

STATISTICAL ANALYSES OF PSYCHIATRIC PATIENT RETURN TIMES IN THE V.A. HOSPITAL SYSTEM

MIKE WEST, VIRIDIANA LOURDES & JIM BURGESS

AUGUST 16, 1999

This report summarises investigations of VA hospital quality monitor data sets providing information about *return times* of individuals treated in psychiatric care, designated as the VA's *Monitor 20* care area. We report on data explorations, models and analyses of information for all individuals in the M20 care area of the VA hospital system, with a primary focus on estimating and understanding components of variability in patient return times. First, we discuss data analyses and modelling exercises on the system-wide data restricted to a single year, 1997. We summarise our development of discrete duration data analyses whose primary focus is to understand and estimate the extent to which effects specific to individual hospitals (facilities, stations) enter into the description of observed variations in return times. We do this in parallel logistic regression models that take into account several key categorical covariates related to the socio-demographic characteristics and VA medical history of individuals. Following this, we extend the discussion to the full ten years of data available, 1988-1997 inclusive, and explore an elaborated framework that links year-specific submodels in a longitudinal framework. This work extends prior work with aggregated data at the hospital level alone; the extensions to the individual level analyses are significant technically, and, as discussed below, identify several interesting covariate related issues as well as leading to inference on hospital specific effects in the context of informative covariates.

Mike West and Viridiana Lourdes are in the Institute of Statistics & Decision Sciences, Duke University (<http://www.stat.duke.edu>), and Jim Burgess is in the VA Management Science Group, Bedford MA.

Contents

1	Introduction	0
2	M20FY97 Data	1
3	Discrete Duration Models	2
3.1	Model form and specification	2
3.2	Screening covariates	5
4	Formal Analysis of FY97 using Selected Model	8
4.1	Model Completion and Computation	8
4.2	Baseline duration model parameters	9
4.3	Random effects for four selected hospitals	10
4.4	Age covariate	11
4.5	DRG covariate	11
4.6	Marital status covariate	12
4.7	Priority covariate	12
4.8	Diagnosis covariate	13
4.9	Gender covariate	13
4.10	Race covariate	13
4.11	Exploring all Hospital Effects	14
4.12	VA Regions	14
5	Multi-Year Models: Discussion and Model Form	15
5.1	Scope and Data Structure	15
5.2	Regression Model	16
5.3	Model Structure for Random Effects	17
6	Multi-Year Models: Analysis of M20FY1988-1997	18
6.1	Implementation	18
6.2	Baseline duration model parameters	19
6.3	Random effects for selected hospitals	19
6.4	Age covariate	20
6.5	DRG covariate	20
6.6	Priority covariate	20
6.7	Diagnosis covariate	21
6.8	Gender covariate	21
6.9	Dependence parameters ϕ_i and associated hyperparameters	21
6.10	Variability parameters w_i and associated hyperparameters	21
7	Possible Further Work	22
8	References	22
9	Graphical Summaries of Analysis: FY97 Analyses	24
10	Graphical Summaries of Analysis: 10 Year Study	43

1 Introduction

The study of VA quality monitor time series in West and Aguilar (1997), Aguilar and West (1998) and West, Aguilar and Lourdes (1998) concerned annual quality measures over the years 1987-1995. That study developed statistical analyses of patient “return for follow-up” data at a highly aggregate level: data studied were care-area and hospital specific annual quality measures. The current report explores similar issues though at the finest level of detail, studying patterns of variability in individual patient return times categorised by care area, hospital and several other socio-demographic and medical history covariates. As with the earlier studies, the key motivating concern is to evaluate differences in return time distributions that are specific to individual hospitals, now in the context of a range of possible explanatory variables.

Our study here focuses on the *general psychiatric* care area, Monitor 20 or M20. We study the individual level data on an annual basis, exploring variations in return times classified by calendar year. We begin with an in-depth study of the data in just one year, 1997. Across the entire hospital system, and for all patients discharged from initial care in this area during the calendar year 1997, the raw outcome of “quality” is the reported time from that point until return to follow-up care. The notion that delayed returns or long return times represent “poor quality” in this area of care underlies the interest in this data base. Interest lies in exploring the relationships between return times and a moderate number of available covariates, with a primary focus on understanding and explaining differences in return time distributions between hospitals. The Hospital factor is the primary covariate of interest, the others being secondary. However, in order to isolate hospital specific effects we need models that adjust for the covariates and so will be exploring differences in return times that are related to some of the other covariates.

Section 2 presents summaries of *M20FY97* (Monitor 20, financial year 1997) data, and discusses some of the finding from our preliminary exploratory analyses. Section 3 introduces the discrete duration models finally adopted, and reviews the process used to explore and isolate relevant subsets of covariates for further study. Section 4 gives in-depth discussion of several related analyses using these models, with substantive comments and summary inferences on the relative importance and effects of several of the categorical covariates in explaining structure in the observed return times, as well as formal summaries for the idiosyncratic effects of each of the hospitals.

Following this, Section 5 extends the focus to multiple years, developing more general models that relate hospital-specific effects across calendar 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 6. Section 7 briefly indicates and some potential further investigations. Appended are all graphical summaries of analysis.

2 M20FY97 Data

The M20FY97 data set includes a total of 35,368 individuals treated at 140 hospitals (stations/facilities) each with information on several covariates. We focus only on individuals, or “cases,” classified as outpatients. From the original data base and following discussion with VA personnel and preliminary/exploratory data analyses, we constructed a modified data base with return time data and categorical covariates; the key covariates are listed in Table 1. The primary record for each “case” includes:

1. Case identifier and the recorded return time in days, measured from the day of discharge from initial visit to the day of return to follow-up care. There are a total of 35,368 cases in the data base.
2. Hospital/station number, an integer label identifying the VA facility. There are a total of 140 facilities with patients recorded in the M20 care area in this data base.
3. Age factor, classifying cases as in Age group 1 ($\text{age} \leq 44$ years), 2 ($45 \leq \text{age} \leq 64$ years), and 3 ($\text{age} \geq 65$ years).
4. DRG factor, classifying cases as in DRG (Diagnostic risk group) 1 – 4 (DRG labels 434, 435, 436 and 437).
5. Marital status, classifying cases as of Marital status 1 (married), 2 (SDW, separated - widowed - divorced), and 3 (UN, unknown - never married).
6. Priority code status, based on a means test indicator that defines eligibility priority codes for use of VA services, classifying cases as in Priority group 1 (AN), 2 (AS) and 3 (other).
7. Gender, classifying cases as 1 (male) and 2 (female).
8. Race or national origin, classifying cases as 1 (White), 2 (Black, not Hispanic), 3 (Hispanic) and 4 (other, including Asian, American Indian and unknown).
9. Diagnosis, classifying cases into one of 11 groups, labelled 1 – 11 and associated with the principal medical diagnosis of the case, as follows:
 - 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

- 11: Non-substance abuse psychoses
10. A few additional categorical covariates relating to socio-economic and military service history of individuals, and region of country in which the facility is classified.

Table 1 provides summary frequencies of individuals classified according to a moving “cut-off” on the return time scale. 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 (hospital numbering serves only to provide labels for hospitals).

Table 1 gives the flavour of patterns of variability in the (discretised) marginal return time distributions categorised by each level of several 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, with 28.71% of all individuals returning at a time greater than 367 days; these are regarded as uninformatively censored cases that are essentially 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). Our formal confirmatory analyses below are based on binomial regression models whose outcomes are 1/0 according to whether or not an individual’s return time is less than/greater than a specified threshold; the thresholds studied are the cut-off values in Table 1, and the regression models include the covariates listed in the table. As we mention below, the remaining covariates in the data base have been explored too, and found to be essentially uninformative about patterns of variation in return times in the context of models that account for the listed covariates.

Several exploratory analyses of the return time data were undertaken at a first stage, to generate familiarity with the large and complicated data set and to provide initial indications of possibly relevant covariates. These analyses explored regressions of aggregated return times on several of the covariates. These included models of natural logs of median of groups of return times, where groups were defined by, for example, the covariates Hospital, Age and DRG. These analyses served, in addition, to focus attention on the specific grouping of data according to the covariates. Follow-on discussions with VA personnel used these studies to refine the groupings and led to the specific levels of all covariates as described and displayed in Table 1.

3 Discrete Duration Models

3.1 Model form and specification

Consider an individual patient with return time T measured from the origin $t = 0$ of release from first treatment visit. 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.

Return time thresholds:	$t = 1$	$t \leq 7$	$t \leq 14$	$t \leq 21$	$t \leq 30$	$t \leq 367$
All data ($n: 35,368$)	14.7	38.0	46.8	51.6	56.0	71.3
Hospital:						
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
Age group:						
≤ 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
DRG:						
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
Marital Status:						
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
Priority:						
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
Diagnosis:						
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
Gender:						
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
Race:						
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

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

$$p(t) = Pr(T \leq t) \quad (1)$$

in which

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

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 $\theta(t)$. By convention we take the base level of each categorical covariate as level 1, and set the regression parameter to zero at the base level, treating the effects of the covariates as fixed effects to be estimated. This applies to all covariates except Hospital, where we introduce hospital specific random effects with a population prior distribution to be estimated. For each hospital i we label the hospital effects $\epsilon_i(t)$ relative to this cut-off time t . Then the logit of the return probability at cut-off time t for individuals in hospital i is $\beta_0(t) + \epsilon_i(t)$ for individuals who fall in the base level of all other covariates included in the model. At other levels of covariates, we simply add the corresponding covariate effects. For example, in the final model chosen for extensive analysis, we include all covariates in Table 1. Then we have, in hospital i ,

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

Hence the parameter vector is

$$\theta(t)' = (\epsilon_1(t), \dots, \epsilon_{140}(t); \delta_2(t), \delta_3(t); \gamma_1(t), \dots, \zeta_3(t), \zeta_4(t)),$$

and the individual's x vector simply selects out the relevant effects parameters. This is a "main effects only" model, there being no terms representing interactions between covariates. There are substantive and empirical reasons for this, as we discuss below.

The notation makes explicit the fact that we allow for all effects to depend on the specified cut-off time t . We study a set of independent analyses of models with cut-offs

$t = 1, t = 7, t = 14, t = 21, t = 30$ and $t = 367$. Our analyses are independent though, of course, $p(t)$ must, for any individual and any set of specified covariates, be an increasing function of time t . However, it suffices to analyses models separately for each of these cut-offs to address the key questions of:

- The patterns of hospital specific random effects $\epsilon_i(t)$, compared across hospitals for a given cut-off, and across cut-offs;
- The patterns of covariate (fixed) effects, and how such patterns may vary with cut-off.

The model on the odds scale implies that, for any times $t > s$ and for an individual with regression vector x ,

$$\frac{p(t)}{1 - p(t)} = \pi(t, s) \frac{p(s)}{1 - p(s)}$$

where

$$\log(\pi(t, s)) = \beta_0(t) - \beta_0(s) + x'(\theta(t) - \theta(s)).$$

Here $\pi(t, s)$ is the *odds ratio*, just the ratio of the odds on a return time no greater than t to the odds on a return time no greater than s .

A very special case arises if all the covariate effects are constant, i.e., $\theta(t) = \theta$ for all t . It then follows that, for any times $t > s$, $\log(\pi(t, s)) = \beta_0(t) - \beta_0(s)$. Hence the odds on the individual returning before time t is that of returning before time $s < t$ adjusted by the multiplicative factor $\pi(t, s)$ and, in this special case, this factor does not depend on the covariates. The odds ratio depends only on the difference between the baseline effects $\beta_0(t) - \beta_0(s)$, and the model is then a discrete *proportional odds model*. The question as to whether or not this is a sensible simplification for the M20FY97 data is addressed below.

In all models, note that a positive value of a covariate effect corresponds, at the specified cut-off, to increased return probabilities relative to the base level of that covariate. High return probabilities, consistent with higher “quality” in the M20 care area, result from larger, positive effects. Larger negative effects correspond to lower return probabilities, hence to longer return times and lower quality of care. These correspondences should be borne in mind in exploring and interpreting graphical and numerical summaries of model analyses.

3.2 Screening covariates

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

Identify as a *model* any specified regression model with a chosen set of covariates, and suppose the set of such models is indexed by $k = 1, \dots, H$. In our study, models differ in the specific covariates they select as “in” the model, and we label the models M_k for $k = 1, \dots, H$. As there are 10 possible covariates for inclusion, there are $2^{10} = 1,024$ possible models. Theoretically, assuming a uniform prior probability of $1/H$ on each possible model,

Bayes' theorem implies that the posterior probability of model k by Bayes' theorem is

$$p(M_k|D) = \frac{p(D|M_k)}{\sum_{i=1}^H p(D|M_i)} \quad (4)$$

where D is the observed data and $p(D|M_k)$ represents the marginal data density under model M_k . 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 the standard *Bayesian Information Criterion (BIC)*-based posterior probabilities

$$p(M_k|D) \approx \frac{\exp\left(-\frac{1}{2}BIC_k\right)}{\sum_{i=1}^H \exp\left(-\frac{1}{2}BIC_i\right)}$$

where the *BIC* measure for model k is $BIC_k = Dev_k - df_k \log(N)$; here Dev_k is the classical deviance of model M_k , N is the total number of observations, and df_k is the degrees of freedom associated with M_k (i.e., N minus the number of regression parameters in the model). If H is large, the direct evaluation of the sum over all models is not feasible. In our situation the number of possible models is equal to the number of subset of covariates, this is $2^{10} = 1,024$ models, and the computation is easily accessible.

The approximate posterior model probabilities were computed this way for all possible combinations of covariates. This screening analysis assumed that the Hospital effects are fixed effects, rather than random, to simplify the computations. Tables 2 and 3 provide summaries of the results, identifying key covariates and giving summaries of the approximate posterior probabilities and BIC measures for several of the most relevant models. We performed the full screening analysis twice for each possible return time cut-off: once using only uncensored data (results in Table 2), and secondly using all the data (results in Table 3). The tables indicate, by the entries X in each column, which covariates are selected in a set of chosen models and analyses, represented by the rows. For example, the first three rows of Table 2 refer to the analyses of the uncensored data alone with models having return time cut-off $t = 1$. Three models are indicated there: the first has approximate posterior probability of 96.5%, and is that model based on covariates Hospital, DRG, Age, Priority and Diagnosis, but excluding covarites Marital Status, Gender and Race.

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. In analyses using only the uncensored data, the covariates Hospital, DRG and Diagnosis are always selected, and Age and Priority generally are (they are excluded only in a few low probability models). Gender and Marital Status are present in some models with low posterior probability. 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 seem to be only of very marginal relevance, if they are at all relevant; however, we choose to include them in formal analysis as specific interests exist in quantifying their effects, however small and subtle they may be.

Return time	Hospital	DRG	Age	Marital Status	Priority	Diagnosis	Gender	Race	Post. Prob	BIC
$t = 1$	×	×	×		×	×			96.50	-124812.93
	×	×	×		×	×			2.39	-124805.53
	×	×	×		×	×	×		1.11	-124804.01
$t \leq 7$	×	×	×		×	×			100.00	-119399.98
$t \leq 14$	×	×	×		×	×			100.00	-119985.11
$t \leq 21$	×	×	×		×	×			94.86	-121003.07
	×	×	×		×	×			2.78	-120996.02
	×	×	×		×	×			1.38	-120994.61
	×	×	×		×	×	×		0.98	-120993.93
$t \leq 30$	×	×	×		×	×			79.94	-122626.23
	×	×	×		×	×			13.97	-122622.74
	×	×	×		×	×	×		4.53	-122620.49
	×	×	×		×	×	×		0.89	-122617.23
	×	×	×		×	×	×		0.68	-122616.68

Table 2: Selected covariates using *BIC* based on uncensored data

Return time	Hospital	DRG	Age	Marital Status	Priority	Diagnosis	Gender	Race	Post. Prob	BIC
$t = 1$	×	×	×		×	×			97.65	-167137.68
	×	×	×	×	×	×			1.28	-167129.01
	×	×	×		×	×	×		1.07	-167128.65
$t \leq 7$	×	×	×		×	×			98.60	-158916.50
	×	×	×	×	×	×			1.40	-158907.99
$t \leq 14$	×	×	×		×	×			100.00	-157769.12
$t \leq 21$	×	×	×		×	×			99.09	-157536.78
	×	×	×		×	×	×		0.91	-157527.40
$t \leq 30$	×	×	×		×	×			96.71	-157713.02
	×	×	×		×	×	×		3.29	-157706.26
$t \leq 367$	×	×	×		×	×			100.00	-160494.80

Table 3: Selected covariates using *BIC* based on all data

Finally, we have explored posterior mean estimates and maximum likelihood estimates of the primary Hospital effects across several analyses with the models selected as displayed in the tables. It is of real note that, across various models with different subsets of covariates, these 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.

4 Formal Analysis of FY97 using Selected Model

4.1 Model Completion and Computation

As mentioned above, our analysis strategy involved fitting the full binomial logistic model of equation (3) to the entire data set, and repeating the analysis in separate studies based on cut-offs at $t = 1, 7, 14, 21, 30$ and 367 respectively. This gives us good coverage of sections of the return time axis, while also focusing explicitly on immediate returns ($t = 1$) and censored/failed returns $t > 367$. We note the complete parallel with the prior studies of West and Aguilar (1997), Aguilar and West (1998) and West, Aguilar and Lourdes (1998), that involved similar models at the single cut-off of $t = 30$ but using highly aggregated data at the hospital level, i.e., not involving individual level covariates. The current models may therefore be viewed as very substantial refinements of that study, with significant complications technically, in terms of dealing with a large and complex data base, and computationally. In addition, our current study expands the scope to several cut-offs on the return time scale, so providing access to information about hospital and covariate effects at fine levels of detail that may be relevant in assessment and interpretation in connection with possible VA policy questions.

Each specific analysis involved completing the model specification with prior distributions for covariate effects and hyperparameters, and then computation of posterior distributions. Within each analysis, this has the following components:

- The hospital specific random effects are assumed drawn from a normal population model

$$\epsilon_i(t) \sim N(0, w(t)^2)$$

where the cut-off dependent standard deviation $w(t)$ 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 methods to iteratively simulate from the full joint posterior distribution of

$$\{\epsilon_i(t), i = 1, \dots, 140; \theta(t), w(t)\},$$

producing large Monte Carlo samples for summary inferences on all parameters and effects.

In addition to exploring summaries for the basic parameters, we trivially compute similar summaries for functions of the parameters, such as individual level return probabilities $p(t)$ from equations (1) and (3). The summary graphs display a range of such posterior summaries, and are discussed in detail in the following subsections. We explain the structure of the graphs in the context of summary inferences on the baseline parameters $\beta_0(t)$ in the next subsection. First, however, we note that, in addition to analysing the model across the specified set of cut-offs, all such analyses are performed twice:

- One set of analyses uses only individuals whose return times are less than 367 days,
- the second set of analyses uses all the data, including those censored at 368 days.

Our interest in performing these two sets of analyses is a response to the fact that a large number of cases are censored at 367 days, and therefore we are concerned to explore possible differences in summary inferences in analyses with and without the censored cases. As it turns out, we do find some important differences, as we shall note below and that require interpretation and investigation.

The summary graphs relevant to each of the covariates have essentially the same format. For each covariate, we display posterior inferences for the effects parameters for each level or group of the variable on both the logit and probability scales. Thus we display inferences for the actual effects $\epsilon_i(t)$ for each of the 4 selected hospitals, and later for all 140 hospitals, for the Age covariate effects $\delta_2(t)$ and $\delta_3(t)$, for the Diagnosis effects $\xi_2(t), \dots, \xi_{11}(t)$, and similarly for the other covariates. The displays present approximate posterior 95% intervals for each of the parameters, with posterior medians and quartiles indicated. The intervals are all presented as vertical lines in frames designed so that it is straightforward to make comparisons across levels of the covariates. In addition, each frame displays these intervals from each of the independent analyses corresponding to different cut-offs t , and with each of these cut-off specific analyses performed separately for the data with and without the censored cases. We distinguish the analyses of the two data sets using green colouring for intervals based on the full data set. Further, inferences for each covariate includes separate displays of similar inferences mapped to the implied differences on a probability scale. At the base level of all covariates, the baseline return probability at cut-off t is simply $p_0(t) = 1/(1 + \exp(-\beta_0(t)))$; moving to level j of any one of the covariates shifts this to $p_j(t) = 1/(1 + \exp(-\beta_0(t) - \tau_j(t)))$ where $\tau_j(t)$ represents the corresponding effect for that level of the chosen covariate (e.g., $\tau_j(t) = \delta_j(t)$ for Age group j , $\tau_j(t) = \xi_j(t)$ for Diagnosis group j , and so forth.) The second set of graphs provide posterior median and interval estimates of the implied *differences in return probabilities* relative to baseline, namely $p_j(t) - p_0(t)$ for each level j of each of the covariates, for each of the chosen cut-offs t , and from analysis of both the full and censored data sets.

4.2 Baseline duration model parameters

Refer to Figure 1 and consider first the posterior intervals displayed in black. In the upper frame, these are intervals for $\beta_0(t)$ in each of the five independent analyses with

$t = 1, 7, 14, 21$ and 30 , and in the restricted data set ignoring the censored return times (greater than 367). The X symbols indicate posterior medians. The lower frame displays the corresponding intervals, again with medians marked, for the baseline return probabilities $p_0(t) = 1/(1 + \exp(-\beta_0(t)))$, i.e., 5 points on the cumulative distribution function for return times, with the interpretation as that of return times at an “average” hospital ($\epsilon_i = 0$) and for individuals in the base levels of all other covariates. Reading off approximate values here indicates about an 15-19% probability of return in exactly 1 day, about a 50-57% probability of return within 7 days, about a 63-68% probability of return within 14 days, about a 72-76% probability of return within 21 days, and about a 79-82% probability of return within 30 days.

(As an aside, note that $p_0(t)$ must increase as a function of t and the graph indicates that, although the analyses were not linked to enforce monotonicity.)

Now consider the second set of intervals in each frame, those coloured green. These present the same inferences for $\beta_0(t)$ and $p_0(t)$ values, but now based on the second set of analyses using all the data, so including the additional cases censored at $t = 367$. Notice that, as should be expected, the intervals and medians for $p_0(t)$ in this case indicate values close to the raw frequencies in the first row of Table 1. Otherwise, the main question is how these result compare with those based on the restricted data set. It is clear and obviously expected that the intervals support rather lower values, i.e., lower baseline “quality” levels, when the study includes the censored cases – meaningful numbers of individuals are censored.

4.3 Random effects for four selected hospitals

Refer to Figure 2 where similar intervals and estimates of the random effects for the 4 selected hospitals are displayed. The colour coding again differentiates between the two analyses (with and without censored cases), and the intervals in the upper frame are those for the $\epsilon_i(t)$ parameters directly. The lower frame displays similar intervals and estimates for the implied differences in return time probabilities relative to baseline. i.e., for $p_i(t) - p_0(t)$ where $p_i(t) = 1/(1 + \exp(-\beta_0(t) - \epsilon_i(t)))$ for each i and t .

Some points to note here are as follows:

- There are general differences between hospitals and the differences are consistent across analyses. These specific hospitals would be ranked (from low quality to higher quality) as 21, 82, {139, 118}. Of these, 21 is clearly below average, with effects $\epsilon_i(t)$ clearly negative.
- The nature and magnitude of the implied differences on the probability scale are clear from the lower graph in the figure.
- There is very clear evidence that the effects represent non-proportional odds structure in the return time regression model for at least one of the hospitals. Consider hospital 139, and note that the clear differences in inferences for this hospital’s effect $\epsilon_i(1)$ compared to the others $\{\epsilon_i(7), \dots, \epsilon_i(30)\}$. The parameter $\epsilon_i(1)$ is obviously higher than the rest, indicating that the probability of return in exactly one day for this hospital is increased by more, relative to the average, than is the probability of return at later days. For the remaining cut-offs at this hospital, the intervals overlap so the differences are unclear; the decreasing pattern of the posterior medians as we

move through $t = 7, 14, ..$ is suggestive, however, and would imply that the increased quality, in terms of increased probability of a return, exhibits a diminishing effect at later times. The same may be said for hospital 118.

- For hospitals 82 and 21, the intervals are suggestive of a constant effect across cut-offs, i.e., $\epsilon_i(t) \approx \epsilon_i$ for these hospitals.
- The results are consistent between the two data sets (with and without censored cases) for all but hospital 21. Here the effects are apparently lower in the analysis of the full data set, consistent with the raw frequencies that indicate over 60% censored cases at this facility.

Figure 3 provides posterior intervals and medians for the population standard deviation $w(t)$ in each of the analyses. The major point to note here is the apparent decreasing pattern as we move through increasing values of the cut-off t . This is particularly significant in moving from $t = 1$ onward. This implies that variability in the distribution of hospital specific effects is significantly more marked at $t = 1$ than at later times, and tends to decrease with increasing t . This implies, and is implied by, variations in at least some of the $\epsilon_i(t)$ over time, as just exemplified. It is also consistent with features evident in the table of frequencies: the differences in raw frequencies across the 4 selected hospitals are less marked at higher cut-offs.

4.4 Age covariate

Refer to Figure 4 for intervals and estimates of the fixed effects $\delta_j(t)$ for the Age covariate at levels $j = 2$ (45-64) and $j = 3$ (65 or over). Recall that these are referenced to the effect of 0 for Age group 1 (less than 45). The lower frame displays similar intervals and estimates for the implied differences in return time probabilities relative to baseline. i.e., for $p_j(t) - p_0(t)$ where $p_j(t) = 1/(1 + \exp(-\beta_0(t) - \delta_j(t)))$ for each j and t .

We note the following:

- We see some evidence of non-proportional odds behaviour, with group 2 effects being less than zero for thresholds $t = 1, 7$ and 14 , but not obviously different from 0 for higher values of t . This corresponds to initially decreased quality, in terms of lower return probabilities, at early times for Age group 2 relative to the younger group, but that these differences disappear after 3 weeks.
- 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 results are consistent between the two data sets (with and without censored cases) in Age group 2, but clearly not in Age group 3. In this older category, the effects are apparently lower in the analysis of the full data set, consistent with the raw frequencies that indicate larger numbers of older individuals as non-returners.

4.5 DRG covariate

Refer to Figure 5 for intervals and estimates of the fixed effects $\gamma_j(t)$ for the DRG categories $j = 2$ (DRG 435), $j = 3$ (DRG 436) and $j = 4$ (DRG 437), relative to the effect of 0 at

level $j = 1$ (DRG=434). The lower frame displays similar intervals and estimates for the implied differences in return time probabilities relative to baseline. i.e., for $p_j(t) - p_0(t)$ where $p_j(t) = 1/(1 + \exp(-\beta_0(t) - \gamma_j(t)))$ for each j and t .

We note the following:

- We note a decreasing trend with t in the posterior median in DRG 435. Though there is considerable uncertainty, this is suggestive of a non-proportional odds effect and would imply a persistent decrease in return probability with threshold at later times.
- In DRG 436 and 437, there are clear differences between estimated effects for $t = 1$ relative to later cut-offs. These DRG categories have generally higher return probabilities than the rest, with the very significant exception of the immediate returns, $t = 1$. At $t = 1$, the return probabilities in these two groups are essentially consistent with the base level DRG 434. By comparison, DRG 435 is clearly above the base level at $t = 1$.

4.6 Marital status covariate

Refer to Figure 6 for intervals and estimates of the fixed effects $\kappa_j(t)$ for the Marital Status categories $j = 2$ (SDW) and $j = 3$ (UN) relative to the effect of 0 at level $j = 1$ (M). The lower frame displays similar intervals and estimates for the implied differences in return time probabilities relative to baseline. i.e., for $p_j(t) - p_0(t)$ where $p_j(t) = 1/(1 + \exp(-\beta_0(t) - \kappa_j(t)))$ for each j and t . We note the following:

- 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.
- There is some evidence of increased probabilities of return immediately, $t = 1$, in both groups 2 and 3 relative to that among Married individuals, and perhaps a minor increase up to $t = 7$ among Separated - Divorced - Widowed individuals.

4.7 Priority covariate

Refer to Figure 7 for intervals and estimates of the fixed effects $\eta_j(t)$ for the Priority categories $j = 2$ (AS) and $j = 3$ (other) relative to the effect of 0 at level $j = 1$ (AN). The lower frame displays similar intervals and estimates for the implied differences in return time probabilities relative to baseline. i.e., for $p_j(t) - p_0(t)$ where $p_j(t) = 1/(1 + \exp(-\beta_0(t) - \eta_j(t)))$ for each j and t . We note the following:

- 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.

4.8 Diagnosis covariate

Refer to Figure 8 for intervals and estimates of the fixed effects $\xi_j(t)$ for Diagnosis categories $j = 2, \dots, 11$ relative to the effect of 0 at level $j = 1$. The lower frame displays similar intervals and estimates for the implied differences in return time probabilities relative to baseline. i.e., for $p_j(t) - p_0(t)$ where $p_j(t) = 1/(1 + \exp(-\beta_0(t) - \xi_j(t)))$ for each j and t . For clarity, the graphs in Figure 9 redisplay the effects separately for the analyses of the two data sets (with and without censoring), and on a different scale. We note the following:

- 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.

4.9 Gender covariate

Refer to Figure 10 for intervals and estimates of the fixed effects $\chi_2(t)$ for the Gender category $j = 2$ (female) relative to the effect of 0 at level $j = 1$ (male). The lower frame displays similar intervals and estimates for the implied differences in return time probabilities of women versus men, i.e., for $p_2(t) - p_0(t)$ where $p_2(t) = 1/(1 + \exp(-\beta_0(t) - \chi_2(t)))$ for each t . We note the following:

- 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, however, as reflected by the intervals displayed. Only at $t = 30$ in the analysis excluding censored data is the effect really significant.
- At $t = 1$ the effect $\chi_2(1)$ is rather uncertain, though the posteriors do give more weight to negative values, which would indicate a small decrease in the probability of an immediate return for women relative to men.

4.10 Race covariate

Refer to Figure 11 for intervals and estimates of the fixed effects $\zeta_j(t)$ for the Race categories $j = 2$ (Black), $j = 3$ (Hispanic) and $j = 4$ (Asian, American Indian and others) relative to the effect of 0 at level $j = 1$ (White). The lower frame displays similar intervals and estimates for the implied differences in return time probabilities relative to baseline. i.e., for $p_j(t) - p_0(t)$ where $p_j(t) = 1/(1 + \exp(-\beta_0(t) - \zeta_j(t)))$ for each j and t . We note the following:

- Consider first the analysis of the data excluding censored cases (black intervals). None of the effects are really significant. There is a suggestion of lower return probabilities in the ‘‘Other’’ category, but with very high uncertainty. There is no evidence at all for a difference with respect to White in the Hispanic category. The effects in the ‘‘Black’’

category, while not really significant, do evidence a monotonic decreasing pattern in the posterior medians that is at least suggestive of a small but persistent decrease in return probabilities, relative to Whites, at later return times. The median estimates alone are consistent with slightly higher return probabilities for Blacks relative to Whites at the early return times up to 7 days, though the magnitude of the difference is practically insignificant.

- In the analysis of all the data, including censored times, the effects for Blacks are inferred as more consistent across cut-offs, and 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.

4.11 Exploring all Hospital Effects

Some final graphs display approximate 95% intervals, with medians marked, for the Hospital specific random effects $\epsilon_i(t)$ in all analyses of the uncensored data. This is mainly to indicate the kinds of additional inferential summaries that may be drawn from the work, and parallels similar displays in earlier related studies (West and Aguilar 1997), Aguilar and West 1998, and West, Aguilar and Lourdes 1998). However, the graphs also serve to reinforce 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. The relevant graphs are Figures 12 to 16 inclusive, running through 5 analyses with return time cut-offs $t = 1, 7, 14, 21$ and 30. The hospitals have a common ordering across figures, chosen simply as the order of the posterior medians on the effects in the analysis with $t = 1$, i.e., that of Figure 12. As we move through analyses with higher cut-offs, the appearance is no longer monotonic, indicating that the relative quality levels of hospitals is indeed quite variable with cut-off. Indications of uncertainty are also clear in these graphs, highlighting hospitals with smaller samples sizes that lead to wider intervals.

4.12 VA Regions

The 140 hospitals studies here are clustered into 22 geographical/administrative VISN regions (Veterans Integrated Service Network). To provide an initial graphical examination of possible patterns of variation in the hospital specific effects by VISN region, Figures 17 and 18 display the posterior medians of each of the $\epsilon_i(t)$ by region. As with all other graphs, we present these results from the independent analyses with the various return time cut-offs, and separately in the two cases based on uncensored data and then all data. Each region has a small number of hospitals, and the corresponding median estimates of region-specific $\epsilon_i(t)$ are dot-plotted above region indicator, with a zero line indicated. The regions are ordered by the median of estimates within region based on the analysis with return time cut-off $t = 1$ and using only the uncensored data. This gives a useful graphical view of structure within region, though we note that the posterior uncertainty about the $\epsilon_i(t)$ is not reflected here. There is little structure here; most regions have hospital effects scattered about the zero line, suggestive of no appreciable region-specific effects. We do see

that the two extreme regions tend to have all estimated effects above or below zero (i.e., above or below average return probabilities), but again this does not account for the estimation uncertainty about the $\epsilon_i(t)$, so the significance is unclear. To more formally assess the effects of this uncertainty, we may compute, for each region, the posterior probability that the hospitals within that region all have $\epsilon_i(t) > 0$. This is an easy calculation in our simulation-based analysis, and leads to the conclusion that for none of the regions is this posterior probability greater than 0.5; most are much smaller. This supports the conclusion that there is no serious pattern of low/high return times within specific regions.

5 Multi-Year Models: Discussion and Model Form

5.1 Scope and Data Structure

We now develop the 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 our 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 with years. We introduce slightly different time series structures, however, in order to adequately capture aspects of heterogeneity across the 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 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 $t = 30$, all data. Before describing the model extensions, we note some data and covariate selection issues.

First, the multi-year models are applied to a total of 136 hospitals, compared to the 140 used in the FY1997 analysis, to include only hospitals in which there are patients in the M20 area each year. Second, we make some minor revisions to the selection and specification of categorical covariates, based on the results of the FY1997 analysis above and following further discussions with VA personnel. Specifically, the model adopts the following changes to covariate specifications:

1. Hospital/station, a total of 136 facilities with patients recorded in the M20 area.
2. Age factor, refined to classify cases into just two groups: Age groups 1 (age \leq 64 years) and 2 (age \geq 65 years).
3. DRG factor with 4 levels, unchanged.
4. Priority code status with 3 groups, unchanged.
5. Gender, 1 (male) and 2 (female), unchanged.
6. Diagnosis, 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.

7. Marital status and Racial groups are not now used.

We finally note some summary sample sizes and indications of the scale of the data analysis undertaken with the full 10 years of data. The original VA data files comprise about 40MB of recorded data. Following very extensive data exploration and reorganisation (in both SAS and S-Plus), the data is reduced in the context of the covariates indicated above to a total of 463,015 individual releases in this care area across the hospital 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×Age×DRG×Priority×Diagnosis×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.

5.2 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

$$\text{logit}(p_r) = \beta_{0,r} + \epsilon_{i,r} + \begin{cases} \delta_{d,r}, & \text{for Age group } d = 2, \\ \gamma_{g,r}, & \text{for DRG level } g = 2, 3, 4, \\ \eta_{e,r}, & \text{for Priority level } e = 2, 3, \\ \xi_{x,r}, & \text{for Diagnosis group } x = 2, 3, \\ \chi_{c,r}, & \text{for Gender group } c = 2 \text{ (women)}, \end{cases} \quad (5)$$

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.

5.3 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 our 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 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

$$\epsilon_{i,r} = \phi_i \epsilon_{i,r-1} + \omega_{i,r} \quad (6)$$

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

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

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

$$\epsilon_{i,r} \sim N(\epsilon_{i,r} | 0, w_i^2) \quad (8)$$

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

This dependence structure parallels that in our prior work, referenced above, though with the extensions noted above: the parameters (ϕ_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 (ϕ_i, u_i) across hospitals i that completes the full model structure.

We assign an hierarchical model structure to the hospital-specific parameters (ϕ_i, u_i) . Assuming these to be exchangeable parameters drawn from a hospital-population prior delivers a class of Bayesian hierarchical models: modelling the (ϕ_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 ϕ_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 , ϕ_i and u_i are independent. Second, the population of dependence parameters appears as a random sample from a beta distribution,

$$\phi_i \sim Be(\phi_i | a\mu, a(1 - \mu)) \quad (9)$$

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

$$u_i^{-2} \sim Ga(u_i^{-2} | c, c\rho) \quad (10)$$

where the hyperparameter ρ 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, 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. One very positive 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.

6 Multi-Year Models: Analysis of M20FY1988-1997

6.1 Implementation

Posterior analysis uses Markov chain Monte Carlo methods to iteratively simulate from the full joint posterior distribution of all model effects and hyperparameters. This extends the previous, single-year analysis to the full 10 years, 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. These graphs are presented in formats analogous to those of the above single-year study, though now extended over the full 10 years of data. We report the analysis of the full data set including the censored cases (return times exceeding one year), and, as mentioned earlier, at the chosen 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 1997 alone are completely consistent and in agreement with the analysis of that year alone, reported above, as is to be expected.

6.2 Baseline duration model parameters

Refer to Figure 19 for the posterior intervals and estimates for the baseline parameters $\beta_{0,r}$ in each of the 10 years $r = 1, \dots, 10$. In the upper frame, these are intervals for $\beta_{0,r}$, the general increasing trend with years is apparent, and the lack of overlap of intervals in the last couple of years indicate significant changes. The lower frame displays the corresponding intervals for the baseline return probabilities $p_{0,r} = 1/(1 + \exp(-\beta_{0,r}))$. This plot 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 as follows:

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

The graphs also displays these observed values (logistic transforms in the upper frame) as points next to the corresponding interval estimates.

The pattern over the years should and does generally agree with that of the inferred baseline parameters. (The fact that these data summaries generally exceed the baseline parameters is simply a result of the fact that the majority of the observations lie in covariate groups with generally positive effects; a key covariate is DRG, for which most of the effects are positive – and in higher DRG levels large and positive – across the years, as illustrated below. Almost 80% of the data lies in these three higher DRG levels.)

This pattern of change over the years reflects hospital system-wide effects on 30-day return time probabilities. The effects of system-wide policy changes and common management practices presumably underlie some of the changes and apparent improvements in 30-day returns.

6.3 Random effects for selected hospitals

Refer to Figures 20 and 21 for the posterior intervals and estimates of the random effects $\epsilon_{i,r}$ for 5 selected hospitals; these are the 4 hospitals selected in the FY1997 analysis above, plus a fifth, hospital number #140. The five lower frames in Figure 20 displays intervals and estimates for the implied differences in return time probabilities relative to baseline, i.e, for $p_{i,r} - p_{0,r}$ where $p_{i,r} = 1/(1 + \exp(-\beta_{0,r} - \epsilon_{i,r}))$ for each hospital i and year r .

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 1997 implied by the large and positive effect in that year. The additional hospital, number 140, has experienced return probabilities much lower than the norm during the first nine years, but has seen a very marked increase also in 1997, with a magnitude of improvement that exceeds that of 118. Changes in the difference in return probabilities relative to baseline from 1996 to 1997 are from about +2% to +13% for hospital 118, and from about -14% to +24% for hospital 140.

6.4 Age covariate

Refer to Figure 22 for intervals and estimates of the fixed effects $\delta_{2,r}$ for the Age covariate at level $j = 2$ (65 or over). The lower frame displays similar intervals and estimates for the implied differences in return time probabilities relative to baseline. i.e., for $p_{2,r} - p_{0,r}$ where $p_{2,r} = 1/(1 + \exp(-\beta_{0,r} - \delta_{2,r}))$ for each r .

Within each year, the effect is clearly negative consistent with generally lower 30-day return probabilities for patients over 65 years of age. Further, the graphs show a mild though significant decreasing pattern over the 10 year period, consistent with a deterioration in 30-day returns in the older Age group.

6.5 DRG covariate

Refer to Figures 23 and 24 for intervals and estimates of the fixed effects $\gamma_{j,r}$ for the DRG categories $j = 2$ (DRG 435), $j = 3$ (DRG 436) and $j = 4$ (DRG 437), relative to the effect of 0 at level $j = 1$ (DRG=434). The three lower frames in Figure 23 displays similar intervals and estimates for the implied differences in return time probabilities relative to baseline. i.e., for $p_{j,r} - p_{0,r}$ where $p_{j,r} = 1/(1 + \exp(-\beta_{0,r} - \gamma_{j,r}))$ for each j and r .

Figure 23 clearly indicates meaningful year-to-year variations in the DRG effects within each DRG category, and an apparently persistent deterioration in return probabilities in DRG groups 436 and 437 over 1994-1997. Figure 24 provides a picture of the relative levels across DRG categories over the years.

6.6 Priority covariate

Refer to Figures 25 and 26 for intervals and estimates of the fixed effects $\eta_{j,r}$ for the Priority categories $j = 2$ (AS) and $j = 3$ (other) relative to the effect of 0 at level $j = 1$ (AN). The two lower frames in Figure 25 displays similar intervals and estimates for the implied differences in return time probabilities relative to baseline. i.e., for $p_{j,r} - p_{0,r}$ where $p_{j,r} = 1/(1 + \exp(-\beta_{0,r} - \eta_{j,r}))$ for each j and r .

Figure 25 shows 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. 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 1989-92/3, but a sharp drop-off in 1997.

6.7 Diagnosis covariate

Refer to Figures 27 and 28 for intervals and estimates of the fixed effects $\xi_{j,r}$ for Diagnosis categories $j = 2$ (psychoses) and $j = 3$ (Abuse) relative to $j = 1$ (Dependence). The two lower frames in Figure 27 displays similar intervals and estimates for the implied differences in return time probabilities relative to baseline. i.e., for $p_{j,r} - p_{0,r}$ where $p_{j,r} = 1/(1 + \exp(-\beta_{0,r} - \xi_{j,r}))$ for each j and r .

The effect in the Psychoses group is generally negative and consistent with lower return probabilities than in either the Dependence or Abuse groups.

6.8 Gender covariate

Refer to Figure 29 for intervals and estimates of the fixed effects $\chi_{2,r}$ for the Gender category $j = 2$ (female) relative to the effect of 0 at level $j = 1$ (male). The lower frame in Figure 29 displays similar intervals and estimates for the implied differences in return time probabilities of women versus men, i.e., for $p_{2,r} - p_{0,r}$ where $p_{2,r} = 1/(1 + \exp(-\beta_{0,r} - \chi_{2,r}))$ for each r .

Over the full 10 year span of the data, the effect is positive, consistent with generally higher 30-day return probabilities among women than men. The levels are somewhat lower during the later years, however.

6.9 Dependence parameters ϕ_i and associated hyperparameters

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, μ) in the hospital population model for the ϕ_i . Parameter μ 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 μ are 0.82-0.85-0.88, indicating high correlation structure generally across hospitals. Parameter a measures the dispersion of the actual, hospital-specific ϕ_i values about μ . A high value indicates the ϕ_i are very tightly distributed around μ , 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 ϕ_i parameters across hospitals. This is confirmed in Figure 30 where we display posterior estimates and intervals for the ϕ_i for all 136 hospitals. We present the hospitals ordered according to increasing values of the posterior medians of the ϕ_i .

The concordance illustrated here suggests that we could cut-back to a simpler model in which the ϕ_i taken the common value μ . 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 ϕ_i has confirmed the consistency across hospitals in the dependence structure between years.

6.10 Variability parameters w_i and associated hyperparameters

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, ρ) in the hospital

population model for the hospital-specific innovation variances u_i^2 . Parameter ρ 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 ρ are 0.04-0.05-0.06. Parameter c measures the dispersion of the actual, hospital-specific innovation variances u_i^2 values about ρ . 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 Figure 31. Rather than displaying posterior summaries for the u_i^2 themselves, we prefer to focus on the more naturally interpretable variances $w_i^2 = u_i^2 / (1 - \phi_i^2)$. Parameter w_i is the standard deviation of the marginal distribution of $\epsilon_{i,r}$ in any year r , so directly relates to how variable the hospital i effects are likely to be. Figure 31 displays posterior estimates and intervals for the standard deviations w_i , with hospitals in the order chosen for Figure 30. We see some variation in both estimates and uncertainties (interval widths) across the hospitals. The five hospitals noted earlier are highlighted in Figure 31; here we see that hospital 140 clearly stands out as a case of relatively high variability, consistent with the earlier discussion about the significantly varying ϵ parameters in this hospital.

7 Possible Further Work

Further work aims to extend this study in several ways. Some aspects for consideration include:

- Putting up posterior summaries for all hospitals and all years, perhaps on a web site for access by VA personnel based on queries with hospital/station numbers.
- Exploration of model diagnostics based on posterior predictive residual analyses.
- Understanding the issues associated with the censored return times.
- Highlighting other inferential uses of these models in response to specific, VA policy-driven questions.
- Analysis of multi-year data with different return time cut-offs.

8 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, pp287-302.
- Burgess, J.F., Christiansen, C.L., Michalak, S.E., and Morris, C.N. (1996) Risk adjustment and economic incentives in identifying extremes using hierarchical models: A profiling application using hospital monitors, *Manuscript*, Management Science Group, U S Department of Veterans Affairs, Bedford MA.
- 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.
- Raftery, A.E. (1995) Bayesian Model Selection in Social Research (with Discussion). In *Sociological Methodology 1995* (eds: P.V. Marsden), Cambridge, Mass: Blackwells, pp111-196.
- Volinsky, C (1996) BIC.GLM S-PLUS functions for Bayesian model averaging and variable selection for generalized linear models. Software available at Statlib.

9 Graphical Summaries of Analysis: FY97 Analyses

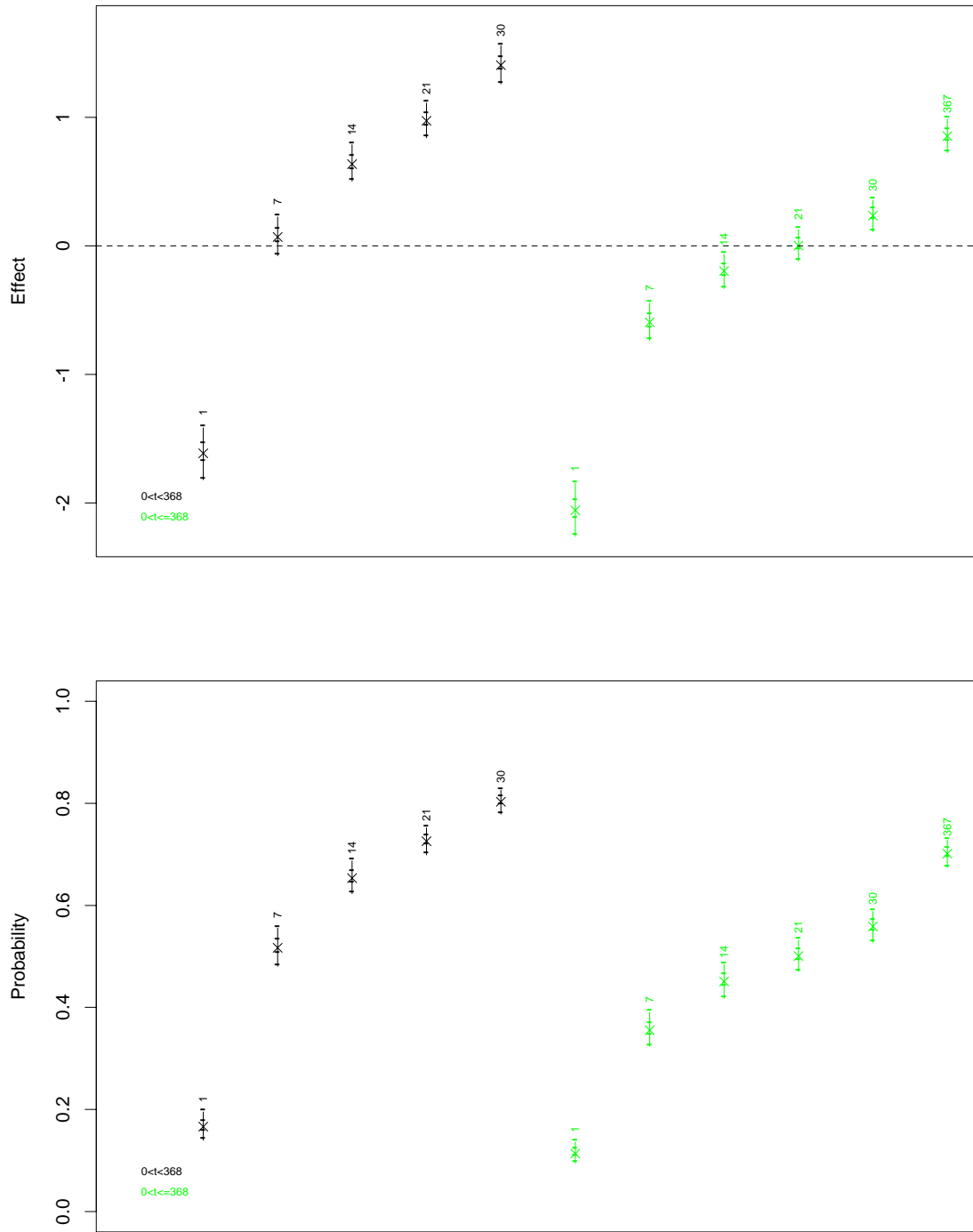


Figure 1: Intervals for baseline duration model parameters $\beta_0(t)$ (upper frame) and the implied baseline return time probabilities $p_0(t)$ (lower frame).

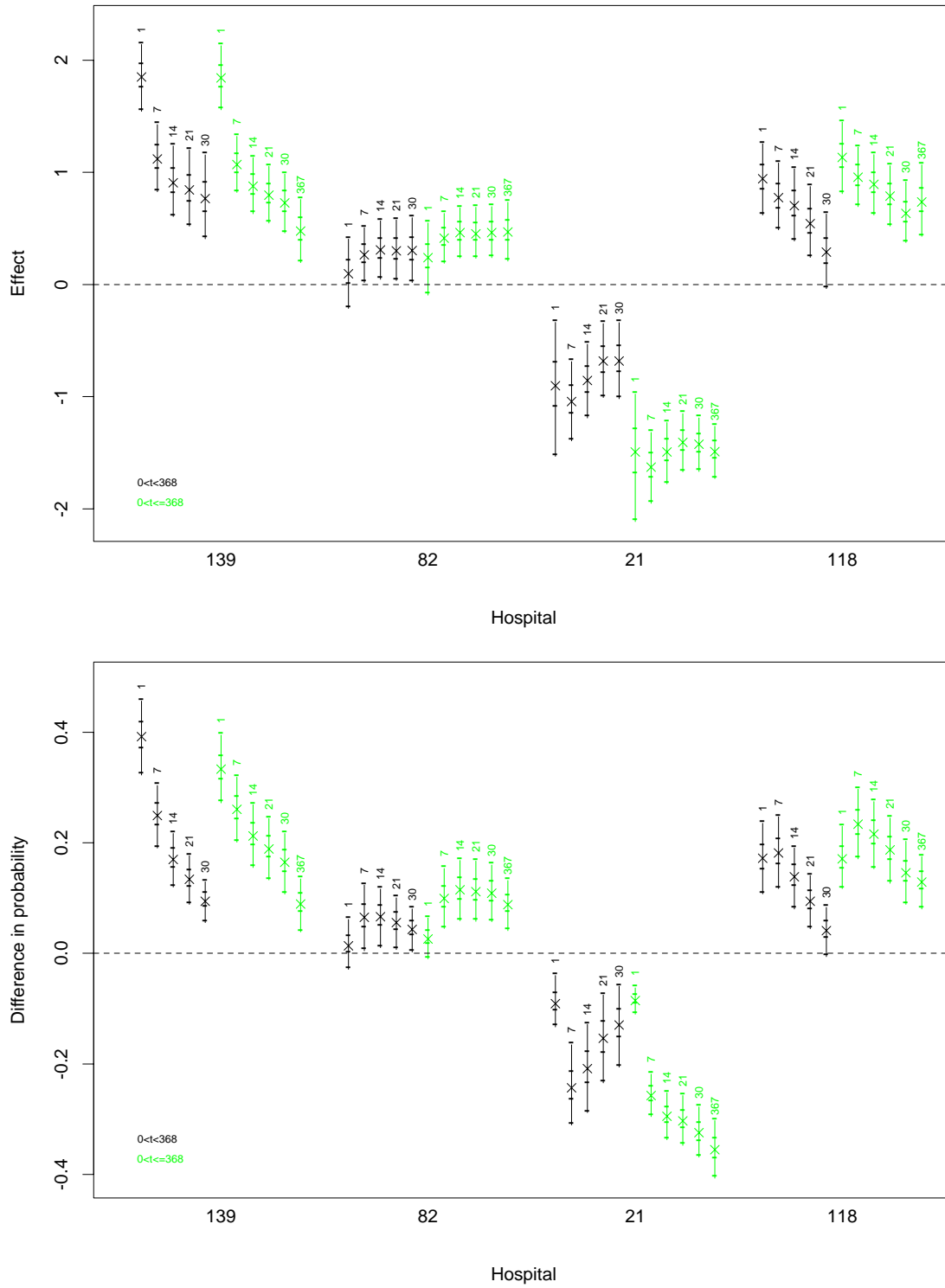


Figure 2: Intervals for hospital specific random effects ϵ_i for four selected hospitals indicated by station numbers. Graphs display the effects ϵ_i (upper frame) and the differences in implied return time probabilities relative to the baseline (lower frame).

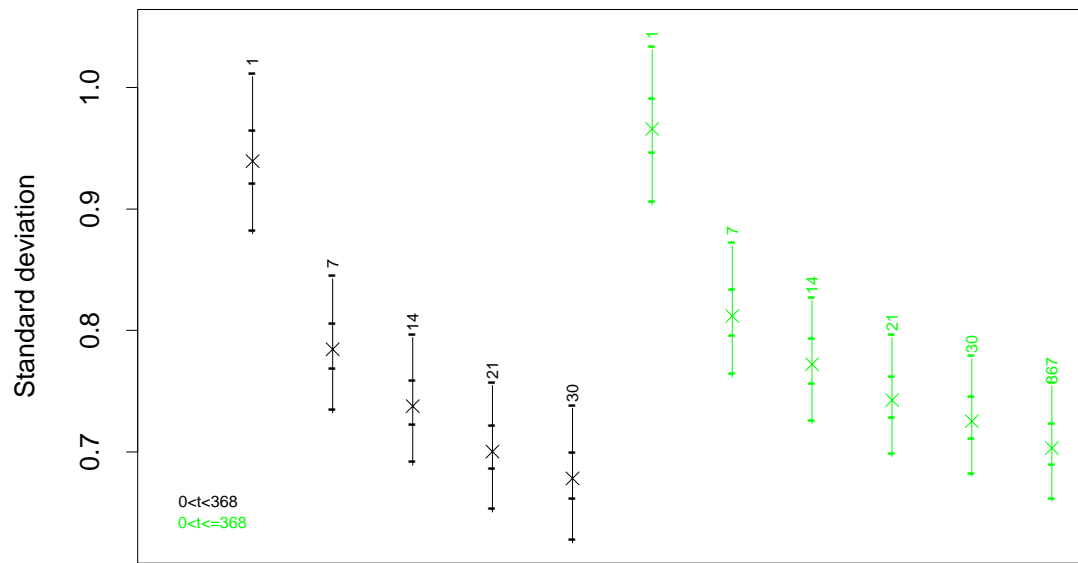


Figure 3: Intervals for the standard deviation $w(t)$ of the population distribution of hospital specific random effects ϵ_i .

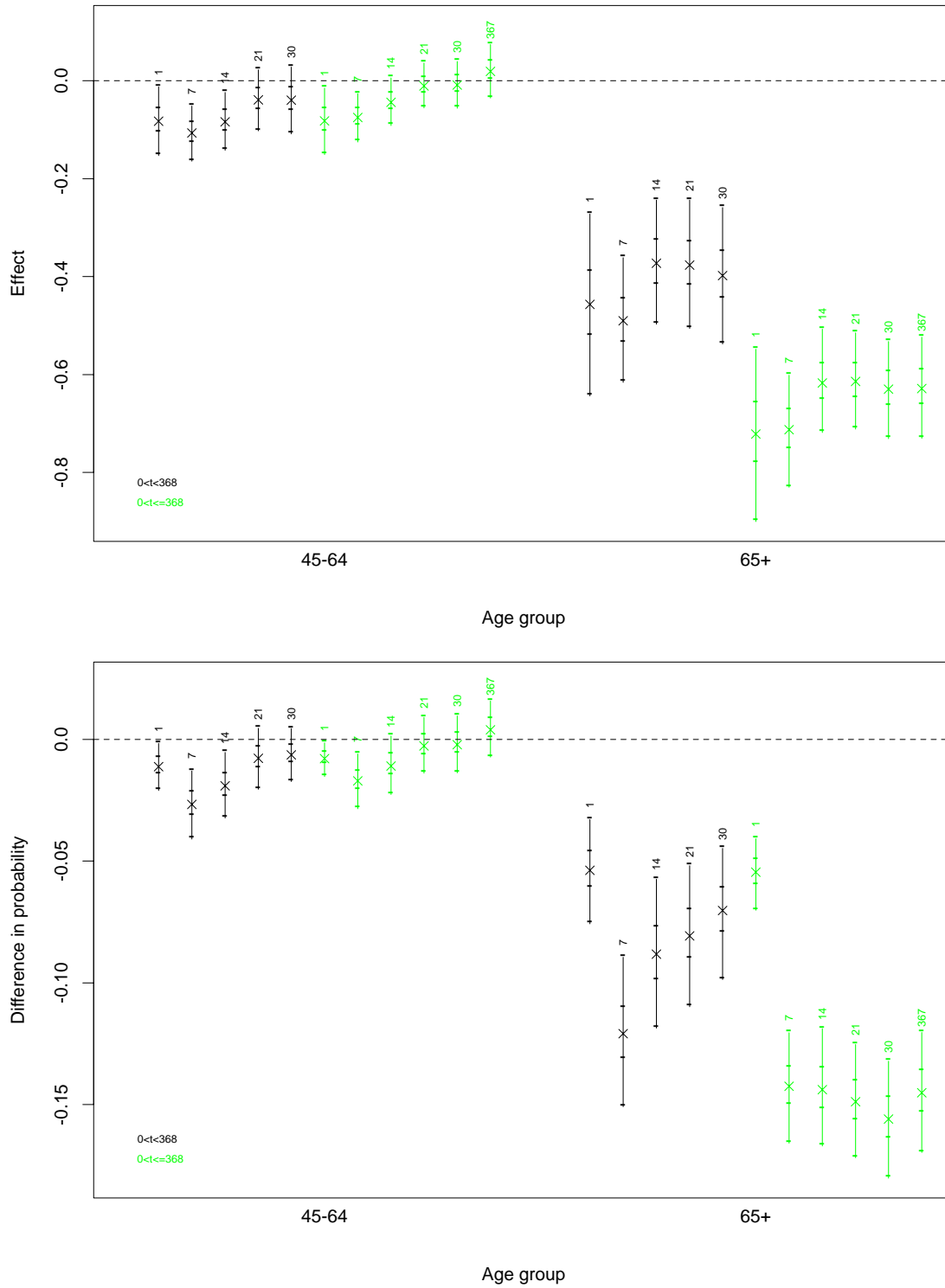


Figure 4: Intervals for Age group effects $\delta_j(t)$ (upper frame) and the differences in implied return time probabilities relative to the baseline (lower frame).

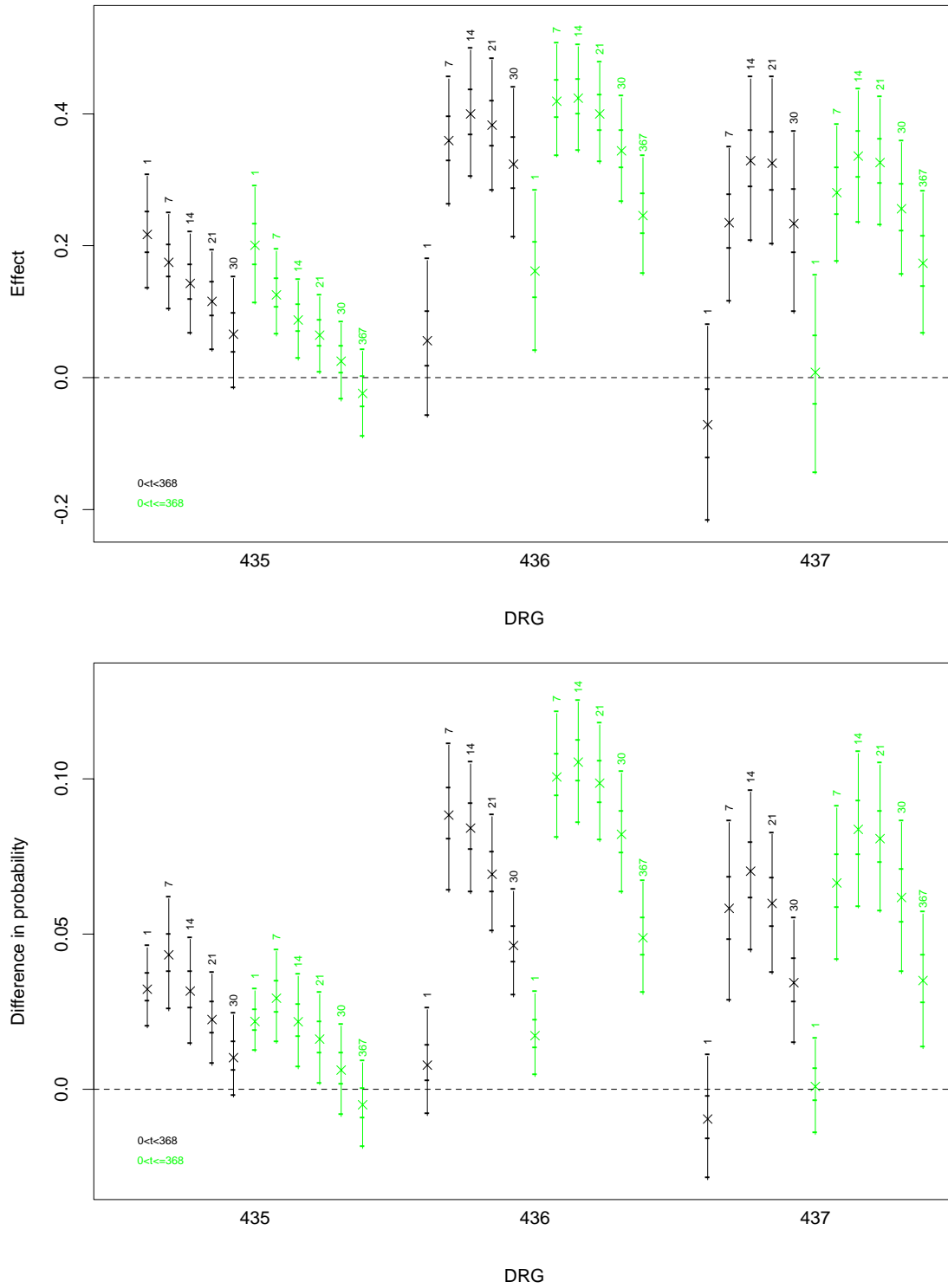


Figure 5: Intervals for DRG effects $\gamma_j(t)$ (upper frame) and the differences in implied return time probabilities relative to the baseline (lower frame).

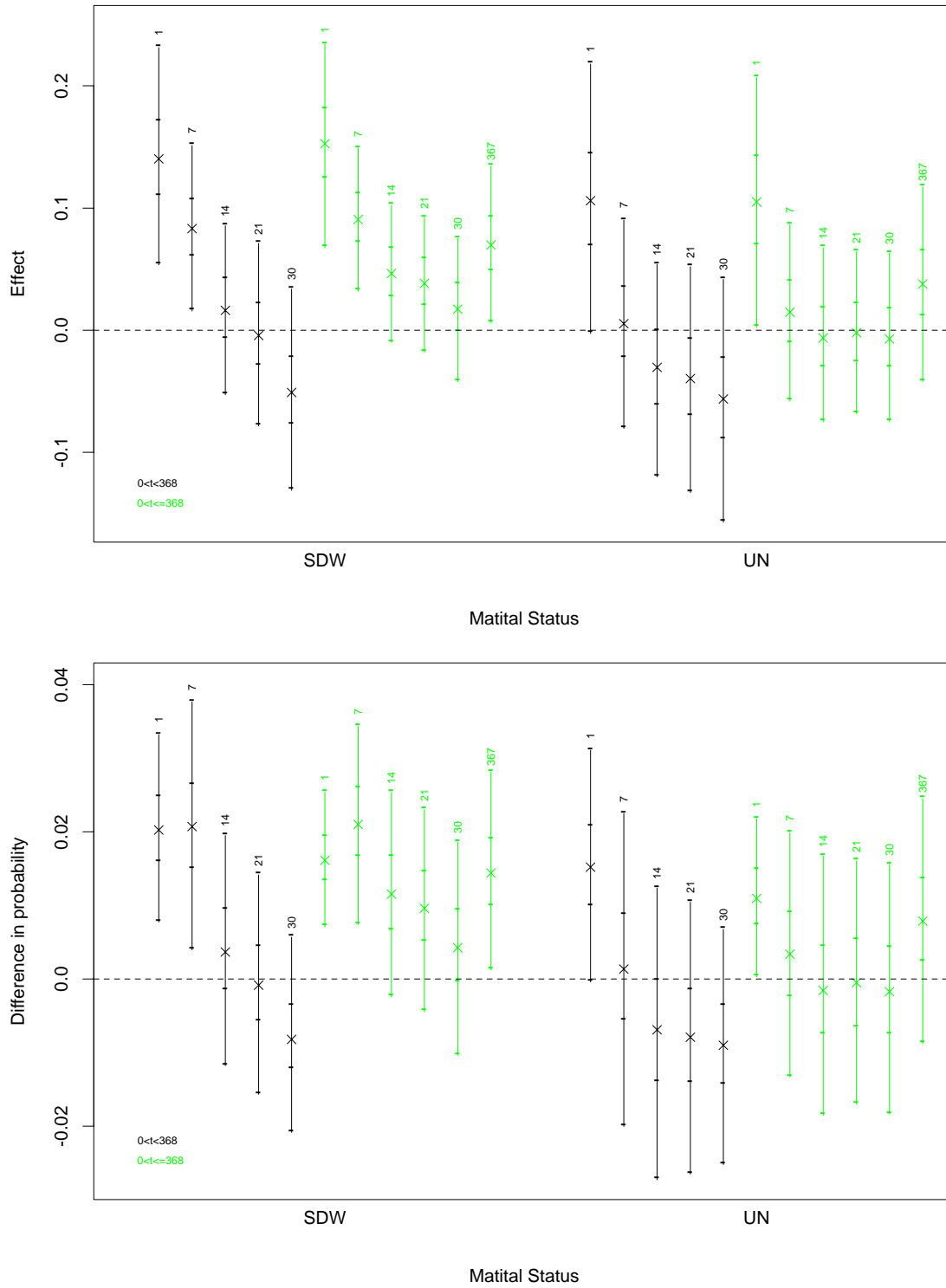


Figure 6: Intervals for Marital Status group effects $\kappa_j(t)$ (upper frame) and the differences in implied return time probabilities relative to the baseline (lower frame).

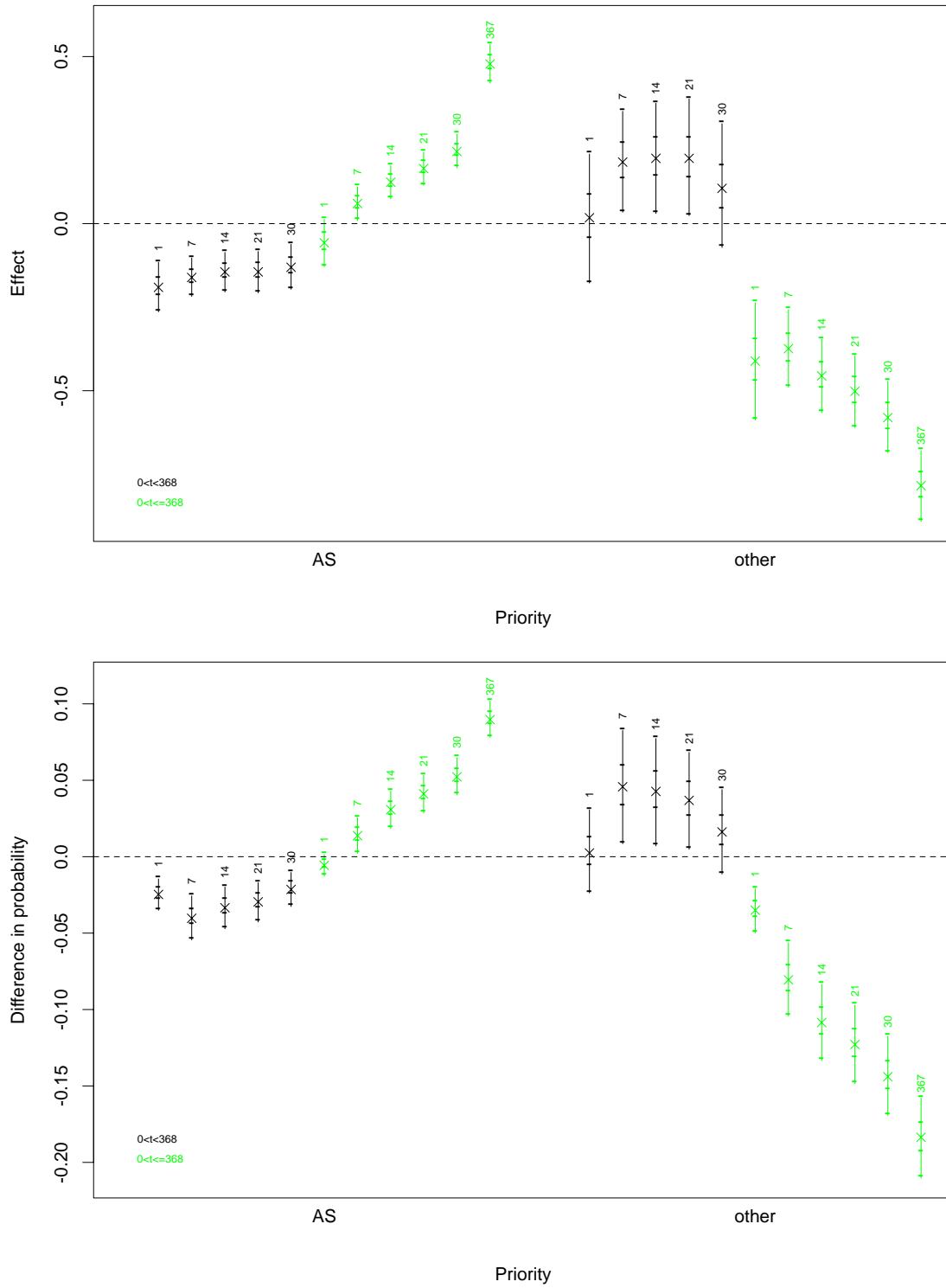


Figure 7: Intervals for Priority (Means) group effects $\eta_j(t)$ (upper frame) and the differences in implied return time probabilities relative to the baseline (lower frame).

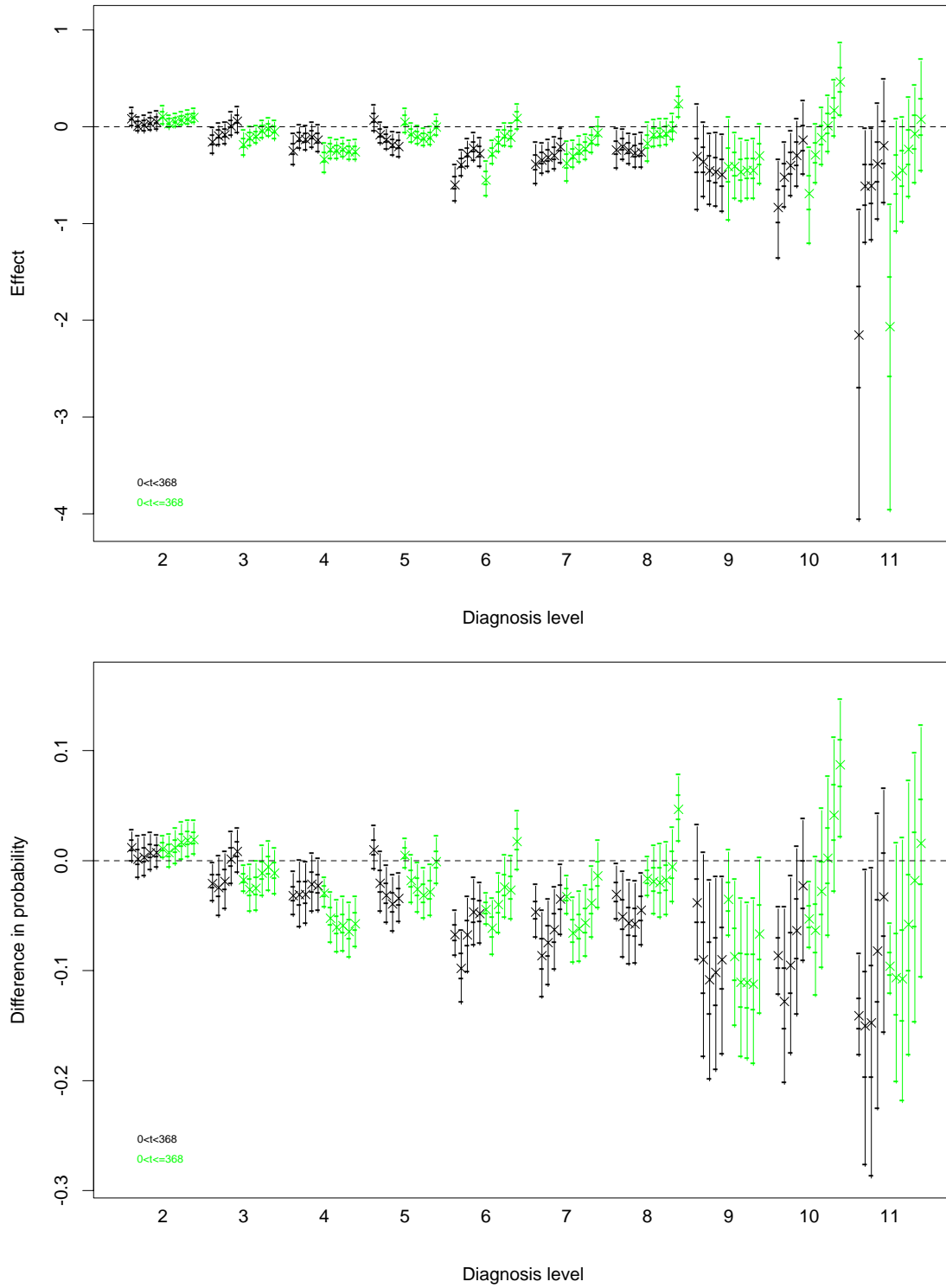


Figure 8: Intervals for Diagnosis group effects $\xi_j(t)$ (upper frame) and the differences in implied return time probabilities relative to the baseline (lower frame).

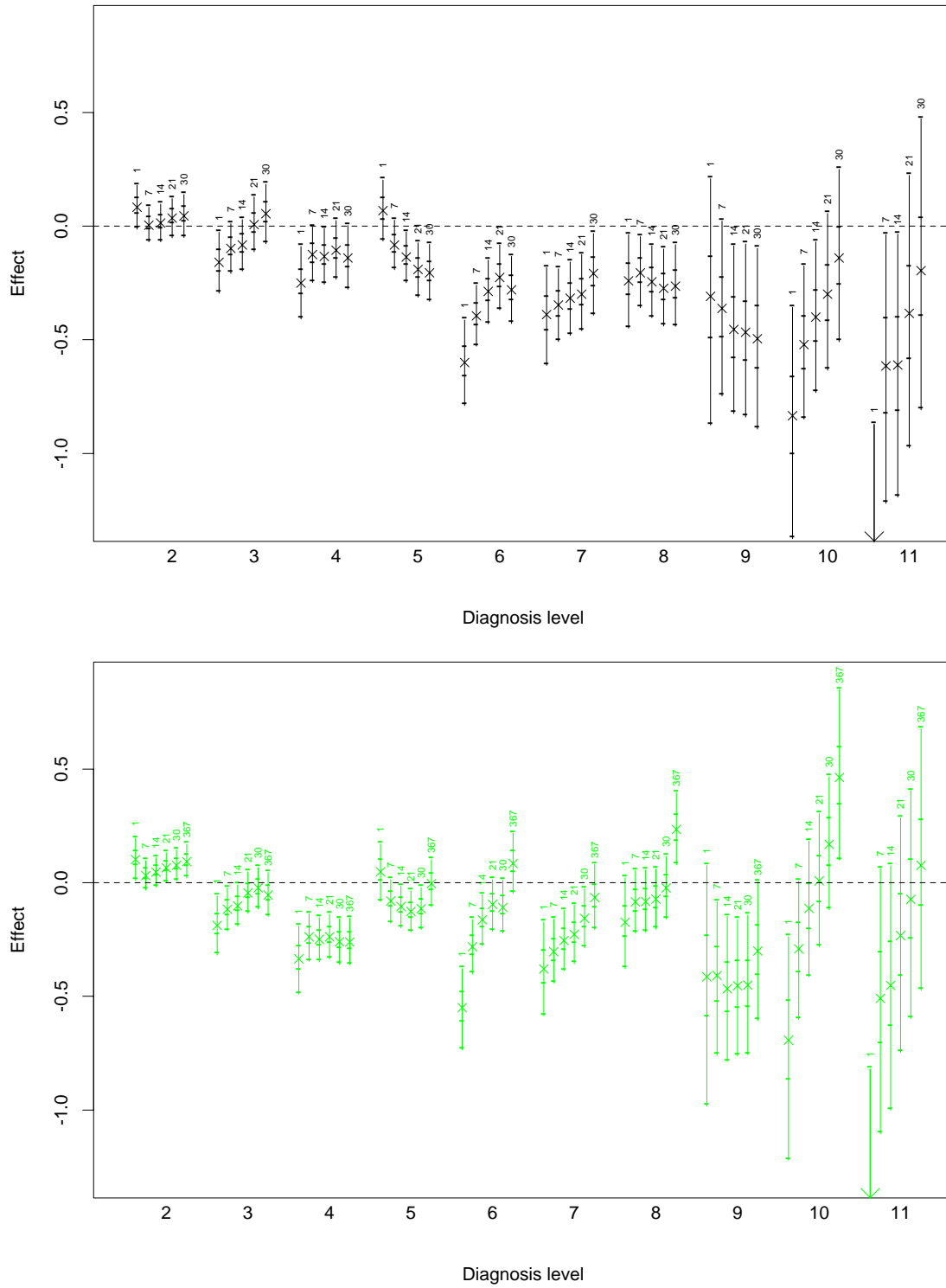


Figure 9: Intervals for Diagnosis group effects $\xi_j(t)$ based on analysis of only uncensored data (upper frame) and on all data (lower frame).

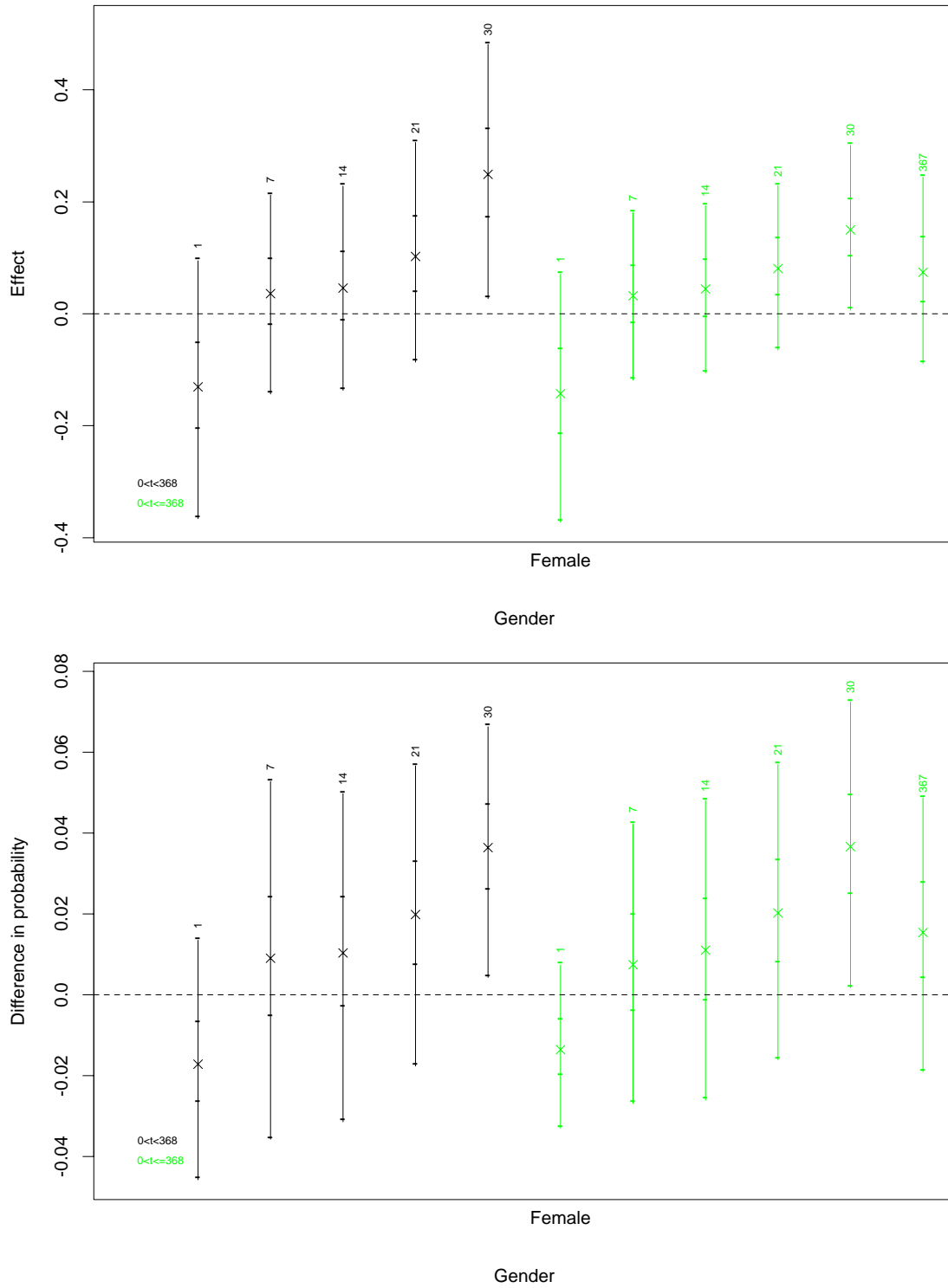


Figure 10: Intervals for Gender effects $\chi_j(t)$ (upper frame) and the differences in implied return time probabilities relative to the baseline (lower frame).

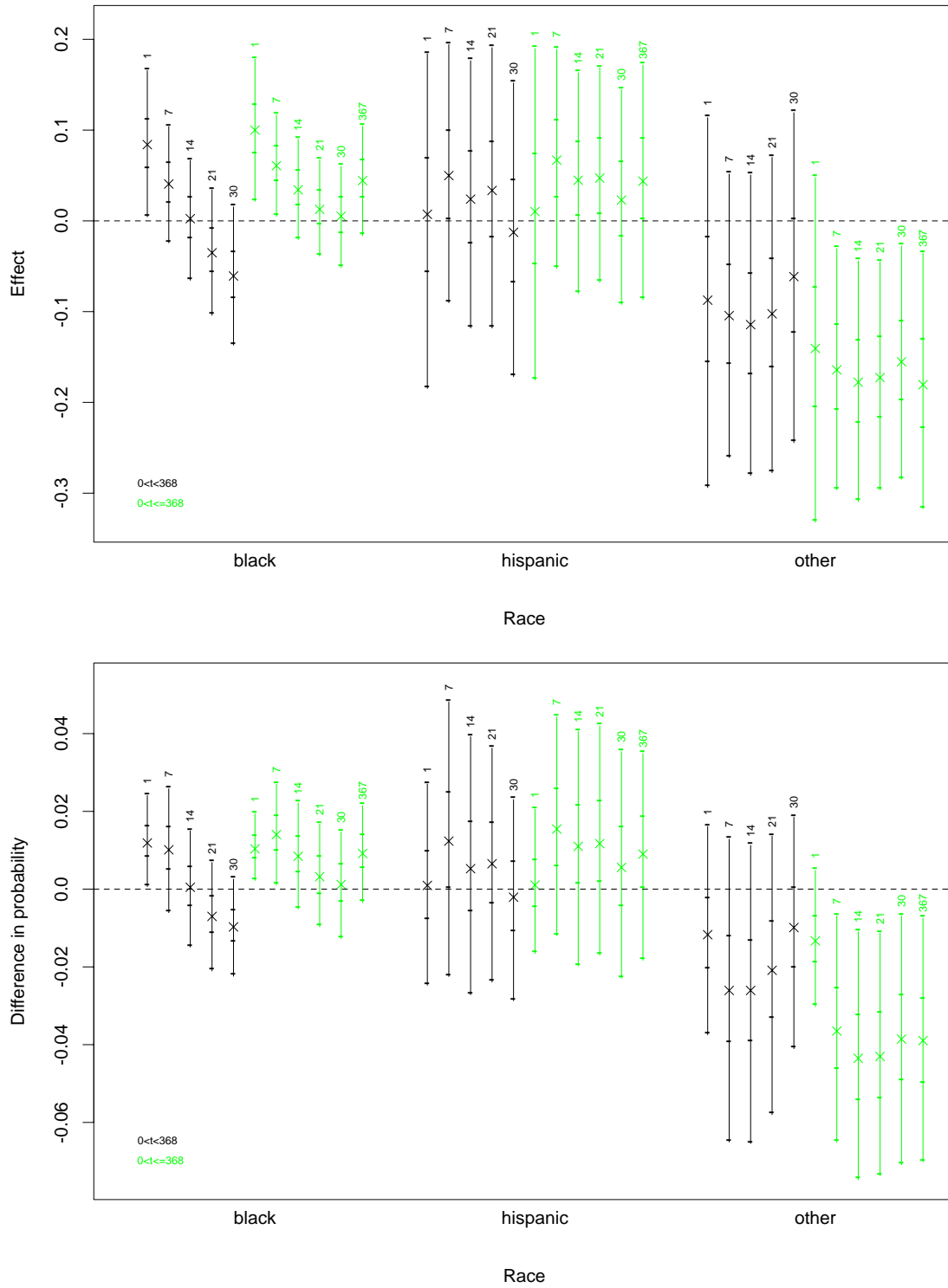


Figure 11: Intervals for Racial effects $\zeta_j(t)$ (upper frame) and the differences in implied return time probabilities relative to the baseline (lower frame).

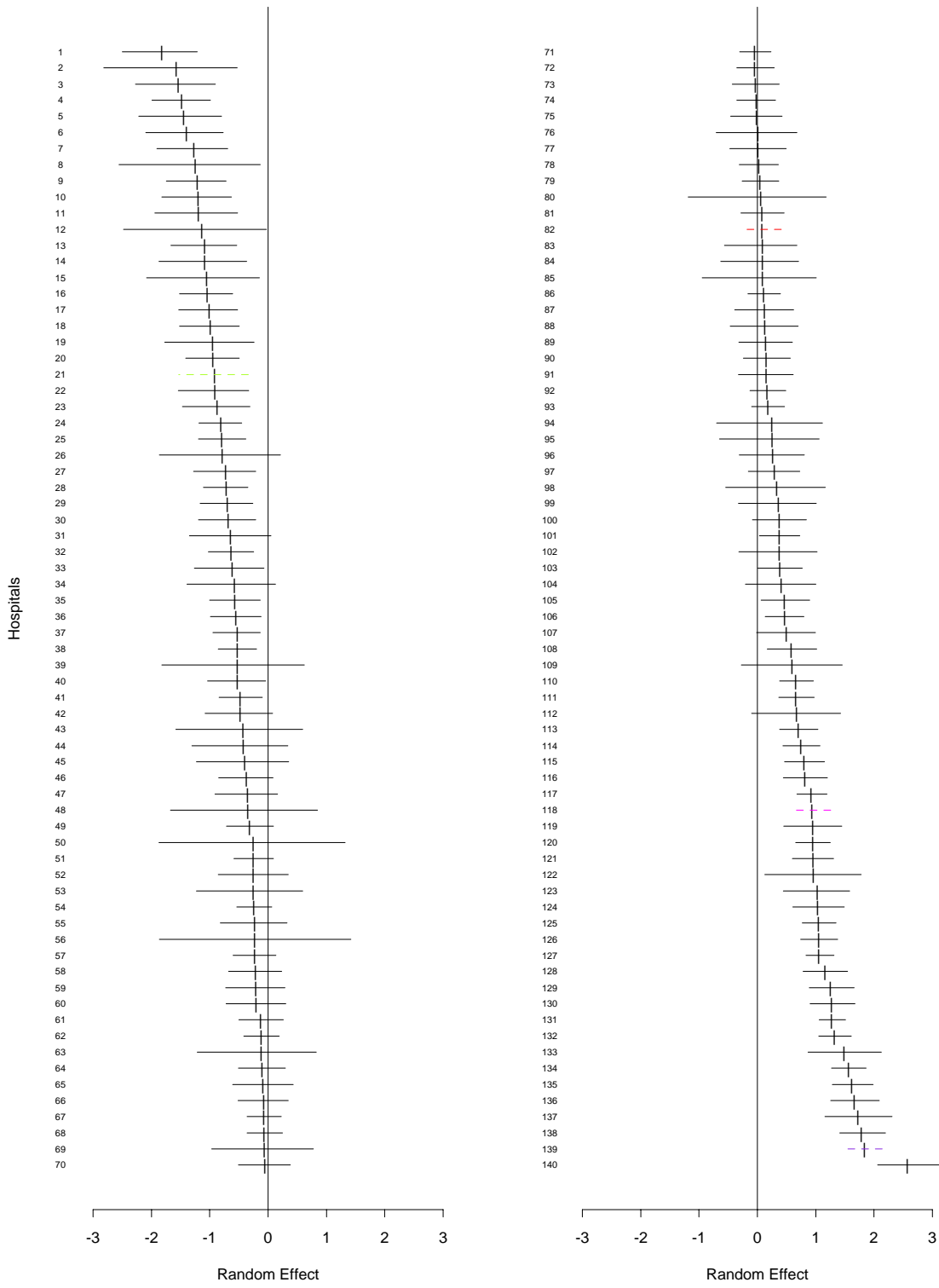


Figure 12: Intervals for Hospital random effects ϵ_i in analysis with cut-off $t = 1$. Hospitals are ordered by the posterior medians of the effects.

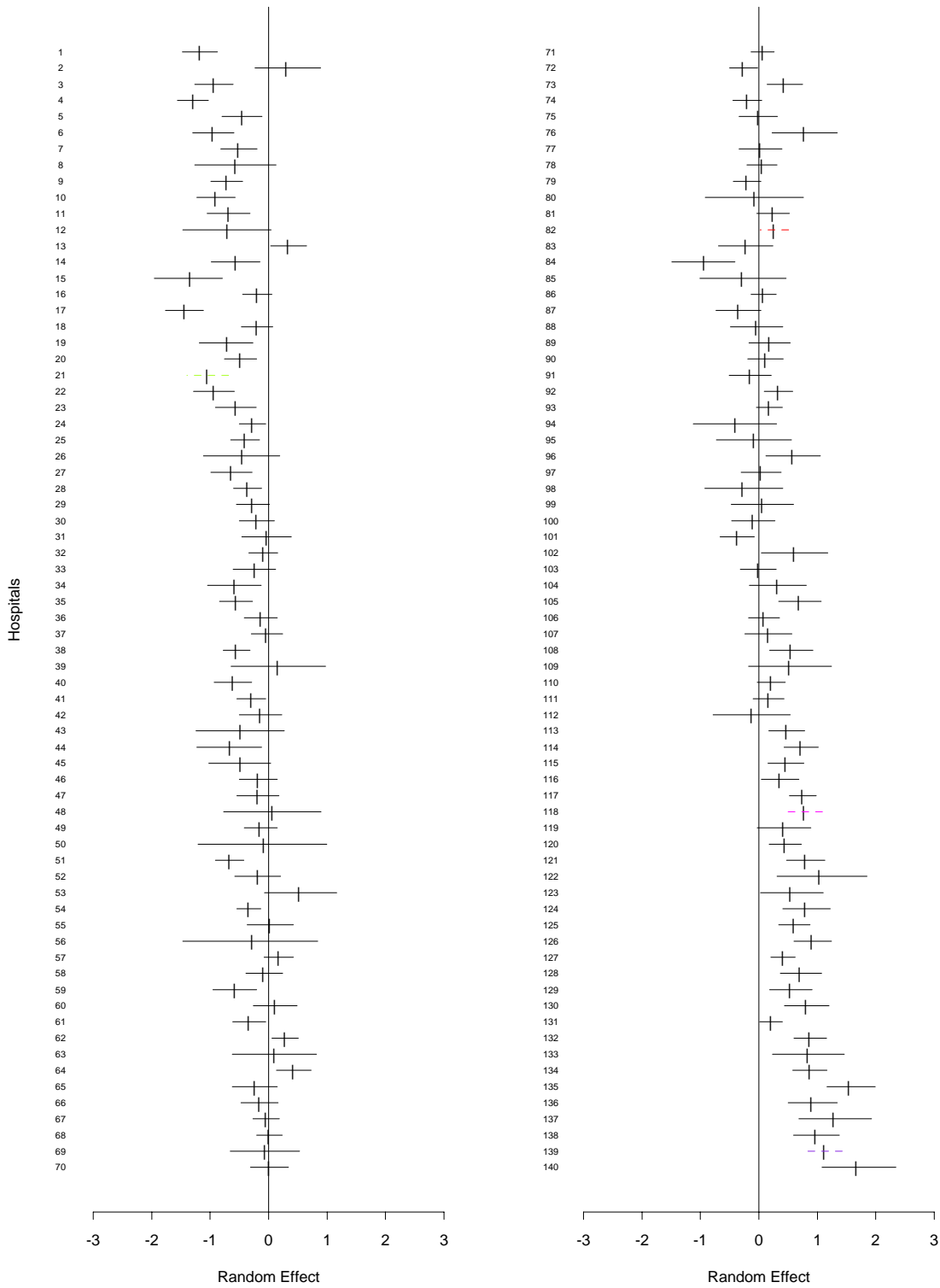


Figure 13: Intervals for Hospital random effects ϵ_i in analysis with cut-off $t \leq 7$. Hospitals are in the same order as in Figure 12.

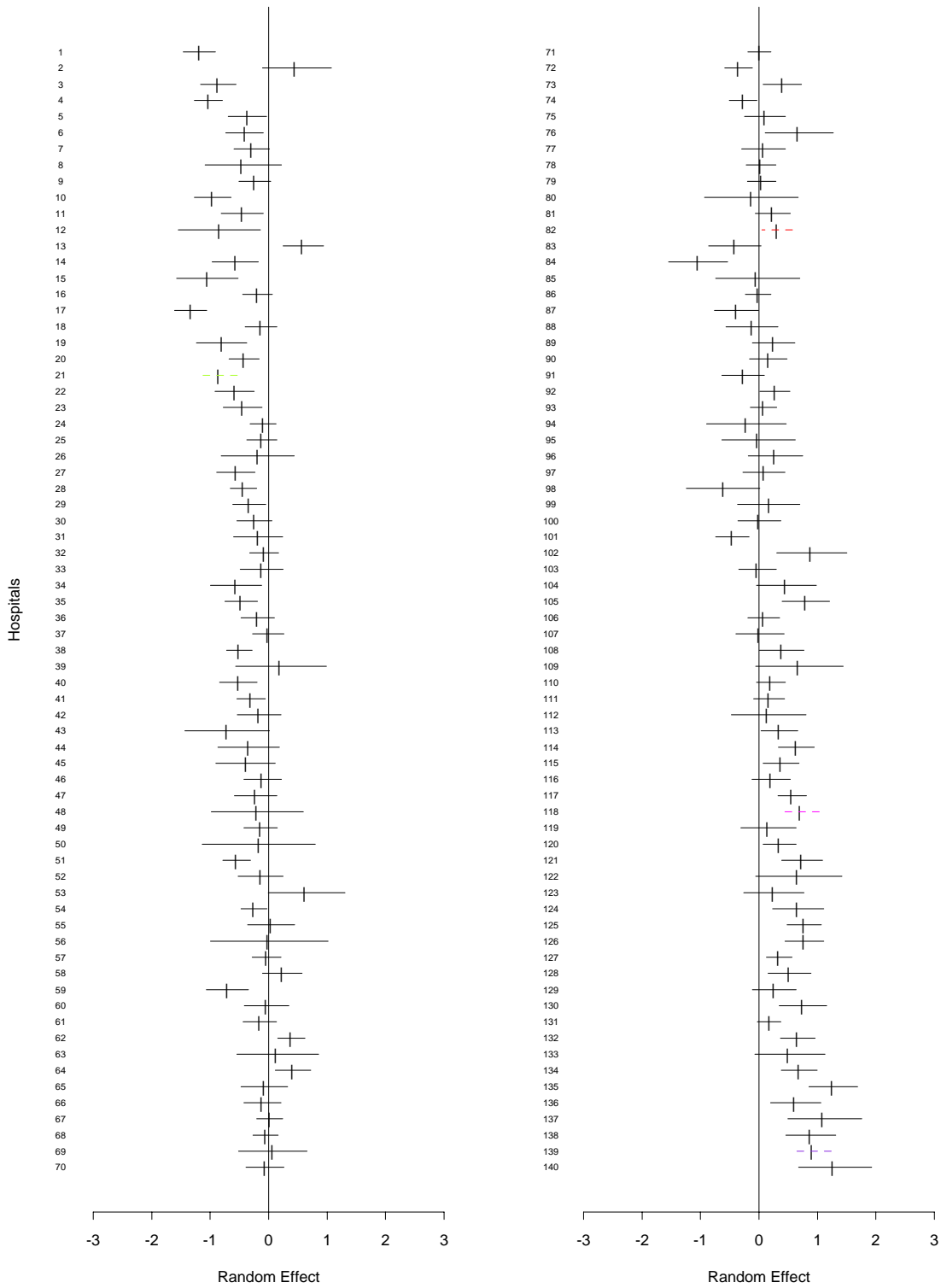


Figure 14: Intervals for Hospital random effects ϵ_i in analysis with cut-off $t \leq 14$. Hospitals are in the same order as in Figure 12.

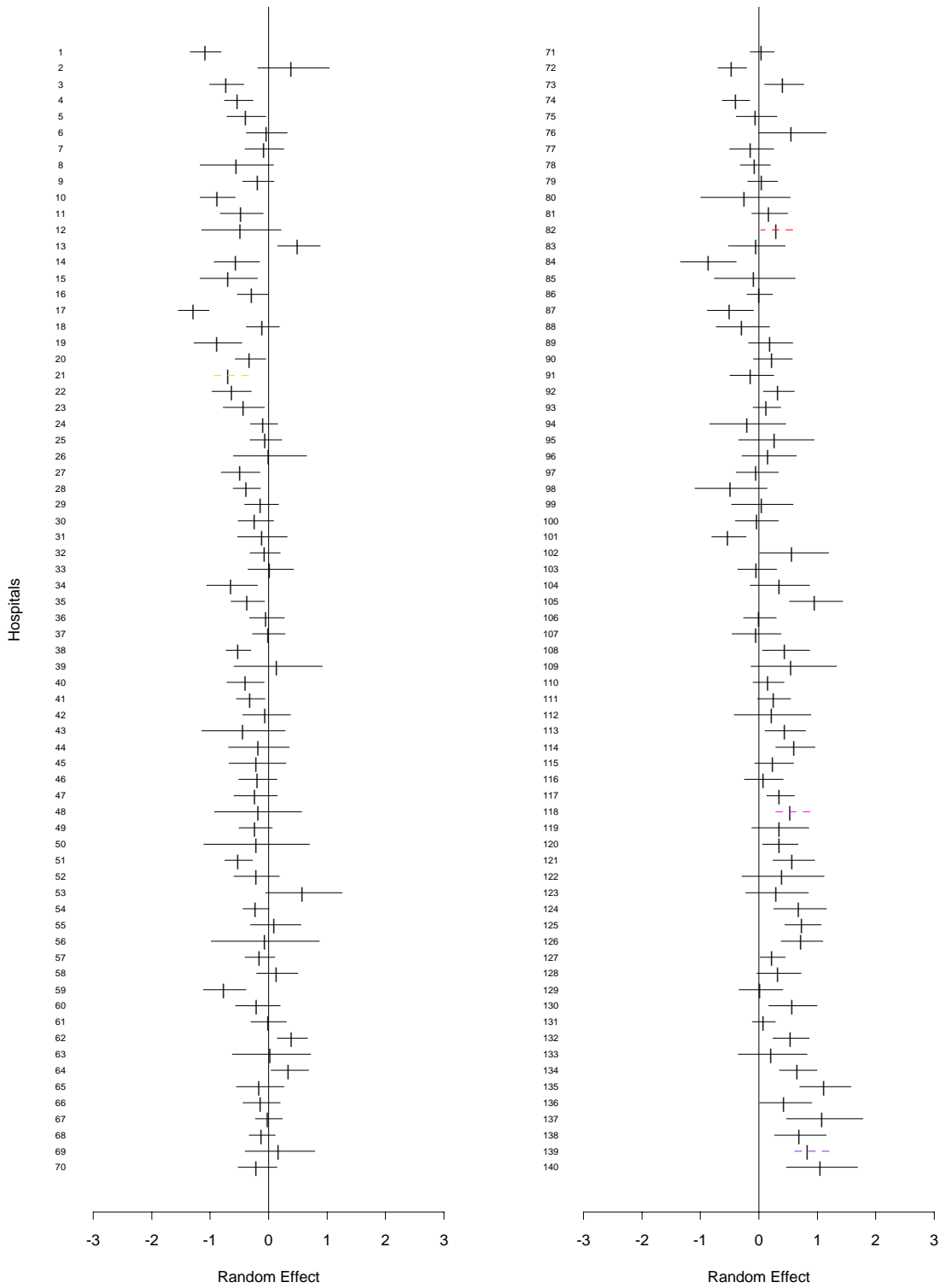


Figure 15: Intervals for Hospital random effects ϵ_i in analysis with cut-off $t \leq 21$. Hospitals are in the same order as in Figure 12.

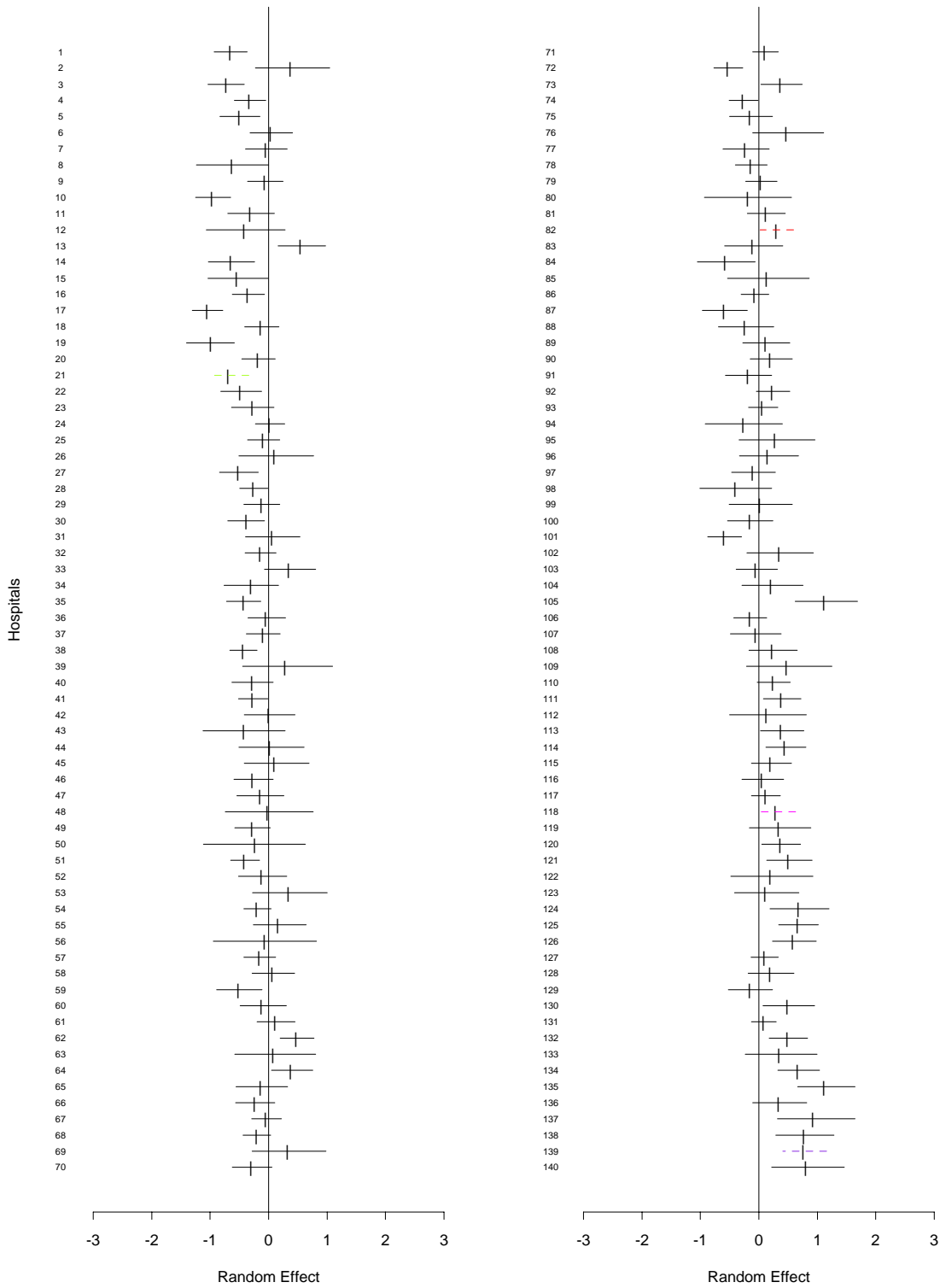


Figure 16: Intervals for Hospital random effects ϵ_i in analysis with cut-off $t \leq 30$. Hospitals are in the same order as in Figure 12.

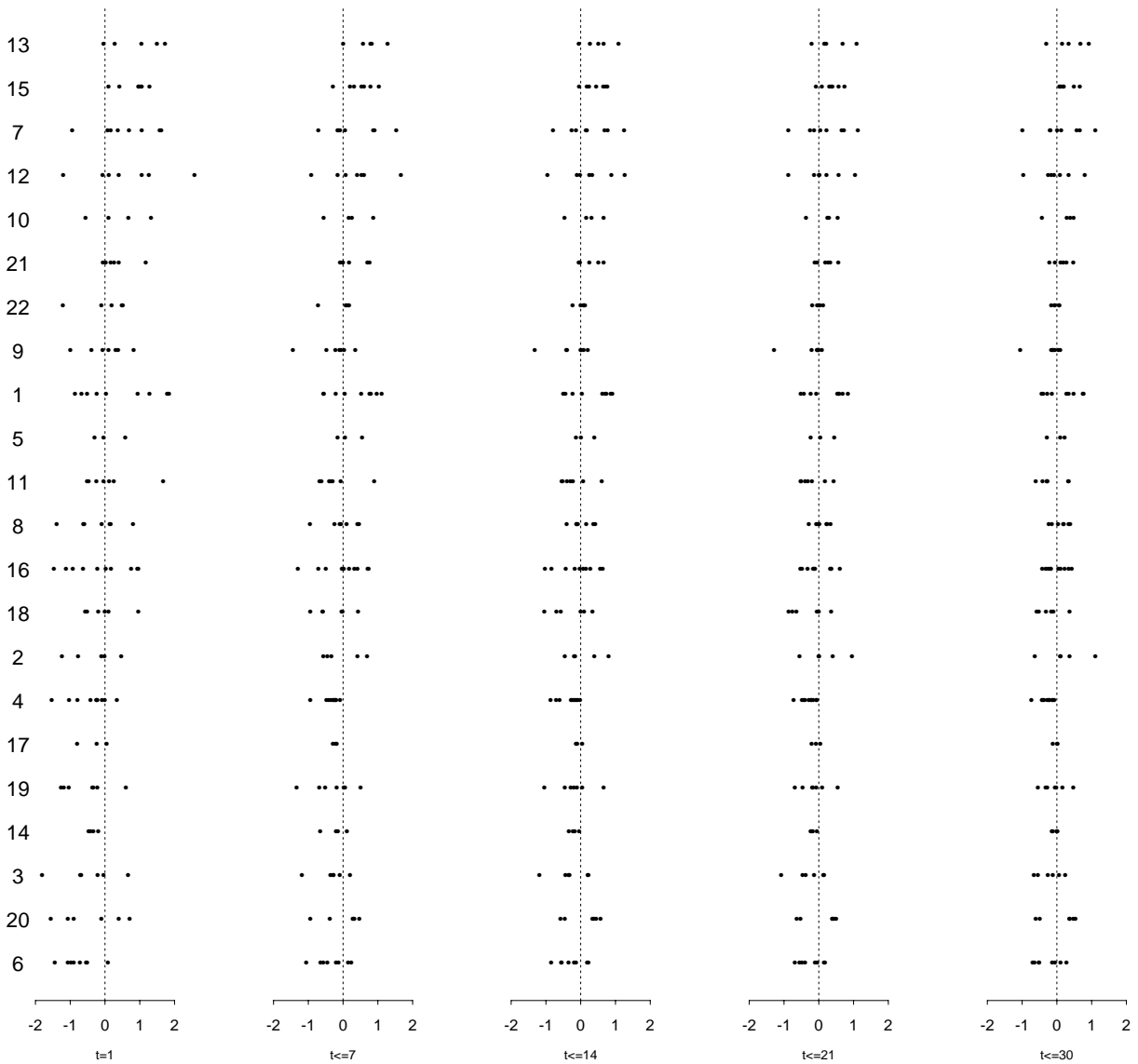


Figure 17: Posterior medians of Hospital random effects $e_i(t)$ classified by VISN Region, based on analysis of only uncensored data. The 22 regions are ordered by the posterior median of the estimated effects from the analysis with cut-off $t = 1$ using only uncensored data.

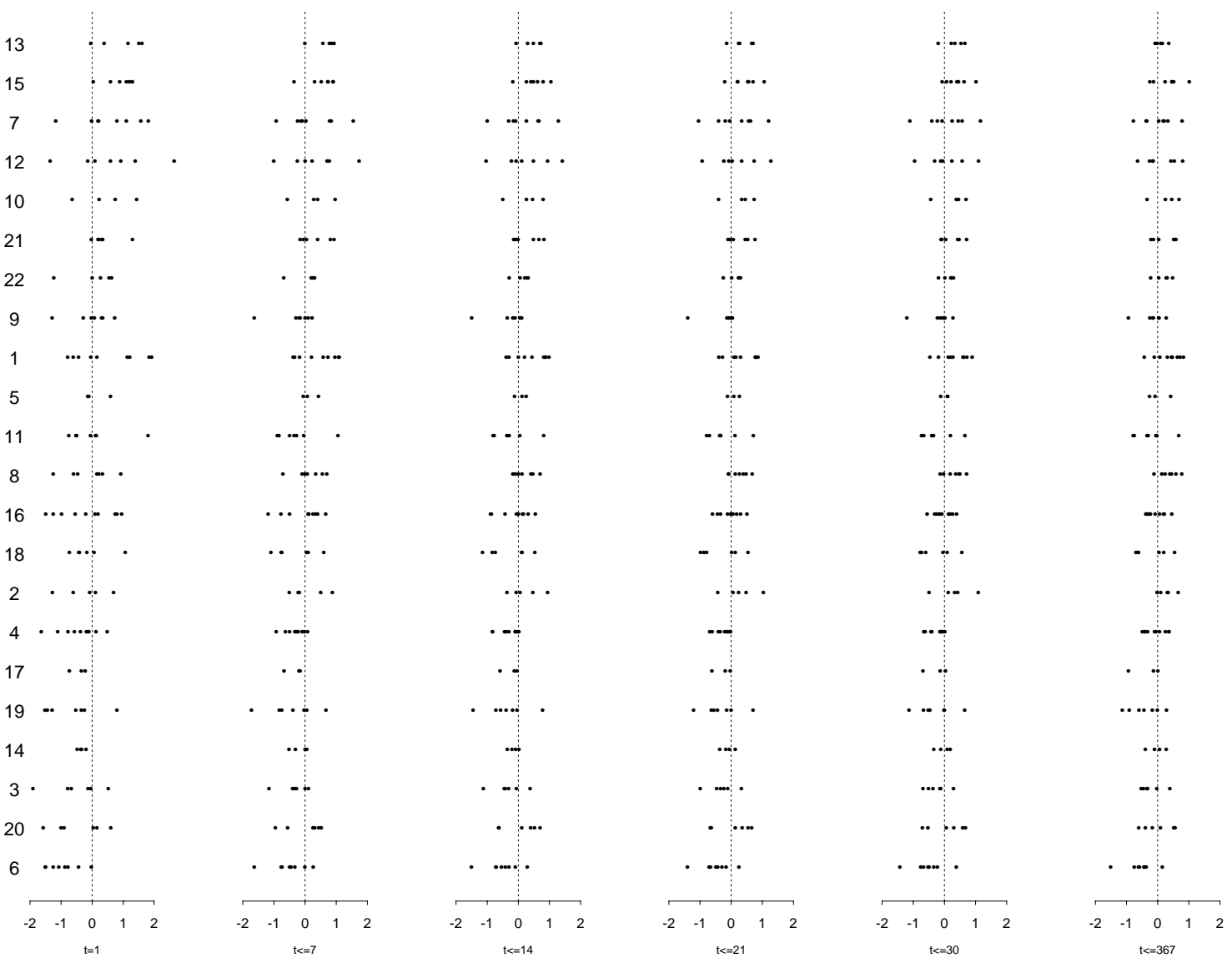


Figure 18: Posterior medians of Hospital random effects $e_i(t)$ classified by VISN region, now based on analysis of all the data. The Regions are ordered as in Figure 17.

10 Graphical Summaries of Analysis: 10 Year Study

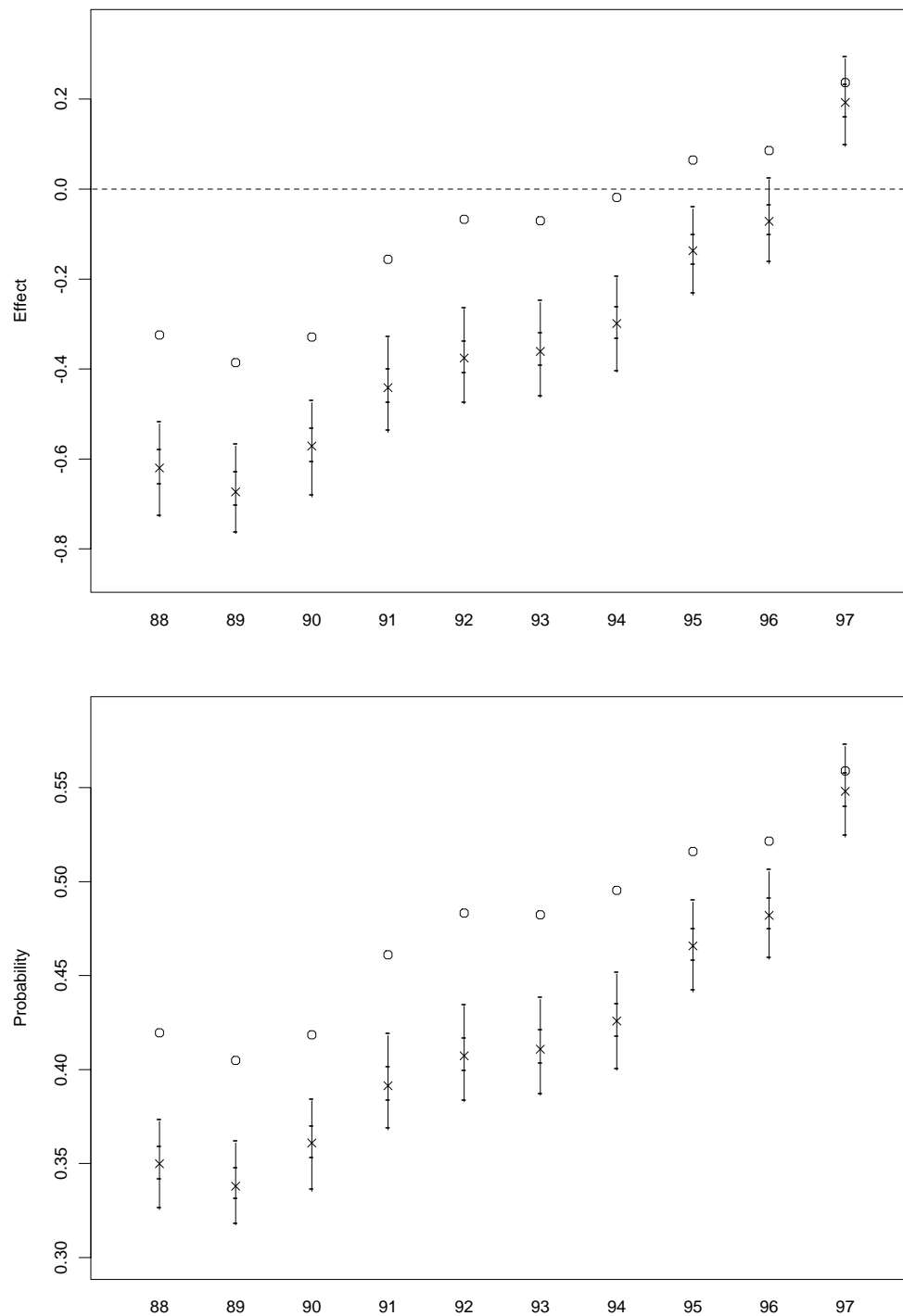


Figure 19: Intervals for baseline duration model parameters $\beta_{0,r}$ (upper frame) and the implied baseline return time probabilities $p_{0,r}$ (lower frame) over years $r = 1988, \dots, 1997$. The points correspond to the overall proportions returning within 30 days from the raw data.

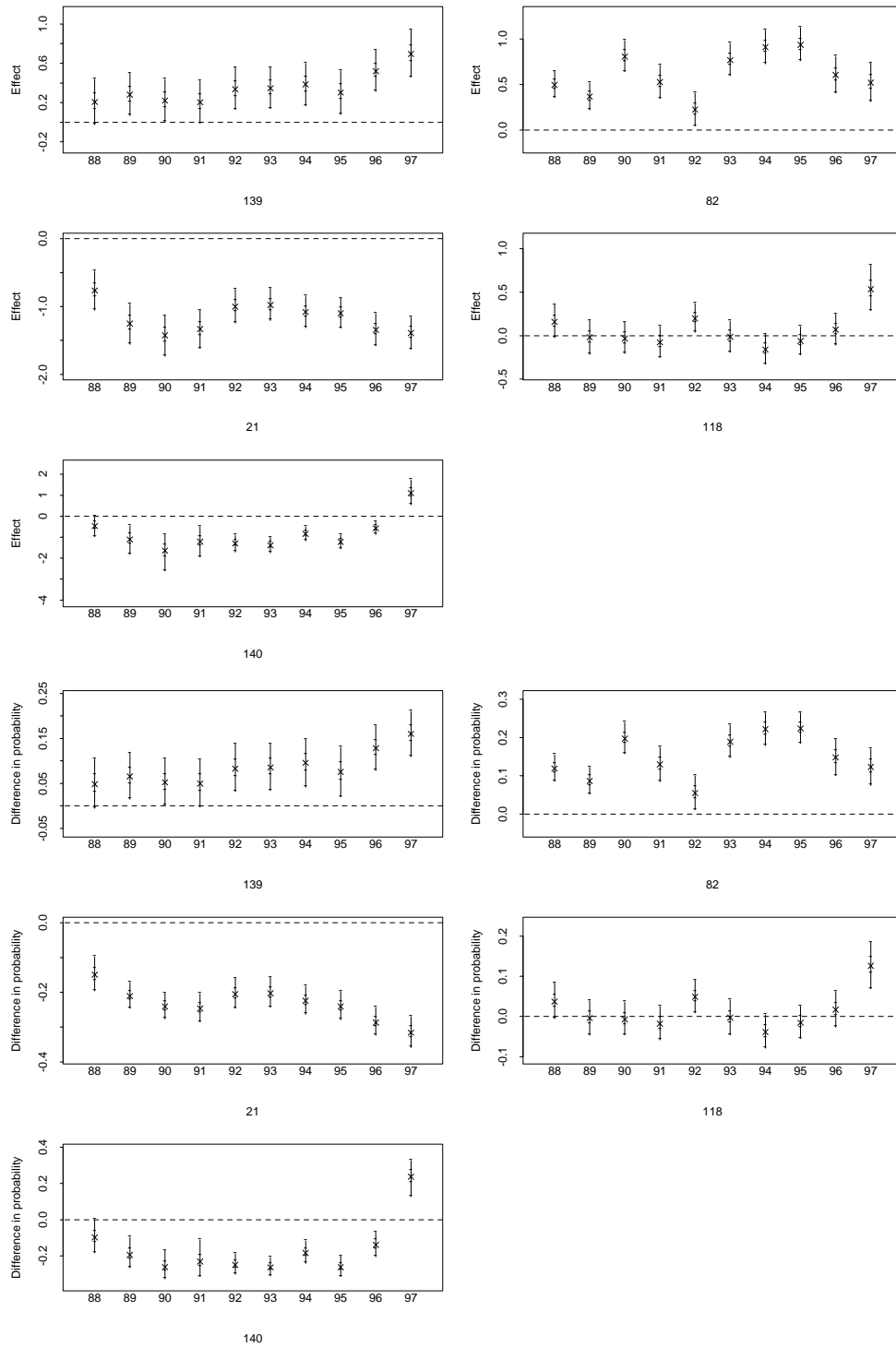


Figure 20: Intervals for hospital-specific random effects $\epsilon_{i,r}$ over years $r = 1988, \dots, 1997$ in multi-year analysis with cut-off $t = 30$ days. Intervals are given for the four hospitals selected in the earlier, single year analysis, as labelled, plus one additional hospital (#140). Graphs display the effects $\epsilon_{i,r}$ (five upper frames) and the differences in implied return time probabilities relative to the baseline (five lower frames).

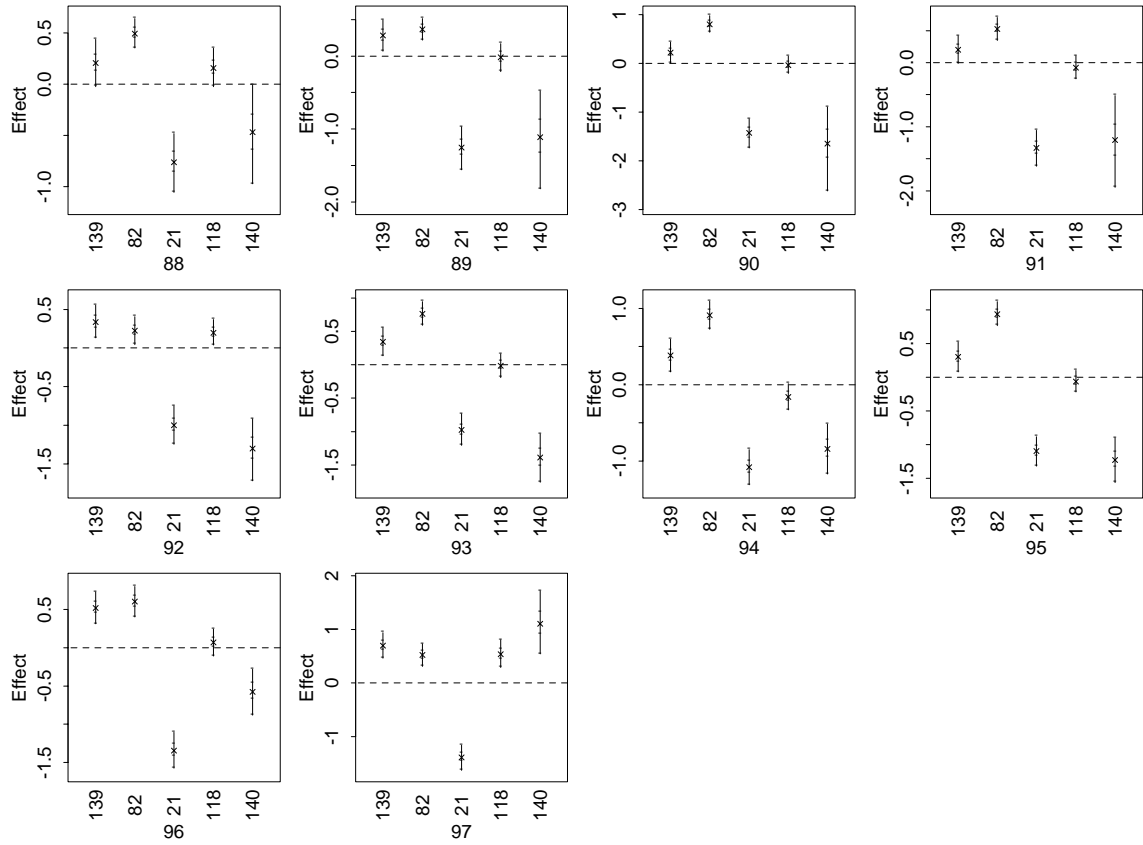


Figure 21: Intervals for hospital specific random effects $\epsilon_{i,r}$ comparing the five selected hospitals within each of the 10 years.

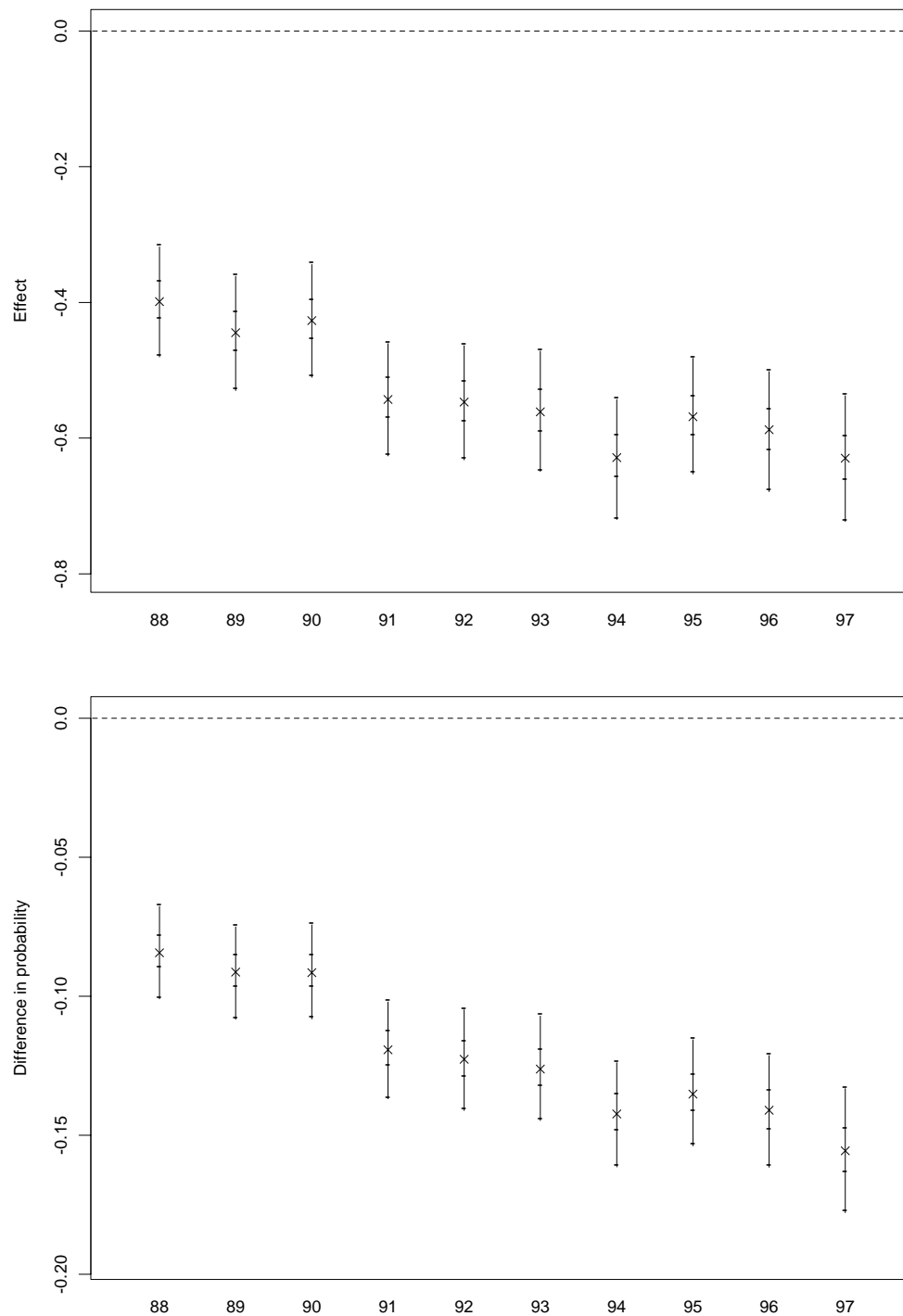


Figure 22: Intervals for Age group effects $\delta_{2,r}$ (upper frame) and the differences in implied baseline return time probabilities relative to the baseline (lower frame) over years $r = 1988, \dots, 1997$ in multi-year analysis with cut-off $t = 30$ days.

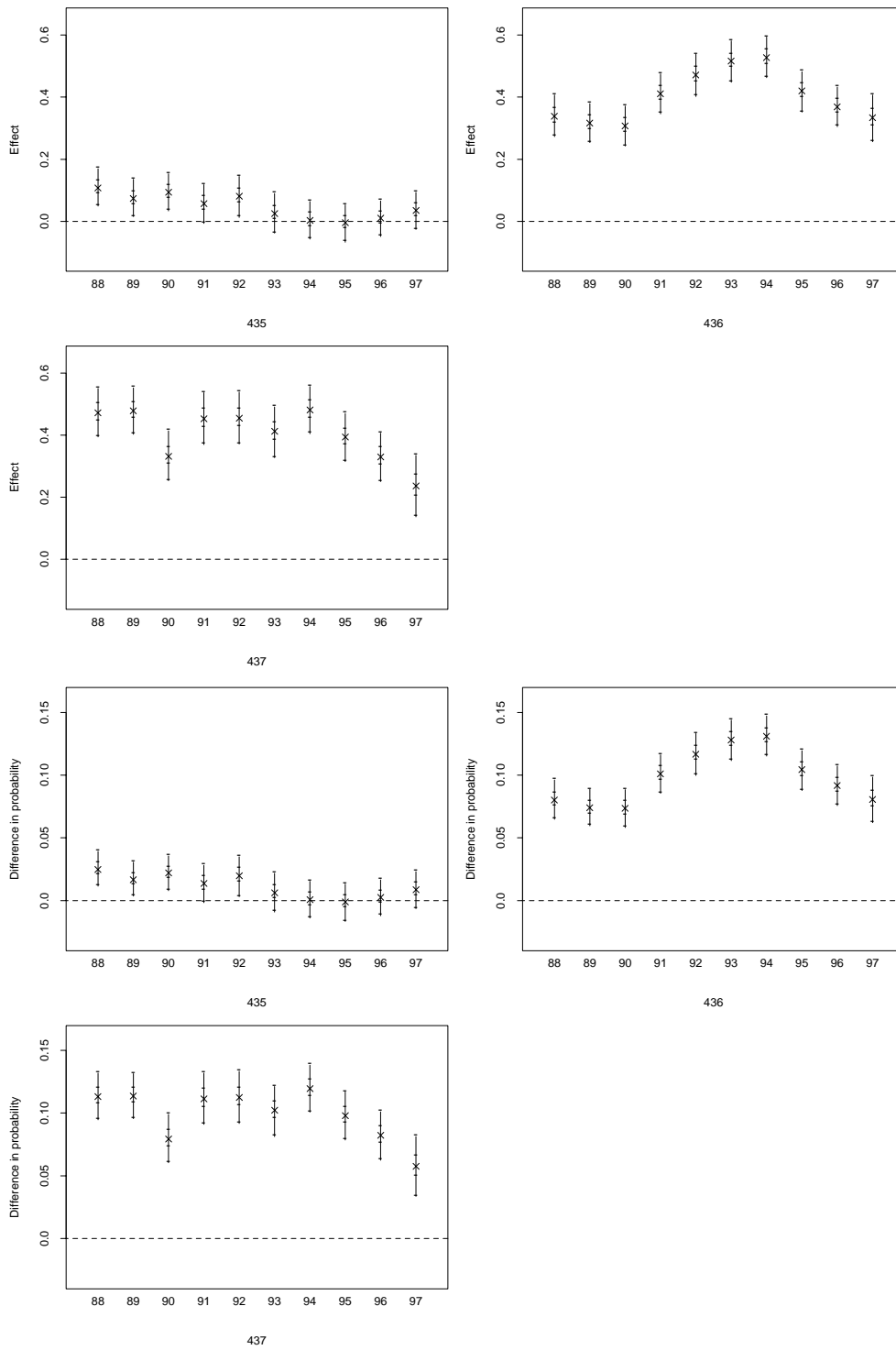


Figure 23: Intervals for DRG group effects $\gamma_{j,r}$ (three upper frames) and the differences in implied baseline return time probabilities relative to the baseline (three lower frames) over years $r = 1988, \dots, 1997$.

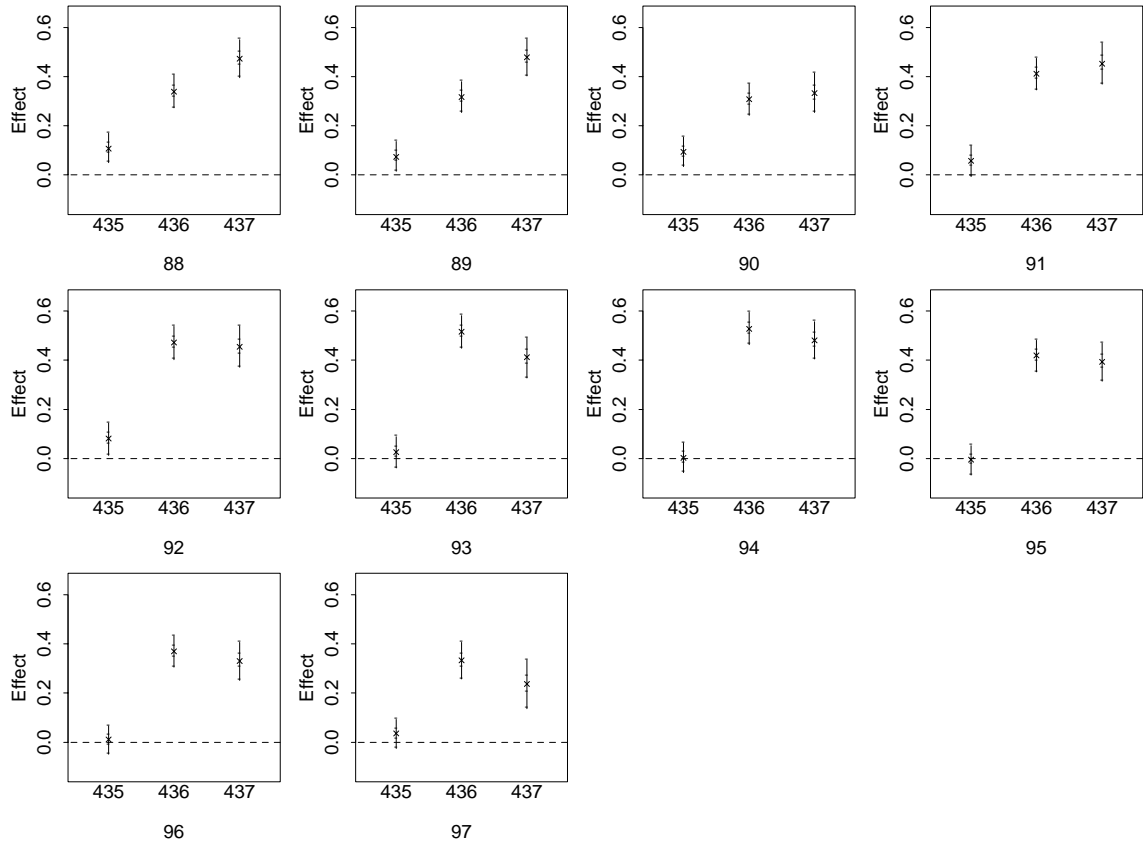


Figure 24: Intervals for DRG group effects $\gamma_{j,r}$ across DRG categories j , displayed for each of the 10 years.

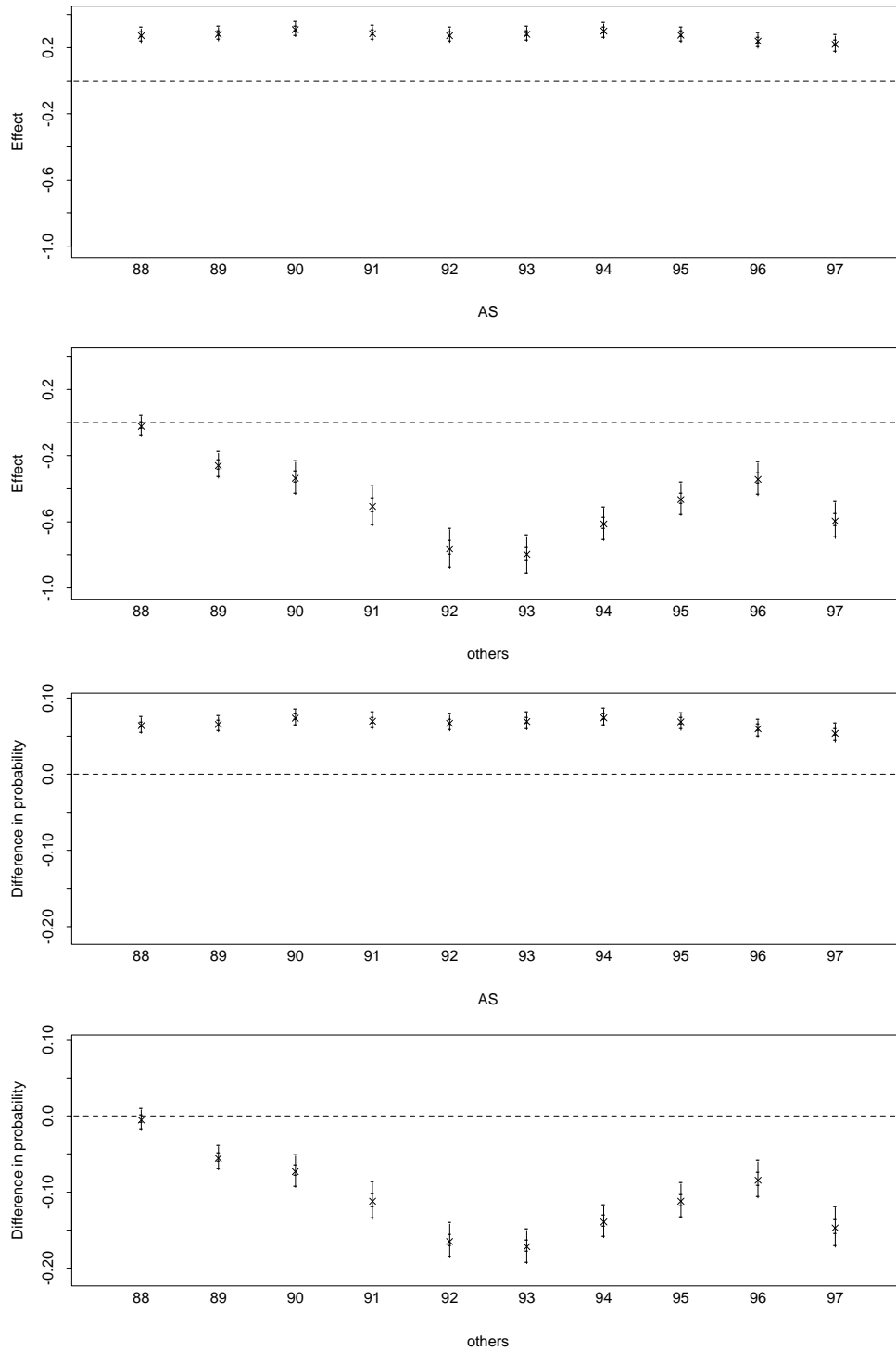


Figure 25: Intervals for Priority (Means) group effects $\eta_{j,r}$ (two upper frames) and the differences in implied baseline return time probabilities relative to the baseline (two lower frames) over years $r = 1988, \dots, 1997$.

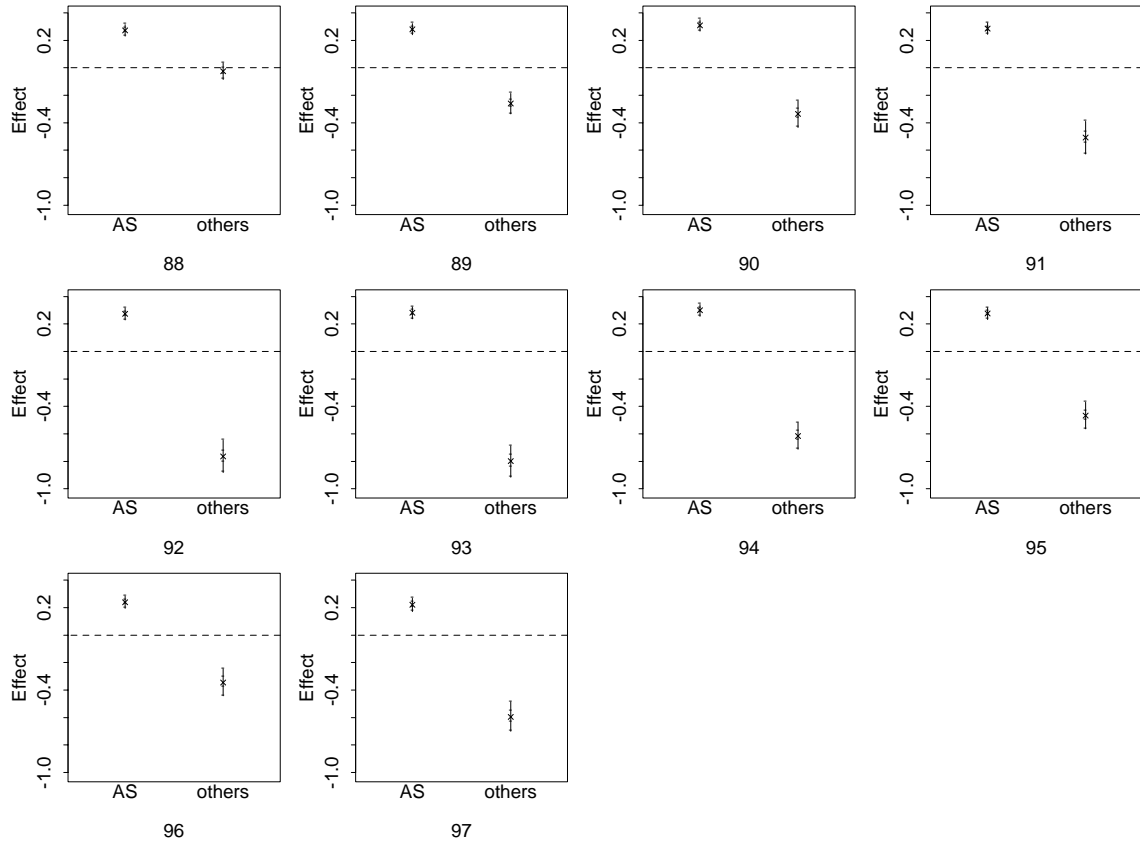


Figure 26: Intervals for Priority (Means) group effects $\eta_{j,r}$ across categories, displayed for each of the 10 years.

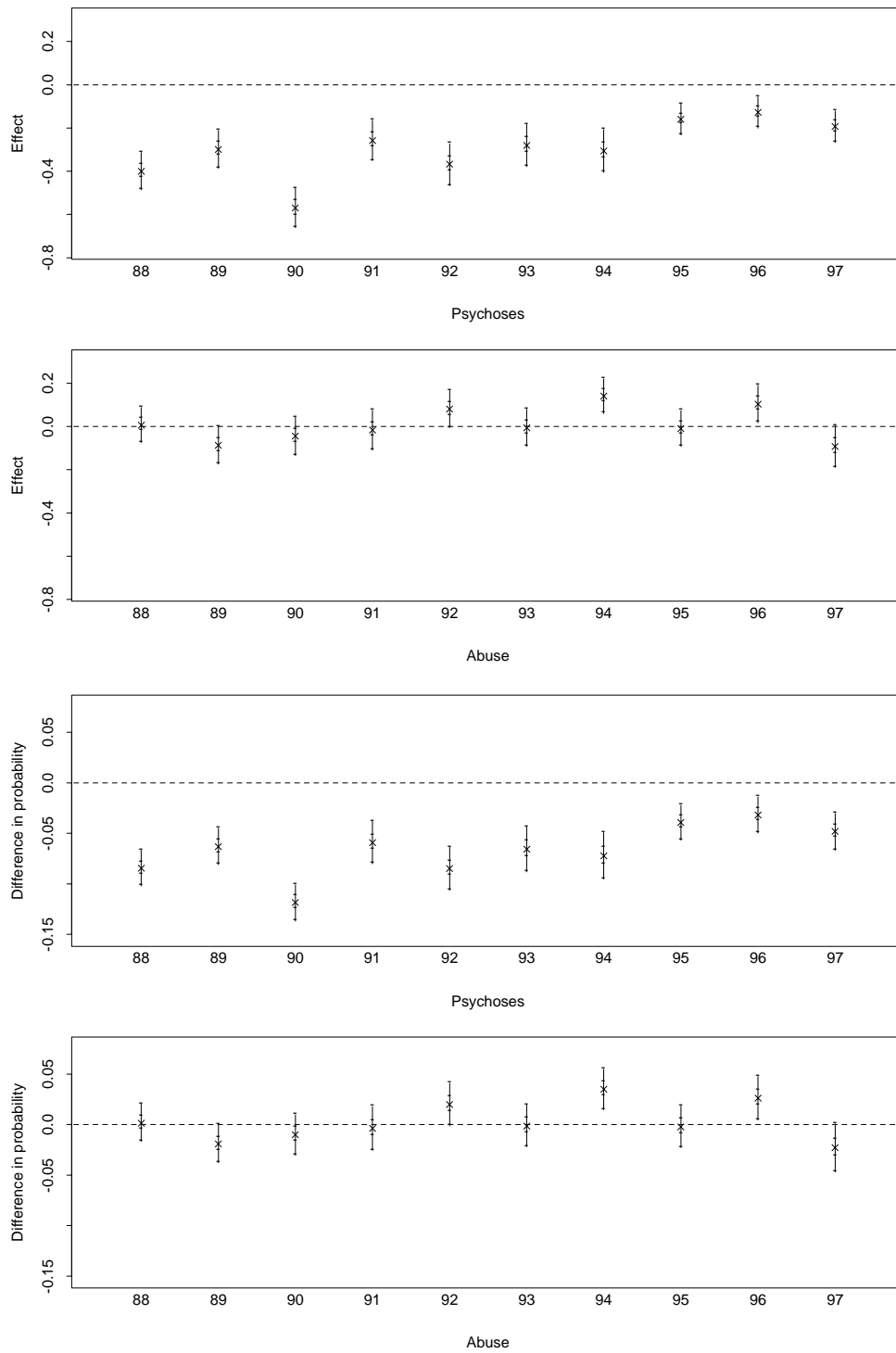


Figure 27: Intervals for Diagnosis group effects $\xi_{j,r}$ (two upper frames) and the differences in implied baseline return time probabilities relative to the baseline (two lower frames) over years $r = 1988, \dots, 1997$.

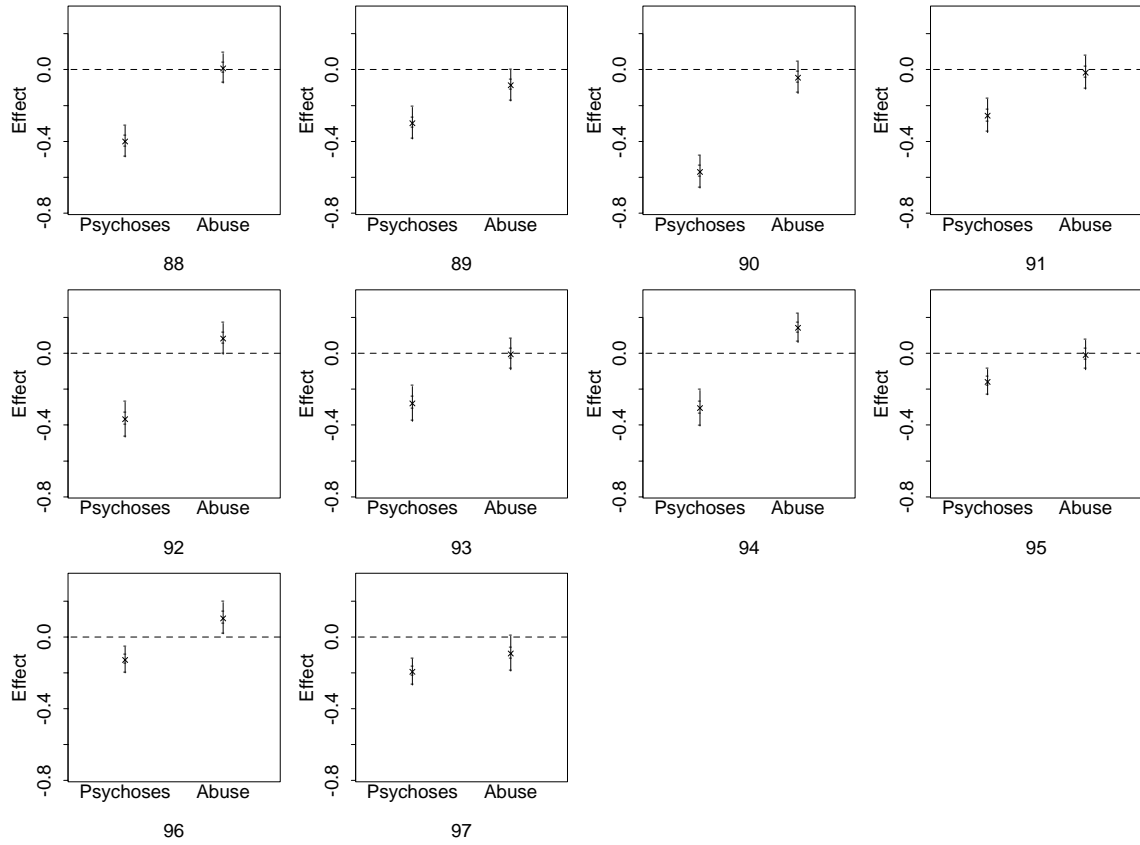


Figure 28: Intervals for Diagnosis group effects $\xi_{j,r}$ across categories, displayed for each of the 10 years.

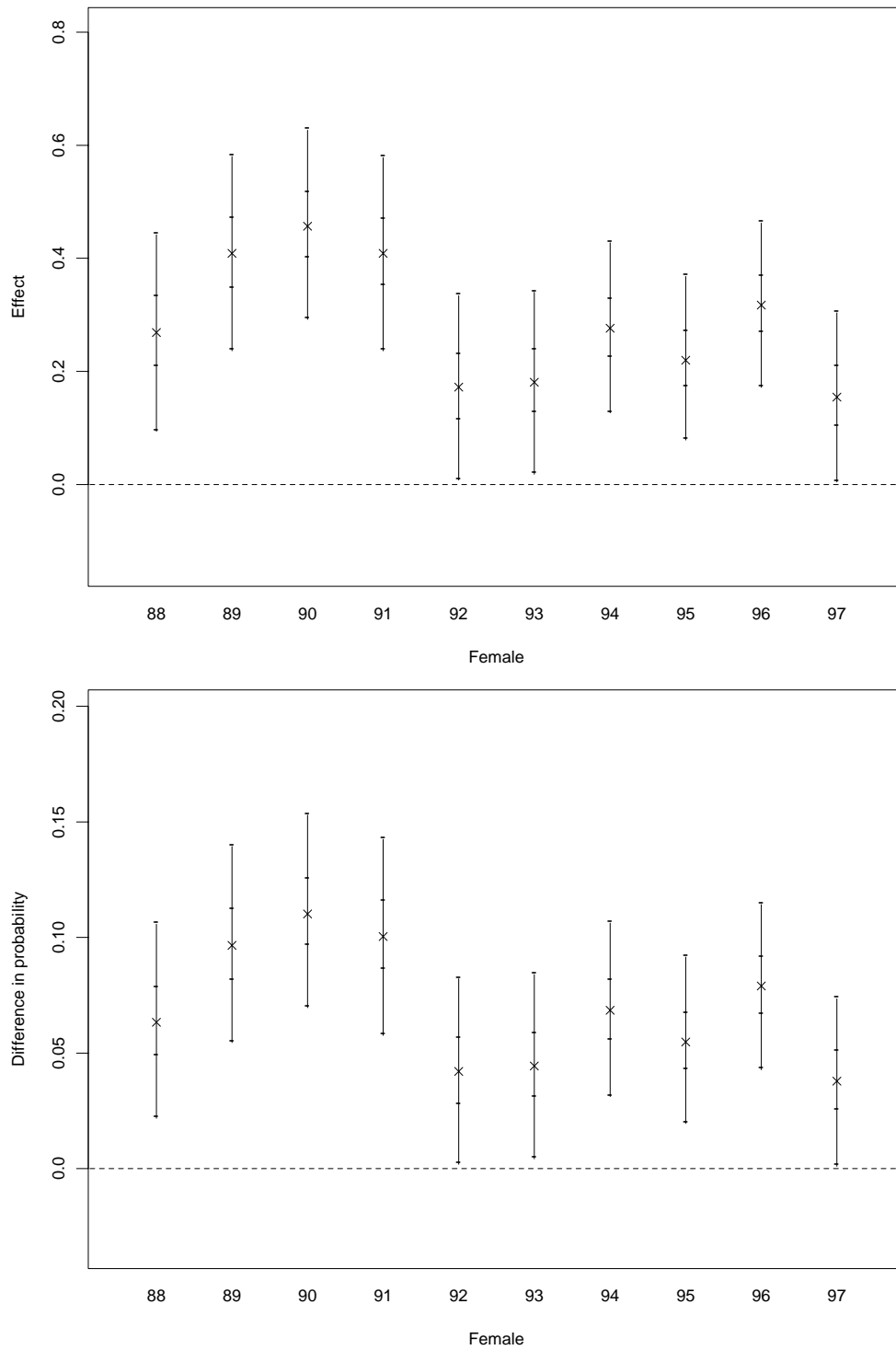


Figure 29: Intervals for Gender effects $\chi_{2,r}$ (upper frame) and the differences in implied baseline return time probabilities relative to the baseline (lower frame) over years $r = 1988, \dots, 1997$.

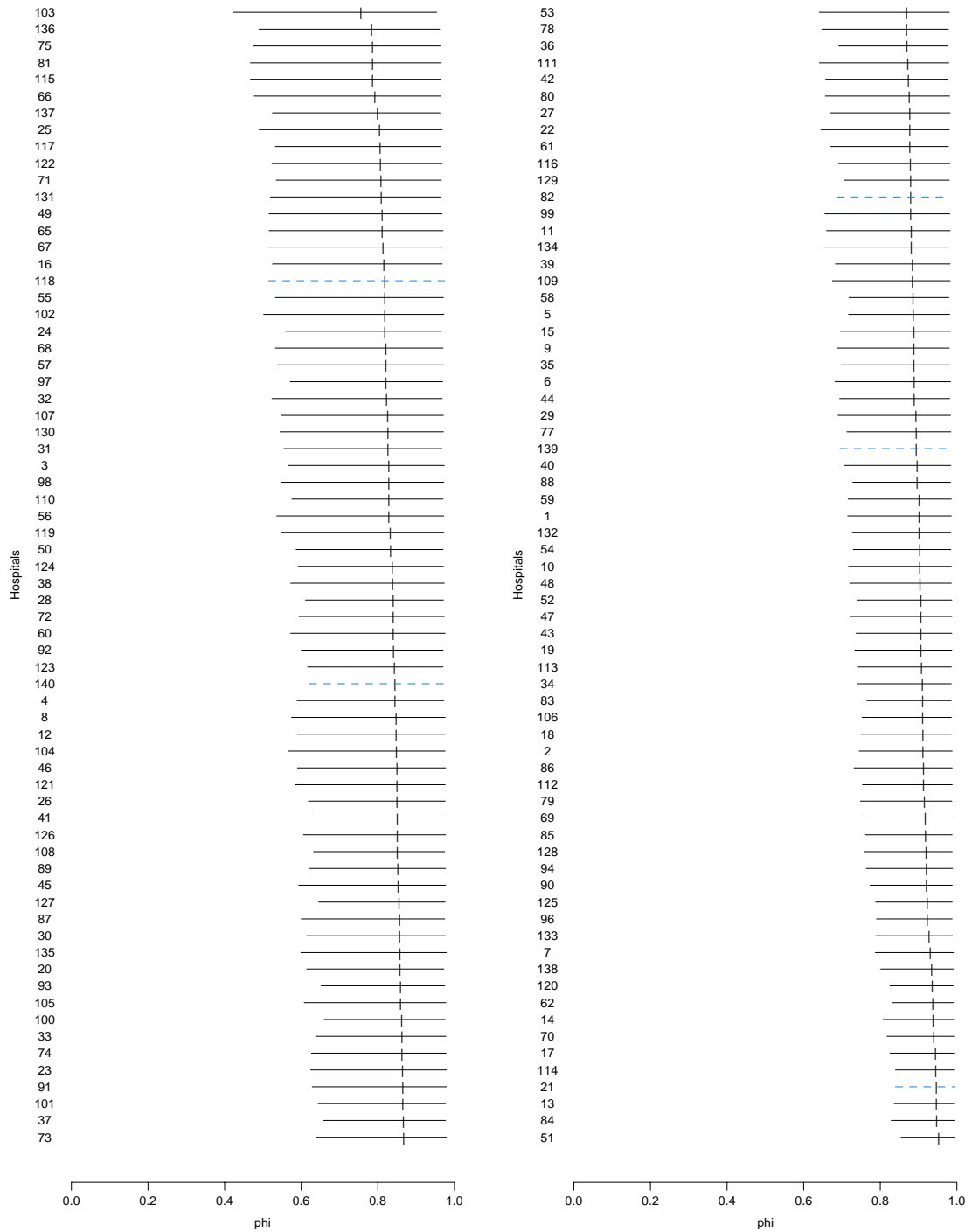


Figure 30: Intervals for the hospital-specific dependence parameters ϕ_i of all 136 hospitals, ordered by posterior medians. The five hospitals selected earlier are highlighted.

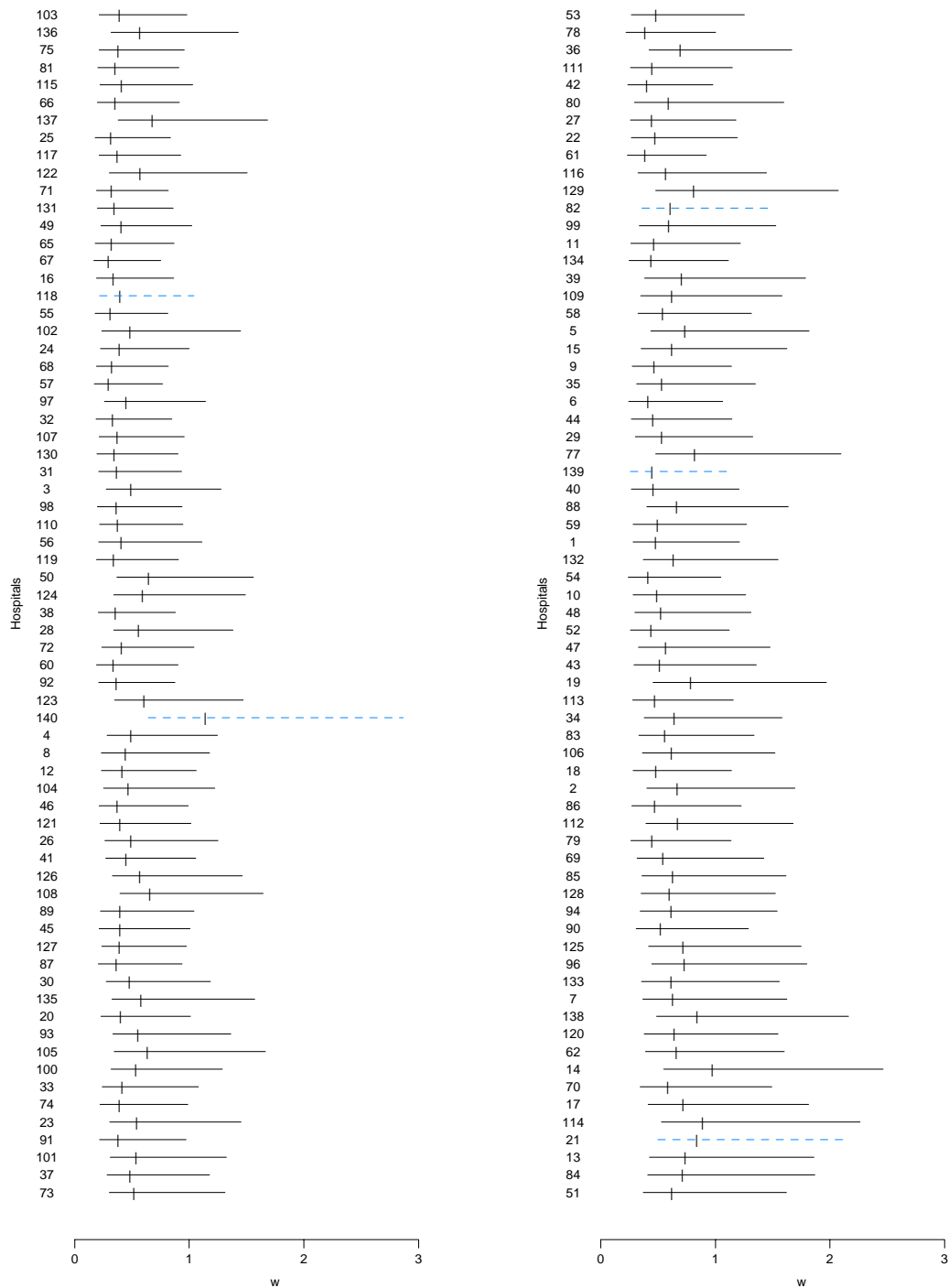


Figure 31: Intervals for standard errors w_i for all 136 hospitals, with hospitals ordered as in Figure 30. The five hospitals selected earlier are highlighted.