Bayesian Analysis (2007) **2**, Number 3, pp. 591–610

Bayesian Hierarchical Multiresolution Hazard Model for the Study of Time-Dependent Failure Patterns in Early Stage Breast Cancer

Vanja Dukić* and James Dignam†

Abstract. The multiresolution estimator, developed originally in engineering applications as a wavelet-based method for density estimation, has been recently extended and adapted for estimation of hazard functions (Bouman et al. 2005, 2007). Using the multiresolution hazard (MRH) estimator in the Bayesian framework, we are able to incorporate any *a priori* desired shape and amount of smoothness in the hazard function. The MRH method's main appeal is in its relatively simple estimation and inference procedures, making it possible to obtain simultaneous confidence bands on the hazard function over the entire time span of interest. Moreover, these confidence bands properly reflect the multiple sources of uncertainty, such as multiple centers or heterogeneity in the patient population. Also, rather than the commonly employed approach of estimating covariate effects and the hazard function separately, the Bayesian MRH method estimates all of these parameters jointly, thus resulting in properly adjusted inference about any of the quantities.

In this paper, we extend the previously proposed MRH methods (Bouman et al. 2005, 2007) into the hierarchical multiresolution hazard setting (HMRH), to accommodate the case of separate hazard rate functions within each of several strata as well as some common covariate effects across all strata while accounting for within-stratum correlation. We apply this method to examine patterns of tumor recurrence after treatment for early stage breast cancer, using data from two large-scale randomized clinical trials that have substantially influenced breast cancer treatment standards. We implement the proposed model to estimate the recurrence hazard and explore how the shape differs between patients grouped by a key tumor characteristic (estrogen receptor status) and treatment types, after adjusting for other important patient characteristics such as age, tumor size and progesterone level. We also comment on whether the hazards exhibit non-monotonic patterns consistent with recent hypotheses suggesting multiple hazard change-points at specific time landmarks.

Keywords: Multiresolution models, Bayesian survival analysis, hazard estimation

*Department of Health Studies, University of Chicago, Chicago, IL, mailto:vdukic@health.bsd.uchicago.edu

†Department of Health Studies, University of Chicago, Chicago, IL, mailto:jdigman@health.bsd.uchicago.edu

1 Introduction

In survival analysis, because the hazard function $h(t)$ often exhibits unstable behavior making it difficult to reliably discern patterns of change or make comparisons between groups, aggregates of the hazard over time are more frequently used. The cumulative hazard $H(t)$, or more commonly, functions of $H(t)$ such as survival or cumulative incidence functions, are used for summary and inference on failure risk. While useful and easy to interpret for most purposes, these summaries can partially obscure important patterns in the hazard of failure over time. Alternatively, examination of the hazard function itself in detail can reveal important properties of the failure process (Aalen and Gjessing 2001). Generally, some type of smoothed estimate of the hazard function is used to characterize its shape, which is indicative of how failure risk (in the population) changes with respect to some time origin. While a variety of approaches toward hazard estimation have been proposed, methodological challenges remain for both estimation and associated statistical inferential procedures. In particular, flexible estimation and modeling approaches are needed, because in contrast to the hazard functional form in most parametric survival models, the hazard may exhibit complex non-unimodal shape with 'change-points' that may reveal important information about the process under study.

In this article, we investigate the hazard of disease recurrence among women treated for breast cancer and followed over several years. The data originate from large multicenter randomized clinical trials evaluating the effect of hormonal or cytotoxic chemotherapy treatment agents administered after surgery (referred to generally as adjuvant therapy) in women with early stage breast cancer. As we describe in the next section, there is considerable interest in the patterns of recurrence hazards after breast cancer diagnosis and initial treatment, both to gain biological insights and to better manage the disease in a clinical setting. We accommodate the biologically plausible situation whereby different subgroups of women with specific disease features may have distinct functional forms of failure hazard that are non-proportional to each other, by constructing a joint model for all separate subgroup hazards, while keeping the effects of some factors (such as age or tumor size) common across all strata and proportional within each stratum.

To model the recurrence hazards, we apply the semiparametric multi-resolution hazard (MRH) estimator recently presented in work by Bouman and colleagues (Bouman et al. 2005, 2007). Employing a piece-wise constant prior for the hazard rate which is constructed in a tree-based and self-consistent manner, the MRH approach permits flexible modeling with the ability to incorporate a variety of *a priori* assumptions about the shape and smoothness of hazard functions in each of several defined strata. Furthermore, Bayesian modeling allows us to easily address specific hypotheses concerning the timing of peaks in the risk of failure and how these may differ with respect to key biologic and clinical parameters in breast cancer.

In the next section, we describe some of the key questions of interest in modeling of the breast cancer recurrence hazard, and the data source for this study. In Section 3 we review the basics of the multiresolution hazard model, and present the extension to the

hierarchical multiresolution (HMRH) setting. In Section 4 we provide technical details of the Markov chain Monte Carlo (MCMC) model implementation. Section 5 presents the analysis. We conclude with a discussion of the findings in relation to the broader literature on breast cancer hazards as well as plans for future work on this problem.

2 Recurrence Risk after Surgery for Early Stage Breast Cancer

Over the past few decades, there has been appreciable progress in therapeutic strategies for early stage (i.e., localized and operable, as opposed to metastatic) breast cancer, with a well-developed array of treatment options. Due to increased screening vigilance and disease awareness, currently over 75% of women diagnosed have early stage tumors. Despite this progress, the clinical course of breast cancer after diagnosis remains heterogeneous from patient to patient and thus highly unpredictable for individuals. Thus, a significant clinical challenge is deciding which and how much adjuvant therapy is needed, and determining the magnitude of recurrence risk over extended post-treatment follow-up. Characteristics that prospectively identify which women are at greater or lesser risk of treatment failure are needed to aid in individually tailoring therapy for optimal disease management. Answers may also lie partly in gaining a better understanding of the intermediate and long-term clinical course of the disease, identifying patterns that can portend time periods of increased recurrence risk.

Apparent patterns in breast cancer recurrence hazard are readily observable from large cohorts of patients systematically followed over time. It is well known that risk of recurrence remains elevated for a long period of time after initial diagnosis and tumor removal, and there is longstanding interest in the the prospect of "cure" after sufficient time tumor-free has been achieved (Berg and Robbins 1966). Some long-term follow-up studies have suggested that a finite but lengthy"dormancy period" exists (possibly over 20 years), whereby tumor recurrences may still appear (Gordon 1990; Demicheli et al. 1996; Karrison et al. 1999). Studies examining the shape of the recurrence hazard consistently show a sharp peak 12-24 months after initial diagnosis and treatment, followed by a decline over time, although the hazard remains persistently elevated relative to individuals in the population never having had breast cancer (Saphner et al. 1996; Hess et al. 2003). This pattern stands in sharp contrast to the recurrence hazard for several other major cancers (e.g., colon, lung), where most recurrences appear within the first five years after discovery and removal of the primary tumor, followed by a period in which the hazard of recurrence and death from the disease resemble that of the population at large.

2.1 Data: Randomized Clinical Trials for Early Stage Breast Cancer

The National Surgical Adjuvant Breast and Bowel Project (NSABP) is a U.S. National Cancer Institute sponsored and funded multi-center cancer clinical trials group that has investigated a spectrum of treatments for breast and colorectal cancers since the

late 1950s. The group consists of more than 5,000 participating physicians, nurses, and other research specialists located at over 1,000 medical centers, university and community hospitals, oncology practice groups, and health maintenance organizations in North America. The group has enrolled more than 60,000 women and men in clinical trials for breast and colorectal cancer.

While use of systemic adjuvant therapy began in the mid 1970s for women with tumors that had spread to the axillary lymph nodes, those with so-called lymph node negative breast cancer continued to be treated by surgery only, as they were considered at sufficiently favorable prognosis so as not to require further interventions. Because a significant fraction of such women *do* suffer recurrence and eventually die from the disease, beginning in the 1980's the NSABP began a series of clinical trials among patients with axillary lymph node negative breast cancer. These trials were serially designed, beginning with comparisons surgery alone to post-surgical hormonal or cytotoxic adjuvant therapy regimens, and following benefits seen for the latter, subsequently comparing different adjuvant therapy regimens. These studies provide a unique view to the long-term prognosis of women with early stage breast cancer, and are an ideal data source for examining factors influencing the hazard of disease recurrence.

A key determinant of both expected prognosis and potential choice of adjuvant therapy type is the presence and quantity of estrogen receptors (ER) on the tumor. From 1982-1988, patients were accrued according to ER status into one of two trials conducted in parallel: Protocol B-13 randomized 760 patients with ER-negative (ER-) tumors to no further treatment after surgery (384) or to 12 cycles of cytotoxic chemotherapy treatment with methotrexate and 5-fluorouracil (376). Protocol B-14 randomized 2,892 patients with ER-positive (ER+) tumors to placebo (1,453) or the estrogen antagonist drug tamoxifen (1,439) after surgery. Primary findings were first obtained in 1989, showing a significant reduction in breast cancer recurrence risk for patients receiving the adjuvant therapy regimens (Fisher et al. 1989b,a). Longer follow-up eventually revealed a survival advantage for those who received adjuvant therapy (Fisher et al. 1996b,a). Further details of the trial designs and findings can be found in the published primary reports. Follow-up continues to date, with mean follow-up of over 15 years and over 900 recurrence events observed.

Primary endpoints for the trial were overall survival, defined as time from surgery to death from any cause, and disease-free survival, defined as time to first breast cancer recurrence at any local, regional, or distant anatomic site, occurrence of a tumor in the opposite breast, occurrence of other second primary cancers, or death prior to these events (that is, time to first event of any kind). In this study, we model the cause-specific hazard for breast recurrence, defined as time to breast cancer recurrence, treating the other event types as censored observations. We do this because modeling the cause-specific hazard for breast cancer events only may have more clinical relevance, and furthermore, with the exception of endometrial cancer, which occurs in less than 1.5% of patients but is more frequent among women taking tamoxifen, hazard rates for non-breast cancer events are essentially equal between the two treatment groups. As in previous studies modeling these data (Fisher et al. 1989c; Bryant et al. 1997), we examine tumor size, tumor progesterone receptor level, menopausal status, and age as

covariates potentially associated with the recurrence hazard.

3 The Multiresolution Hazard Model

In this section, we review the multiresolution approach to modeling hazard rate in a semiparametric Bayesian context developed and described in more detail in Bouman et al. (Bouman et al. 2005, 2007). As will be explained below, the method relies on the clever tree-based construction of the prior for the hazard rate, that ultimately yields a resolution-invariant and self-consistent prior for an arbitrarily fine piece-wise constant approximation to the hazard rate. The parameterization of the prior tree uniquely defines not only the prior expectations of the hazard rate in each of the intervals, but it also determines the amount of correlation and smoothness in hazard between the intervals.

3.1 Approaches to Hazard Modeling

One of the most common approaches to assessing the impact of factors on the hazard of failure is the Cox proportional hazard model. In this model individual covariates X affect baseline hazard $h_{\text{base}}(t)$ via $\exp(\mathbf{X}'\boldsymbol{\beta})$ (Cox 1972). This approach is readily adapted to modeling the cause-specific hazard (Prentice et al. 1978). In the typical application of the Cox model, covariate effects for the relative hazard are the primary focus, with the baseline hazard function treated as a nuisance parameter. As we have indicated, however, interest in our particular study here lies precisely in the estimation of the hazard functions. Specifically, we are interested in the estimation of a separate hazard rate function within certain strata based on treatment type and tumor characteristics, so that we might compare the shape of the hazards, while simultaneously estimating and performing inference about other covariate effects that are reasonably assumed to be common over strata.

Non-parametric methods for extracting hazard function estimates have been proposed for the Cox model (Gray 1990, 1992), primarily for the purpose of performing model diagnostics (e.g., changes in the effects of covariates over time) and correct functional forms for covariates in relation to failure hazard. A more general hazard regression approach that involves partitioning the time axis more specifically focused on hazard estimation within covariate strata (Gray 1996). There are many other approaches and variations (see Andersen et al. (1993) for detailed review), many of which provide estimates of functionals ($S(t)$, etc.) after the model is fit, but most still focus on covariates and model checking, rather than on efficient hazard function estimation *per se*. We discuss the properties and justify the MRH approach to hazard modeling more in the next section.

3.2 Multiresolution Model for the Baseline Hazard

Multiscale models for estimation of a discretized intensity function were developed for astrophysics applications by Kolaczyk (Kolaczyk 1999; Nowak and Kolaczyk 2000). Details of how this methodology has been adapted to hazard estimation are summarized in the following, while a more extensive discussion of its theoretical properties can be found in Bouman et al. (2005, 2007).

In summary, the multiresolution hazard (MRH) approach yields an estimate of the baseline (i.e., estimate from which covariate-specific curves can be generated) survival function based on the multiresolution baseline hazard estimate. It consists of first choosing the "time resolution" – a set of time points $\{t_0, t_1, ..., t_J\}$ – and then estimating the underlying baseline survival at those points, $S_{\text{base}}(t_j)$, based on the the cumulative hazard $H_{\text{base}}(t)$ and its discrete increments $d_j \equiv H_{\text{base}}(t_j) - H_{\text{base}}(t_{j-1}) = \int_{t_{j-1}}^{t_j} h_{\text{base}}(s)ds$, where $h_{\text{base}}(t)$ represents the baseline hazard rate at time t. For those times t such that $t_{j-1} < t < t_j, j = 1, \ldots, J$, a piecewise-constant hazard rate is assumed.

The MRH model is thus a semiparametric model which is able to estimate the baseline hazard rate $h_{\text{base}}(t)$ at the resolution times t_j, along with covariate effects $\boldsymbol{\beta}$. For convenience, we set the number of time intervals J equal to 2^M, where $M > 0$. In general, these intervals need not be of equal length – one can choose any resolution, though in practice we would recommend one such that there are multiple failure times observed in almost all intervals. Furthermore, the resolution should be chosen so that the average (or total) hazard rates within its intervals are clinically meaningful. However, in cases when prior information and clinical input about the resolution are vague, the optimal number of intervals could be chosen via model selection criteria, such as the DIC (Spiegelhalter et al. 2002). Note that the failure times after t_J become right-censored at t_J; hence, J should also be chosen so that a relatively small fraction of failure times is censored as a result.

After fixing the resolution, the discretized hazard is modeled in a way that allows us to incorporate the prior belief about the shape and smoothness of the true underlying hazard function. Following the notation in Bouman et al. (Bouman et al. 2005, 2007), we denote the total cumulative baseline hazard $H(t_J)$ as $H_{0,0}$, and the hazard increments d_1 as $H_{M,0}$, d_2 as $H_{M,1}, \ldots$, and d_J as $H_{M,2^M-1}$. We then build the multiresolution hazard tree by recursively defining $H_{m-1,p} = H_{m,2p} + H_{m,2p+1}$, for $m = 1, \ldots, M$, and $p = 0, \ldots, 2^{m-1} - 1$. We refer to m as the level of resolution and p as the position within that level. Thus, at the top of this hazard tree we have the total cumulative hazard $H(t_J)$, which we split into finer components with each additional level of the tree, until we finally end at the the bottom of the tree (the highest level of resolution) with the hazard increments $d_1, \ldots, d_J$. If we further define $R_{m,p} \equiv H_{m,2p}/H_{m-1,p}$, we can parametrize the hazard increments by $H_{0,0}$ and the "splits" $R_{1,0}, \ldots, R_{M,2^{M-1}-1}$ (denoted $\mathbf{R}_{m,p}$). For example, when $M = 3$ (implying $J = 8$) we have: $d_1 = H_{0,0}R_{1,0}R_{2,0}R_{3,0}$, $d_2 = H_{0,0}R_{1,0}R_{2,0}(1 - R_{3,0}), \ldots, d_8 = H_{0,0}(1 - R_{1,0})(1 - R_{2,1})(1 - R_{3,3})$.

It is important to note that the piecewise-constant hazard assumption has been em-

ployed multiple times in the literature (for example, Walker and Mallick (1997). However, the uniqueness of the MRH model lies in its clever construction of the tree-based prior for a piece-wise constant function, so that the prior essentially does not depend on the final resolution level M (i.e., it is invariant to the height of the tree). More precisely, integrating out higher-level parameters, one would obtain the exact same prior as if that level and its parameters had simply not been considered in the first place. To aid with the understanding of the MRH model, a simple diagram of the two-level multiresolution prior is given in Figure 1.

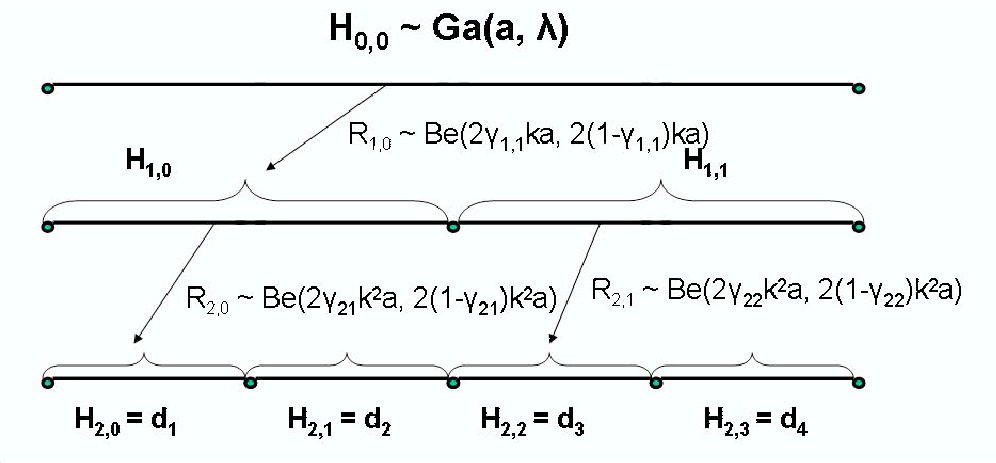

Figure 1: Diagram illustrating the multiresolution prior for the hazard rate function, with 2 levels (i.e., with resolution 2).

As in Nowak and Kolaczyk (Nowak and Kolaczyk 2000), we place a beta prior on each $R_{m,p}$ and a gamma prior on H. The shape parameters of each of these beta priors and the hyperparameters for H determine the prior expectations of the hazard increments, $d_j^*, j = 1, \ldots, J$. Furthermore, to allow for extra smoothness in the multiresolution prior, Bouman et al. (2007) introduce a multiplier for the shape parameters of the beta priors at each additional level of the hierarchy, denoted by k. Proceeding in this fashion, the priors for H and each $R_{m,p}$ when $M = 3$ and $J = 8$ are:

$$\begin{aligned} H \sim & \quad \mathcal{G}a(a, \lambda), \\ R_{1,0} \sim & \quad \mathcal{B}e(2\gamma_{1,0}ka, 2(1-\gamma_{1,0})ka), \\ R_{2,p} \sim & \quad \mathcal{B}e(2\gamma_{2,p}k^2a, 2(1-\gamma_{2,p})k^2a), \quad p = 0, 1 \\ R_{3,p} \sim & \quad \mathcal{B}e(2\gamma_{3,p}k^3a, 2(1-\gamma_{3,p})k^3a), \quad p = 0, 1, 2, 3. \end{aligned}$$

Note that under this prior structure, $E(R_{m,p}) = \gamma_{m,p}$, which, because H and $\mathbf{R}_{m,p}$ are independent *a priori*, easily allows one to choose $\gamma_{m,p}$, λ and a so that the prior expectation of the baseline hazard in each time interval j is any value d_j^* desired. This

particular formulation of the gamma-beta tree also determines the prior correlation of the d_j as a function (among other things) of k and a. Specifically, when $k = 0.5$, the baseline hazard increments d_j are *a priori* independent gamma random variables. Choosing k less or greater than 0.5 yields, respectively, negative (rougher hazard) or positive (smoother hazard) prior correlation among the d_j's (Bouman et al. 2005, 2007). Smoother hazard may in particular be employed in problems with much censoring. It is also possible to treat k as a hyperparameter and estimate it jointly with other parameters.

It is well known that MRH models, because they are based on a tree-like structure, may have a blocky correlation pattern; for example, it is possible that two neighboring hazard increments are less correlated than those further apart which happen to share more ancestral split parameters. Bouman et al. (Bouman et al. 2005) propose placing a hyperprior on the hazard H shape parameter a to even out the prior correlations among hazard increments and bypass this rather counterintuitive property.

3.3 Hierarchical Multiresolution Hazard Model

It may sometimes be of interest or necessity to relax the proportional hazards assumption, permitting different baseline hazards in particular groups. These strata-specific hazards may be treated as fixed, or as random (infinite dimensional) strata-specific parameters. This could be desirable in particular when we have data from multiple centers or multiple studies, when one needs to allow for differences in baseline hazards due to unobserved covariate processes, or to account for correlation within subjects from the same strata.

In breast cancer, it is well-known that women with ER- and ER+ tumors have different expected prognosis due to association of ER with both tumor pathology and clinical characteristics such as patient age (Hess et al. 2003). Because we are primarily interested in estimating the hazard shapes, we wish to avoid imposing any proportionality constraint on ER in the model. In any case, Figure 2, showing recurrence-free survival curves by ER and treatment, clearly illustrates deviation from proportional hazards between ER- and ER+ patients. While proportionality appears to hold better between treatment groups within ER categories (Fig. 2), we also wish to permit a different hazard shape according to treatment type, as biologic hypotheses concerning the action of adjuvant therapy would suggest the possibility of different shapes (Skipper 1971). Thus, we define strata defined by the ER by treatment group combinations (i.e., ER- surgery only, ER- chemotherapy, ER+ surgery only (placebo), ER+ tamoxifen).

More specifically, in this model we allow the hazard for each of the strata to be *a priori* an independent and identically distributed random MRH variable. For each stratum $s, s \in \{1, 2, 3, 4\}$, we draw the stratum cumulative hazard H_s from a $\mathcal{G}a(a_s, \lambda_s)$, and then draw the stratum set of splits $\mathbf{R}_{m,p,s}$. For this reason we will call this model the hierarchical MRH or the HMRH for short.

To complete the prior for the hazard rate, we place a zero-truncated Poisson (ZTP) hyperprior with mean μ_a on each a_s: $e^{-\mu_a}{\mu_a}^a/[a!(1-e^{-\mu_a})]$, and we allow each scale

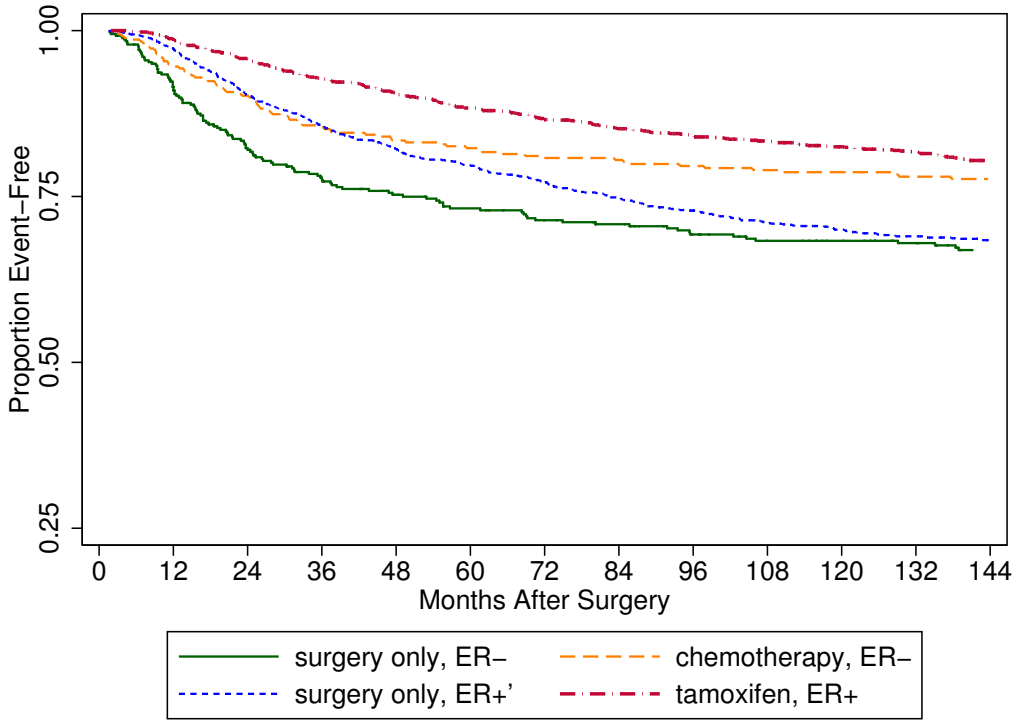

Figure 2: Kaplan-Meier estimates of recurrence-free survival among early stage breast cancer patients from NSABP clinical trials. Failures continue to occur in all groups many years after initial diagnosis and surgery.

parameter λ_s, the parameter of the total cumulative hazards $H_s(t_J)$, to follow an exponential distribution with mean μ_λ. The parameters k_s can be also given exponential priors with mean μ_k, but they could also be fixed if specific smoothness (positive or negative correlation) is desired *a priori*.

The continuous-time complete-data likelihood is based on the proportional-hazards model: $h_s(t|\mathbf{X}, \boldsymbol{\beta}) = \exp(\mathbf{X}'\boldsymbol{\beta})h_{\text{base},s}(t)$, where s denotes the stratum group. For a patient i in stratum s, if the failure time $T_{i,s} \in [0, t_J]$ is observed without censoring, the likelihood function is:

$$L_{i,s}(\boldsymbol{\beta} \mid T_{i,s}, \mathbf{X}_{i,s}) = \exp(\mathbf{X}'_{i,s}\boldsymbol{\beta})h_{\text{base},s}(T_{i,s})S_{\text{base},s}(T_{i,s})^{\exp(X'_{i,s}\boldsymbol{\beta})}. \quad (1)$$

When an observation is right-censored, i.e., $T_{i,s} > t_{cens}$ for some $t_{cens} \leq t_J$, we have:

$$L_{i,s}(\boldsymbol{\beta} \mid T_{i,s}, \mathbf{X}_{i,s}) = S_s(t_{cens}|\mathbf{X}_{i,s}, \boldsymbol{\beta}) = S_{\text{base},s}(t_{cens})^{\exp(X'_{i,s}\boldsymbol{\beta})}. \quad (2)$$

Here, each h_s is a priori distributed as an MRH variable, with independent total hazard and split parameters. The vector of covariates $\mathbf{X}_{i,s}$ contains patient and disease characteristics (age, tumor size, tumor progesterone receptor level, etc.) for patient i in stratum s.

4 Markov chain Monte Carlo

The Gibbs sampling algorithm used to simulate the parameter posterior and its sequence of full conditional posterior distributions for parameters from all strata is outlined below. The details of the simpler model are provided in Bouman et al. (2005, 2007).

The likelihood for patient i in stratum s, whose failure (or censoring) time is $T_{i,s}$, the censoring indicator $\delta_{i,s}$, and the covariates $\mathbf{X}_{i,s}$, is

$$L(T_{i,s} \mid \delta_{i,s}, \boldsymbol{\beta}, H_s, \mathbf{R}_{m,p,s}, \mathbf{X}_{i,s}) = \left[\exp(\mathbf{X}'_{i,s}\boldsymbol{\beta}) h_{\text{base},s}(T_{i,s})\right]^{\delta_{i,s}} \exp(-\exp(\mathbf{X}'_{i,s}\boldsymbol{\beta}) H_{\text{base},s}(\min(T_{i,s}, t_J))). \quad (3)$$

The log-likelihood for all $N = \sum_{s=1}^{4} N_s$ patients is thus:

$$\delta' \left[\mathbf{X}\boldsymbol{\beta} + \mathbf{F_s}\boldsymbol{\Pi}\tilde{\mathbf{R}}\right] - \sum_{s=1}^{4}\sum_{i=1}^{N_s} \exp(\mathbf{X}'_{i,s}\boldsymbol{\beta}) H_{\text{base},s}(\min(T_{i,s}, t_J)) \quad (4)$$

where δ is the vector of censoring indicators for all patients, $\mathbf{X}$ is the $N \times L$ matrix of covariates and $\boldsymbol{\Pi}$ is the $2^M \times (2^{M+1} - 1)$ multiresolution tree matrix. In that matrix, the (i, j) element is 1 when $j = 1$ or $i \in [1 + (j \bmod 2^{\mathrm{m}}), \ldots, 2^{\mathrm{M-m}} + (\mathrm{j} \bmod 2^{\mathrm{m}})]$, $\mathrm{m} = \lfloor \log_2(\mathrm{j}) \rfloor$, and 0 otherwise. $\tilde{\mathbf{R}}$ is the multiresolution log-parameter vector $(\log(H), \log(R_{1,0}), \log(1 - R_{1,0}), \ldots, \log(R_{M,2^M-1}), \log(1 - R_{M,2^M-1}))$. $\mathbf{F}$ is an $N \times 2^M$ matrix for which the (i, j) element is 1 if the ith patient (among all patients in all strata put together) has $T_i \in (t_{j-1}, t_j]$, and 0 otherwise; patients with $T_i > t_J$ have $F_{i,j} = 0, j = 1, \ldots, J$ (see Bouman et al. (2005, 2007) for details).

The Gibbs sampler steps (Geman and Geman 1984) for the parameters H_s, $R_{m,p,s}$, λ_s, a_s, and k_s, for all strata s are the same:

1. Sample H_s from its full conditional density:
   $\pi(H_s|\lambda_s, a_s, \mathbf{R}_{m,p,s}) = \mathcal{G}a\left((a_s + \sum_{i=1}^{N_s} \delta_{i,s}), 1/\left[(1/\lambda_s) + \sum_{i=1}^{N_s} \mathcal{F}(T_{i,s})\right]\right)$,
   with mean $\mu = (a_s + \sum_{i=1}^{N_s} \delta_{i,s})/\left[(1/\lambda_s) + \sum_{i=1}^{N_s} \mathcal{F}(T_{i,s})\right]$,
   where $\mathcal{F}(T_{i,s}) = H_s(\min(T_{i,s}, t_J))/H_s(t_J)$ is a function of the $T_{i,s}$ and $\mathbf{R}_{m,p,s}$.
2. Sample each $R_{m,p,s}$ from the full conditional $\pi(R_{m,p,s}|k_s, a_s, H_s)$
3. Sample k from $\pi(k_s|a_s, \mathbf{R}_{m,p,s}, \boldsymbol{\beta})$, λ_s from $\pi(\lambda_s|H_s, a_s, \boldsymbol{\beta})$, and a_s from $\pi(a_s|H_s, \mathbf{R}_{m,p,s}, \lambda_s, \boldsymbol{\beta})$.
4. Sample $\boldsymbol{\beta}$ from $\pi(\boldsymbol{\beta}|H_s, \mathbf{R}_{m,p,s})$.

Similarly as in Bouman et al. (2005, 2007), the full conditional posterior distributions for H_s are gamma, while the full conditional distributions for each $R_{m,p,s}$ and $\boldsymbol{\beta}$ are log-concave and therefore easy to sample from (using Gilks and Wild (1992) algorithm, for example). On the other hand, the full conditional distributions for hyperparameters λ_s, a_s, and k_s are in general more difficult to sample from as they are not log-concave. We recommend following Bouman et al. (2005) who use the rejection Metropolis sampling (see Gilks et al. (1995)).

5 Analysis of the Recurrence Hazard after Breast Cancer

5.1 Covariate Effect Estimation

The 16- and 32-bin multiresolution model with the "flat" prior hazard rate for each stratum (with all $\gamma_{m,p,s}$ and all k_s set to 0.5), was fit using output from Gibbs sampler chains with 12,000 iterations each, with the first 2,000 iterations of each chain discarded as burn-in. Every 5^{th} iteration was retained to reduce correlation between adjacent draws. The Gelman-Rubin diagnostics, performed separately for each parameter, were used to establish convergence.

Table 1: Posterior Credible Intervals for Predictor Effects, 16-bin model

	Tumor Size	Standardized PGR	Standardized Age
2.5%	0.0148	−0.151	−0.281
50%	0.0199	−0.066	−0.217
97.5%	0.0251	0.007	−0.149

Table 1 gives marginal 95% posterior credible intervals for covariates considered. Tumor size was measured in millimeters (mm), ranging from 0 to 60mm. Progesterone receptor concentration (PGR) was standardized using the sample mean of 139.17 and standard deviation of 294.59. Age was standardized using the sample mean of 53.27 and standard deviation of 10.42 years. Larger tumor size and younger age at diagnosis were found associated with increased recurrence hazard: within each stratum, an increase of 10.42 years (1 standard deviation) in age resulted approximately in 20% reduction, while an increase of 1mm in tumor size resulted in approximately 2% increase in recurrence hazard. A higher concentration of progesterone receptors on the tumor is weakly indicative of lesser failure risk. While menopausal status is generally an important factor in breast cancer prognosis, here it was only marginally associated with recurrence hazard after stratification by estrogen receptor status and inclusion of age at diagnosis in the model, and so was omitted from further consideration. The direction and magnitude of these covariate effects are essentially consistent with other prognostic factor studies of these patients.

5.2 Hazard Function Estimation

Figure 3 displays by treatment/ER strata the median posterior estimates for the 16-bin baseline hazard increments (corresponding to constant 11.3-month hazard function values). This 16-dimensional vector is a discrete approximation to the baseline hazard rate $h(t)$, estimated via the hazard increments $d_j = \int_{t_{j-1}}^{t_j} h(s)ds$. In Figure 4, we show a smoothed version of the same hazard estimates. Smoothing was performed via the median-spline method, with a hazard value of zero included at time zero for each stratum, to reflect the fact that patients are considered cancer-free immediately after surgery and essentially do not fail until some time has elapsed.

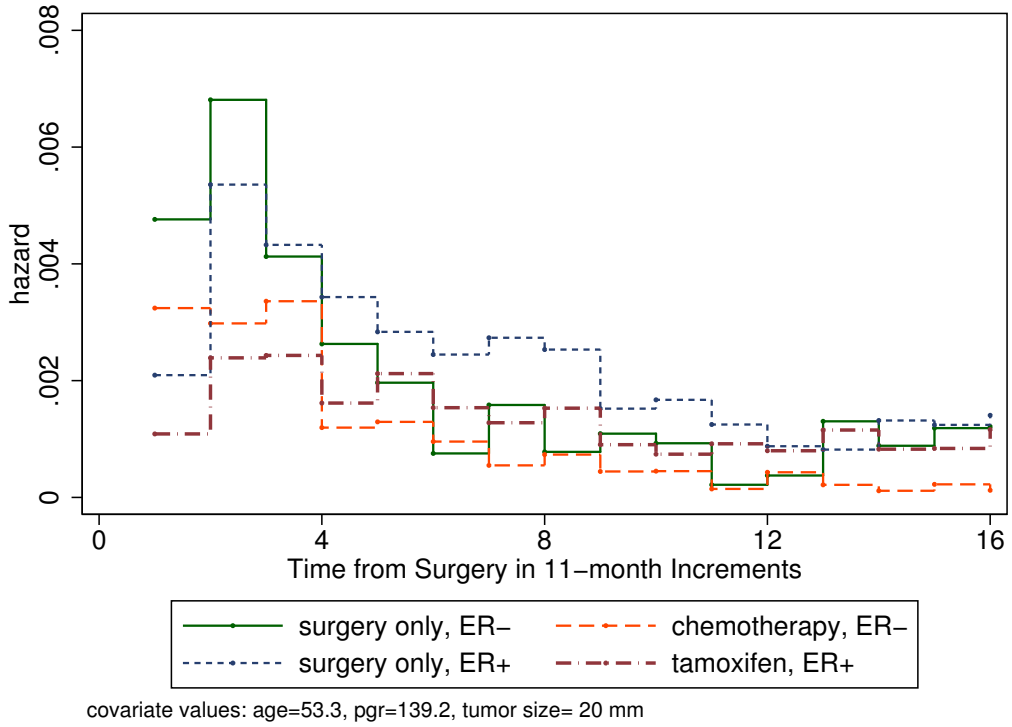

Figure 3: Discrete recurrence hazard increments for the 16-bin HMRH model. Horizontal sections represent the hazard value within the approximate 11-month increment in time from surgery.

Several notable features are seen. First, all four groups have the distinctive hazard peak around 12-24 months after surgery, with the ER- groups experiencing the peak a bit earlier. This pattern is similar to that noted by Hess et al. (2003) in their study of recurrence hazards by ER status. Second, the peak is greatest for the ER- patients

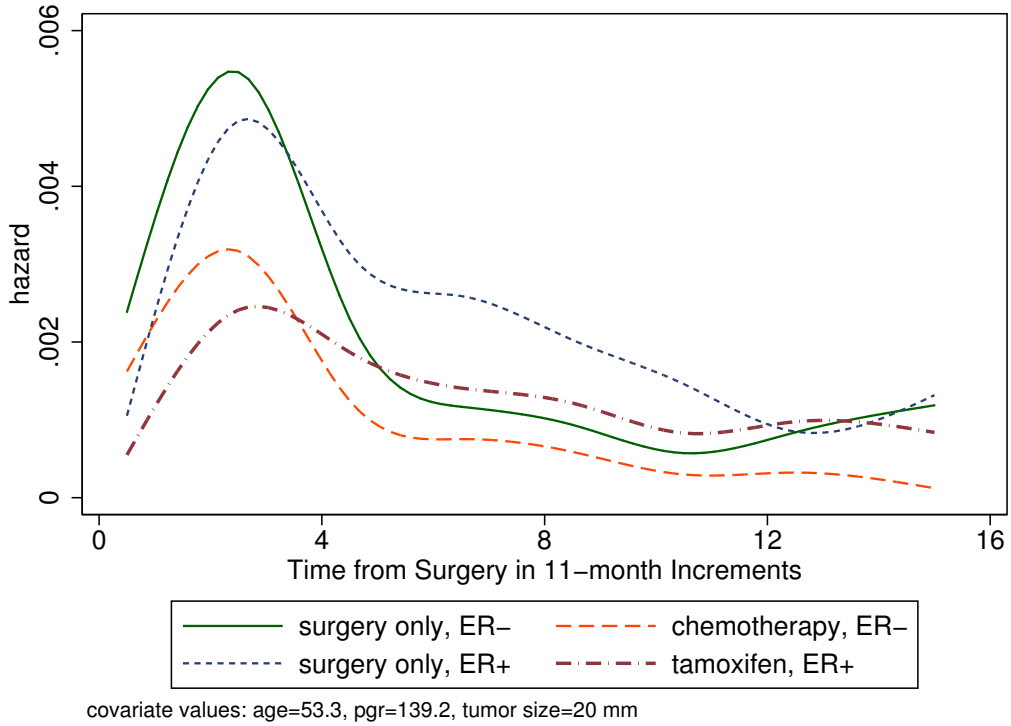

Figure 4: Smoothed recurrence hazards for the 16-bin model.

receiving surgery only, and is substantially reduced by chemotherapy, being lower than untreated patients from the more favorable ER+ group. The lowest peak is among ER+ patients randomized to tamoxifen. Interestingly however, at longer follow-up times the hazard in this group is no lower than that of chemotherapy treated and even untreated ER- patient groups, both of which have smaller hazard than untreated ER+ patients. Ultimately, the ER- chemotherapy treated group has the lowest recurrence hazard.

Figure 5 shows pointwise posterior credible intervals, based on 2.5% and 97.5% estimated posterior percentiles, for the four strata. With the exception of time points around the hazard peak, credible intervals for the four hazard estimates tend to overlap, particularly at longer follow-up times. Thus, we currently cannot reliably conclude whether there is a crossover of failure hazard among the groups at later time points. We should note that in other analyses collapsing across one or the other stratification factors (ER or treatment) and treating ER or treatment as covariates in modeling, large treatment effects within ER groups were apparent, while differences between ER groups within treatment modalities (surgery, adjuvant) were large initially but attenuated over time, substantively violating the proportional hazards assumption, as is evident in Fig-

ure 2.

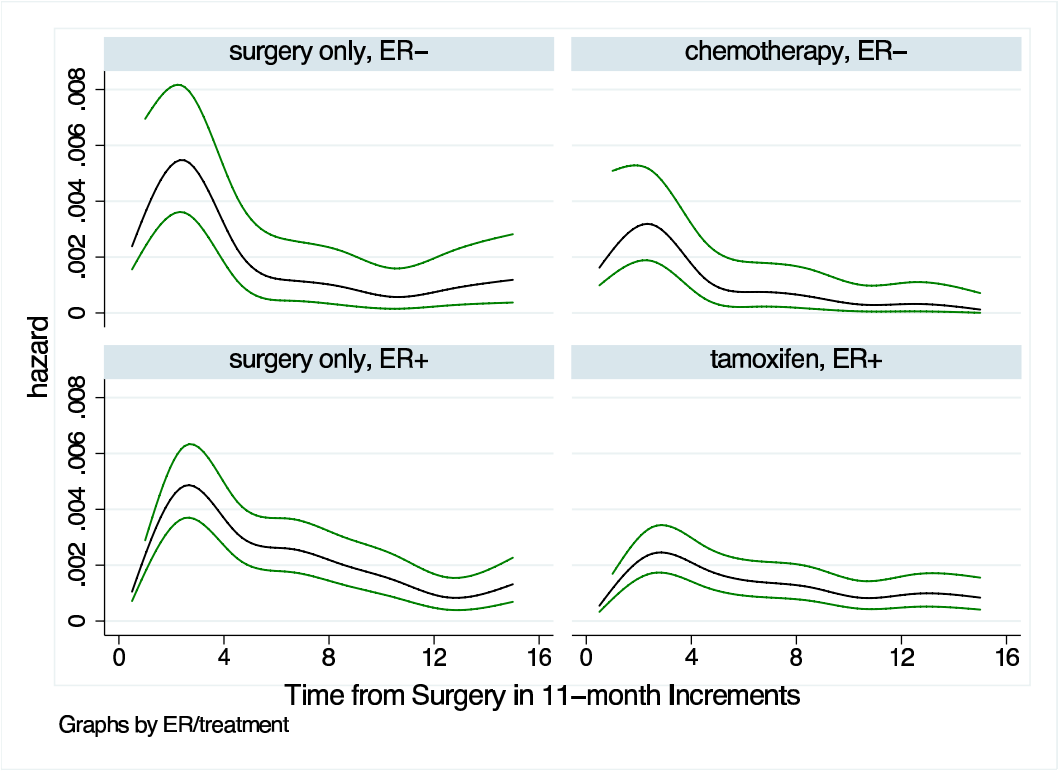

Figure 5: Credible intervals for the recurrence hazards from Figure 4 by ER and treatment strata.

5.3 Sensitivity Analysis

We now turn to the question of robustness of our model results. An alternate model was based on 32 time intervals of length just under 6 months each, instead of the 16 intervals of approximate length of 1 year. The purpose of this sensitivity check was twofold: first, to compare the estimates of the hazard rates for each of the four groups under these alternative resolutions; and second, to assess the impact of additional resolution level onto the estimates of the covariates (age, tumor progesterone receptor level, and tumor size). Note that in this data set, we could not reasonably use a finer resolution than the 32-bin choice for two reasons: 1) although this is a large cohort of patients, prognosis is relatively favorable and so failures are few if the intervals (bins) are very small, and 2) follow-up visits do not happen more frequently than once every 3 to 6 months, and while recurrence events can occur in continuous time, recurrence events that are not clinically apparent will be detected at these visits, causing clustering of events (and consequently

artificial periodicity in the hazard function estimate) if very small time bins are used.

Table 2: Posterior Credible Intervals for Predictor Effects, 32-bin model

	Tumor Size	Standardized PGR	Standardized Age
2.5%	0.0150	−0.147	−0.281
50%	0.0200	−0.066	−0.215
97.5%	0.0250	0.009	−0.147

Based on the invariance properties of the MRH prior, one can expect that some of that invariance is preserved in the posterior as well; and we observe this to a large degree. For example, a 16-bin hazard estimate obtained by fitting the 32-bin model and then aggregating the neighboring hazard increments is almost indistinguishable from the original 16-bin model hazard estimate. In addition, hazard rate plots using the 32-bin model and the 16-bin model are very similar as well; as expected, the 32-bin estimates are slightly more variable, but all hazard shapes are very similar (not shown).

With respect to covariate effect estimates, the effects of increasing the resolution is minimal. Compared to Table 1, the effects shown in Table 2 differ by a negligible amount.

6 Discussion

We have illustrated the application of a flexible extension of the familiar Cox proportional hazards model to jointly estimate covariate effects and separate hazard rate functions for several patient strata. This approach allows us to incorporate covariate effects and perform inference related to shapes and change-points in the hazard over time, our primary interest in the problem of recurrence after early stage breast cancer. The estimation and examination of the hazard functions directly reveals important patterns not readily apparent from quantities such as the survival functions. However, the hazard function remains a difficult quantity to draw robust inference from, as even in this large dataset, estimates suggest potentially important differences in shape, but variability estimates preclude any definitive conclusions pending additional analyses, as discussed below.

The observation that among those patients with initially higher risk disease (i.e., those with ER- tumors), the fraction escaping the early failure risk go on to have substantially lower long-term failure risk than those with initially more favorable prognosis (ER+ tumors), has significant implications in both clinical management and considerations regarding further developments of adjuvant therapies. In fact, there has been much recent interest in the development of 'switching' strategies whereby women with ER+ tumors discontinue tamoxifen and begin use of other hormonal treatments, in order to extend and improve on the benefit of this treatment modality. Currently, little

is known about what factors might be key to optimizing the switching strategy. For ER- patients, newer chemotherapy and molecularly targeted agents that act on specific tumor vulnerabilities may offer the best opportunity for a *bona fide* cure once early failure is avoided.

In addition to the well-known initial wave of failures following surgery, in recent years investigators have suggested that additional reproducible patterns are manifest in the recurrence hazard in the intermediate to long-term follow-up period. The notion of a bimodal or "double-peaked" recurrence hazard has been proposed, where after the first period of increased failure risk at 1-5.3.0 years post-surgery, the decline in failure hazard is followed by a second peak centered roughly around 8 years (Demicheli et al. 1996, 2001; Baum et al. 2005). A number of cancer biologic hypotheses have been put forth regarding the meaning and cause of a possible double-peaked failure pattern. For example, it has been conjectured that growth kinetics perturbed by surgery may contribute to the first wave in failure hazard, while heterogeneous disseminated tumor cells that require more time to become established may account for the latter peak in failure (Demicheli et al. 2001; Baum et al. 2005). This idea may seem to harken back to the naive concept that cancer surgery "spreads" cancer, but the influence of surgery on growth kinetics does have foundation in substantive biologic theory (Fisher et al. 1983). However, before any such interpretations of the hazard shape can be made or gain further credibility, more rigorous analytic methods such as those proposed here must be applied. Furthermore, it may be difficult to uniquely ascribe such a pattern to specific biologic phenomena, because other circumstances, such as the existence of patient mixtures due to unrecognized factors present at diagnosis or apparent hazard spikes caused by clustering of failures in time due to discovery of subclinical disease around certain time landmarks (e.g., mandatory 5-year post-diagnosis screen) would be expected to produce similar patterns. Nonetheless, this intriguing concept merits further investigation.

Our future work on this problem will involve the inclusion of data from trials conducted subsequent to those included here. As the trials are designed in a hierarchical fashion, these studies share some treatment arms in common with the current data, but also include newer treatment regimens. Extension of the model to more data sources will involve incorporation of 'trial' effects to allow for heterogeneity among common treatment arms. The inclusion of additional data will permit a more thorough exploration of changes in the hazard over time and more robust inference (including "collapsibility" of neighboring intervals in some regions), due to the considerably larger sample size that will result in narrower bounds on estimated hazards. This analysis will also have greater biologic and clinical relevance as we explore more recent drug regimens designed to reduce recurrence risk in women with breast cancer.

References

Aalen, O. and Gjessing, H. (2001). "Understanding the shape of the hazard rate: A process point of view." *Statistical Science*, 16: 1–22. 592

Andersen, P., Borgan, O., Gill, R., and Keiding, N. (1993). *Statistical Methods Based on Counting Processes*. Berlin: Springer–Verlag. 595

Baum, M., Demicheli, R., Hrushesky, W., and Retsky, M. (2005). "Does surgery unfavourably perturb the "natural history" of early breast cancer by accelerating the appearance of distant metastases?" *European Journal of Cancer*, 41: 508–515. 606

Berg, J. and Robbins, G. (1966). "Factors influencing short and long term survival of breast cancer patients." *Surgical and Gynecologic Obstetrics*, 122: 1311–1316. 593

Bouman, P., Dignam, J., Dukic, V., and Meng, X. (2007). "A multiresolution hazard model for multi-center survival studies: Application to tamoxifen treatment in early stage breast cancer." *Journal of the American Statistical Association*, in press. 591, 592, 595, 596, 597, 598, 600, 601

Bouman, P., Dukic, V., and Meng, X. (2005). "Bayesian multiresolution hazard model with application to an AIDS reporting delay study." *Statistica Sinica*, 15: 325–357. 591, 592, 595, 596, 598, 600, 601

Bryant, J., Fisher, B., Gunduz, N., Costantino, J., and Emir, B. (1997). "S-phase fraction combined with other patient and tumor characteristics for the prognosis of node-negative, estrogen-receptor-positive breast cancer." *Breast Cancer Research and Treatment*, 51: 239–253. 594

Cox, D. (1972). "Regression models and life tables." *Journal of the Royal Statistical Society - Series B*, 34: 187–220. 595

Demicheli, R., , Valagussa, P., and Bonadonna, G. (2001). "Does surgery modify growth kinetics of breast cancer micrometastases?" *British Journal of Cancer*, 85: 490–492. 606

Demicheli, R., Abbattista, A., Miceli, R., Valagussa, P., and Bonadonna, G. (1996). "Time distribution of the recurrence risk for breast cancer patients undergoing mastectomy: further support about the concept of tumor dormancy." *Breast Cancer Research and Treatment*, 41: 177–185. 593, 606

Fisher, B., Constantino, J., Redmond, C., Poisson, R., Bowman, D., Couture, J., Dimitrov, N., Wolmark, N., Wickerham, D., and Fisher, E. (1989a). "A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors." *The New England Journal of Medicine*, 320: 479–484. 594

Fisher, B., Dignam, J., Bryant, J., DeCillis, A., Wickerham, D., Wolmark, N., J, J. C., Redmond, C., Fisher, E., Bowman, D., Deschenes, D., Dimitrov, N., Margolese, R., Robidoux, A., Shibata, H., Terz, J., Paterson, A., Feldman, M., Farrar, W., Evans, J., and Lickley, H. (1996a). "Five versus more than five years of tamoxifen therapy for breast cancer patients with negative lymph nodes and estrogen receptor positive tumors." *Journal of the National Cancer Institute*, 88: 1529–1542. 594

Fisher, B., Dignam, J., Mamounas, E., Costantino, J., Wickerham, D., Redmond, C., Wolmark, N., Dimitrov, N., Bowman, D., Glass, A., Atkins, J., Abramson, N., Sutherland, C., Aron, B., and Margolese, R. (1996b). "Sequential methotrexate and fluorouracil for the treatment of node-negative breast cancer patients with estrogen receptor-negative tumors: eight-year results from National Surgical Adjuvant Breast and Bowel Project (NSABP) B-13 and first report of findings from NSABP B-19 comparing methotrexate and fluorouracil with conventional cyclophosphamide, methotrexate, and fluorouracil." *Journal of Clinical Oncology*, 14: 1982–1992. 594

Fisher, B., Gunduz, N., and Saffer, E. (1983). "Influence of the interval between primary tumor removal and chemotherapy on kinetics and growth of metastases." *Cancer Research*, 43: 1488–1492. 606

Fisher, B., Redmond, C., Dimitrov, N., Bowman, D., Legault-Poisson, S., Wickerham, D., Wolmark, N., Fisher, E., Margolese, R., and Sutherland, C. (1989b). "A randomized clinical trial evaluating sequential methotrexate and fluorouracil in the treatment of patients with node-negative breast cancer who have estrogen-receptor-negative tumors." *The New England Journal of Medicine*, 320: 473–478. 594

Fisher, B., Redmond, C., Wickerham, D., Wolmark, N., Bowman, D., Couture, J., Dimitrov, N., Margolese, R., Legault-Poisson, S., and Robidoux, A. (1989c). "Systemic therapy in patients with node-negative breast cancer. A commentary based on two National Surgical Adjuvant Breast and Bowel Project (NSABP) clinical trials." *Annals of Internal Medicine*, 111: 703–712. 594

Geman, S. and Geman, D. (1984). "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6: 721–741. 600

Gilks, W., Best, N., and Tan, K. (1995). "Adaptive rejection Metropolis sampling." *Applied Statistics*, 44: 455–472. 601

Gilks, W. and Wild, P. (1992). "Adaptive rejection sampling for Gibbs sampling." *Applied Statistics*, 41: 337–348. 601

Gordon, N. (1990). "Application of the theory of finite mixtures for the estimation of 'cure'." *Statistics in Medicine*, 9: 397–407. 593

Gray, R. (1990). "Some diagnostic methods for Cox regression models through hazard smoothing." *Biometrics*, 46: 93–102. 595

— (1992). "Flexible methods for analyzing survival data using splines, with application to breast cancer prognosis." *Journal of the American Statistical Association*, 87: 942–951. 595

— (1996). "Hazard rate regression using ordinary nonparametric regression smoothers." *Journal of Computational and Graphical Statistics*, 5: 190–207. 595

Hess, K., Pusztai, L., Buzdar, A., and Hortobagyi, G. (2003). "Estrogen receptors and distinct patterns of breast cancer relapse." *Breast Cancer Research and Treatment*, 78: 105–118. 593, 598, 602

Karrison, T., Ferguson, D., and Meier, P. (1999). "Dormancy of mammary carcinoma after mastectomy." *Journal of the National Cancer Institute*, 91: 80–85. 593

Kolaczyk, E. (1999). "Bayesian multiscale models for Poisson processes." *Journal of the American Statistical Association*, 94: 920–933. 596

Nowak, R. and Kolaczyk, E. (2000). "A statistical multiscale framework for Poisson inverse problems." *IEEE Transactions on Information Theory*, 46: 1811–1825. 596, 597

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). "The Analysis of Failure Times in the Presence of Competing Risks." *Biometrics*, 34: 541–554. 595

Saphner, T., Tormey, D., and Gray, R. (1996). "Annual hazard rates of recurrence for breast cancer after primary therapy." *Journal of Clinical Oncology*, 14: 2738–2746. 593

Skipper, H. (1971). "Kinetics of mammary tumor cell growth and implications for therapy." *Cancer*, 28: 1479–1499. 598

Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002). "Bayesian measures of model complexity and fit (with discussion)." *Journal of the Royal Statistical Society - Series B*, 64: 583–616. 596

Walker, S. and Mallick, B. (1997). "Hierarchical Generalized Linear Models and Frailty Models with Bayesian Nonparameteric Mixing." *Journal of the Royal Statistical Society - Series B*, 59: 845–860. 597

Acknowledgments

Support for this research was partially provided by the Susan G. Komen Breast Cancer Foundation and Public Health Service grant NCI-U10-CA-69651 from Institutes of Health, Department of Health and Human Services.