A Nonparametric Bayesian Model for Inference in Related Longitudinal Studies

Peter Müller, Gary L. Rosner

University of Texas M.D. Anderson Cancer Center, Houston, U.S.A.

Maria De Iorio

Imperial College, London, U.K.

Steven MacEachern

The Ohio State University, Columbus, U.S.A.

Summary. We discuss a method for combining different but related longitudinal studies to improve predictive precision. The motivation is to borrow strength across clinical studies in which the same measurements are collected at different frequencies. Key features of the data are heterogeneous populations and an unbalanced design across three studies of interest. The first two studies (CALGB 8881 and 9160) are phase I studies with very detailed observations on a relatively small number of patients. The third study (CALGB 8541) is a large phase III study with over 1500 enrolled patients, but with relatively few measurements made on each patient. Patients receive different doses of several drugs in the studies, with the phase III study containing significantly less toxic treatments. Thus, the main challenges for the analysis are to accommodate heterogeneous population distributions and to formalize borrowing strength across the studies and across the different treatment levels. We describe a hierarchical extension over suitable semiparametric longitudinal data models to achieve the inferential goal. A nonparametric random effects model accommodates the heterogeneity of the patient population. A hierarchical extension allows borrowing strength across different studies and different levels of treatment by introducing dependence across these nonparametric random effects distributions. Dependence is introduced by building an ANOVA-like structure over the random effects distributions for different studies and treatment combinations. Model structure and parameter interpretation are similar to standard ANOVA models. Instead of the unknown normal means as in standard ANOVA models, however, the basic objects of inference are random distributions, namely the unknown population distributions under each study. The analysis is based on a mixture of Dirichlet process (MDP) model as the underlying semiparametric model.

Key words: Dirichlet process; nonparametric Bayesian inference; limited sampling;

1. Introduction

Studies on novel therapeutic drugs progress from early toxicity studies (called phase I) through small studies examining activity (phase II), ending with large, randomized clinical trials comparing new treatments to placebo or current standard therapy (phase III). Phase I studies enroll a small number of patients and monitor them very closely to ensure the patients' safety and to learn about the drug's pharmacokinetics. In contrast, the later phase III studies involve large numbers of patients and are often conducted as multi-center randomized clinical trials. The complexity of organizing such a large scale study prevents one from obtaining as detailed information on individual patients as was obtained in the earlier phase I studies. In this paper we propose a hierarchical semiparametric Bayesian population model that allows us to combine the relative strengths of the early phase studies with many measurements on few patients and the phase III study with few measurements on many patients.

In the motivating example, the clinical outcome of interest is myelosuppression, a common side effect of anticancer drug therapy. Myelosuppression is a profound lowering of a person's bone marrow activity leading to a reduction in the number of platelets, red blood cells, and white blood cells in the blood. White blood cells are an important component in the human immune system, and severly lowered white blood cell counts (WBCs) put the patient at risk of possibly fatal infection. Inference about myelosuppression is complicated by heterogeneity in the effect of anticancer therapy among patients. For example, some patients can tolerate standard doses of the drugs, while other patients experience more side effects. Furthermore, there are different ways that people report the extent of myelosuppression, such as the nadir count or the number of days the patient's WBCs are below some threshold value. If one could model the entire longitudinal profile of a patient's WBCs over time, one could provide inference for any desired summary of interest. Unfortunately, the WBCs are often not measured frequently enough to allow one to model the profiles without imposing constraints. Without being able to model the outcome of interest with adequate precision, one cannot hope to achieve a meaningful inference concerning the effect of therapy among the patient population. One possible way to improve the precision of the inference is to borrow strength from the more detailed information collected as part of some smaller clinical studies. The modeling framework should include informative priors to accomodate information about the different nature of the patient populations.

In this report, we are concerned with inference about myelosuppression from a large multi-center randomized clinical trial in which the focus was on clinical efficacy and not on frequent measurement of WBCs. This trial, taken by itself, cannot provide us with the detailed longitudinal model of a patient's WBC needed for inference. To develop the

longitudinal model, we turn to two earlier studies in which a small number of patients were examined intensively. The complementary information in the two types of study allows us to both develop a longitudinal model for a patient's WBC and assess the impact of treatment on a large number of patients.

The general problem that we address here is how to combine the information in qualitatively different studies to make effective inference. The tool that we use is a hierarchical Bayesian population model, which is ideally suited to analyze a collection of studies in a coherent fashion. The Bayesian model allows us to pass along both the qualitative form of the longitudinal model and the information about the parameters in the model, conveyed by the prior distribution and likelihood. This last information is essential, since the parameters in the longitudinal model are only weakly identifiable from the data in the in the multi-center trial. To implement the analysis, we set up a hierarchical model with submodels for each study and with a non-exchangeable prior probability model that reflects the different nature of the early phase and phase III studies.

Besides the hierarchical modeling, a second important element of the proposed model is the use of a semiparametric population model. In our analysis, we have found a need to move beyond the traditional parametric hierarchical models, as there is known population heterogeneity that cannot be described in a simple parametric model. An additional complication arises from the fact that drugs and doses are not all the same in the three studies we analyze. This heterogeneity between studies and within studies is a common feature of many biomedical data and an important theme in population models. Bayesian nonparametric methods have been proposed for population models to accomodate population heterogeneity and to relax distributional assumptions and restrictive models. Without the additional hierarchical structure across related studies, such approaches have been discussed in Kleinman and Ibrahim (1998b,a), Müller and Rosner (1997), Walker and Wakefield (1998), and Tomlinson and Escobar (1999), among others. All use variations of Dirichlet process (DP) models to define flexible nonparametric models for an unknown random effects distribution. DP models are by far the most widely used nonparametric Bayesian model, mainly because of computational simplicity and because computational complexity of posterior simulation in DP models is essentially dimension independent, allowing consideration of possibly high dimensional random effects distributions. See, for example, Ferguson (1973) or Antoniak (1974) for basic properties of the DP model. A random measure G generated by a DP is almost surely discrete. In many applications this discreteness is inappropriate. DP mixture models avoid the discreteness by defining a random measure F as a mixture of some continuous kernel $f(x|\mu)$ with respect to a discrete DP measure G, i.e., $F(\cdot) = \int f(\cdot|\mu)\,\mathrm{d}G(\mu)$. Such models are known as DP mixtures (MDP) (Escobar, 1988; MacEachern, 1994; Escobar and West, 1995). See Walker et al. (1999) for a recent review of nonparametric Bayesian

methods in general. The use of semiparametric longitudinal models for repeated measurement data when patient heterogeneity is a concern is a common theme also in non-Bayesian modeling. Recent examples are Davidian and Gallant (1993), Zhang et al. (1998), Brumback and Rice (1998), and Rice and Wu (2001). See Davidian and Giltinan (1995) for a review.

Joint inference for several related population studies, as needed for the case study discussed in this article, requires that one tie together both modeling strategies: hierarchical models and nonparametric inference. While hierarchical extensions are standard modeling practice in the context of traditional parametric models, such extensions are rarely used with nonparametric models. The main reason is that the infinite dimensional nature of the parameters complicates the definition of hierarchical model extensions. Early developments appear in Cifarelli and Regazzini (1978), who propose an approach based on introducing the hierarchy on the prior parameters for the nonparametric submodels. They use DP submodels and define a joint distribution across all submodels by assuming a regression for some (finite dimensional) prior parameters of the DP models. More recently, the same strategy is used, among others, in Mira and Petrone (1996), Giudici et al. (2003), and Carota and Parmigiani (2002). In this article we pursue an alternative approach. We use one underlying DP model and define nonparametric random effects distributions for each submodels by considering DP mixture models with different mixture kernels for each submodel. The resulting structure can alternatively be described as a DP mixture of ANOVA models, since the mixture kernels differ by an ANOVA-like decomposition of the means across factor levels. A description of a wider class of DP mixture of ANOVA models, including a discussion of other application areas and theoretical issues, appears in De Iorio et al. (2004).

We give a detailed description of the data in the next section. Section 3 describes the nonlinear model for WBC over time and the MDP prior for the patient-specific random effects distributions. The hierarchical extension of the model is presented in Section 4. Results of our hierarchical model data analysis are discussed in Section 5. We conclude with a short discussion.

The discussion of the non-parametric model features is restricted to two subsections, 3.2 and 4.3. A reader with a primary interest in the joint data analysis for related longitudinal studies might consider skipping these sections at first reading. Without subsections 3.2 and 4.3 the paper still provides a complete discussion of an appropriate hierarchical random effects model.

2. Data

The Cancer and Leukemia Group B (CALGB), a cooperative group of university hospitals funded by the U.S. National Cancer Institute to conduct studies relating to cancer therapy,

carried out a large randomized study, CALGB 8541 (Wood et al., 1994). A large number of hospitals and clinics participated in the study of the three chemotherapeutic regimens of the drugs cyclophosphamide (CTX), doxorubicin, and 5-fluorouracil. This clinical trial enrolled and randomized 1572 women with breast cancer to evaluate and compare the clinical benefits of three different treatment regimens for women with stage II, non-metastatic breast cancer after surgery. The three treatment arms contained the same three drugs but differed in dose and dose intensity. A side effect of the three drugs used in this phase III study (CTX, doxorubicin, and 5-fluorouracil) is to lower a patient's WBC, leaving the patient with a compromised immune system and at risk of serious infection. Not all patients react to these drugs the same way, however, and knowing what contributes to this heterogeneity may help the medical oncologists individualize therapy. Therefore, it would be helpful to learn about the change in the WBCs over time for women treated with this three-drug regimen.

In the clinical trial, the women received their chemotherapy every four weeks. The most dose-intense regimen caused the most myelosuppression, and we focus attention on this group of 513 women and analyze their WBCs during the first cycle (28 days) of treatment. The trial protocol prescribed four 28-day cycles with CTX at 600 mg/m^2, doxorubicin at 60 mg/m^2, and 5-Flurouracil at 600 mg/m^2. CTX and Doxorubincin were given on day 1, and 5-Fluorouracil was given on days 1 and 8 of each cycle. Because of the logistic difficulties associated with drawing blood from the many patients at the many clinics and hospitals that participated in the study, this phase III clinical trial required collecting blood count measurements only once a week. In the data, there are between 1 and 4 WBC measurements per patient, with an average of 3 measurements per patient. Blood count measurements occurred at roughly the same time for each patient, relative to the start of therapy (i.e., days 1, 8, 15, 22). There are too few data points with which to fit a model having much precision for interpolating between blood sample times to estimate, for example, nadir WBCs or other measures of myelosuppression.

We introduce information on the time course of WBCs in response to anticancer drug therapy to strengthen the inference. This information comes from data collected on individuals who participated in two related earlier phase studies. These studies (CALGB 8881 and CALGB 9160) collected WBCs three or four times a week for each patient, with between 4 and 25 repeated measurements per patient during their first course of therapy. We borrow strength from these studies to make more precise inference on the change in blood counts over time in the phase III breast cancer clinical trial.

CALGB 8881 was a phase I study carried out by the Cancer and Leukemia Group B (CALGB) to determine the highest dose of the anti-cancer agent CTX one can safely deliver every two weeks in an outpatient setting (Lichtman et al., 1993). The drug is known to cause a drop in WBCs. Therefore, patients also received GM-CSF, a colony stimulating factor

given to spur regrowth of blood cells (i.e., for hematologic support). The protocol required fairly extensive monitoring of patient blood counts during treatment cycles. The number of measurements per patient varies between 4 and 18, with an average of 13. The investigators treated cohorts of patients at different doses of the agents. Six patients each were treated at the following combinations (CTX, GM-CSF) of CTX (in g/m^2) and GM-CSF (in $\mu g/kg$): (1.5, 10), (3.0, 2.5), (3.0, 5.0), (3.0, 10.0) and (6.0, 5.0). Cohorts of 12 and 10 patients, respectively, were treated at dose combinations of (4.5, 5.0) and (4.5, 10.0). Hematologic toxicity was the primary endpoint.

CALGB 9160 built upon the experience gained in study 8881 and used the doses of CTX and GM-CSF considered practical after completion of 8881 (Budman et al., 1998). The goal of 9160 was to evaluate the ability of the drug amifostine to lessen the toxic effects of relatively high-dose CTX, since amifostine had been shown in some studies to reduce some of the toxic side effects of radiation therapy (Spencer and Goa, 1995) and cancer chemotherapy. CALGB 9160 randomized 46 patients to receive amifostine or not, along with CTX (3 g/m^2) and GM-CSF (5 $\mu g/kg$). Since the addition of amifostine did not appear to provide added benefit or detriment, we will not consider amifostine in our subsequent analyses, choosing instead to combine the two randomized groups of patients. The number of measurements per patient varies between 10 and 25, with an average of 15.

Figures 1 and 2 show some typical patients from the three studies. Note the disparate sampling frequencies for the three studies. The sparse data in study 8541 alone would not allow one detailed inference about the hematologic profiles. The mean curves shown in the figures are the posterior estimated mean log WBC profiles for these patients using the model proposed in the next section of this paper.

3. The Longitudinal Data Model

3.1. A Random Effects Model

We first describe a nonlinear empirical model for characterizing WBCs over time for one study. Also, for the moment we ignore regression on different treatment levels x_i. Introducing a regression on covariates will be straightforward in the extended probability model for multiple studies.

Let y_{ij} denote the response of patient i on (known) day t_{ij}. For each patient, the response $y_i = (y_{i1}, \ldots, y_{in_i})$ follows a typical "bath tub" pattern, starting with an initial base line, followed by a sudden drop in WBC at the beginning of chemotherapy, and eventually a slow S-shaped recovery. In Müller and Rosner (1997) we studied inference for 8881 alone. We use a non-linear regression in the form of a piecewise linear and logistic curve to model the

typical profile:

$$y_{ij} \sim N\left[f(\theta_i, t_{ij}), \sigma^2\right]. \tag{1}$$

The mean function $f(\theta, t)$ is parameterized by a vector of random effects $\theta = (z_1, z_2, z_3, \tau_1, \tau_2, \beta_1)$:

$$f(\theta, t) = \begin{cases} z_1 & t < \tau_1 \\ rz_1 + (1-r)g(\theta, \tau_2) & \tau_1 \le t < \tau_2 \\ g(\theta, t) & t \ge \tau_2 \end{cases} \tag{2}$$

where $r = (\tau_2 - t)/(\tau_2 - \tau_1)$ and $g(\theta, t) = z_2 + z_3/\{1 + \exp[2.0 - \beta_1(t - \tau_2)]\}$. The intercept in the logistic regression was fixed at 2.0 after finding in a preliminary data analysis that a variable intercept did not significantly improve the fit.

The non-linear regression (2) fixes the sampling distribution of the observable data conditional on a random effects vector θ_i. In the next level of the hierarchy we assume a prior probability model $H(\theta_i)$, that is, a random effects distribution. A traditional and technically convenient choice is a multivariate normal random effects model

$$H(\theta_i) = N(\mu, S). \tag{3}$$

The remaining discussion, except for subsections 3.2 and 4.3 could be based on this model. All of of the desired inference and the further model extensions to multiple studies and study-specific covariates can be formalized in the context of (3). In fact, we will introduce all extensions first in the context of this simple multivariate normal model. Thus, if a reader were not primarily interested in the semi-parametric aspect of the inference, subsections 3.2 and 4.3 could be skipped at first reading.

3.2. A Non-parametric Random Effects Model

Several considerations lead us to work with a generalization of the multivariate normal random effects distribution (3). The random effects distribution needs to accommodate the heterogeneity in the population and allow for outliers, clustering and over-dispersion. At the same time, the model should not be overly complex and should still allow computationally efficient implementation of full posterior inference. Ideally the model should be a natural generalization of a traditional multivariate normal random effects distribution. Based on these considerations, we use a mixture of normals model. Let $\mathrm{N}(x; m, S)$ denote a multivariate normal distribution for the random variable x with moments (m, S), and let $\delta(x)$ denote a point mass at x. We assume the prior model $H(\theta_i) = \sum w_h \, \mathrm{N}(\theta_i; \mu_h, S)$. This distribution can be interpreted as a mixture with respect to the mixing measure $G = \sum_{h=1}^{\infty} w_h \, \delta(\mu_h)$:

$$H(\theta_i) = \int \mathrm{N}(\theta_i; \mu, S) \, \mathrm{d}G(\mu). \tag{4}$$

We complete the model with with a nonparametric prior on the unknown mixing measure $G = \sum_{h=1}^{\infty} w_h \, \delta(\mu_h)$,

$$G \sim \mathrm{DP}(M, G^o). \tag{5}$$

Here $\mathrm{DP}(M, G^o)$ denotes a Dirichlet process (DP) model (Ferguson, 1973; Antoniak, 1974). The DP is a probability model on distributions. It requires two parameters, a scalar precision parameter M, and a mean measure $G^o = E(G)$. One of the key properties of the DP is the discreteness of a random measure $G \sim \mathrm{DP}(\cdot)$. Above, we have implicitly made use of this by writing G as a sum of point masses. Sethuraman (1994) shows that a random distribution $G \sim \mathrm{DP}(M, G^o)$ can be constructively defined as follows ("stick breaking").

$$\frac{w_h}{\prod_{i<h}(1-w_i)} \overset{iid}{\sim} \mathrm{Be}(1, M) \text{ and } \mu_h \overset{iid}{\sim} G^o, \quad h = 1, 2, \ldots. \tag{6}$$

The weights w_h, conditional on $w_i, i < h$, are generated by rescaled Beta distributions. The point masses μ_h are an i.i.d. sample from the base measure. The stick breaking representation (6) will be used to introduce the hierarchical extension in the next section.

Longitudinal data models with semiparametric, DP mixture random effects distributions have been used in Müller and Rosner (1997) and Kleinman and Ibrahim (1998a). Similar mixture models as in (4) and (5) with a DP prior on the mixture parameters have been successfully used in many recent applications. See, for example, Escobar and West (1998) or MacEachern and Müller (2000) for a review of related models.

4. Hierarchical Extension

4.1. Multiple Studies

Equations (1) – (5) define a model for inference in one study. Joint analysis of all three studies requires a hierarchical extension across studies. Denote with $y_{ki} = (y_{kij}, \; j = 1, \ldots, n_{ki})$ the responses of patient i in study k. We define a probability model for y_{ki}, $i = 1, \ldots, I_k$, using equations (1) and (3) with an additional subscript k in θ_{ki} and H_k. This includes a random effects distribution H_k for study k, i.e., $\theta_{ki} \sim H_k$. Recall that θ_{ki} is a 6-dimensional random effects vector of parameters in the non-linear regression (2).

A natural extension of (3) to multiple studies is as a multivariate ANOVA model with study specific factors. Let $d_1 = (1, 1, 0)$, $d_2 = (1, 0, 1)$ and $d_3 = (1, 0, 0)$. Let $A = [m, \mathrm{ST}_1, \mathrm{ST}_2]$ denote a (6×3) matrix. Here m, ST_1 and ST_2 denote the columns in A. We replace (3) by

$$\theta_{ki} \sim \mathrm{N}(A \, d_k, S), \tag{7}$$

i.e., $H_k = \mathrm{N}(A \, d_k, S)$. The columns in A are interpreted as an overall mean effect vector m and study-specific offset vectors ST_k. We add an identifiability constraint $\mathrm{ST}_3 \equiv 0$ for an

effect (not included in A) for the large study CALGB 8541. We fix ST_3 for convenience, since it is the largest data set.

An important aspect of model (7) is the opportunity to introduce prior information about the different nature of the three studies. This is implemented as an informative prior $F^o(m, ST_1, ST_2)$ for the ANOVA effects. For example, the third study, CALGB 8541, calls for a much lower dose of the anticancer agents than the earlier phase studies 8881 and 9160. Let a denote the (18×1) column vector defined by stacking the columns of A. We define F^o as $F^o(a) = N(a^o, C)$ and complete the model with a hyperprior

$$a^o \sim N(\alpha, I),\ S^{-1} \sim p(S^{-1}) \text{ and } C \sim p(C) \tag{8}$$

using standard conditionally conjugate priors for S and C. See Section 5 for the specific choices in our implementation. Let α_{ST1} denote the subvector of the hypermean α corresponding to the ST_1 column of the ANOVA matrix A, and analogously for α_{ST2}, etc. By fixing nonzero α_{ST1} and α_{ST2} we define a prior probability model with non-zero expected study-specific offsets. This allows us to explicitly include prior information about the lower toxicity of the treatment in the third study.

4.2. Treatment Covariates

So far, we have only defined a one-way ANOVA over the three studies. Generalizing the model to include covariates is straightforward. One adds appropriate columns to the matrix A of ANOVA effects. In particular, we use this opportunity to include a regression on treatment levels within study 8881. Considering all treatments as categorical, this requires that we add one main effect for each possible treatment level. In study 8881, there were seven treatment groups spread across three levels of GM-CSF and four levels of CTX, as described in Lichtman et al. (1993) and summarized in Section 2. We excluded the lowest level of GM-CSF from the analysis, since only one cohort of 6 patients was studied under this level of GM-CSF, simplifying the analysis. We extend the matrix A of ANOVA effects to $A = [m, \mathrm{ST}_1, \mathrm{ST}_2, \mathrm{CTX}_1, \mathrm{CTX}_3, \mathrm{CTX}_4, \mathrm{GM}_2]$ to include additional columns for each treatment level, with CTX_1 corresponding to the 1st level of CTX, GM_2 corresponding to high GM-CSF, ST_1 for a patient in study 1, and ST_2 for a patient in study 2. Main effects CTX_2, GM_1 and ST_3 are set to zero to ensure identifiability. Recall that the random effects vectors θ_{ki} are 6-dimensional. Thus all main effects ST_2, etc., are 6-dimensional column vectors and A is a $(6 \times q)$ matrix of ANOVA effects with $q = 7$ here. The design vector d_{ki} now becomes patient specific, defined to select the appropriate effects from matrix A for patient i in study k. Besides a higher dimension for A and redefined design vectors d_{ki}, model (7) remains unchanged.

4.3. Dependent Non-parametric Random Effects Distributions

We first consider the extension of the non-parametric random effects model (4) and (5) to multiple studies. This requires a joint probability model for the non-parametric random effects distributions (4) defined for each study, $k = 1, 2$ and 3. Let $H_k(\theta) = \sum w_h \, \mathrm{N}(\mu_{kh}, S)$ denote the random effects distribution under study k. In anticipation of the following discussion, we add a study-specific index k only for the point masses μ_{kh} but not for the weights w_h. We want to define this extension in such a way that the marginal model for H_k remains unchanged. We achieve this by replacing (7) by a mixture of ANOVA model.

$$\theta_{ki} \sim \int \mathrm{N}(\theta_{ki}; A\, d_k, S)\, dF(A) = \sum_{h=1}^{\infty} w_h \, \mathrm{N}(A_h\, d_k, S), \tag{9}$$

with $F = \sum w_h \, \delta(A_h)$ generated from a DP, $F \sim \mathrm{DP}(M, F^o)$. The prior F^o defined in (8) becomes the base measure in the DP prior. The hyperprior on S and C remains unchanged. An additional hyperprior $p(M)$ is added to complete the model.

The interpretation of the columns in A remain unchanged as overall mean effect m and study-specific offsets ST_k. Recognizing (9) as a mixture model extension of a standard ANOVA model highlights the model structure. Model (9) is a nonparametric mixture of ANOVA models, using a DP prior for the unknown mixing measure. Consider the following construction to see that the joint model (9) leaves the marginal distribution for H_k unchanged. Let $G^o(\theta)$ denote the random distribution defined by drawing $A_h \sim F^o$ and setting $\mu_{kh} = A_h \, d_k$. The implied marginal distribution for H_k remains a DP mixture of normals as before, with DP base measure G^o, as in (4) and (5).

Extension to treatment covariates is straightforward by adding additional columns in A_h, paralleling the discussion in 4.2. As a mixture of ANOVA models, (9) inherits structure and modeling tools from standard ANOVA models. The design can be further extended to include interactions, nested effects, etc., as desired. Order constraints among effects can be introduced when appropriate. Contrasts and hypotheses related to equality of effects can be explored on the basis of the full posterior distribution.

5. Implementation and Results

Model (9) is a DP mixture model. Posterior simulation in such models is well understood. Specific algorithms for posterior Markov chain Monte Carlo (MCMC) simulation are given, for example, in Neal (2000) or MacEachern and Müller (1998). We implemented posterior simulation for the data described in Section 2. Let $B \sim Wish(\nu, B_0)$ denote a Wishart distributed random matrix B with ν degrees of freedom and parameterized such that $E(B) = \nu B_0$. Let $\mathrm{IG}(a, b)$ denote an inverse Gamma distribution with parameters (a, b). We assume $S^{-1} \sim Wish(\nu_0, \Phi_0/\nu_0)$, $\sigma^2 \sim \mathrm{IG}(a_\sigma/2, b_\sigma/2)$, and $M \sim \mathrm{Ga}(0.05, 0.05)$. We use $\nu_0 = 12$,

$a_\sigma = 4.25$, $b_\sigma = 1.25$ and $\Phi^{-1} = S_0$, the empirical variance-covariance matrix of the maximum likelihood estimates for θ_{ki}, computed separately for each patient. We make one more simplifying assumption. We take z_1, the initial base line in (2) out of the ANOVA model to reduce the dimension of the A matrix. Let z_{1ki} denote the z_1 element for patient i in study k, and let θ_{ki} denote the vector of the remaining 5 random effects. We replace the normal kernel in (9) by

$$N\left[\begin{pmatrix} z_{1ki} \\ \theta_{ki} \end{pmatrix}; \begin{pmatrix} 2 \\ Ad_{ki} \end{pmatrix}, S\right],$$

For the study-specific offsets in the base measure, we assume $\alpha_{ST1} = (-2.5, 2.0, -3, -7, 0.3)$ and $\alpha_{ST2} = \alpha_{ST1}$. See the discussion after (8) for a definition of the study-specific offsets. The offsets were chosen based on the informed prior judgment of a clinician who was asked about likely nadir levels, delay in the initial decline of WBC, and day and rate of recovery. The clinician is familiar with the treatments, but does not know about the CALGB studies or the data.

Figures 3 and 4a show some elements of the posterior estimated distribution of the ANOVA parameters, i.e., $p(F|y)$ in (9). Since F is high dimensional, we plot appropriate summaries only. The summaries relate to particular factor levels. Specifically, let Fd denote the random distribution of the ANOVA effects selected by a design vector d, and let $d_m = (1, 0, 0, \ldots, 0)$ and $d_{ST2} = (0, 0, 0, 1, 0, \ldots, 0)$ denote the design vectors that select only m, only ST_2, etc. Consider the random distributions Fd_m, Fd_{ST2}, etc. These are still 6-dimensional distributions, requiring further simplification to allow illustration. We choose the following problem-specific graphical representation. We show a distribution Fd indirectly by plotting the implied non-linear regression. Formally, let $f(\theta, t)$ denote the non-linear regression given in (2). We compute the posterior quantiles for $f(\theta, t)$ implied by $Fd(\theta)$ and plot them against t. In other words, the figures show pointwise posterior quantiles corresponding to the indicated ANOVA effect.

The two panels in Figure 5 show posterior predictive mean profiles for a new patient from study 2 and 3, respectively. The figures includes pointwise 95% predictive intervals. The method allows for differential posterior uncertainty across factor levels, as seen in the figure. Despite the fact that there are few WBCs for patients in study 3, one can make predictions about the time course of a future patient's WBCs.

All inference is based on posterior and posterior predictive simulation. Thus it is possible to report posterior, or posterior predictive, probabilities for any event of interest. In particular, we can find the posterior predictive distribution for any function of the predicted WBC profile for a new patient. For example, we can predict the nadir WBC, a commonly used measure of the degree of myelosuppression, for a future patient in any of the studies. Figure 6 shows the predictive distribution for the nadir WBC for a future patient across the

3 studies. For the first study, we predict the profile of a patient who will receive 3.0 gm/m^2 CTX and 5 $\mu g/kg$ GM-CSF. There is more variability in the nadir prediction for study 3, compared to the other 2, but the nadirs are generally higher in study 3, reflecting the fact that the treatment is less toxic than in the other studies. The relatively low predictive uncertainty displayed in the figure may be an artifact of the compression to lower values predicted for the nadirs in studies 1 and 2, since the WBCs must be non-negative. One can produce similar figures and statistics for other summary measures of a future patient's profile for inference.

Finally, Figure 4b compares posterior predictive mean profiles for a future patient in study 3 under the proposed model versus inference using data from study 3 only. We used identical prior assumptions in both analyses. Without incorporating the information from the earlier studies, prediction is much more uncertain about the time of the nadir count and the start of the recovery, as would be expected for inference conditional on the sparse phase III data only. In contrast, the predicted WBC profile based on the hierarchical model is more consistent with what one would expect to see for patients receiving anticancer chemotherapy and has reasonable predictive precision. The implications for clinically important summaries of the profiles are significant. For example, consider the number of days that the mean WBC is below a critical level of WBC= 1000. Under the hierarchical model we find an estimated posterior mean of 5.15. In contrast, using data from study 3 only we find a posterior mean of 1.04. The large difference is due to the fact that the relatively few observations under study 3 do not allow precise inference about the day of recovery.

The effect of a treatment regimen can be described directly, as in Figures 3 through 6, or it can be described in comparison to the effect of another treatment regimen. For example, consider the four different levels of CTX in study 1. One can show the posterior distribution of the difference, relative to CTX equal 3.0 g/m^2, in the number of days that WBC is below 1000. These posterior distributions can be estimated as histograms corresponding to imputed differences evaluated during the MCMC posterior simulation, and allow one to assess the change in risk to the patient at a higher or lower dose. Since the study of interest, CALGB 8541, uses only one treatment combination we did not implement this in the example.

6. Discussion

We have discussed a modeling technique for combining nonparametric inference across related studies. More generally, we view this as a model for arrays of random distributions. We use the DP model as marginal model for each of the random distributions of the submodels. The submodels are embedded in one large joint model that takes the form of a

DP mixture of ANOVA models. It contains the marginal submodels through multiplying by the appropriate design vector. We exploit this model to define inference across related, non-exchangeable studies. We focused on applications combining early phase studies with frequent measurements on few patients with another study having many patients but with fewer measurements per patient.

An important feature of our model is its ability to allow for the different doses, different dose intensities, and even different drugs appearing in the three studies. We capture information about these differences in our prior distribution, behavior that would be difficult to mimic with a classical analysis of the data.

Although our models are tailored to the particular cancer therapy case study we discuss here, the approach is valid in any application in which one combines disparate studies, especially studies that differ in how extensively they examine groups of "experimental units." Such situations routinely arise in drug development and other areas of biomiedical research. The same feature also arises in a much wider variety of contexts in which the data are often described as having been collected at different resolutions by design. An application where we see particular value to these models is population studies of pharmacokinetics. Such studies would benefit from the ability to combine information across disparate populations, within which different levels of data collection have taken place. The larger sample sizes afforded by appropriately combining studies would provide better opportunities to discover differences that can be explained by patient characteristics, including genetics, concurrent medications (drug-drug interactions), age, etc. Similarly, the question posed by limited-sampling methods would be rephrased as a search for the best small set of sampling times to use in conjunction with existing data rather than the best set of sampling times to use in isolation from existing data. There are many other specific problems that will benefit from these models.

A limitation of the model is the need to use a conjugate base measure and mixing kernel (the normal mixing kernel in (9)). Models with non-conjugate base measure and mixing kernel could still be estimated using one of the approaches for non-conjugate DP mixture models discussed in Neal (2000), MacEachern and Müller (1998) or Walker and Damien (1998). The high dimensional nature of the ANOVA effects matrix, however, is likely to complicate computations. In the current implementation, we rely on conjugacy. An interesting alternative characterization of the probability model is as a special case of the dependent DP (DDP) introduced in MacEachern (1999). We develop this perspective in De Iorio et al. (2004).

Acknowledgment

We thank the Cancer and Leukemia Group B (CALGB), a cooperative group of university hospitals funded by the U.S. National Cancer Institute, for the permission to use the data. Research was partially supported by NIH under grant 1R01CA75981-01 and NSF under grant DMS-0072526.

References

Antoniak, C. E. (1974) Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics*, **2**, 1152–1174.

Brumback, B. A. and Rice, J. A. (1998) Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion p976-994). *Journal of the American Statistical Association*, **93**, 961–976.

Budman, D., Rosner, G., Lichtman, S., Miller, A., Ratain, M. and Schilsky, R. (1998) A randomized trial of wr-2721 (amifostine) as a chemoprotective agent in combination with high-dose cyclophosphamide and molgramostim (gm-csf). *Cancer Therapeutics*, **1**, 164–167.

Carota, C. and Parmigiani, G. (2002) Semiparametric regression for count data. *Biometrika*, **89**, 265–281.

Cifarelli, D. and Regazzini, E. (1978) Problemi statistici non parametrici in condizioni di scambialbilita parziale e impiego di medie associative. *Tech. rep.*, Quaderni Istituto Matematica Finanziaria, Torino.

Davidian, M. and Gallant, A. R. (1993) The Nonlinear Mixed Effects Model with a Smooth Random Effects Density. *Biometrika*, **80**, 475–488.

Davidian, M. and Giltinan, D. (1995) *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall.

De Iorio, M., Müller, P., Rosner, G. L. and MacEachern, S. N. (2004) An ANOVA model for dependent random measures. *Journal of the American Statistical Association*, to appear.

Escobar, M. (1988) *Estimating the means of several normal populations by estimating the distribution of the means*. Ph.D. thesis, Yale University.

Escobar, M. D. and West, M. (1995) Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, **90**, 577–588.

— (1998) Computing nonparametric hierarchical models. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (eds. D. Dey, P. Müller and D. Sinha), 1–22. New York, NY, USA.

Ferguson, T. S. (1973) A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, **1**, 209–230.

Giudici, P., Mezzetti, M. and Muliere, P. (2003) Mixtures of dirichlet process priors for variable selection in survival analysis. *Journal of Statistical Planning and Inference*, **111**, 101–115.

Kleinman, K. and Ibrahim, J. (1998a) A semi-parametric bayesian approach to the random effects model. *Biometrics*, **54**, 921–938.

Kleinman, K. P. and Ibrahim, J. G. (1998b) A semi-parametric Bayesian approach to generalized linear mixed models. *Statistics in Medicine*, **17**, 2579–2596.

Lichtman, S. M., Ratain, M. J., Echo, D. A., Rosner, G., Egorin, M. J., Budman, D. R., Vogelzang, N. J., Norton, L. and Schilsky, R. L. (1993) Phase I trial and granulocyte-macrophage colony-stimulating factor plus high-dose cyclophosphamide given every 2 weeks: a Cancer and Leukemia Group B study. *Journal of the National Cancer Institute*, **85**, 1319–1326.

MacEachern, S. N. (1994) Estimating normal means with a conjugate style Dirichlet process prior. *Communications in Statistics: Simulation and Computation*, **23**, 727–741.

— (1999) Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*, 50–55. Alexandria, VA: American Statistical Association.

MacEachern, S. N. and Müller, P. (1998) Estimating mixture of dirichlet process models. *Journal of Computational and Graphical Statistics*, **7**, 223–239.

— (2000) Efficient mcmc schemes for robust model extensions using encompassing dirichelt process mixture models. In *Robust Bayesian Analysis* (eds. F. Ruggeri and D. R. Insua). New York, NY, USA.

Mira, A. and Petrone, S. (1996) Bayesian hierarchical nonparametric inference for change-point problems. In *Bayesian Statistics 5* (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith). Oxford: Oxford University Press.

Müller, P. and Rosner, G. (1997) A bayesian population model with hierarchical mixture priors applied to blood count data. *Journal of the American Statistical Association*, **92**, 1279–1292.

Neal, R. M. (2000) Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, **9**, 249–265.

Rice, J. A. and Wu, C. O. (2001) Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, **57**, 253–259.

Sethuraman, J. (1994) A constructive defnition of the dirichlet process prior. *Statistica Sinica*, **2**, 639–650.

Spencer, C. M. and Goa, K. L. (1995) Amifostine. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic potential as a radioprotector and cytotoxic chemoprotector. *Drugs*, **91**, 1001–1031.

Tomlinson, G. and Escobar, M. (1999) Analysis of densities. *Tech. rep.*, University of Toronto.

Walker, S. and Damien, P. (1998) Sampling methods for bayesian nonparametric inference involving stochastic processes. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (eds. D. Dey, P. Müller and D. Sinha), 243–254. New York, NY, USA.

Walker, S., Damien, P., Laud, P. and Smith, A. (1999) Bayesian nonparametric inference for distributions and related functions (with discussion). *Journal of the Royal Statistical Society, Series B*, **61**, 485–527.

Walker, S. and Wakefield, J. (1998) Population models with a nonparametric random coefficient distribution. *Sankhya, Series B*, **60**, 196–214.

Wood, W., Budman, D., Korzun, A., Cooper, M., Younger, J., Hart, R., Moore, A., Ellerton, J., Norton, L., Ferree, C., Ballow, A., Ill, E. and Henderson, I. (1994) Dose and dose intensity of adjuvant chemotherapy for stage ii, node positive breast cancer. *New England Journal of Medicine*, **330**, 1253–1259.

Zhang, D., Lin, X., Raz, J. and Sowers, M. (1998) Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, **93**, 710–719.

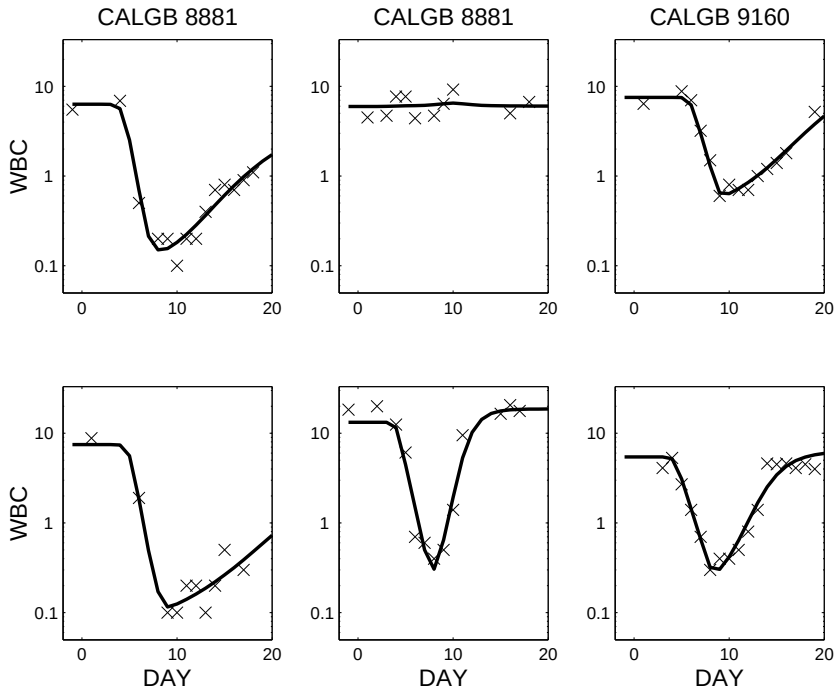


Fig. 1. Some selected patients from studies 1 (CALGB 8881) and 2 (CALBG 9160). The crosses indicate the data. The curves are posterior fits under model (1) and (2), with the mixture of ANOVA prior model from Section 4.3. Note the heterogeneity across patients, with some patients showing no decline, slower recovery, delayed recovery, etc.

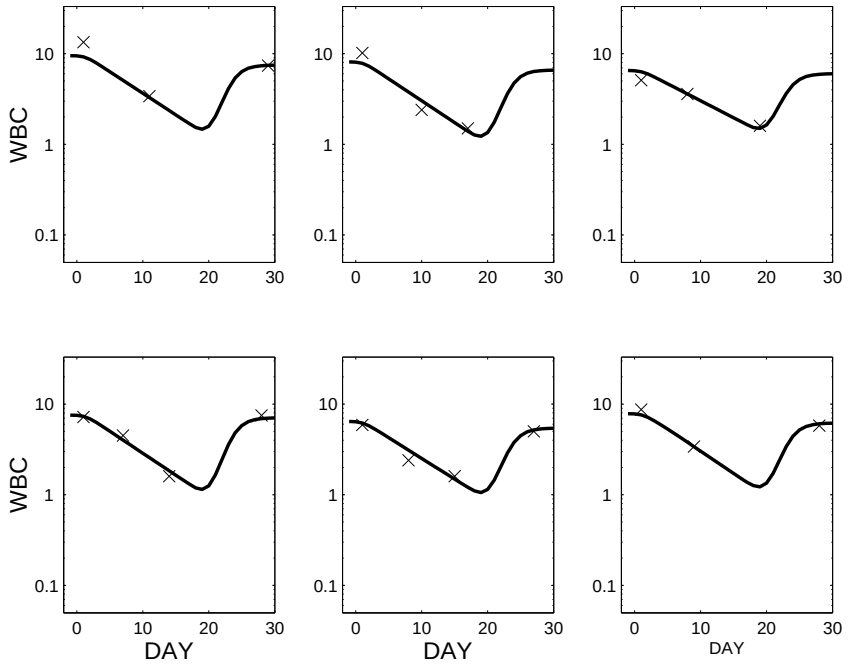


Fig. 2. Some selected patients from study 3 (CALGB 8541). The crosses indicate the data. The curves are posterior fits under model (1) and (2), with the mixture of ANOVA prior model from Section 4.3.

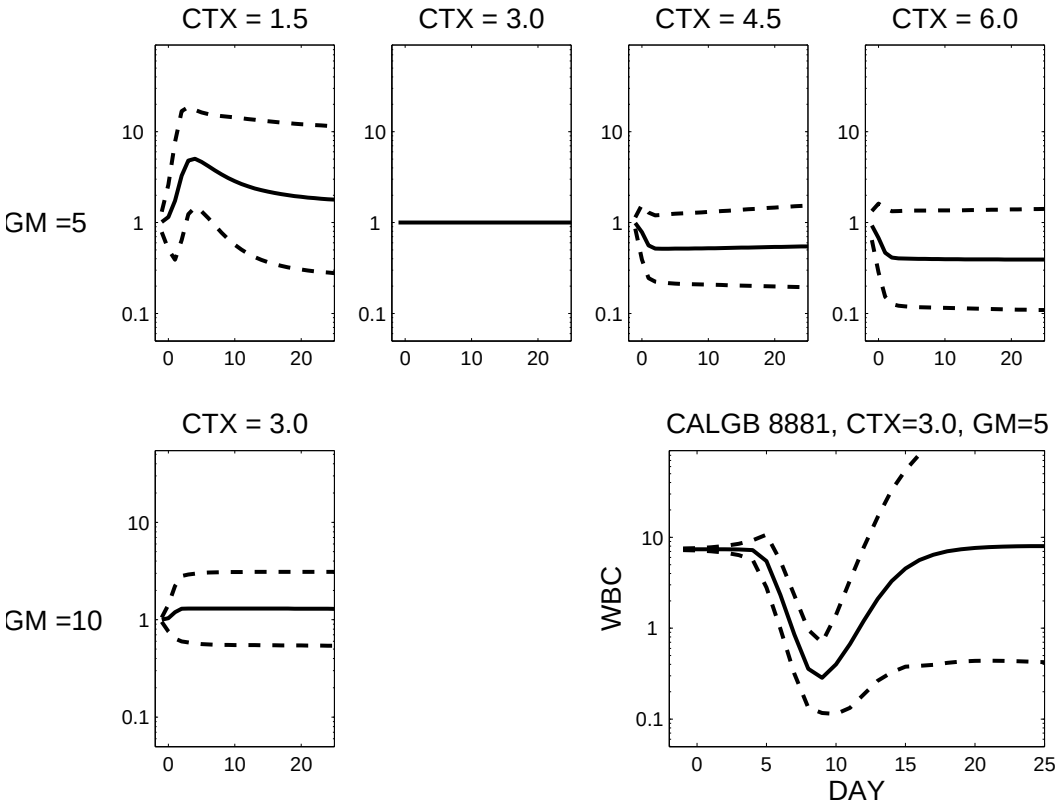
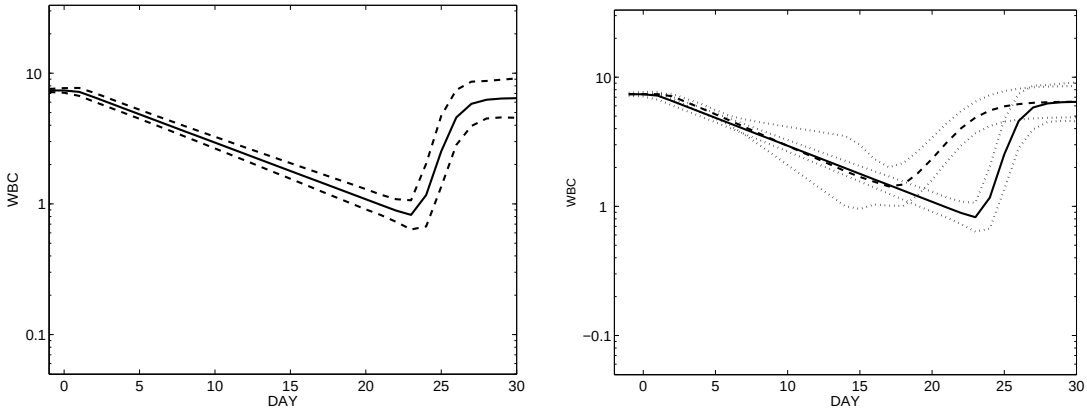


Fig. 3. Study 1: Posterior estimated distributions corresponding to the ANOVA effects $m + ST_1$ (right lower corner), CTX_1 (top left), ..., CTX_3 (top right), GM_2 (left bottom). The figure shows the pointwise posterior quantiles (2.5% and 97.5%, i.e., 95% highest posterior density intervals) of the non-linear regression curve $f(\theta, t)$ corresponding to the indicated ANOVA effect. For example, the random effects distribution for a patient with $(CTX = 4.5, GM = 5, ST = 1)$ is defined as the convolution of the random effects distribution for m, for ST_1, CTX_3 and GM_1. The figure summarizes these 6-dimensional distributions by showing the corresponding mean profiles. See the text for more discussion.



(a) Hierarchical model

(b) Hierarchical model (solid) and study 3 data only (dashed)

Fig. 4. Study 3: Posterior estimated mean profile for a future patient from study 3. Figure (a) shows the pointwise posterior quantiles of the non-linear regression curve $f(\theta, t)$ corresponding to the ANOVA effect m (=$m + \mathrm{ST}_3$, since $\mathrm{ST}_3 = 0$). See the text for a formal explanation. The dased curves show pointwise 95% highest posterior density intervals. The right panel (b) compares inference under the hierarchical model (solid line) with inference using study 3 data only (dashed line). The dotted curves show pointwise 95% highest posterior density intervals.

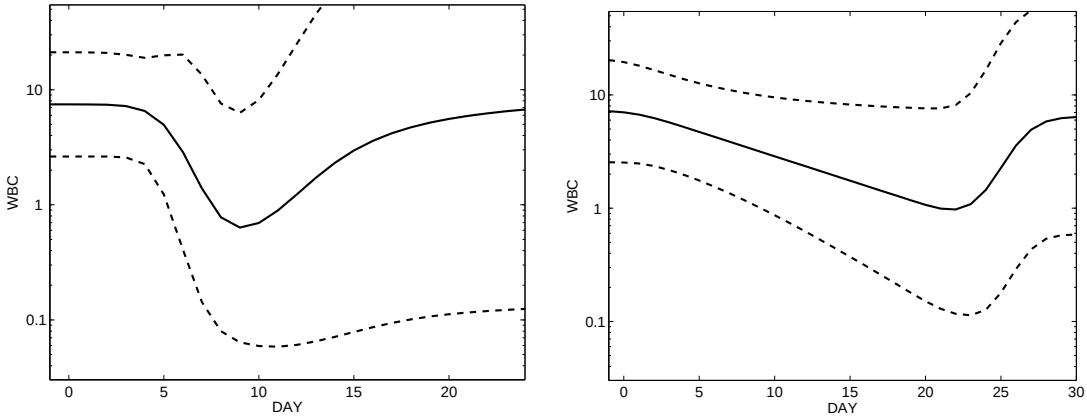


Fig. 5. Posterior predictive mean profiles and 95% predictive uncertainty for new patients in study 2 (left panel) and study 3 (right panel).

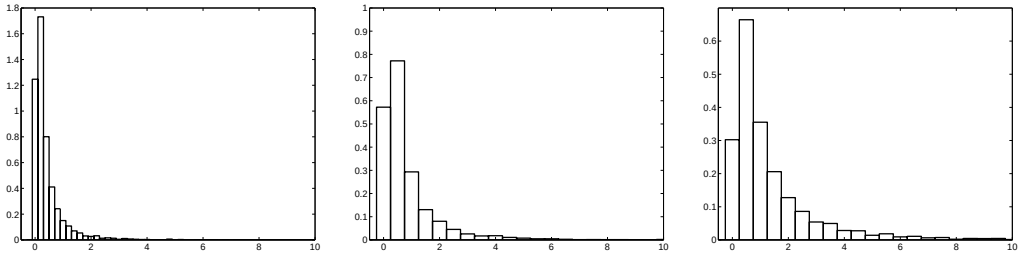


Fig. 6. Posterior predictive distribution of nadir WBCs for a future patient in studies 1, 2 and 3, respectively (in absolute counts). The future patient for study 1 is assumed to be treated at 3.0 gm/m^2 CTX and 5 $\mu g/kg$ GM-CSF.