develop accurate case mix adjustment strategies for substance abuse patients. The goal of case mix adjustment is to make comparisons among provider organizations (e.g., facilities or treatment programs) on an outcome such as readmission while controlling for differences in the mix of patients across providers (Pine and Harperd 1994). Such methods already are being used to inform health
care consumers' decisions regarding where to seek medical care (Hibbard and Jewett 1996). Significant strides have been made in extending these techniques to the field of substance abuse treatment (Peterson et al. 1994), and experts have demonstrated that very different conclusions about providers' relative quality are suggested by case mix-adjusted and non-case mix-adjusted outcomes (Phibbs, Swindle and Recine 1997). This study extends those efforts by allowing for a more accurate prediction of readmission for VA substance abuse patients.

In recent decades, researchers have used a variety of emerging multivariate techniques to understand the process of health service use among substance abuse patients. Through these studies, scientists, clinicians, and policy makers have learned about the characteristics of substance abuse patients at risk for readmission and the clinical processes that might be changed to prevent these events. However, most statistical analyses of these problems have shed little light on patients' predicted event-history beyond a singular future time point. In addition, prior efforts have not yielded models that are sufficiently predictive to contribute meaningfully to the policy debate. Though no single component of our analyses breaks unexpected theoretical ground, the combination of approaches we take is original and could be of value for predicting important events such as readmissions.

Here, we used a multivariate Weibull model to estimate failure times. By Weibull, we mean that conditional on values of covariates, there is a Weibull probability mechanism by which, at least approximately, nature has generated the "failure time" (time to readmission) for a subject. A major benefit of this model over the more common Cox model is that readers can estimate directly the predicted failure time given the parameter estimates
and a patient's particular values for the covariates. It is the case here that we want flexible
but simple models; they should enable understanding of where on the time scale they fit
well, and where they do not. Not only because it satisfies these criteria, but also because
estimation of survival is direct and simple from the hazard, do we employ and think well
of our two-parameter, "accelerated failure time" Weibull models.

As we have described, there are four sets of numbers upon which we base inferences
about these Weibull models. The first, and most important, is the vector of min(failure
time, three years) and covariates. The second is the expected failure time based on the
patient's covariates and the model, while the third is the standard deviation for that
patient, again estimated from the model. The last is fundamental to our assessment of
goodness-of-fit: a simulated failure time from a Weibull distribution with subject-specific
parameters.

We present two types of goodness-of-fit measures for these Weibull models based on the
four types of calculations just described. The first type allows us to assess graphically and
quantitatively the reasonableness of the choice of the Weibull model for this application.
The second type allows us to quantify the predictive value associated with various sets of
covariates. For both types of goodness-of-fit, we used Total Mean Squared Error (TMSE)
as a measure of prediction error that characterizes the total amount of deviation between
observed and predicted failure times. Discussions with J. Hilden have convinced us that
functionals of observed, modeled, and simulated minima are not sufficient to quantify
goodness-of-fit. Thus, we also developed the Total Hilden Error (THE) goodness-of-fit
measure to account for variation in prediction error across observations as well as the
total average deviation. Both TMSE and THE have been formulated to accommodate
the censored data that are common in most survival data sets. These measures provide
analysts with benchmarks against which various models can be compared and a means of
characterizing the predictive value of the best model.

Assessing goodness-of-fit for models in prospective mode has not been the usual con-
cern in survival analysis. This may owe to the Cox model as usually implemented having
focus on proportional hazards, as if baseline hazard and overall survival are of secondary
interest. To be fair, the literature is replete with examples of chi-square tests of goodness
of fit being brought to survival analysis in which failure times are categorized into succes-
sive groups or "bins" that typically are defined arbitrarily. In the increasingly common
instances such as ours in which analysts are estimating survival functions using large sec-
ondary databases of scores of thousands of subjects, almost any parametric model fit to
the data will show statistical significance even for meaningless covariates.

When using models such as this, it is critical that one assess the extent to which
the hazards are conditionally Weibull. When subject-specific conditional distributions are
Weibull and the empirical distribution of covariates is not a point mass, then the uniqueness
of Laplace transforms and calculus imply that the unconditional distribution of survival
cannot be Weibull for any choice of parameters. For any correct model of survival, the
expected and observed minima of failure times and censoring times should match those
of subject-specific simulated data. For example, the sum of squared differences between
observed and expected minima should approximately equal the sum of squared differences
between simulated and expected minima, where there is one simulated minimum per sub-
ject chosen with that subject's covariates and according to the model. Thus, it is not enough that the modeled sum of squared differences with observed is smaller than the comparable sum of squares in which each subject is fitted with a group average value. Indeed, as shown in Figures 4c and 4d, even a dramatic reduction may simply signify an incorrect choice of the underlying distribution. When used prospectively to predict readmissions for additional patients, such a model will fail miserably.

We have mentioned both goodness-of-fit and the uniqueness of Laplace transforms as they come into contact in assessing goodness-of-fit. The uniqueness theorem and inversion (which we did not attempt) could enable us to infer what the empirical distribution of linear forms that figure in the Weibull modeling would have to be were the conditional Weibull model correct and our estimation of its parameters accurate. This inferred distribution and the estimated parameters could be combined with the known empirical distribution of covariates to assess goodness-of-fit. However, the exercise would be tedious and extremely difficult computationally. Instead, we took the simpler and more conventional statistical approach of Q-Q comparisons of data simulated from the various conditional Weibull distributions and the observed minima. With abscissa the true empirical quantiles and ordinate what are given from the model-based simulations, the plot is decidedly though only slightly convex. This suggests that the ratio of densities, true to modeled, is decreasing and thus that the model fits better at intermediate rather than early or late times.

As shown in Figure 5, we observed that roughly 20% of all observations contributed more than 80% of the total prediction error using either TMSE or THE as the criterion. We learned that their are two important subgroups of VA substance abuse patients whose
readmissions we were less successful in predicting using a Weibull model and the predictors we had available: (a) patients who are readmitted within the first weeks following their initial discharge; and (b) patients who never return to VA within three years. Specifically, our model over-predicts the readmission times for patients who are readmitted early and under-predicts the readmission times for patients who never come back at all. This pattern of error is clearly depicted by the U-shaped curve in Figures 6a and 6b.

This pattern of errors is consistent with the conditional Weibull model. When the errors across time intervals were plotted using a simulation of the Weibull model, the same U-shaped pattern was observed. Thus, it appears that the prediction problem we are studying is intrinsically more difficult for patients with relatively short and relatively long readmission times. This is, of course, consistent with what we see. Were we to formulate our problem decision theoretically, it is as if the “Bayes risk” of the “Bayes rule” comes mostly from early and late times. It may be that factors other than the covariates available in this data set dominate during these periods. For example, patients’ severity of illness and social support networks, both of which are characterized poorly in this data set, may profoundly affect their experiences during the first days after they leave the hospital. Alternatively, information available at the time of discharge may have little relevance to a patient’s readmission risk several years later.

The bulk of the patients for whom our model predicted poorly were those who never returned to VA following discharge. On average, these patients had an expected readmission time of 1,300 days. Because patients were followed for a maximum of 1,095 days, it remains unknown how many were subsequently readmitted. Ancillary discriminant analy-
ses and recursive partitioning failed to identify reliably who these patients were based on the available information (data not shown). It is likely that some of these patients receive the bulk of their health care outside of the VA and had readmissions that were not observed in our VA data set. For example, eligible Medicare patients receive as much as 37% of their hospital care from non-VA hospitals (Fleming, Fisher, Chang, Bubolz and Malenka 1992). Other patients who never returned may have “beaten the odds” and remained healthy despite the risk factors they carried. Extensions of this study should include the examination of the factors that determine long-time health after an inpatient episode.
REFERENCES


Table 1. Description of the Analytic Sample

<table>
<thead>
<tr>
<th></th>
<th>Substance Abuse Units</th>
<th>Psych. Units</th>
<th>Medical Units</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mental Health Diagnoses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol Abuse</td>
<td>81.2</td>
<td>62.8</td>
<td>52.0</td>
<td>61.9</td>
</tr>
<tr>
<td>Drug Abuse</td>
<td>61.6</td>
<td>44.4</td>
<td>24.4</td>
<td>38.8</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>3.6</td>
<td>34.0</td>
<td>2.2</td>
<td>11.9</td>
</tr>
<tr>
<td>Post Traumatic Stress</td>
<td>3.6</td>
<td>11.7</td>
<td>1.5</td>
<td>5.0</td>
</tr>
<tr>
<td>Depression</td>
<td>6.3</td>
<td>18.3</td>
<td>3.0</td>
<td>8.3</td>
</tr>
<tr>
<td><strong>Sociodemographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married (yes)</td>
<td>22.5</td>
<td>19.7</td>
<td>35.5</td>
<td>27.9</td>
</tr>
<tr>
<td>Service-Connected Disability (yes)</td>
<td>26.3</td>
<td>30.0</td>
<td>47.6</td>
<td>37.5</td>
</tr>
<tr>
<td>Age (mean, SD)</td>
<td>41.1(10.1)</td>
<td>41.6(10.0)</td>
<td>54.3(12.9)</td>
<td>47.5(13.2)</td>
</tr>
<tr>
<td>Distance from Care</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5 miles</td>
<td>21.0</td>
<td>26.1</td>
<td>27.4</td>
<td>25.6</td>
</tr>
<tr>
<td>5 to &lt; 10 miles</td>
<td>21.1</td>
<td>22.1</td>
<td>23.2</td>
<td>22.4</td>
</tr>
<tr>
<td>10 to &lt; 50 miles</td>
<td>37.6</td>
<td>34.3</td>
<td>29.1</td>
<td>32.6</td>
</tr>
<tr>
<td>50+ miles</td>
<td>20.3</td>
<td>17.5</td>
<td>20.3</td>
<td>19.5</td>
</tr>
<tr>
<td><strong>Medical Diagnoses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congestive Heart Failure</td>
<td>0.3</td>
<td>0.2</td>
<td>6.2</td>
<td>3.1</td>
</tr>
<tr>
<td>Malignancy</td>
<td>0.2</td>
<td>0.2</td>
<td>9.2</td>
<td>4.5</td>
</tr>
<tr>
<td>COPD</td>
<td>5.6</td>
<td>4.1</td>
<td>21.3</td>
<td>12.6</td>
</tr>
<tr>
<td>Cirrhosis of the Liver</td>
<td>1.3</td>
<td>0.6</td>
<td>6.7</td>
<td>3.7</td>
</tr>
<tr>
<td><strong>Prior Admission (yes)</strong></td>
<td>37.6</td>
<td>70.3</td>
<td>64.3</td>
<td>60.0</td>
</tr>
<tr>
<td><strong>Number of Admissions</strong></td>
<td>9,718</td>
<td>12,628</td>
<td>20,302</td>
<td>42,648</td>
</tr>
</tbody>
</table>

NOTE:

Cell entries are column percents; COPD = congestive obstructive pulmonary disease.
Table 2. Readmission Times by Unit Type

<table>
<thead>
<tr>
<th></th>
<th>Substance Abuse Units</th>
<th>Psychiatric Units</th>
<th>Medical Units</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean (days)</td>
<td>616</td>
<td>314</td>
<td>327</td>
<td>389</td>
</tr>
<tr>
<td>median (days)</td>
<td>604</td>
<td>104</td>
<td>116</td>
<td>160</td>
</tr>
<tr>
<td>25% Readmitted (days)</td>
<td>147</td>
<td>22</td>
<td>28</td>
<td>33</td>
</tr>
<tr>
<td>75% Readmitted (days)</td>
<td>N/A</td>
<td>473</td>
<td>512</td>
<td>824</td>
</tr>
<tr>
<td>% Censored at 3 years</td>
<td>40</td>
<td>16</td>
<td>17</td>
<td>22</td>
</tr>
</tbody>
</table>
Table 3a. Parameter Estimates and Standard Errors for the Selected Weibull Model

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Beta</th>
<th>SE</th>
<th>Chi-square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mental Health Diagnoses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol Abuse</td>
<td>.3507</td>
<td>.0213</td>
<td>270.7</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Drug Abuse</td>
<td>.3737</td>
<td>.0224</td>
<td>278.5</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>-.4529</td>
<td>.0329</td>
<td>189.3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Post Traumatic Stress</td>
<td>-.2013</td>
<td>.0449</td>
<td>20.1</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Depression</td>
<td>-.1841</td>
<td>.0356</td>
<td>26.8</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Sociodemographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>.2062</td>
<td>.0222</td>
<td>86.7</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Service-Connected Disability</td>
<td>-.2409</td>
<td>.0203</td>
<td>140.2</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Distance from Care</td>
<td>.0005</td>
<td>.0001</td>
<td>16.0</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Age</td>
<td>-.0151</td>
<td>.0057</td>
<td>6.9</td>
<td>.0085</td>
</tr>
<tr>
<td>Distance$^2$</td>
<td>-2.3001*E^-7</td>
<td>5.7210*E^-8</td>
<td>16.2</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Age$^2$</td>
<td>.0001</td>
<td>.0001</td>
<td>5.6</td>
<td>.0179</td>
</tr>
<tr>
<td><strong>Medical Diagnoses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congestive Heart Failure</td>
<td>-.4874</td>
<td>.0536</td>
<td>82.8</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Malignancy</td>
<td>-.7252</td>
<td>.0462</td>
<td>246.4</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Cirrhosis of the Liver</td>
<td>-.3308</td>
<td>.0514</td>
<td>41.4</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>No Medical Diagnoses (yes)</td>
<td>.2310</td>
<td>.0302</td>
<td>58.5</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Prior Admission</strong></td>
<td>-.7371</td>
<td>.0382</td>
<td>373.3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Bedsection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substance Abuse Unit</td>
<td>.6468</td>
<td>.0321</td>
<td>406.0</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Psychiatric Unit</td>
<td>-.2360</td>
<td>.0290</td>
<td>66.3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Length of Stay</td>
<td>.0052</td>
<td>.0005</td>
<td>94.3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Facility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital #1</td>
<td>.4409</td>
<td>.0473</td>
<td>86.9</td>
<td>.0001</td>
</tr>
<tr>
<td>Hospital #2</td>
<td>.1650</td>
<td>.0406</td>
<td>16.5</td>
<td>.0001</td>
</tr>
<tr>
<td>Hospital #3</td>
<td>.3365</td>
<td>.0509</td>
<td>43.6</td>
<td>.0001</td>
</tr>
<tr>
<td>Hospital #4</td>
<td>-.3710</td>
<td>.0347</td>
<td>114.5</td>
<td>.0001</td>
</tr>
<tr>
<td>Hospital #5</td>
<td>.5016</td>
<td>.0361</td>
<td>193.8</td>
<td>.0001</td>
</tr>
<tr>
<td>Hospital #6</td>
<td>.0893</td>
<td>.0446</td>
<td>4.0</td>
<td>.0452</td>
</tr>
<tr>
<td>Hospital #7</td>
<td>.1569</td>
<td>.0416</td>
<td>14.3</td>
<td>.0002</td>
</tr>
<tr>
<td>no medical dx*1st admission</td>
<td>.3627</td>
<td>.0450</td>
<td>64.9</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

$\gamma = .5913$

$\alpha = .0240$
Table 3b. Parameter Estimates and Standard Errors for a Cox Proportional Hazard Model Including the Parameters Used In the Weibull Model

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Beta</th>
<th>SE</th>
<th>Chi-Square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mental Health Diagnoses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol Abuse</td>
<td>.1968</td>
<td>.0126</td>
<td>244.2</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Drug Abuse</td>
<td>.2078</td>
<td>.0132</td>
<td>246.2</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>-.2480</td>
<td>.0195</td>
<td>162.0</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Post Traumatic Stress</td>
<td>-.1044</td>
<td>.0265</td>
<td>15.5</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Depression</td>
<td>-.1013</td>
<td>.0210</td>
<td>23.2</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Sociodemographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>.1142</td>
<td>.0131</td>
<td>76.0</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Service-Connected Disability</td>
<td>-.1296</td>
<td>.0120</td>
<td>115.9</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Distance from Care</td>
<td>.0003</td>
<td>.0001</td>
<td>11.9</td>
<td>.0006</td>
</tr>
<tr>
<td>Age</td>
<td>-.0090</td>
<td>.0034</td>
<td>7.1</td>
<td>.0079</td>
</tr>
<tr>
<td>Distance²</td>
<td>-1.19*E⁻⁷</td>
<td>3.4004*E⁻⁴</td>
<td>12.2</td>
<td>.0005</td>
</tr>
<tr>
<td>Age²</td>
<td>.0001</td>
<td>.0001</td>
<td>6.0</td>
<td>.0140</td>
</tr>
<tr>
<td><strong>Medical Diagnoses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congestive Heart Failure</td>
<td>-.2660</td>
<td>.0317</td>
<td>70.6</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Malignancy</td>
<td>-.4193</td>
<td>.0273</td>
<td>235.8</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Cirrhosis of the Liver</td>
<td>-.1894</td>
<td>.0303</td>
<td>38.9</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>No Medical Diagnoses (yes)</td>
<td>.1257</td>
<td>.0226</td>
<td>49.5</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Prior Admission</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.4040</td>
<td>.0226</td>
<td>320.7</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Bedsection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substance Abuse Unit</td>
<td>.3689</td>
<td>.0189</td>
<td>379.1</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Psychiatric Unit</td>
<td>-.1363</td>
<td>.0171</td>
<td>63.3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Length of Stay</td>
<td>.0030</td>
<td>.0003</td>
<td>91.8</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Facility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital #1</td>
<td>.2480</td>
<td>.0280</td>
<td>78.7</td>
<td>.0001</td>
</tr>
<tr>
<td>Hospital #2</td>
<td>.0929</td>
<td>.0240</td>
<td>15.0</td>
<td>.0001</td>
</tr>
<tr>
<td>Hospital #3</td>
<td>.1859</td>
<td>.0301</td>
<td>38.1</td>
<td>.0001</td>
</tr>
<tr>
<td>Hospital #4</td>
<td>-.2089</td>
<td>.0205</td>
<td>103.8</td>
<td>.0001</td>
</tr>
<tr>
<td>Hospital #5</td>
<td>.2872</td>
<td>.0213</td>
<td>181.4</td>
<td>.0001</td>
</tr>
<tr>
<td>Hospital #6</td>
<td>.0570</td>
<td>.0264</td>
<td>4.7</td>
<td>.0305</td>
</tr>
<tr>
<td>Hospital #7</td>
<td>.0908</td>
<td>.0246</td>
<td>13.7</td>
<td>.0002</td>
</tr>
<tr>
<td>no medical dx*1st admission</td>
<td>.2066</td>
<td>.0266</td>
<td>60.31</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Table 4a. Measures of Fit for the Chosen Weibull Model Measured Using The Mean Squared Error (MSE) Criterion

<table>
<thead>
<tr>
<th></th>
<th>ITMSE</th>
<th>TMSE</th>
<th>IVR</th>
<th>VR</th>
<th>HD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull Model Without Covariates</td>
<td>7.07 x 10^9</td>
<td>7.94 x 10^9</td>
<td>12%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull Model With Covariates</td>
<td>5.63 x 10^9</td>
<td>5.99 x 10^9</td>
<td>20%</td>
<td>25%</td>
<td>6%</td>
</tr>
</tbody>
</table>

Table 4b. Measures of Fit for the Chosen Weibull Model Measured Using The Hilden Error (THE) Criterion

<table>
<thead>
<tr>
<th></th>
<th>ITHE</th>
<th>THE</th>
<th>IVR</th>
<th>VR</th>
<th>HD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull Model Without Covariates</td>
<td>3.40 x 10^7</td>
<td>3.68 x 10^7</td>
<td>8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull Model With Covariates</td>
<td>2.96 x 10^7</td>
<td>3.09 x 10^7</td>
<td>13%</td>
<td>16%</td>
<td>4%</td>
</tr>
</tbody>
</table>

Table 4c. Measures of Fit for A Model Chosen Under a Clearly Incorrect Assumption Regarding the Underlying Hazard (Measured Using the Mean Squared Error Criterion)

<table>
<thead>
<tr>
<th></th>
<th>ITMSE</th>
<th>TMSE</th>
<th>IVR</th>
<th>VR</th>
<th>HD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wrong Model Without Covariates</td>
<td>4.01 x 10^9</td>
<td>12.6 x 10^9</td>
<td>214%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wrong Model With Covariates</td>
<td>2.46 x 10^9</td>
<td>6.89 x 10^9</td>
<td>39%</td>
<td>45%</td>
<td>180%</td>
</tr>
</tbody>
</table>

Table 4d. Measures of Fit for A Model Chosen Under a Clearly Incorrect Assumption Regarding the Underlying Hazard (Measured Using the Hilden Error Criterion)

<table>
<thead>
<tr>
<th></th>
<th>ITHE</th>
<th>THE</th>
<th>IVR</th>
<th>VR</th>
<th>HD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wrong Model Without Covariates</td>
<td>2.56 x 10^7</td>
<td>41.2 x 10^7</td>
<td>1500%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wrong Model With Covariates</td>
<td>1.95 x 10^7</td>
<td>3.80 x 10^7</td>
<td>24%</td>
<td>91%</td>
<td>95%</td>
</tr>
</tbody>
</table>

NOTE:

ITMSE = ideal total mean squared error; TMSE = total mean squared error; IVR = ideal vertical reduction; VR = vertical reduction; HD = horizontal deviation; ITHE = ideal total Hilden squared error; THE = total Hilden error.
FIGURE CAPTIONS

FIGURE 1
Cumulative Distribution of Readmission Times for 42,648 1990 Discharges among VA Hospital Patients with Substance Abuse Disorders. The vertical rise in the curve reflects censored observations.

FIGURE 2
Kaplan-Meier Estimates of the Probability of Readmission for VA Hospital Patients with Substance Abuse Disorders. The solid line represents substance abuse unit discharges. The dashed line represents psychiatric unit discharges. The dotted line represents medical unit discharges. The dotted/dashed line represents discharges from all units.

FIGURE 3
Readmission Hazard Functions for VA Hospital Patients with Substance Abuse Disorders Computed by the Life Table Method. The solid line represents substance abuse unit discharges. The dashed line represents psychiatric unit discharges. The dotted line represents medical unit discharges. The dotted/dashed line represents discharges from all units.

FIGURE 4
Quantile-Quantile (Q-Q) Plot Comparing the Empirical Distribution of Failure Times with Predicted Distributions of Failure Times Computed under Various Assumptions on the Underlying Generating Process. The solid line represents the Q-Q plot for a Weibull simulation and the selected covariates. The dashed line represents the Q-Q plot for a Weibull simulation and no covariates. The dotted line represents the Q-Q plot for a simulation assuming a linear hazard with positive slope.
FIGURE 5

Distribution across Discharges of the Total Prediction Error for the Weibull Model. Discharges are sorted by their relative contribution to the total prediction error. The solid line represents prediction error calculated using the mean square error (MSE) criterion. The dashed line represents prediction error calculated using the Hilden error criterion.

FIGURE 6a

Contribution to Total Mean Squared Error (MSE) as a Function of the Observed Readmission Time. The value $y$ that corresponds to day post-discharge $x$ is the total MSE for patients readmitted on day $x$ divided by the number of patients readmitted on day $x$. The solid line represents observed MSE (i.e., the MSE comparing the observed readmission time and the readmission time predicted by the model and patients' covariate values). The dashed line represents the simulated MSE (i.e., the MSE comparing the observed readmission time and the readmission time predicted using the simulation data set).

FIGURE 6b

Contribution to Total Hilden Error (THE) as a Function of the Observed Readmission Time. The value $y$ that corresponds to day post-discharge $x$ is the THE for patients readmitted on day $x$ divided by the number of patients readmitted on day $x$ (i.e., the Hilden error comparing the observed readmission time and the readmission time predicted by the model and patients' covariate values). The dashed line represents the simulated Hilden error (i.e., the Hilden error comparing the observed readmission time and the readmission time predicted using the simulation data set).
FIGURE 7

Predicted Cumulative Probability of Readmission for Prototypical Groups of Patients. The solid line represents patients younger than 65 years of age discharged from substance abuse units with both alcohol and drug abuse diagnoses and no admission during the previous year; the short dashed line represents patients with identical features but with VA hospitalizations during the past year; the dotted line represents alcoholic patients 65 years of age or older discharged from medical units without diagnoses of drug abuse or schizophrenia and with no VA hospitalizations during the previous year; the long dashed line represents patients younger than 65 years of age discharged from psychiatric units with both cited diagnoses but no admission during the previous year; the dotted/dashed line represents patients similar in all respects to the latter group but who were hospitalized during the past year.
Cumulative Number of Readmissions

Figure 1

Days Post-discharge
Hazard of Readmission
Simulated Quantile Function

Figure 4  Empirical Quantile Function
Figure 5  Discharges Ordered by Decreasing Contribution
Probability of Readmission

Figure 7

Days Post-discharge