

# PROFILING SUBSTANCE ABUSE PROVIDER TRENDS IN HEALTH CARE DELIVERY SYSTEMS

---

JAMES F. BURGESS, JR., VIRIDIANA LOURDES AND MIKE WEST

**Abstract.** In profiling of health care providers, distinguishing extreme behavior from random variation with incomplete risk adjustment requires multiple observations on each profiling unit. Unlike typical health care delivery system studies, the US Department of Veterans Affairs (VA) information system does provide longer time series of data at individual, institutional and system levels. This information resource allows us to develop incisive profiling analyses that isolate and highlight system wide improvements and institution-specific profiles in the context of risk adjustments using several covariates. This is illustrated here in the context of substance abuse care. One common process monitor for systems delivering substance abuse care is follow-up outpatient care within a certain number of days after inpatient substance abuse discharges. The VA system provides ten years of such data, at the individual level, and we employ this to build hierarchical models that profile providers within the system. Our models use logistic regression, longitudinal random effects models at the individual patient level, combined with simple time series models of institutional effects across years. This structure effectively captures variability across hospitals within each year as well as systematic dependencies within hospitals from year to year. Analysis depends on Markov chain Monte Carlo methods to derive posterior inferences for all parameters. Results indicate significant system wide improvement in the monitor in addition to large amounts of variation in this improvement across medical centers. Covariates such as age of patient, VA treatment priority, and diagnoses (where psychotic patients have lower return rates) help to illustrate important potential new health policy interventions and the outcomes of previous interventions.

**Key Words:** Bayesian analysis, Profiling, Hierarchical Models, Longitudinal Data

Jim Burgess is in the VA Management Science Group, Bedford, MA, and has a faculty appointment in the Health Services Department of the Boston University School of Public Health. Mike West is at the Institute of Statistics & Decision Sciences, Duke University. Viri Lourdes is in the Department of Statistics and Actuarial Sciences, at the Instituto Tecnológico Autónomo de México (ITAM). We thank two referees for helpful comments. Opinions and conclusions expressed do not necessarily reflect official positions of the US Department of Veterans Affairs.

*Contact author:* James F. Burgess, Jr., VA Medical Center (518/MSG), 200 Springs Road, Bedford, MA 01730. *tel/fax:*(781) 687-2488/2376, *e-mail:* burgess@world.std.com

## 1 Introduction

Health care profiling, still a relatively new field, is increasingly reflecting the need for standardization of health system data and the collection of institutional information repeatedly over time to allow time series analysis of providers. Profiling providers over time is needed to account for a number of major concerns reflected in the profiling literature. First, risk adjustment is necessarily incomplete, yet bias trends in unadjusted risk are likely to persist over time. As a result, profiling providers against themselves over time while drawing on the power of other providers for statistical analysis is likely to be fruitful. In particular, it facilitates the identification of deviations in outcomes from existing behavior trends. Second, small sample issues that exist within providers (a restricted population of patients) and between providers (a small number of providers gathering information from compatible information systems) are mitigated to some extent by appropriately combining information across panels of providers over time.

Two possible approaches to profiling for quality of care are based on comparing patient outcomes or comparing processes of care across providers. Patient homogeneity issues and the challenge of defining and measuring outcomes have led to a general trend toward profiling well-defined processes of care that are considered to be vital for clinical quality of care. In the mental health arena, outcomes can be particularly difficult to define across broad populations, so process-based measures such as those we employ here have been omnipresent. The substance abuse area we study here represents a subset of mental health discharges, the four DRGs relating to substance abuse. Among managed care organizations, a key set of process measures has evolved from the National Committee for Quality Assurance (NCQA) and their Health Plan Employer Data and Information Set (HEDIS). One of the Effectiveness of Care measures in HEDIS is for *Follow-Up After Hospitalization for Mental Illness*. The HEDIS measure is designed to track the percentage of health plan members who received outpatient follow-up care within 30 days of hospital discharge for selected mental health disorders. Since the early 1990's the US Department of Veterans Affairs (VA) has been tracking a similar series of psychiatric process measures as part of a larger set of profiling monitors (Burgess, Christiansen, Michalak, and Morris, 2000). In this situation, risk adjustment is used to account for diagnosis and other patient level differences. Initial VA efforts in this area focused on DRG based risk adjustment of the psychiatric discharges and the substance abuse discharges. It was recognized, however, that more detailed analysis based and other patient level differences also might be important. The profiling models presented here also employ the same HEDIS standard – 30 day return rate for outpatient follow-up care – specifically applied to substance abuse discharges.

The VA operates over 170 hospitals as part of one of the largest US health care delivery systems. The mission of this system has been to provide health care to US military veterans with a priority on poor veterans (Category AN) and those with service connected injuries (Category AS). Intense interest has arisen in performance monitoring and profiling VA facilities because of needs to aid and inform decision making in allocation of the \$20 billion budget of the VA (Lehner, *et al.*, 1996), to improve accountability of providers insulated from competitive market forces, and to integrate health care service networks within and outside the VA system. Substance abuse is a key focus area in the VA system; 136 of the VA hospitals had active inpatient substance abuse units in every fiscal year between 1988 and 1997. In FY 1997 alone, a total of 35,368 individuals were treated at those 136 hospitals;

the figure in FY 1988 was over 50,000. The VA information system provides information on several covariates for each individual over this period of ten years. Hence the data base provides opportunity to study profiling issues at the individual level and over a meaningful time period for policy purposes.

The study of VA quality monitor time series for psychiatric discharges began with a series of analyses in West and Aguilar (1997), Aguilar and West (1998) and West, Aguilar and Lourdes (1998) that focused on annual quality measures over the years 1987-1995. These studies developed statistical analyses of patient “return for follow-up” data at a highly aggregate level: data studied were hospital specific annual quality measures. This paper explores similar themes at the patient level – still profiling hospital level providers, but accounting for several other socio-demographic and medical history covariates in addition to risk adjustment at the level of the DRG of the inpatient stay. As with the earlier studies, the key motivating concern is to evaluate differences in outpatient returns for care within 30 days, now in the context of a range of possible individual level explanatory variables.

### 1.1 Primary Aims

The primary point of this paper is to demonstrate the technical feasibility of time series hierarchical modeling for profiling, and to illustrate the usefulness to administrators and clinicians of the numerical and graphical provider profiling summaries which can be derived from such models. In analyses not reported in detail in this paper, we studied the individual level data on an annual basis, exploring variations in return times classified by calendar year of the inpatient discharge. In particular, the focus of this exploratory analysis was an assessment of the usefulness of the 30 day cut-off for “adequate” follow-up care that has been commonly used, not only in HEDIS and by VA but by other health care systems. Despite the general reductions in return time over the sample period, the conclusion of that exploratory analysis was that using the 30 day return period for all substance abuse diagnoses was appropriate. Over this time period in VA, as in other health care systems, the practice of substance abuse and psychiatric care was changing substantially. Much care was now being provided on an outpatient basis which formerly was provided on an inpatient basis, motivated by changing incentive systems. In addition, inpatient stays for substance abuse and psychiatric care were getting shorter, and more of an emphasis was being placed on avoidance of readmissions. Over the 1988-97 period, the VA also was incrementally implementing a capitated budgeting system which now provides high-powered incentives for reductions in per person utilization. A primary concern in this situation is the impact on delivered quality of care, though this measure obviously represents only one dimension of quality. Traditionally, follow-up care has been linked to notions of continuity of care which may be more likely in turn to yield favorable patient outcomes. In the analysis in this paper we specifically focus on follow-up care in the four DRGs that comprise the substance abuse area of care.

For the 10 year analysis of fiscal years 1988-1997, we investigate the following concerns. The effort to incorporate full sets of individual level covariates and model the variation hierarchically across patients and hospitals imposes extensive analysis burdens, so a fundamental question is whether the individual level covariates matter. The hierarchical time series model that relates individual and hospital level effects is a novel and relevant technical development. Existing literature attempting time series profiling (McClellan and Staiger,

1999) resorts to strong simplifying assumptions and fixed effect patient level regressions which are not integrated into the hierarchical model directly. The technical aim here is to account for individual level variation in an integrated way, to see if it resolves problems of extra-binomial variation at the hospital level with which previous analyses had to contend. Next, the central aim of profiling providers is to use the model results in ways that are useful to managers and policy analysts. We select five hospitals, more or less at random, to study at length to assess whether changes in performance over time at the hospital level can be identified at statistically significant levels. Finally, at a national policy level for VA hospitals, there are two primary questions to assess. First, are there differences in hospital level variation in performance across the 136 hospitals with inpatient substance abuse programs over this ten year period? Second, as inpatient care patterns have shifted to outpatient settings, have VA policy efforts to improve case mix adjusted returns for follow-up been successful at the aggregate level?

Section 2 presents some descriptions of the data using summaries of fiscal year 1997 information, discusses some findings from our preliminary exploratory analyses using discrete duration models to assess different potential cut-offs, and reviews the process used to explore and isolate relevant subsets of covariates for further study.

Following this, Section 3 describes the primary focus on the full ten year panel, developing more general models that relate hospital-specific effects across fiscal years. This longitudinal framework is relevant in connection with our interests in the patterns of change from year-to-year in quality levels at the institutional level, and allows for hospital-level effects to be systematically related year-to-year through time series models. This extends our previous work with longitudinal hierarchical models for more highly aggregate data (West and Aguilar 1997, Aguilar and West 1998, West, Aguilar and Lourdes 1998). We explore the analysis of the full 10 years of data available, 1988-1997 inclusive, and report a range of summary conclusions in Section 4. Section 5 briefly indicates some potential further investigations.

## 2 Data Overview and Summary of Preliminary Analysis

Of the 170+ hospitals in the US VA health care delivery systems, 136 of the 140 hospitals with any inpatient substance abuse activity had active inpatient substance abuse units in each of the fiscal years between 1988 and 1997, inclusive. From the original data base and preliminary/exploratory data analyses, we constructed a modified data base with return time data and categorical covariates at the individual level for all individuals admitted in this ten year period for this slightly circumscribed group of 136 hospitals. A summary of this, with key covariates, appears in Table 1 for the last year of the sample, Fiscal Year 1997 (FY97).

The primary record for each “case” includes diagnostic, process of care, VA eligibility, and demographic characteristics on each individual. For FY97, there are a total of 35,368 cases in the data base. There is a case identifier with the recorded return time in days  $t$ , measured from the day of discharge from initial visit to the day of return to follow-up care. The Hospital/station number is an integer label running from 1 to 140 identifying the VA facility. There are a total of 140 facilities with patients recorded in the substance abuse inpatient care area though we only use the 136 with complete annual data in the analysis. The age factor classifies cases as in Age group 1 ( $\text{age} \leq 44$ years), 2 ( $45 \leq \text{age} \leq 64$ years),

and 3 (age  $\geq$  65years). The DRG factor classifies cases in the four substance abuse DRGs: (1) 434: Alcohol/Drug abuse with complications or comorbidities; (2) 435: Alcohol/Drug abuse without complications or comorbidities; (3) 436: Alcohol/Drug dependency with rehabilitation therapy and (4) 437: Alcohol/Drug dependency with rehab. and detoxification. Marital status classifies cases by 1 (married), 2 (SDW, separated - widowed - divorced), and 3 (UN, unknown - never married). Priority code status, based on a means test indicator that defines eligibility priority codes for use of VA services, classifies cases as in Priority group 1 (AN), 2 (AS) and 3 (other – lower priority). Gender classifies cases as 1 (male) and 2 (female). Race or national origin classifies cases as 1 (White), 2 (Black, not Hispanic), 3 (Hispanic) and 4 (other, including Asian, American Indian and unknown). Diagnosis, classifying cases into one of 11 groups, labelled 1 – 11, and associated with the principal medical diagnosis code of the case from the ICD-9-CM: (1) Chronic alcohol dependency; (2) Other drug dependency; (3) Acute alcohol dependency; (4) Alcoholic psychoses; (5) Opiate dependency and combinations; (6) Drug psychoses; (7) Alcohol abuse; (8) Drug abuse; (9) All non-mental health diagnoses; (10) Other disorders and (11) Non-substance abuse psychoses. A few additional categorical covariates relating to socio-economic and military service history of individuals, and region of country in which the facility is located also were studied.

Table 1 provides summary frequencies of individuals classified according to a moving cut-off on the return time scale  $t$ . Each row of the table is an estimate of a discretised version of the marginal distribution function of return times in the specific covariate group of that row. In addition, the total numbers of individuals in each sub-category are indicated by the entries  $n: \cdot$  in each row. The table includes such frequencies for 4 hospitals selected from the full 140 for initial profiling focus (hospital numbering serves only to provide labels for hospitals).

Table 1 illustrates some patterns of variability in the (discretised) marginal return time distributions categorised by each level of potential covariates. For example, in the full data base some 14.67% of the total of 35,368 individuals had a return time of exactly 1 day, 38.03% returned within 7 days, 46.78% returned within 14 days, and so forth. The return times are truncated at 367 (one year, chosen in general to allow for leap years of 366 days, though FY97 is not a leap year), with 28.71% of all individuals returning at a time greater than 367 days or not at all; these are regarded as uninformatively censored cases and essentially as non-returners. The table provides separate breakdowns of these cut-off specific percentages for each of the above categorical covariates, and for 4 selected hospitals. For example, 15.90% of all the 16,255 individuals in Age group 1 (less than 45 years of age) have return times of exactly 1 day, compared to only 7.63% in Age group 3 (greater than 65 years of age). We conducted formal binomial regression models across the specified thresholds according to whether or not an individual's return time is greater than or less than the threshold, accounting for the covariates.

A raw outcome of “quality” is the reported time from the inpatient substance abuse discharge to a return for follow-up outpatient substance abuse care. Since 30 days could be seen as an arbitrary cut-off, the first level of analysis was to explore the impact of different potential cut-offs on the FY97 data, with the hospital factor as the primary covariate of interest. However, in order to isolate hospital specific effects, we need models that adjust for the covariates and so we explored differences in return times that are related to some of the other covariates. We studied a set of independent analyses of models with cut-offs

Return time thresholds:	$t = 1$	$t \leq 7$	$t \leq 14$	$t \leq 21$	$t \leq 30$	$t \leq 367$
<b>All data</b> ( $n: 35,368$ )	14.7	38.0	46.8	51.6	56.0	71.3
<b>Hospital:</b>						
139 ( $n: 314$ )	46.8	62.1	66.2	69.4	72.3	80.3
82 ( $n: 400$ )	16.3	49.5	60.0	64.8	69.3	82.8
21 ( $n: 373$ )	3.0	11.8	18.5	23.3	26.3	39.4
118 ( $n: 271$ )	30.6	59.4	66.8	69.0	70.1	84.5
<b>Age group:</b>						
$\leq 44$ ( $n: 16,255$ )	15.9	40.1	48.5	52.9	57.3	71.9
45-64 ( $n: 16,963$ )	14.4	37.9	47.0	52.2	56.8	72.5
65+ ( $n: 2,150$ )	7.6	23.3	32.0	36.3	40.5	57.1
<b>DRG:</b>						
434 ( $n: 7,404$ )	13.6	35.0	43.7	48.9	54.2	71.2
435 ( $n: 15,407$ )	16.5	38.0	45.9	50.4	54.8	70.5
436 ( $n: 9,422$ )	13.4	40.4	50.3	55.2	59.4	72.5
437 ( $n: 3,135$ )	12.3	38.3	48.1	52.6	56.6	71.9
<b>Marital Status:</b>						
M ( $n: 7,463$ )	12.4	35.2	44.7	49.8	54.5	69.2
SDW ( $n: 20,527$ )	15.3	39.0	47.5	52.2	56.5	72.0
UN ( $n: 7,378$ )	15.2	38.2	46.9	51.6	56.2	71.5
<b>Priority:</b>						
AN ( $n: 23,617$ )	15.5	38.5	46.9	51.4	55.6	69.9
AS ( $n: 10,078$ )	13.6	38.5	48.7	54.4	59.9	78.3
Other ( $n: 1,673$ )	9.9	28.4	34.2	37.1	39.2	48.2
<b>Diagnosis:</b>						
1. ( $n: 17,787$ )	14.5	38.4	47.3	52.0	56.3	70.8
2. ( $n: 5,829$ )	17.0	42.0	51.0	55.8	60.1	74.4
3. ( $n: 2,937$ )	16.3	37.6	45.6	51.0	55.7	69.7
4. ( $n: 2,425$ )	12.4	33.0	40.6	45.2	48.9	64.3
5. ( $n: 2,309$ )	15.9	38.4	46.3	50.1	54.7	72.4
6. ( $n: 1,505$ )	10.7	33.9	44.5	50.4	54.8	75.0
7. ( $n: 1,112$ )	11.6	31.9	40.7	46.0	52.5	70.5
8. ( $n: 1,006$ )	14.6	39.2	47.2	51.9	57.4	77.3
9. ( $n: 194$ )	8.8	25.3	32.0	36.6	41.2	61.3
10. ( $n: 196$ )	9.7	33.2	45.4	52.6	60.7	80.1
11. ( $n: 68$ )	2.9	26.5	35.3	44.1	52.9	73.5
<b>Gender:</b>						
Male ( $n: 34,566$ )	14.7	38.0	46.7	51.5	56.0	71.3
Female ( $n: 802$ )	13.2	39.3	48.0	53.2	58.9	71.3
<b>Race:</b>						
White ( $n: 20,432$ )	14.2	37.3	46.1	51.1	55.6	70.6
Black ( $n: 12,095$ )	16.1	40.0	48.4	52.9	57.1	72.9
Hispanic ( $n: 1,662$ )	11.3	34.4	44.7	50.4	54.5	69.8
Other ( $n: 1,179$ )	13.3	36.1	44.5	49.1	53.9	68.7

Table 1: Cumulative empirical frequency distributions of return times at specific thresholds, categorised by levels of several primary covariates

$t = 1, t = 7, t = 14, t = 21, t = 30$  and  $t = 367$ . Our analyses are independent though, of course, the return probability must, for any individual and any set of specified covariates, be an increasing function of time  $t$ .

## 2.1 Discrete Duration Models

### 2.1.1 Model Form and Specifications

Consider an individual patient with return time  $T$  measured from the origin  $t = 0$  of the hospital discharge. For any set of chosen covariates, represent by  $x$  the column vector of values of the covariates for this individual. Since all of the candidate covariates are categorical,  $x$  is a vector of binary dummy variables. We consider regression models that describe a discretised representation of the underlying continuous return time distributions. Choose a specific time point  $t$ , and consider the event that  $T \leq t$ . For this individual, we adopt a general logistic regression model for the return probability

$$p(t) = Pr(T \leq t) \tag{1}$$

in which

$$\text{logit}(p(t)) = \beta_0(t) + x'\theta(t) \tag{2}$$

where  $\text{logit}(p) = \log(p/(1-p))$ ,  $\beta_0(t)$  represents a baseline return probability at time  $t$ , and  $\theta(t)$  is a regression parameter column vector relative to the specified covariate structure in  $x$ . As we consider all possible covariates, we consider a range of possible elements of  $\theta(t)$ .

### 2.1.2 Choosing the Covariates and Return Cutoff for the Analysis

Initial work involved exploratory Bayesian modelling as a screening analysis to identify potentially relevant covariates from the set of ten key variables discussed above under the logistic regression in (2). This used approximate Bayesian model selection methods, based on the use of approximate Bayes' factor and corresponding model probabilities (e.g., Raftery, 1995). The study here used a slight modification of code by Volinsky (1996).

A standard asymptotic approximation to these posterior probabilities provides for simplified computations that are useful in preliminary screening for covariates. This approximation, based on the Laplace method for integrals that is a standard tool in Bayesian asymptotics, delivers standard *Bayesian Information Criterion (BIC)*-based posterior probabilities. The approximate posterior model probabilities were computed for all possible combinations of covariates under model (2) for FY97. This screening analysis assumed that the Hospital effects are fixed effects, rather than random, to simplify the computations. In the analyses based on all the data, relevant selected models include the covariates Hospital, DRG, Age, Priority and Diagnosis, across all possible return time cut-offs. The most important case for Gender is at return time cut-off  $t \leq 30$ , when there is a small but non-negligible posterior probability that this covariate be included. Marital Status is included in some lower probability models in the analyses using all data, but only at return time cut-offs  $t = 1, 7$ .

Based on this exploratory screening analysis, we conclude that covariates Hospital, DRG, Age, Priority and Diagnosis should be included in models for more formal study.

The Gender, Marital Status and Race factors appear to be only of marginal relevance; however, we choose to include the Gender factor in the full panel analysis as well. It is of real note that, across various models with different subsets of covariates, point estimates of Hospital effects are very stable indeed, varying negligibly with different models. This is most reassuring, as it indicates that inferences about these primary effects will be robust to the issue of whether or not to include marginally interesting covariates, and also that unmodelled interaction effects are likely small and may be safely ignored.

Our preliminary analysis strategy involved fitting the full logistic model of equation (2) to the FY97 data and repeating the analysis in separate studies based on different cut-offs. Including the covariates displayed in Table 1 and the random effects, equation (2) is:

$$\text{logit}(p(t)) = \beta_0(t) + \epsilon_i(t) + \begin{cases} \delta_d(t), & \text{for Age group } d = 2, 3, \\ \gamma_g(t), & \text{for DRG level } g = 2, 3, 4, \\ \kappa_k(t), & \text{for Marital status group } k = 2, 3, \\ \eta_e(t), & \text{for Priority level } e = 2, 3, \\ \xi_x(t), & \text{for Diagnosis group } x = 2, \dots, 11, \\ \chi_c(t), & \text{for Gender group } c = 2 \text{ (women)}, \\ \zeta_z(t), & \text{for Racial group } z = 2, 3, 4. \end{cases} \quad (3)$$

Here all parameters are cut-off specific,  $\delta_d, \gamma_g, \kappa_k, \eta_e, \xi_x$  and  $\zeta_z$  are fixed effects associated with each level  $d, g, k, e, x, c, z$  of the covariates, and  $\epsilon_j$  is the random effect associated to hospital  $j$ . These hospital specific random effects are assumed drawn from a normal population model  $\epsilon_i(t) \sim N(0, \omega(t)^2)$  where the cut-off dependent standard deviation  $\omega(t)$  represents the dispersion in effects across the VA system.

Very vague but proper priors are adopted for all fixed effects parameters, the effects of all other covariates. Specifically, the elements of  $\theta(t)$  have independent, zero-mean normal priors with variances of 1000, and the hospital precision parameter  $1/w(t)^2$  has a gamma prior with shape and scale parameters both equal to 0.001. Posterior analysis uses Markov chain Monte Carlo (MCMC) methods to iteratively simulate from the full joint posterior distribution of  $\theta(t)$  and  $w(t)$  producing large Monte Carlo samples for summary inferences on all parameters and effects. Separate analyses were performed for each chosen cut-off  $t$ .

### 2.1.3 Analysis Results

Together, hospital level comparisons and the declining variances as outpatient care return time is increased support two commonly held reasons for using the 30 day cut-off in profiling these substance abuse processes. First, while the 30 day cut-off appears to be arbitrary, it is less noisy than other potential cut-off times since it appears to limit the effects of unobserved heterogeneity at the individual level. Also, 30 day cut-off comparisons between facilities generally give similar answers as those from using other cut-off values. Thus, we found that the 30 day cut-off appears to be a reasonable compromise across the different substance abuse DRGs and proceeded to the final profiling analysis using the traditional focus on that 30 day return cut-off.

A closer look at the covariates also sharpens our use of these factors in the full ten year panel. The older individuals have uniformly lower probabilities of return at all stages, and the effect seems to be roughly constant with respect to cut-off. The DRG is capturing high

level differences in treatment patterns and relative needs for quick returns across different principal DRGs. We hypothesized that marital status for substance abuse patients would affect the efficacy of the support network which might make follow up appointments and guarantee that they are kept. Yet, there are no major differences between the groups  $j = 2, 3$ , except perhaps at the early return times  $t = 1, 7$ . The overall effects are small in terms of their impact on return probabilities and so this covariate is dropped from the full panel analysis.

VA priority levels are hypothesized to have two possible effects. First, higher priority veterans may be given better access to the system and priority in getting favorable appointments, although once veterans are accepted for care they are supposed to be treated identically. The priority level also affects the likelihood of leaving the system to receive private sector care, specifically the “other” category having a higher probability of leaving and a lower probability of returning. However, once censored cases with return times greater than one year are excluded, this effect should be significantly less. In analysis of the data excluding the censored cases, Priority group AS has consistently lower return probabilities than the rest, and the “other” group is consistently higher except at  $t = 1$ . The estimated effects generally appear constant, consistent with a proportional odds structure, with that one exception. In analysis of the data including the censored cases, conclusions are quite different. Priority group AS evidences increasing return probabilities that are higher than average apart from at  $t = 1$ . The “other” group has dramatically lower return probabilities, and they appear to decrease at higher return times. This substantially supports our inferences about censoring in the data. Censored individuals may well be receiving their follow up substance abuse care in the private sector, unobserved by our data and model. We still choose to employ all of the data in the panel analysis since we cannot distinguish between true failures to follow up from those who are followed up outside the VA system. In any case, the policy implication of discharging a substance abuse patient from an inpatient stay remains: the VA retains a responsibility for follow up care.

We know that within DRG case mix can be very important in any profiling analysis, and here we employ ICD-9-CM code groups to attempt to capture this. There is much higher uncertainty about the Diagnosis effects at the higher levels  $j = 9, 10$  and  $11$  than the rest, due to smaller sample sizes in these groups (see Table 1). There is some evidence of non-proportional odds behaviour in several of the Diagnosis categories, though this is not obviously highly significant nor uniform across categories. The categories with higher labels ( $j = 5, 6, \dots$ ), tend to generally negative effects and return probabilities lower than the earlier categories relative to the chronic alcohol dependency reference group. Given the small sample sizes in many diagnosis groups we use these results to merge similar diagnosis groups together in the full panel analysis as described below.

For gender, we hypothesize two potentially offsetting effects. In general, women can be more compliant with follow up care effects; however, the dominance of male patients in the VA system can encourage women to shy away from VA care or return to private sector care. There is a suspicion of a generally increased effect with longer return times, consistent with small increases in return probabilities for women relative to men at later times. The effects are rather uncertain, though. Only at  $t = 30$  in the analysis excluding censored data is the effect really significant; however, since it is this cut-off we use in the panel analysis this covariate is included.

In general (e.g., Burgess and DeFiore, 1994), non-whites are more likely to use VA

outpatient services; however, we might hypothesize that conditional on admission whites might follow up after discharge more effectively. Moreover, investigation of the race covariate brings up another important issue in profiling – the concept of “allowable” differences. We wish to avoid giving hospitals “credit” for racial differences in follow up care when those differences may be induced by local discriminatory practices. Nevertheless, for statistical reasons it is desirable to account for all measurable factors which explain variation in the model. As a result, we proceeded with our investigation. In the analysis of all the data, including censored times, the suggestion is that of very small positive values – consistent with very slightly higher return probabilities than Whites – though with a fair degree of uncertainty. The effects for Hispanics remain essentially the same as for Whites. The effects in the Other group are, however, relatively lower, consistent with larger numbers of non-returners in this group, and the effects here do appear generally significant. Given the high degree of variation and the desire not to credit hospitals for racial differences in any case, this covariate also was dropped from the panel analysis.

Finally, we explored approximate 95% intervals for the Hospital specific random effects in analyses for all of the potential return times. This analysis reinforced the conclusion that some or many of the hospital effects are indeed different at different return time cut-offs  $t$ , consistent with a general non-proportional odds structure. As we moved through analyses with higher cut-offs, the relative quality levels of hospitals is indeed quite variable with cut-off, illustrating the extreme effect with  $t = 30$ . Indications of uncertainty also are clear in these graphs, highlighting that hospitals with smaller sample sizes lead to wider intervals. Further details of this exploratory work are available in the PhD thesis of the second author (Lourdes, 2000).

### 3 Panel Models: Discussion and Model Form

#### 3.1 Scope and Data Structure

We now develop the hospital profiling analysis to encompass multiple years of return time data, taking into account dependencies between hospital-specific effects from year-to-year. Here we build on previous work with aggregate data (West and Aguilar 1997, Aguilar and West 1998, West, Aguilar and Lourdes 1998) and develop simple time series model components that relate the hospital-specific effects between years, while maintaining the same natural random effects/hierarchical model within years. We introduce slightly different time series structures; however, in order to adequately capture aspects of heterogeneity across the VA hospital system.

We study the full 10 years of return time data, 1988-1997 inclusive. Note that the prior works referenced above have similar objectives but analyse only very highly aggregated data, namely the total numbers of patients with return times at specific cut-offs; they do not consider covariates other than the hospital. The current study is therefore a refinement of these previous works, aiming to assess similar issues but now using data at the patient level.

The model adopted is again a logistic regression with outcomes classified by return times below or exceeding a specified cut-off. The data analysis reported below is that based on the return time cut-off  $t = 30$ , using all data. This most closely approximates the HEDIS measure and other similar substance abuse follow up models generally in use.

Before describing the model extensions, we note some additional data and covariate selection issues.

We make some minor revisions to the selection and specification of categorical covariates, based on the results of the FY97 analysis as described briefly above. Further details on the FY97 analysis are available from the authors. In summary, the model adopts the following changes to covariate specifications: For the hospital/station, a total of 136 facilities with patients recorded in the substance abuse area for all ten years are used. The Age factor is refined to classify cases into just two groups: Age groups 1 (age  $\leq$  64years) and 2 (age  $\geq$  65years). The DRG factor with 4 levels is unchanged. The Priority code status with 3 groups is unchanged. The Gender factor, with 1 (male) and 2 (female), is unchanged. The diagnosis is refined to classify cases into one of just 3 groups: (1) a *Dependence* group that combines the original groups 1,2,3 and 5; (2) a *Psychoses* group that combines the original groups 4,6,9,10 and 11 and (3) an *Abuse* group that combines the original groups 7 and 8.

Following very extensive data exploration and reorganisation, the data is reduced in the context of the covariates indicated above to a total of 463,015 individual substance abuse discharges across the 136 profiled hospitals in the VA system in the 10 years, with annual numbers as follows:

1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
52,837	51,451	48,514	46,163	47,054	45,175	47,325	46,438	43,062	34,996

With our chosen categorical covariate structure, we have a total of 19,584 cells in the cross-classification (Hospital  $\times$  Age  $\times$  DRG  $\times$  Priority  $\times$  Diagnosis  $\times$  Gender). Of these, there are a total of 7,848 cells that are non-empty for at least one year, so reducing the problem to a more manageable effective sample size of 7,848 conditionally binomial responses per year. The regression model now described applies to structure the binomial probabilities for these responses.

### 3.2 Panel Regression Model

We adapt the basic logistic regression model of equations (2) and (3) to multiple consecutive years labelled  $r = 1, \dots, 10$ , as follows. As noted above, we focus here on a specific, chosen return time cut-off  $t$ , and from here on drop the explicit indication of cut-off in the notation for all model parameters. We do this simply for notational clarity, and it should be borne in mind that all parameters are cut-off specific.

In each year  $r$ , an individual treated in hospital  $i$  has return time less than the specified cut-off  $t$  (in days) with logit probability

$$\text{logit}(p_r) = \beta_{0,r} + \epsilon_{i,r} + \begin{cases} \delta_{d,r}, & \text{for Age group } d = 2, \\ \gamma_{g,r}, & \text{for DRG level } g = 2, 3, 4, \\ \eta_{e,r}, & \text{for Priority level } e = 2, 3, \\ \xi_{x,r}, & \text{for Diagnosis group } x = 2, 3, \\ \chi_{c,r}, & \text{for Gender group } c = 2 \text{ (women)}, \end{cases} \quad (4)$$

where, as usual, covariate parameters in the first group/level of the corresponding factor are constrained to be zero, i.e.,  $\delta_{1,r} = \gamma_{1,r} = \eta_{1,r} = \xi_{1,r} = \chi_{1,r} = 0$ .

Note that all parameters are indexed by  $r$ , the year in question. Hence the model so far is just 10 copies of the earlier model, one for each year, with the slightly modified covariate

specifications. We note and stress that our models do not relate the population parameters  $\beta_{0,r}$  over the years, so these baseline quality levels are unconstrained in how they vary between years. This neutral standpoint is adopted in order to “let the data speak” about any patterns of variation, or lack thereof, in the baselines. We adopt the same attitude to the effects parameters for all covariates with the exception of the Hospital effects, similarly viewing the covariate effects as nuisance parameters to be estimated but not anticipating systematic structure in the effects over time.

### 3.3 Model Structure for Random Effects

The model extensions developed relate to the hospital-specific effects  $\epsilon_{i,r}$  across years, in a way similar to prior work with aggregated data (West and Aguilar 1997, Aguilar and West 1998, West, Aguilar and Lourdes 1998). The specific structure adopted involves simple time series panel models to incorporate the view that the  $\epsilon_{i,r}$  are expected to remain relatively stable within each hospital from year to year, while allowing for unexplained sources of variability at the hospital level that may induce random changes. Each  $\epsilon_{i,r}$  term is modelled as a simple AR(1) time series over the years, and we find that this structure captures much of the random-effects variability across hospitals within each year as well as the systematic dependencies within hospitals from year-to-year. Inference on the correlation between years leads, as we shall see, to generally high degrees of correlation between years, as might be expected.

The dependence model allows for both between-year correlations and within year variability to be hospital-specific. For hospital  $i$ , we take the standard AR(1) model

$$\epsilon_{i,r} = \phi_i \epsilon_{i,r-1} + \omega_{i,r} \quad (5)$$

for years  $r = 2, \dots, 10$ , where  $\phi_i$  is the correlation parameter (in  $(-1, 1)$ ) and the  $\omega_{i,r}$  are independent *innovations* distributed as

$$\omega_{i,r} \sim N(\omega_{i,r} | 0, u_i^2) \quad (6)$$

for some innovations variance  $u_i^2$ . The AR(1) model is such that, for all  $r$  and including the first year  $r = 1$ , the implied *marginal* distribution of hospital effects within the year is simply

$$\epsilon_{i,r} \sim N(\epsilon_{i,r} | 0, w_i^2) \quad (7)$$

with marginal variance  $w_i^2 = u_i^2 / (1 - \phi_i^2)$ .

This dependence structure parallels that in prior work, referenced above, though with the extensions noted above: the parameters  $(\phi_i, u_i)$  are hospital specific, so allowing for variations in both systematic dependency and overall levels of variation in the  $\epsilon_{i,r}$  across hospitals. Our prior work had assumed common parameters across hospitals. We comment further on this assumption in summarising the data analyses. First, however, we describe the Bayesian hierarchical model for the  $(\phi_i, u_i)$  across hospitals  $i$  that completes the full model structure.

We assign an hierarchical model structure to the hospital-specific parameters  $(\phi_i, u_i)$ . Assuming these to be exchangeable parameters drawn from a hospital-population prior delivers a class of Bayesian hierarchical models: modelling the  $(\phi_i, u_i)$  as a random sample

from a population prior implies that the resulting random effects  $\epsilon_{i,r}$  follow a common marginal distribution within each year. It is important to maintain this view of hospitals as being exchangeable once we have corrected for covariates while providing flexibility to assess differing degrees of dependency in random effects (through the randomly differing  $\phi_i$ ) and levels of contribution of the systematic component of variation (through the randomly differing  $u_i$ ) across hospitals. The specific prior models we adopt have the following form. First, across hospitals  $i$ ,  $\phi_i$  and  $u_i$  are independent. Second, the population of dependence parameters appears as a random sample from a beta distribution,

$$\phi_i \sim Be(\phi_i | a\mu, a(1 - \mu)) \quad (8)$$

where the underlying average dependence level is represented by the hyperparameter  $\mu$ , and variations among the  $\phi_i$  are determined by the precision hyperparameter  $a$ . These two hyperparameters are to be estimated, along with the  $\phi_i$  themselves. Third, the population of dispersion parameters  $u_i^2$  is modelled as a random sample from an inverse gamma distribution, with

$$u_i^{-2} \sim Ga(u_i^{-2} | c, c\rho) \quad (9)$$

where the hyperparameter  $\rho$  represents an underlying average dispersion level, and variations among the  $u_i$  are determined by the precision hyperparameter  $c$ . These two hyperparameters are to be estimated, along with the  $u_i$  themselves.

This completes the model description. Before proceeding to data analyses, we comment on a further model extension that has been explored and then discarded. In our prior work (West and Aguilar 1997, Aguilar and West 1998, West, Aguilar and Lourdes 1998) we studied overall numbers of patients returning/non-returning within each hospital/year, in the way that HEDIS and other standard profiling measures do, with no accounting at all for individual-level covariates. In modelling hospital random effects in logistic regressions for that aggregated data, we found that the AR(1) structure did not fully account for the levels of overall (extra-binomial) variation apparent in the data. We appropriately catered for this by elaborating the model to include additional, residual or “idiosyncratic” random effects within each hospital and year. In the notation here, that would involve replacing the  $\epsilon_{i,r}$  terms with a sum of two terms, say  $\epsilon_{i,r} + \nu_{i,r}$ , where the  $\epsilon_{i,r}$  follow our AR(1) models and the new, residual terms are simply noise,  $\nu_{i,r} \sim N(\nu_{i,r} | 0, v^2)$  independently. We have indeed explored this model extension in the current study. The conclusions are that  $v$  and the  $\nu_{i,r}$  are quite small, and essentially negligible compared to the levels of variation in the  $\epsilon_{i,r}$ . As a result, we cut-back from this more general model to that detailed above. An extremely important general point to note is that this rejection of the extended model found necessary for the aggregated data can be taken as an indication that we have indeed adequately explained most of the additional variability observed in the aggregate level study through the use of the several categorical covariates at this individual level. Or, more generally, it suggests that this level of risk adjustment is highly desirable in HEDIS or other profiling models in order to put observed units on a level playing field.

## 4 Panel Models: Profiling Analysis for FY88-97

### 4.1 Implementation

Posterior analysis uses Markov chain Monte Carlo (MCMC) methods to simulate iteratively from the full joint posterior distribution of all model effects and hyperparameters. This extends the single-year analysis to the full 10 year panel, so that we obtain posterior inferences about the baseline parameters  $\beta_{0,r}$  and the parameters representing the effects of all covariates specific to each year  $r = 1, \dots, 10$ . In addition, the analysis produces posterior samples for the hospital-specific random effects  $\epsilon_{i,r}$ . The posterior analysis also now includes the set of new parameters  $\{\phi_i, u_i : i = 1, \dots, 136\}$  and the corresponding hyperparameters  $(a, \mu; c, \rho)$ . All graphical summaries discussed below are based on a large Monte Carlo sample from the full posterior for all these quantities. We report the analysis of the full data set including the censored cases (return times exceeding one year), and, as mentioned earlier, at the return time cut-off of  $t = 30$  days.

We now briefly review analysis results as summarised in the accompanying graphs. We note that the results for parameters in FY97 alone are completely consistent and in agreement with the analysis of that year alone, as is to be expected. The model also is found to be feasible, with the patient characteristics accounting for important levels of unexplained hospital level variation.

### 4.2 Baseline Duration Model Parameters

Figure 1 displays posterior intervals and estimates for the baseline return time probabilities  $p_{0,r} = 1/(1 + \exp(-\beta_{0,r}))$  in each of the ten years  $r = 1, \dots, 10$ . The general increasing trend with years is apparent, and the lack of overlap of intervals in the last couple of years indicate important changes in those particular years. In particular, the change from FY96 to FY97 is the most distinct, though it does not represent the largest absolute improvement in expected rates of follow up. This corresponds with the time period where management efforts in the VA system were put in place aimed at reducing psychiatric and substance abuse inpatient days and discharges, along with an emphasis on continuity of care.

Thus, these results account for these major shifts in the setting of care and still permit comparisons over time on a level playing field. Figure 1 also indicates an increase in the probability of return within 30 days at an “average” hospital from around 35% in 1990 to nearly 55% in 1997. From the raw data sets, the crude aggregate proportions of returners in each year in terms of percent returning within 30 days are:

Year:	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
Percent:	42.0	40.5	41.9	46.1	48.3	48.2	49.5	51.6	52.2	55.9

These observed values are also displayed in the graph as points next to the corresponding interval estimates. Their pattern over the years mainly agree with that of the inferred baseline parameters. That they are higher than the baseline simply reflects the fact that the majority of the observations lie in covariate groups with generally positive effects. For example, results discussed later reveal that for FY92, the only categories with negative effects are 65+ in Age group and Others in Priority group with 6% and 3% of the FY92 data lying in these levels respectively. This pattern of change over the years reflects hospital system-wide effects on 30-day return time probabilities. The effects of system-wide policy

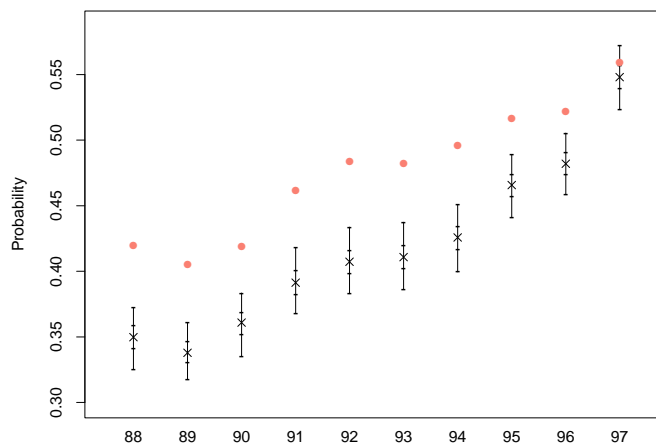


Figure 1: 95% posterior intervals for baseline return time probabilities  $p_{0,r}$  over years  $r = 1, \dots, 10$ . The points correspond to the overall proportions returning within 30 days from the raw data.

changes and common management practices are clear in these improvements in 30-day returns. The primary point of the analysis, though, is to allow us to put hospitals with diverse case mix differences on a level playing field for comparison. In particular, we note that the norm across years improves; however, we can look at the performance of any individual hospital against that shifting norm.

### 4.3 Random Effects for Profiling Selected Hospitals

Figure 2 represents the central profiling result from the paper in the form shared with VA managers: posterior intervals for the random effects  $\epsilon_{i,r}$  for 5 selected hospitals. These hospital effects display diverse patterns over the years. Hospitals 139 and 82 have apparently positive effects across all 10 years, consistent with higher return probabilities than the norm. Hospital 21 is well below the norm consistently across the years, with no evidence of improvement in recent years. Hospital 118 has tended to vary mildly about the system-wide norm, but has seen a marked increase in return probability in FY97 implied by the large and positive effect in that year. Hospital 140 has experienced return probabilities much lower than the norm during the first nine years, but has seen a very marked increase also in FY97, with a magnitude of improvement that exceeds that of 118.

### 4.4 Analysis of Covariates

Figures 3 to 6 present posterior intervals of the fixed effects. Figure 3 displays posterior intervals of the fixed effects  $\delta_{2,r}$  for the Age covariate at level  $d = 2$  (65+). The elderly group effect is clearly negative in each year, and the graph shows a mild though significant decreasing pattern over the ten year period, consistent with a deterioration in 30-day returns in the older Age group. While this effect is definitely a policy concern, it also may be related to private sector payment incentives in the Medicare program, as a number of

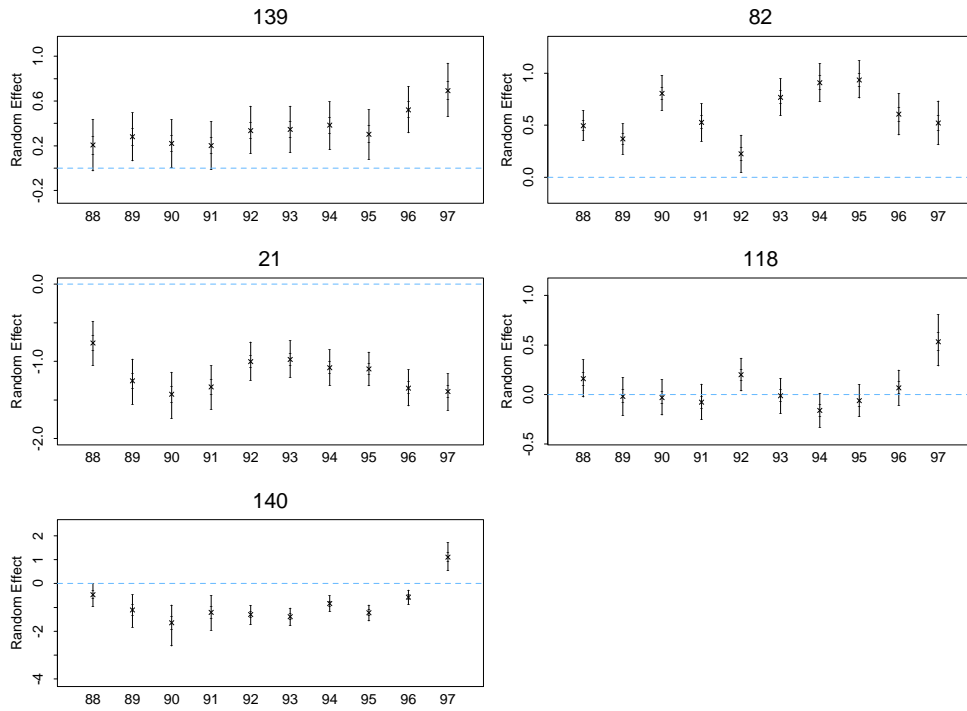


Figure 2: 95 % posterior intervals for hospital-specific random effects  $\epsilon_{i,r}$  over years  $r = 1, \dots, 10$  in multi-year analysis with cut-off  $t = 30$  days. Profiling intervals are given for five selected hospitals.

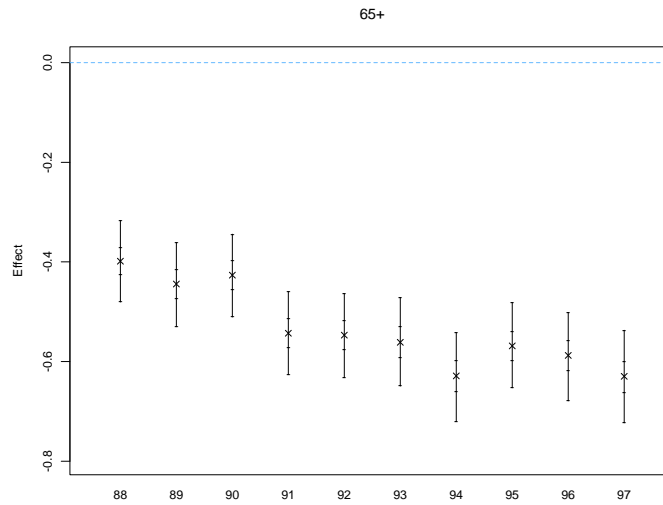


Figure 3: 95% posterior intervals for Age group effects  $\delta_{2,r}$  over years  $r = 1, \dots, 10$  relative to the age group under 65.

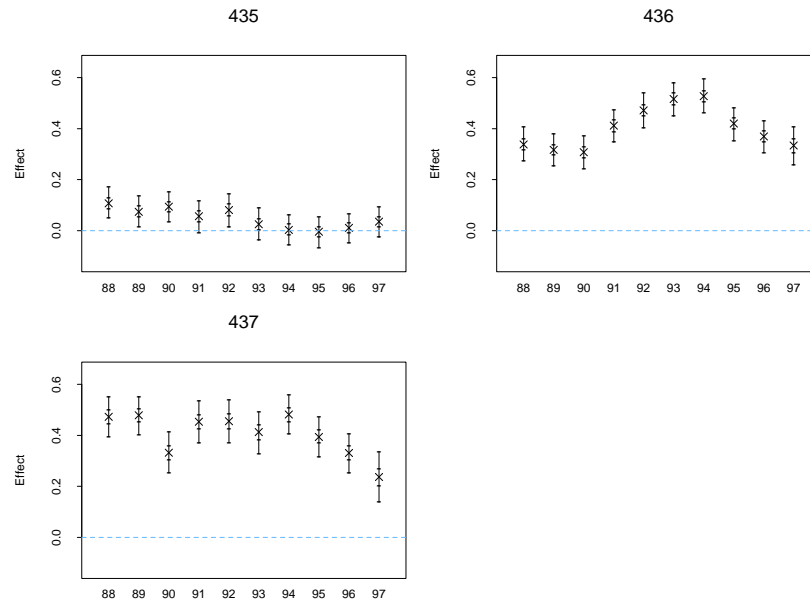


Figure 4: 95% posterior intervals for DRG group effects  $\gamma_{g,r}$  over years  $r = 1, \dots, 10$  relative to DRG 434.

selection effects may be occurring. Elderly veterans who also are eligible for Medicare may find it easier to have outpatient substance abuse follow up care covered in the private sector than to be admitted for inpatient care. However, with changes in the VA system beginning to mirror private sector psychiatric and substance abuse treatment patterns, this effect should be lessening, not increasing. Or they may have been admitted to get started on expensive psychiatric and substance abuse drugs that they continue to receive on an outpatient basis at VA prices (either free to the veteran or with a nominal co-pay depending on their eligibility class). Then they may seek follow up outpatient visits in the private sector, which would be reimbursable through Medicare. This explanation appears more likely as the increasing prices of many psychiatric and substance abuse pharmaceutical treatments can stretch elderly budgets, whether by direct payment or insured through Medigap or Medicare+Choice plans purchased by the elderly. Burgess and DeFiore (1994) show that elderly veterans are who have higher VA eligibility are less likely to purchase Medigap supplements.

Figure 4 provides intervals and estimates of the fixed effects  $\gamma_{g,r}$  for DRG. As can be seen from the graph, DRG 436 and 437 have positive effects and meaningful year-to-year variations in the DRG effects within each DRG category. There is also an apparently persistent deterioration in return probabilities in DRG groups 436 and 437 over FY94-FY97. This actually is more a result of relative improvements in DRG 434, the reference group for this covariate, as well as DRG 435.

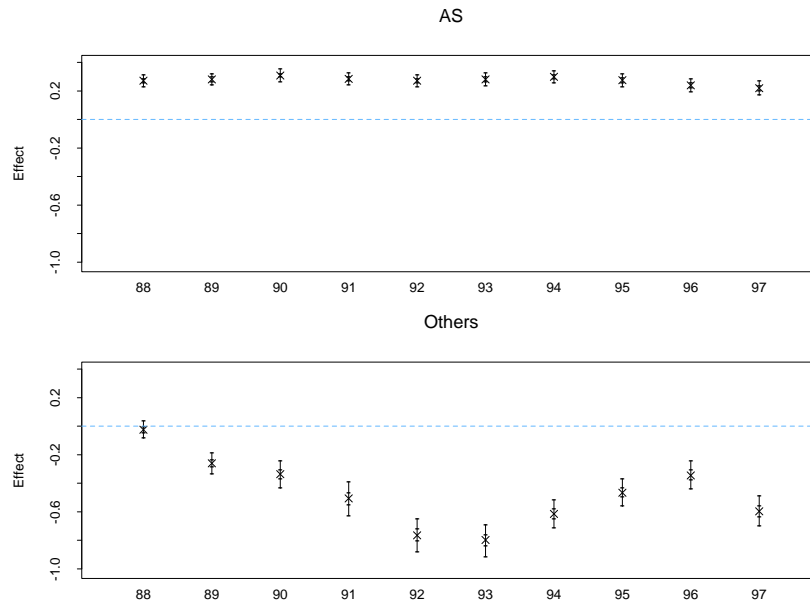


Figure 5: 95% posterior intervals for Priority group effects  $\eta_{e,r}$  over years  $r = 1, \dots, 10$  relative to Priority group AN.

Plots in Figure 5 show posterior intervals of the fixed effects  $\eta_{e,r}$  for the Priority variable. Priority eligibility for using VA health care services is obtained either from having service connected injuries (AS) or from having income and wealth below a means test threshold (AN). The plots suggest that the effects in Priority group AN are accurately estimated, very stable and positive over the years, indicating a significant and sustained relative level of 30-day return probability. These poorer veterans are most likely to have the fewest alternatives for follow up treatment. By contrast, the effect in Priority group “others” is quite variable and generally negative; there is an indication of improvement in later years following deterioration during FY89-92/3, but a sharp drop-off in FY97. The time trends here correspond to national budget policies to balance the budget that mandated a constant nominal budget at \$17 billion from 1996 through 2002 (which has been broken for FY00 with a budget increment of \$1.7 billion) and started to put downward pressure on utilization in the non-priority category. For profiling purposes, this documents the impact of this overall effect on outpatient follow up after these substance abuse discharges.

The ICD-9-CM diagnosis groups used in the model are very broad in order to obtain sample sizes large enough to recognize effects. Figure 6 displays posterior intervals of the fixed effects  $\xi_{x,r}$  for the Diagnosis variable. The effect in the Psychoses group is generally negative and consistent with lower return probabilities than in either the Dependence or Abuse groups. The psychosis patients are most likely to have psychological impairments that would make follow up care more difficult. As a policy matter, there is some evidence that efforts to improve returns for this group has been successful. However, these improvements have not implied an increment on the return probabilities.

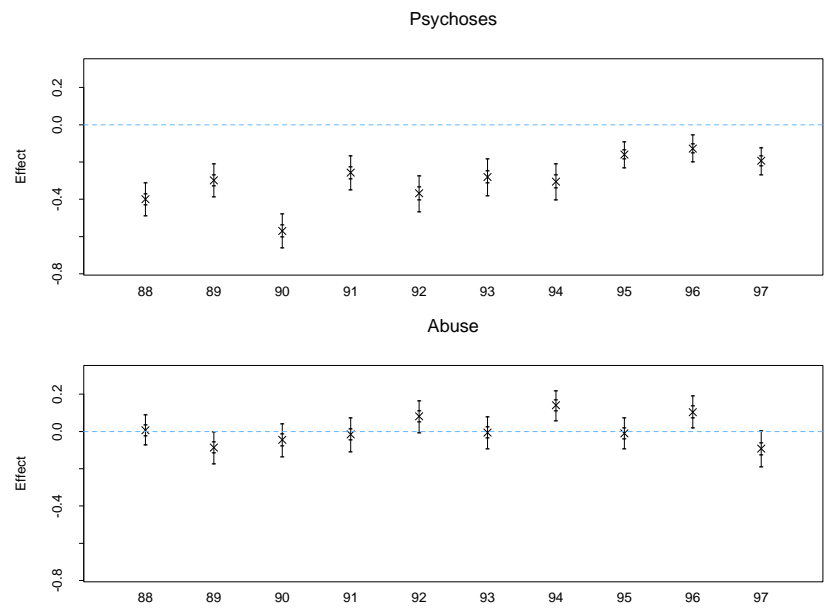


Figure 6: 95% posterior intervals for Diagnosis effects  $\xi_{x,r}$  over years  $r = 1, \dots, 10$ , psychoses and abuse groups relative to dependence category.

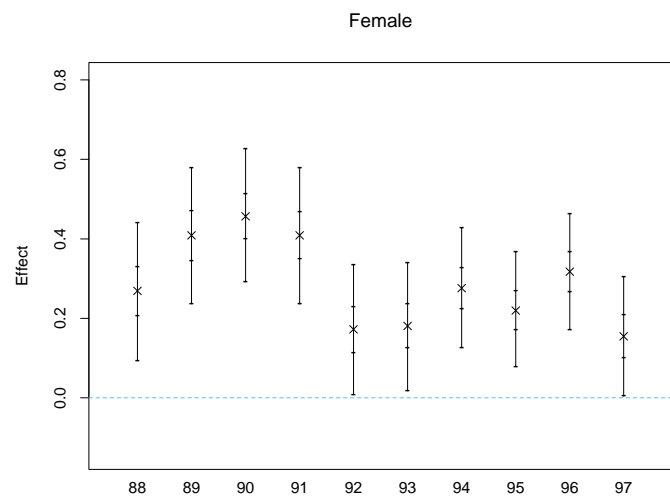


Figure 7: 95% posterior intervals for female group  $\chi_{2,r}$  over years  $r = 1, \dots, 10$  relative to male effect.

Figure 7 shows posterior intervals of the fixed effects  $\chi_{2,r}$  for the Female category relative to the Male effect. Effects are always positive over the full ten year span of the data. However, the levels are somewhat lower for FY92-93 and FY97.

#### 4.5 Posterior Inferences from the Random Effects Model

We now turn to posterior inferences on the parameters of the hierarchical time series model for the hospital effects, beginning with the dependence parameters. First we mention the hyperparameters  $(a, \mu)$  in the hospital population model for the  $\phi_i$ . Parameter  $\mu$  represents a system-wide average value for year-to-year correlations between hospital effects within hospitals. The approximate posterior median and end-points of a 95% interval for  $\mu$  are 0.82-0.85-0.88, indicating a high correlation structure generally across hospitals. Parameter  $a$  measures the dispersion of the actual, hospital-specific  $\phi_i$  values about  $\mu$ . A high value indicates the  $\phi_i$  are very tightly distributed around  $\mu$ , lower values indicate more variability. The approximate posterior median and end-points of a 95% interval for  $a$  are 6.67-10.62-16.68; this range of fairly high values indicates that there is actually rather little variability in the  $\phi_i$  parameters across hospitals. This is confirmed in Figure 8 where we display posterior estimates and intervals for the  $\phi_i$  for all 136 hospitals. We present the hospitals ordered according to increasing values of the posterior medians of the  $\phi_i$ .

The concordance illustrated here suggests that we could cut-back to a simpler model in which the  $\phi_i$  takes the common value  $\mu$ . We have reanalysed the data with such a model and find that the inferences on all model parameters and effects as summarised in the displayed graph here are basically unchanged, as should be expected. Exploring the more general hierarchical model for the  $\phi_i$  has confirmed the consistency across hospitals in the dependence structure between years. This supports the use of the model for profiling individual hospitals.

We discuss similar issues related to the variability parameters of the hierarchical time series model for the hospital effects. First we mention the hyperparameters  $(c, \rho)$  in the hospital population model for the hospital-specific innovation variances  $u_i^2$ . Parameter  $\rho$  represents a system-wide average value for innovation variance, though is not of primary interest; for the record, the approximate posterior median and end-points of a 95% interval for  $\rho$  are 0.04-0.05-0.06. Parameter  $c$  measures the dispersion of the actual, hospital-specific innovation variances  $u_i^2$  values about  $\rho$ . A high value of  $c$  would indicate that the  $u_i^2$  take similar values, whereas lower values of  $c$  indicate more variation in the  $u_i^2$  across hospitals. The approximate posterior median and end-points of a 95% interval for  $c$  are 2.06-3.07-4.73; this range of rather low values indicates that there is a fair degree of heterogeneity in the actual set of 136  $u_i^2$  quantities across hospitals. This is confirmed in the bottom panel of Figure 8. The plot shows some variation in both estimates and uncertainties across the hospitals. The five hospitals noted earlier are highlighted, hospital 140 clearly stands out as a case of relatively high variability, consistent with the earlier discussion about the significantly varying  $\epsilon_{j,r}$  parameters in this hospital.

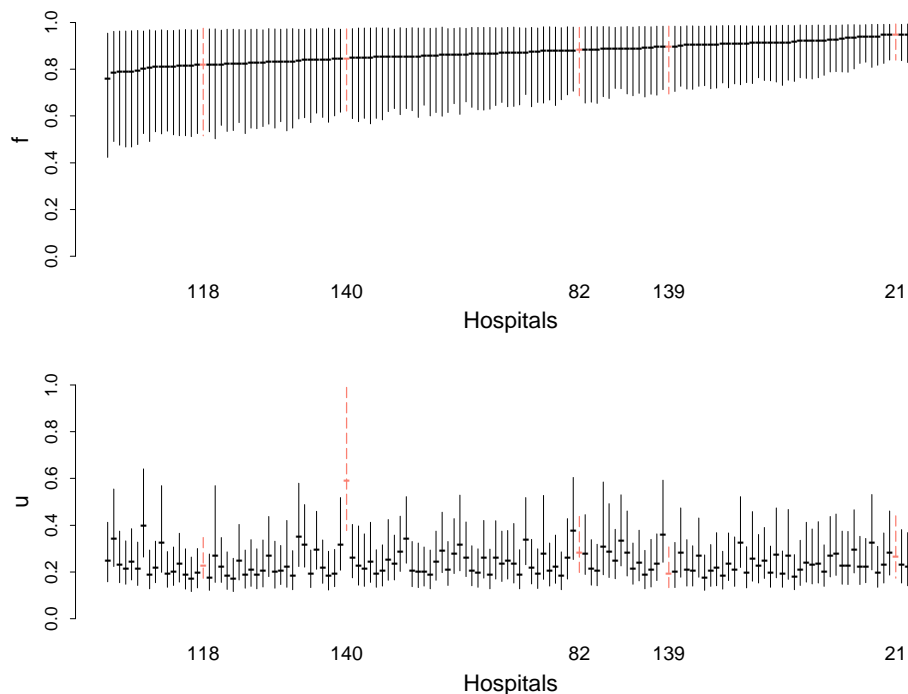


Figure 8: 95% posterior intervals for the hospital-specific dependence parameters  $\phi_j$  (top) and intervals for the hospital-specific innovation variance  $u_j$  (bottom) for all 136 hospitals. Hospitals ordered by posterior median of  $\phi_j$ . The five hospitals selected earlier are highlighted.

## 5 Conclusions and Further Research

This project is part of a series of profiling efforts by VA (Burgess, Christiansen, Michalak, and Morris, 2000; West and Aguilar, 1997; Aguilar and West, 1998; West, Aguilar and Lourdes, 1998) using hierarchical modeling techniques. A focus of these efforts has been to communicate profile results to managers to provide feedback on implemented policies and to effect quality improvements. Experience with all of these projects suggests that posterior probabilities and graphical distribution presentations such as those presented here are much easier for managers to comprehend than “significance tests” and parameter estimates or odds ratios derived from regression models. Time series profiles are shown to be useful in adding statistical power to profiling analyses and in identifying changes in outcome trends.

For this particular substance abuse process example, we illustrate how the prevailing 30-day outpatient return following inpatient discharge standard appears to be a reasonable compromise measure when including all substance abuse discharges. However, we also illustrate the importance of including individual level covariates in a hierarchical model when profiling provider units. In the 10 year panel, we document the persistence of hospital specific effects and the effectiveness of profiling individual hospitals both in any given year against the population norm and in changes over time measured against the hospital’s own previous performance. Major policy implications for the VA system include documentation of the effectiveness of policies aimed at improving returns while adjusting for major changes

in inpatient treatment patterns occurring at the same time. Covariates in the model all operate as expected and document probability differences and uncertainty in those differences in probability plots.

This study also represents a significant technical accomplishment as a time series hierarchical profiling model that accounts for individual level case mix variation and hospital level random effects variation over previous cross sectional analyses (e.g., Normand, Glickman, and Gatsonis, 1997). Improvements in Markov Chain Monte Carlo methods are permitting analysis of models with more data and more complexity. Related methods also are being investigated in a number of other similar health care applications (Normand, Frank and McGuire, 1999; Landrum and Normand, 1999; Landrum, Dendukuri, Bronskill, and Normand, 1999) which illustrate a wide range of possible applications – including multi-dimensional profiling and estimating the value of health care treatments. Continuous models which build hierarchical models in explaining cost and utilization differences also are possible as this continues to be a vibrant and active area for further investigation.

Future research and efforts in this area can extend the effectiveness of this work in several ways. Further investigation of impacts of profiling on policy and outcomes also can be productive. Incorporation of decision theory that implements cost-benefit analyses of effort in profiling (both in collection of data and efforts of researchers) can steer efforts in the most productive directions. Given the complexity and heterogeneity of medical care, all case mix adjustment is incomplete while posterior distributions appear to be an effective way of illustrating unexplained variation. A better understanding of these relationships could help to document the usefulness of hierarchical models in profiling. In general, though, continued technological improvements and use of similar models by more teams of researchers will be important to meet the sharply increasing demand for better profiling.

## 6 References

- Aguilar, O. and West, M. (1998) Analysis of hospital quality monitors using hierarchical time series models. In *Case Studies in Bayesian Statistics, Vol 4*, (eds: C. Gatsonis et al), New York: Springer-Verlag, 287–302.
- Burgess, J.F., Christiansen, C.L., Michalak, S.E., and Morris, C.N. (2000) Medical profiling: Improving standards and risk adjustments using hierarchical models. *Journal of Health Economics* 19, 291–309.
- Burgess, J.F. and DeFiore, D.A. (1994) The effect of distance to VA facilities on the choice and level of utilization of VA outpatient services. *Social Science and Medicine* 39, 95–103.
- Landrum, M.B., Dendukuri, N., Bronskill, S. and Normand S.L.T. (1999) Cross-Section Profiles of Medical Providers Using Multiple Quality Indicators. *Manuscript*, Harvard Medical School, Boston, MA. Presented at the 1999 International Conference on Health Policy Research, Santa Monica, CA.
- Landrum, M.B., and Normand, S.L.T. (1999) Applying Bayesian ideas to the development of medical guidelines. *Statistics in Medicine* 18, 117–137.
- Lehner, L.A., Burgess, J.F., Jr., Hults, D., and Stefos, T. (1996) Data and information requirements for the Department of Veterans Affairs resource allocation systems. *Medical Care* 34, MS21–MS30.

- Lourdes, V. (2000) Bayesian modelling and analysis of multivariate time series, with applications in finance and health policy. *PhD Thesis*, Institute of Statistics and Decision Sciences, Duke University. <http://www.stat.duke.edu/thesis/viri.html>
- McClellan, M. and Staiger, D. (1999) The Quality of Health Care Providers. *Manuscript*, National Bureau of Economic Research Working Paper 7327, Cambridge, MA.
- Normand, S.L.T., Frank, R.G. and McGuire, T.G. (1999) Elicitation in Observational Studies: Estimating the Value of Acute Treatments for Depression. *Manuscript*, Harvard Medical School, Boston, MA.
- Normand, S.L.T., Glickman, M.E. and Gatsonis, C.A. (1997) Statistical methods for profiling providers of medical care: issues and applications. *JASA* 92(439), 803–814.
- Raftery, A.E. (1995) Bayesian Model Selection in Social Research (with Discussion). In *Sociological Methodology 1995* (eds: P.V. Marsden), Cambridge, Mass: Blackwells, 111–196.
- Volinsky, C. (1996) BIC.GLM S-PLUS functions for Bayesian model averaging and variable selection for generalized linear models. Software available at Statlib.
- West, M. and Aguilar, O. (1997) Studies of Quality Monitor Time Series: The V.A. Hospital System. *Manuscript*, Management Science Group, U S Department of Veterans Affairs, and *ISDS Discussion Paper #97-22a*, Duke University.
- West, M., Aguilar, O. and Lourdes, V. (1998) V.A. Hospital Quality Monitors: 1988-1997. *Manuscript*, Management Science Group, U S Department of Veterans Affairs, and *ISDS Discussion Paper #98-41*, Duke University.