



Interim recruitment prediction for multi-center clinical trials

SZYMON URBAS*

STOR-i Centre for Doctoral Training, Lancaster University, Lancaster, UK
s.urbas@lancaster.ac.uk

CHRIS SHERLOCK

Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

PAUL METCALFE

Data Science Solutions, AstraZeneca, Cambridge, UK

SUMMARY

We introduce a general framework for monitoring, modeling, and predicting the recruitment to multi-center clinical trials. The work is motivated by overly optimistic and narrow prediction intervals produced by existing time-homogeneous recruitment models for multi-center recruitment. We first present two tests for detection of decay in recruitment rates, together with a power study. We then introduce a model based on the inhomogeneous Poisson process with monotonically decaying intensity, motivated by recruitment trends observed in oncology trials. The general form of the model permits adaptation to any parametric curve-shape. A general method for constructing sensible parameter priors is provided and Bayesian model averaging is used for making predictions which account for the uncertainty in both the parameters and the model. The validity of the method and its robustness to misspecification are tested using simulated datasets. The new methodology is then applied to oncology trial data, where we make interim accrual predictions, comparing them to those obtained by existing methods, and indicate where unexpected changes in the accrual pattern occur.

Keywords: Bayesian prediction modeling; clinical trial recruitment; inhomogeneous Poisson process; model averaging; Poisson-gamma model

1. INTRODUCTION

Efficiently recruiting patients to clinical trials is a critical factor in running clinical trials and hence delivering new medicines to patients as quickly as possible. Late-stage clinical trials are commonly run across many sites, and successfully managing and running trials and subsequent processes require accurate forecasts of trial recruitment.

Early recruitment rates can be high, for example, because patients with the required condition are already available, and rates can then drop once these patients have been recruited. Deterministic approaches and ad

*To whom correspondence should be addressed.

hoc techniques may yield simplified and, often, overly optimistic recruitment timelines, a phenomenon thus dubbed *Lasagna's Law* (Lasagna, 1979). For example, 48% of centers studied by Getz and Lamberti (2013) failed to enroll the required number of patients in the time originally allocated, leading to extensions of the recruitment timelines and the need to bring more centers into the study, which itself is a costly process. The timelines are usually pushed to nearly twice the originally proposed plan. The most frequent reason for trial discontinuation appears to be poor recruitment; out of 253 discontinued trials studied in Kasenda and others (2014), 101 were terminated due to under-recruitment.

This motivates the need for robust statistical methods for modeling and predicting the recruitment to clinical trials at site level. Early detections of possible center underperformance may allow practitioners to swiftly intervene in the operations. It can also provide realistic timelines for the completion of different stages of the trials.

In this work, we introduce a novel flexible framework for effectively modeling and predicting patient recruitment. We will focus on the oncology therapeutic area as it is known for sparse enrollments whose patterns are not sufficiently captured by the state-of-the-art methods (Anisimov and Fedorov, 2007; Lan and others, 2018). Our framework utilizes time-varying recruitment rates whilst also permitting variation between recruitment centers. Inference is based on the set of known center initiation times to date, whilst the prediction is conditional on a set of future initiation times. Past initiation times are known, but typically, whilst there is a plan for future initiation times along with potential contingencies, the actual times are not known precisely in advance. The proposed methodology can be used with user-specified initiation schedules to facilitate the choice between different initiation-time scenarios, or it can be combined with a center-initiation model. Predictions of future recruitment incorporate parameter and model uncertainty, which is essential when data are limited.

Existing methods for predicting recruitment to clinical trials are overviewed in Section 2. Section 3 outlines methods for detecting recruitment rate decay in the multi-center recruitment setting along with result of a Monte Carlo power study. Section 4 introduces the flexible modeling framework, and Section 5 presents a general method for choosing sensible Bayesian parameter priors, along with an appropriate posterior sampling method and diagnostics. A simulation study is presented in Section 6, illustrating the fitting of the model, model validation, and forecasting recruitment using Bayesian model-averaging. In Section 7, the model is fitted to an oncology dataset, and this is followed by a discussion in Section 8.

2. EXISTING METHODS

The first statistical modeling framework for clinical trial recruitment was introduced in Lee (1983), where the recruitment was assumed to be a constant-rate Poisson process, leading to tractable inference based on interim data. Williford and others (1987) built on the model by considering Bayesian inference with conjugate priors. Gajewski and others (2008) and Jiang and others (2015) further explored the effects various prior densities can have on predictions. Time-inhomogeneous accrual was first considered in Piantadosi and Patterson (1987), where the aggregated accrual across all sites was modeled as an inhomogeneous Poisson process with intensity $\lambda(t) = \zeta(1 - \exp\{-\kappa t\})$, $\zeta, \kappa > 0$. Zhang and Long (2010) took a non-parametric approach, using B-splines to model the trends in accrual and using the intensity value at the census time for predictions. Tang and others (2012) proposed a Poisson model with a piece-wise linear intensity which captured aspects of recruitment such as slow initial recruitment and a spike in recruitment close to the end of the trial. For a more thorough review of these as well as other methods, see Heitjan and others (2015). Accrual-only modeling methods do not consider the effect that initiating new centers can have on recruitment trends. For that reason, we shall focus on methods which can take advantage of center-specific recruitment data.

Anisimov and Fedorov (2007) introduced the Poisson-gamma (PG) model of recruitment in a multi-center setting, with the main appeal being the use of random effects for the recruitment rates of centers,

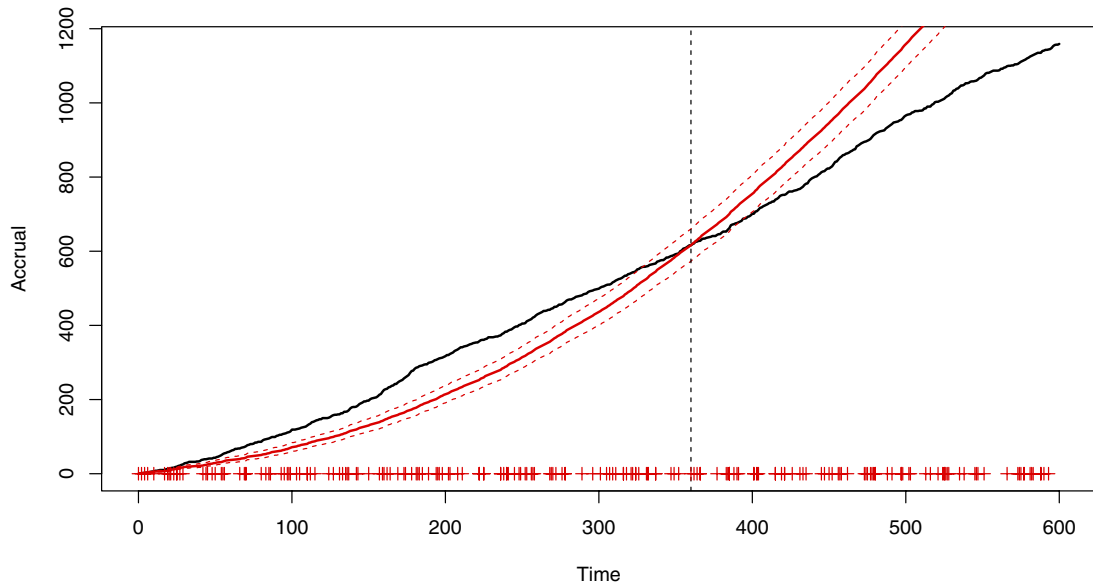


Fig. 1. Accrual (black, solid) with the predictive mean (red, solid) and 95% prediction bands (red, dashed), based on the PG model (2.1) with the census time marked by the vertical, dashed line.

providing a tractable, data-driven prior predictive distribution for recruitment in yet-unopened centers. The model consists of C centers, each recruiting N_c patients over τ_c days, $c = 1, \dots, C$. The framework makes the following distributional assumptions,

$$\begin{aligned} \lambda_c &\sim \text{Gamma}(\alpha, \alpha/\phi), \\ N_c | \lambda_c &\sim \text{Pois}(\lambda_c \tau_c), \end{aligned} \quad c = 1, \dots, C. \quad (2.1)$$

The random effect λ_c is the *recruitment rate* for center c . The rates, and thus the center recruitments, are assumed to be independent conditional on α and ϕ . There are, however, several caveats with the approach taken. The article advocates using the Empirical Bayes approach, that is, maximum likelihood estimation for the hierarchical parameters (α, ϕ) followed by re-estimation of the distribution of random effect λ_c given α , ϕ , and n_c , for each center. A method for obtaining the uncertainty in the hierarchical (α, ϕ) parameters is provided, but this uncertainty is not accounted for when making predictions, leading to overly confident prediction intervals. However, the main issue which could result from employing the model arises from the strong assumption of time-homogeneity of center recruitments, which can lead to underestimations of the time to completion.

Figure 1 shows the accrual in a simulated trial where the rates gradually decay with time as well as the predictive distribution of the PG model fitted at a census time of three-fifths of the total length of the study; the initiation day for each center is marked. The accrual appears to follow a straight line which could initially suggest using a time-homogeneous model. However, new centers are constantly being initiated so that a constant recruitment rate for each center leads to an upward arching trend in accrual. This is encapsulated by the fitted predictive. Here, the accrual is initially badly underestimated and then grossly overestimated after the census time. The apparent “matching” at the census time is due to predictions using re-estimated random-effect distributions.

Lan and others (2018) describe the first multi-center recruitment model in which the rates decrease over time. The model assumes inhomogeneous Poisson for arrivals center c with an intensity of the form

$$\lambda_c(t) = \begin{cases} \lambda_c^o, & t < t_o \\ \lambda_c^o \exp\{-\theta(t - t_o)\}, & t \geq t_o \end{cases},$$

where λ_c^o is a gamma random effect, as in (2.1), and t_o a user-specified parameter and is not estimated as part of the inference. By enforcing the specific intensity-form, the possibilities of time-homogeneous recruitments or even intensity decays with heavier tails are excluded. A more systematic alternative is to start by testing the time-homogeneity assumption.

3. DETECTING TIME-INHOMOGENEITY

Given series of daily center recruitment counts over the recruitment period of τ_c days, $\{N_c(t)\}_{t=1}^{\tau_c}$, $c = 1, \dots, C$, we can test the hypothesis of time-homogeneity. To detect a decay in the rate, we only need to use the sums $X_1^{(c)} = \sum_{t=1}^{\tau_c/2} N_c(t)$ and $X_2^{(c)} = \sum_{t=\tau_c/2+1}^{\tau_c} N_c(t)$ ($c = 1, \dots, C$), whose expectations we denote by $\mu_1^{(c)}$ and $\mu_2^{(c)}$, respectively. Detecting time-inhomogeneity in a single center can be difficult as the infrequent counts will lead to low powers of tests (Krishnamoorthy and Thomson, 2004) (see also Tables 1 and 2). Thus, we combine the recruitments across all centers leading to two counts: $X_1 = \sum_{c=1}^C X_1^{(c)}$ and $X_2 = \sum_{c=1}^C X_2^{(c)}$, and we choose our hypotheses to be $H_0 : \sum_{c=1}^C \mu_1^{(c)} = \sum_{c=1}^C \mu_2^{(c)}$ vs $H_1 : \sum_{c=1}^C \mu_1^{(c)} > \sum_{c=1}^C \mu_2^{(c)}$.

The tests are one-sided as we are only interested in recruitment which decays over time. We consider tests with respect to the following assumptions:

Table 1. Power for likelihood-ratio test

$E[X_1]$	$R = 1$	$R = 0.9$	$R = 0.8$	$R = 0.7$	$R = 0.6$	$R = 0.5$
5	0.06	0.08	0.11	0.15	0.20	0.27
10	0.05	0.08	0.12	0.18	0.26	0.37
20	0.05	0.09	0.17	0.27	0.41	0.58
50	0.05	0.13	0.28	0.50	0.73	0.90
100	0.05	0.18	0.44	0.75	0.94	0.99
200	0.05	0.27	0.68	0.95	1.00	1.00

Table 2. Power for non-parametric bootstrap test

$E[X_1]$	$R = 1$	$R = 0.9$	$R = 0.8$	$R = 0.7$	$R = 0.6$	$R = 0.5$
5	0.04	0.06	0.08	0.11	0.14	0.18
10	0.05	0.08	0.12	0.16	0.24	0.33
20	0.05	0.10	0.16	0.25	0.39	0.57
50	0.05	0.14	0.28	0.48	0.70	0.88
100	0.05	0.18	0.42	0.74	0.93	0.99
200	0.05	0.28	0.67	0.94	1.00	1.00

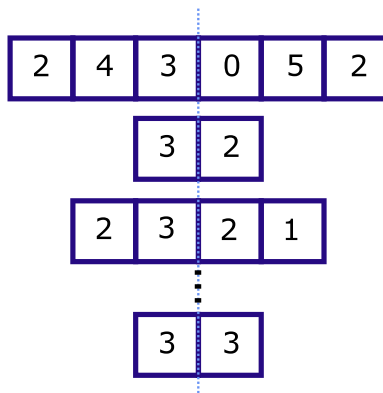


Fig. 2. Site-level count series are all centered and the sum of all the first halves is compared to the sum of second halves.

Assumption 1: For each center $c = 1, \dots, C$, the counts in the first and second halves of that center's recruitment period are independent and have the same distribution, $X_1^{(c)} \stackrel{d}{=} X_2^{(c)}$, with expectation $\mu_1^{(c)}$. Furthermore, the recruitments at each center are independent of each other.

Assumption 2: The patients arrive according to a Poisson process such that $X_1^{(c)}, X_2^{(c)} \sim \text{Pois}(\mu_1^{(c)})$, for some $\mu_1^{(c)}, c = 1, \dots, C$.

Assumption 1 implies that X_1 and X_2 must have the same distributions, with respective expectations $\mu_1 = \sum_{c=1}^C \mu_1^{(c)}$ and $\mu_2 = \sum_{c=1}^C \mu_2^{(c)}$ being equal. Assumption 2 further implies that the distributions must be Poisson. Figure 2 shows the construction of the quantities X_1 and X_2 by aligning the centers of the recruiting periods. The splitting of the series halfway is arbitrary, though splitting it in half (or at least close to this) would theoretically yield the highest power. It assumes that the τ_c are even. However, centers recruiting over odd numbers of days can still be used by removing the middle day observation. This reduces the power of the tests, though the reduction is negligible.

Gu and others (2008) offer a detailed Monte Carlo study of the different methods used for testing for a difference in means of two Poisson variables. Here, we focus on the ones most applicable to the clinical-trial recruitment setting, bearing in mind statistical power and robustness. We identified two methods: the non-parametric bootstrapped test (BST), which is powerful yet robust, and the Poisson likelihood-ratio test (LRT), which makes stronger distribution assumptions to achieve an even higher power. The BST only assumes that the counts in each day are independent and identically distributed (Assumption 1). With this assumption, resampling within each center with replacement, from the original data would still produce a valid sample from the assumed distribution under H_0 . A large number of bootstrap samples is used to simulate the distribution of the difference in two means, which is then used to test the hypothesis. Appendix A of the Supplementary material available at *Biostatistics* online details the sampling procedure for obtaining the distribution and the p -value.

For the LRT, we require Assumption 2, which is already an underlying assumption for the model in Anisimov and Fedorov (2007). Upon aggregation, the two sums follow Poisson distributions, that is, $X_1 \sim \text{Pois}(\mu_1)$ and $X_2 \sim \text{Pois}(\mu_2)$. The likelihood under the null model ($\mu_1 = \mu_2$) is compared to the likelihood under the alternative two-mean model ($\mu_1 > \mu_2$). Here, the likelihood function is

$$L(\mu_1, \mu_2 | x_1, x_2) = \frac{\mu_1^{x_1} \exp\{-\mu_1\}}{x_1!} \frac{\mu_2^{x_2} \exp\{-\mu_2\}}{x_2!}, \quad \mu_1, \mu_2 > 0.$$

We let

$$T_L(x_1, x_2) = \begin{cases} 2[\log L(\hat{\mu}_1, \hat{\mu}_2 | x_1, x_2) - \log L(\hat{\mu}, \hat{\mu} | x_1, x_2)], & \hat{\mu}_1 > \hat{\mu}_2 \\ 0, & \hat{\mu}_1 \leq \hat{\mu}_2 \end{cases},$$

where $\hat{\mu}$ is the MLE under the null, and $\hat{\mu}_1$ and $\hat{\mu}_2$ are the MLEs under the alternative hypothesis. Under the null, we would expect the test statistic $T_L(X_1, X_2)$ to asymptotically be zero half the time with the other half following a χ_1^2 distribution (Robertson and others, 1988). When using the LRT, the simulated significance levels can differ from the pre-specified level when μ values are low. This is due to using the asymptotic χ^2 distribution when calculating the p -value (Gu and others, 2008).

The performance of the two tests was assessed by carrying out a Monte Carlo study. Test powers were estimated using Poisson data with different expectations and ratios, $R = \mu_2/\mu_1$. For the LRT power estimates, 5×10^6 samples were used as the test itself is very computationally cheap. For the BST, 5×10^4 samples were used, with each test using a bootstrapped distribution of size 10^3 . Tables 1 and 2 show the results of the study. The biggest difference in powers occurs for lower expectations, with the LRT outperforming BST. It must be noted, however, that the BST only requires the data to be i.i.d. within each center and thus is robust to violations of the Poisson assumption; if the counts within each center are overdispersed, for example, it does not affect the Type I error.

To exemplify the usefulness of this test, we can consider an interim likelihood ratio test where the expected number of enrollments is 170. This corresponds to $E[X_1] = 100$ and $R = 0.7$, for example, and results in a statistical power of approximately 0.75. Considering many trials require an upward of 500 enrollments, informed decisions can be made relatively early on in the trial.

4. PROPOSED MODEL

We consider a scenario of C centers recruiting patients, with each center c being initiated for τ_c days. The number recruited by center c on day t shall be denoted by $N_c^{(t)}$. We propose the following modeling framework for the multi-center clinical-trial recruitment, based on the inhomogeneous Poisson process,

$$\begin{aligned} \lambda_c^o &\sim \text{Gamma}\left(\alpha, \frac{\alpha}{\phi}\right), & c = 1, \dots, C, \\ N_c^{(t)} &\sim \text{Pois}\left(\lambda_c^o \int_{t-1}^t g(s; \theta) ds\right), & t = 1, \dots, \tau_c, \end{aligned}$$

where g is a non-negative function which dictates the curve-shape of the intensity and θ is a parameter (or parameter vector) associated with the functional form. We use the (α, ϕ) parametrization for the hierarchical gamma distribution as it leads to orthogonality of α and ϕ in the Poisson-gamma model (Huzurbazar, 1950). *A priori*, $E[\lambda_c] = \phi$ and $V[\lambda_c] = \phi^2/\alpha$. For notational simplicity, we define $G(t; \theta) = \int_0^t g(s; \theta) ds$. The likelihood contribution from center c is

$$\begin{aligned} \Pr(\mathbf{N}_c = \mathbf{n}_c | \lambda_c^o, \theta, \tau_c) &= \prod_{t=1}^{\tau_c} \Pr(N_c^{(t)} = n_c^{(t)} | \lambda_c^o, \theta) \\ &= \exp\{-\lambda_c^o G(\tau_c; \theta)\} (\lambda_c^o)^{n_c^{(c)}} \prod_{t=1}^{\tau_c} \frac{[G(t; \theta) - G(t-1; \theta)]^{n_c^{(t)}}}{n_c^{(t)}!}, \end{aligned}$$

where $n_c^{(i)} = \sum_{t=1}^{\tau_c} n_c^{(t)}$. Marginalizing over the random-effect component gives

$$\Pr(\mathbf{N}_c = \mathbf{n}_c | \alpha, \phi, \theta, \tau_c) = \frac{(\alpha/\phi)^\alpha \Gamma(\alpha + n_c^{(i)})}{\Gamma(\alpha)[G(\tau_c; \theta) + \alpha/\phi]^{(\alpha+n_c^{(i)})}} \prod_{t=1}^{\tau_c} \frac{[G(t; \theta) - G(t-1; \theta)]^{n_c^{(t)}}}{n_c^{(t)}!},$$

whence the full likelihood of the model given the recruitment data is:

$$\begin{aligned} L(\alpha, \phi, \theta | \mathbf{n}, \boldsymbol{\tau}) &= \prod_{c=1}^C \Pr(\mathbf{N}_c = \mathbf{n}_c | \alpha, \phi, \boldsymbol{\tau}) \\ &= \frac{(\alpha/\phi)^{C\alpha}}{\Gamma(\alpha)^C} \prod_{c=1}^C \frac{\Gamma(\alpha + n_c^{(i)})}{[G(\tau_c; \theta) + \alpha/\phi]^{(\alpha+n_c^{(i)})}} \prod_{t=1}^{\tau_c} \frac{[G(t; \theta) - G(t-1; \theta)]^{n_c^{(t)}}}{n_c^{(t)}!}. \end{aligned} \quad (4.2)$$

If all the centers had been recruiting for the same amount of time, that is, $\tau_c \equiv \tau \forall c$, then by fixing the integral of $g(t; \theta)$ over τ days we could introduce orthogonality between (α, ϕ) and θ by imposing the normalization: $\int_0^\tau g(t; \theta) dt = \tau$. This generalizes the homogeneous model with $g(t; \theta) = 1$ and leads to the following factorizable likelihood,

$$\begin{aligned} L(\alpha, \phi, \theta | \mathbf{n}, \boldsymbol{\tau}) &= \frac{(\alpha/\phi)^{C\alpha}}{\Gamma(\alpha)^C (\tau + \alpha/\phi)^{(C\alpha+n_\Sigma)}} \prod_{c=1}^C \Gamma(\alpha + n_c^{(i)}) \prod_{t=1}^{\tau_c} \frac{[G(t; \theta) - G(t-1; \theta)]^{n_c^{(t)}}}{n_c^{(t)}!} \\ &= L(\alpha, \phi | \mathbf{n}, \tau) L(\theta | \mathbf{n}, \boldsymbol{\tau}), \end{aligned} \quad (4.3)$$

where $n_\Sigma = \sum_{c=1}^C n_c^{(i)}$.

The factorization means that now the θ parameter describes the shape of the intensity only, and α and ϕ describe the distribution of the magnitude of the integrated intensity, leading to a more interpretable model. Even when centers are not all recruiting for the same length of time, we choose to impose a similar normalization using some representative τ , here $\frac{1}{C} \sum_{c=1}^C \tau_c$. As demonstrated empirically in Section 6, the condition leads to approximate orthogonality even when the centers are initiated uniformly throughout the study.

4.1. Intensity curve-shape

In this work, we will restrict our choice of curve-shape g to parametric forms. The functional form of g is arbitrary and the best choices may depend on the context of the problem. When working with oncology datasets, for each center we observe low-frequency counts which seem to become even less frequent over time but with varying tail behaviors. For this reason, we chose the following curve-shape

$$g_\kappa(t; \theta) \propto \left(1 + \frac{\theta t}{\kappa}\right)^{-\kappa}, \quad t \geq 0, \quad \theta, \kappa > 0. \quad (4.4)$$

The proportionality is used as multiplying g_κ by some positive constant and dividing ϕ by the same constant leads to the same model. The limit as $\kappa \rightarrow 0$ recovers the standard PG model (2.1); and letting

$\kappa \rightarrow \infty$, we obtain an exponential tail. The full (normalized) forms are then

$$g_0(t) \equiv 1, \quad (4.5)$$

$$g_1(t; \theta) = \frac{\theta(1 + \theta t)^{-1}}{\log(1 + \theta \tau)} \tau, \quad (4.6)$$

$$g_\kappa(t; \theta) = \frac{\theta(1 - \kappa)(1 + \theta t/\kappa)^{-\kappa}}{\kappa(1 + \theta \tau/\kappa)^{1-\kappa} - \kappa} \tau, \quad \kappa \notin \{0, 1, \infty\}, \quad (4.7)$$

$$g_\infty(t; \theta) = \frac{\theta \exp\{-\theta t\}}{1 - \exp\{-\theta \tau\}} \tau. \quad (4.8)$$

The associated integrated forms, $G_\kappa(t; \theta)$ are provided in [Appendix B](#) of the [Supplementary material](#) available at *Biostatistics* online.

The flexibility of the model, however, can result in potential identifiability issues. Inference methods, such as maximum likelihood, can run into numerical instabilities when $\kappa \gg 1 > \theta$ or $\kappa < 1 \ll \theta$ (see [Appendix B](#) of the [Supplementary material](#) available at *Biostatistics* online for details). For this reason, we recommend restricting the choice of κ to a discrete set of values; in this work, we use $\{0, 0.5, 1, 2, \infty\}$. This will be elaborated on in [Section 5.3](#).

5. INFERENCE, DIAGNOSTICS, AND PREDICTIONS

We aim to construct a framework which can provide reliable predictions whilst capturing uncertainty in the estimated parameters and in the underlying model itself. We employ the Bayesian paradigm since it naturally incorporates the distribution of the random effects, λ_c , with the uncertainty in the model and the parameter values. However, we note that in some scenarios frequentist methods may be preferred and give a brief outline of how one may employ them in [Appendix C](#) of the [Supplementary material](#) available at *Biostatistics* online.

Given a parametric statistical model, the Bayesian paradigm starts from a prior distribution for the parameters, here denoted $\pi_0(\alpha, \phi, \theta)$ and updates this according to some data, y , to provide a posterior distribution, here denoted by $\pi(\alpha, \phi, \theta|y)$. When multiple parametric models, M_k , $k = 1, \dots, K$, are being considered, the posterior probability for model k , here denoted by $\pi_p(M_k|y)$, may also be calculated. [Section E](#) of the [supplementary material](#) available at *Biostatistics* online provides more details on these quantities; see also [Robert and Casella \(2013\)](#) or [Gelman and others \(2013\)](#), for example.

For the models under consideration for trial-recruitment data, neither the posterior model probabilities nor the posteriors for the parameters for any particular model are tractable, and so we employ importance sampling to obtain Monte Carlo samples $(\alpha_m, \phi_m, \theta_m)$, $m = 1, \dots, M$ from the posterior distribution for any given model, as well as an estimate of $\pi(M_k)$, $k = 1, \dots, K$. [Appendix D](#) of the [Supplementary material](#) available at *Biostatistics* online provides further details of this method, as well as of effective sample size (ESS), a diagnostic which indicates the reliability of the Monte Carlo estimates; see also [Robert and Casella \(2013\)](#) or [Doucet and others \(2001\)](#).

In [Sections 6](#) and [7](#), we carry out inference on $\tilde{\alpha} = \log \alpha$, $\tilde{\phi} = \log \phi$, and $\tilde{\theta} = \log \theta$ since analyses of trial data showed the likelihood in the log-parameters to be more symmetric about the mode, which can make sampling more efficient. For the importance sampling proposal distribution, we use a multivariate t -distribution on four degrees of freedom, with the same mode as the posterior and the shape matrix equal to the inverse Hessian at the posterior mode.

5.1. Prior choices

We base our prior specification on a maximum likelihood meta-analysis of 20 oncology clinical trial recruitment datasets. The trials studied were for seven different types of cancers: ovarian, prostate, breast, small and non-small lung, bladder, and pancreatic. The number of centers ranged from 58 to 244 with a median of 140 and total enrollments ranged from 245 to 4391 with a median of 1035. In all cases, the parameter estimators were close to orthogonal justifying the use of independent priors: $\pi_0(\tilde{\alpha}, \tilde{\phi}, \tilde{\theta}) = \pi_0(\tilde{\alpha})\pi_0(\tilde{\phi})\pi_0(\tilde{\theta})$.

We found that the α parameter does not change much from one study to another. The weakly informative prior $\tilde{\alpha} \sim N(0.2, 2^2)$ sufficiently reflects the distribution of the estimated values.

The ϕ parameter estimates varied by orders of magnitude between studies. The parameter reflects the mean center recruitment and is well identified by the data; it depends upon the catchment region, type of indication and protocol, for example. For this reason, we advocate using a vague prior unless reliable expert knowledge is available. In our analyses, we used the uninformative, proper prior $\tilde{\phi} \sim U(-8, 8)$.

The difference between the homogeneous (4.5) and the inhomogeneous (4.6, 4.7, 4.8) models is the curve-shape parameter θ . Lindley's paradox (Lindley, 1957) warns that assigning θ a vague prior can lower the posterior probabilities of the models that use θ , compared to the model with $\kappa = 0$ which does not use θ . To avoid the paradox, we set an informative but sensible prior by considering the drop off in intensity after some time, t_0 . We let $R_\kappa = g_\kappa(t_0; \theta)/g_\kappa(0; \theta)$ and set $R_\kappa \sim \text{Beta}(a, b)$ a priori, with $a = b = 1.1$ to indicate a lack of information, excepting that this is not a constant intensity model, since this is covered by $\kappa = 0$, and that we do not expect a 100% drop off after a time of t_0 (expert opinion); here we take $t_0 = 4$ months. As R_κ is a monotonic function of θ , we can use a density transform to derive the corresponding prior for θ . If prior information is abundant, be it in the form of historical data or expert knowledge, the beta distribution parameters can be adjusted to reflect this. Given (4.4), the resulting prior density for $\tilde{\theta}$ is given in Appendix E of the Supplementary material available at *Biostatistics* online.

5.2. Predictive distribution

There are two complementary properties for which predictions might be required: the distribution of future recruitments within a set time interval, and the distribution of time until the target number of recruitments is reached. In this section, we focus on the former; details of the latter appear in Appendix F of the Supplementary material available at *Biostatistics* online.

Suppose we are interested in sampling the recruitment, denoted N_c^+ , at some day t^+ by center c . Given samples from the parameter posteriors, we can sample exactly from the posterior predictive for N_c^+ by exploiting the Poisson-gamma conjugacy of the random-effect distribution. The posterior distribution for the λ_c^o random effect for center c is

$$\lambda_c^o | \alpha, \phi, \theta, \mathbf{n}_c, \tau_c \sim \text{Gamma}(\alpha + n_c^{(\cdot)}, \alpha/\phi + G(\tau_c; \theta)) = \text{Gamma}\left(\alpha_c^*, \frac{\alpha_c^*}{\phi_c^*}\right), \quad (5.9)$$

where $\alpha_c^* = \alpha + n_c^{(\cdot)}$ and $\phi_c^* = \phi \times \left(\frac{\alpha + n_c^{(\cdot)}}{\alpha + \phi G(\tau_c; \theta)}\right)$. The predictive distribution for N_c^+ conditional on the random effect is:

$$N_c^+ | \lambda_c^o, \theta \sim \text{Pois}\left(\lambda_c^o \int_{t^+}^{t^+} g(s; \theta) ds\right) = \text{Pois}(\lambda_c^o G_\theta^+), \quad (5.10)$$

where $G_\theta^+ = \int_{t^+}^{t^+} g(s; \theta) ds$.

Marginalizing over the random effect posterior, we arrive at the negative binomial distribution:

$$\Pr(N_c^+ = n | \alpha_c^*, \phi_c^*) = \frac{\Gamma(\alpha_c^* + n)}{\Gamma(\alpha_c^*)n!} \left(\frac{\alpha_c^*}{\alpha_c^* + \phi_c^* G_\theta^+} \right)^{\alpha_c^*} \left(\frac{\phi_c^* G_\theta^+}{\alpha_c^* + \phi_c^* G_\theta^+} \right)^n, \quad n = 0, 1, 2, \dots \quad (5.11)$$

The length of interval to t^+ does not need to be a day and could instead be a week or a month, depending on the context of the application. To obtain the full marginal predictive, we sample the recruitments conditional on parameters sampled from the posterior. For as yet unopened centers, we set $n_c^{(i)} = \tau_c = 0$. For each triplet (or couplet, if $\kappa = 0$) of parameters sampled from the posterior, we sample N_c^+ , $c = 1, \dots, C$, and sum them to obtain a sample from $N^+ | \alpha, \phi, \theta$. The collection of these sums is a sample from the posterior predictive distribution for the model.

If simulations for multiple distinct time periods are required for a given center, c , as needed for the accrual curve for example, then we first sample λ_c^o from its posterior (5.9). We then simulate the Poisson counts for the individual time periods, which are conditionally independent given λ_c^o , from (5.10).

5.3. Model averaging

When predicting the enrollments using a fitted model, we implicitly assume that a single model best reflects reality; however, prediction methods should consider the uncertainty in the models used for inference. We shall, therefore, use model averaging for making predictions, that is, take a weighted average of predictions made by each model. Working in the Bayesian paradigm provides us with an intuitive choice for weights in the form of marginal likelihoods of the models.

$$\Pr(N^+ = n^+ | \mathbf{n}, \boldsymbol{\tau}) = \sum_{k=1}^K \Pr(N^+ = n^+ | \mathbf{n}, \boldsymbol{\tau}, M_k) \pi_p(M_k | \mathbf{n}, \boldsymbol{\tau}),$$

where $\pi_p(M_k | \mathbf{n}, \boldsymbol{\tau}) \propto \pi(\mathbf{n} | \boldsymbol{\tau}, M_k) \pi_0(M_k)$, $k = 1, \dots, K$, with $\pi_0(M_k)$ being prior model probabilities. The averaging framework fits in with the restriction of the shape parameter κ to a discrete space. Each κ value generates an inhomogeneous Poisson-gamma model with the tail behavior of the associated intensity shape. This includes the null ($\kappa = 0$) model as in Anisimov and Fedorov (2007). In this work, we set all prior model probabilities equal.

5.4. Model validation

Before making any statements in regards to the future recruitments, we should validate that the fitted model does indeed capture the true data-generating process sufficiently well. Since the true process is unknown, we compare the observed data to the modal model (the model with the highest posterior probability) fixed at posterior parameter means $(\hat{\alpha}, \hat{\phi}, \hat{\theta})$.

Firstly, we wish to assess that the chosen hierarchical structure is reflected in the data. The distribution of posterior means of the individual random effects should approximately follow the hierarchical Gamma($\hat{\alpha}, \hat{\alpha}/\hat{\phi}$) distribution. A QQ-plot can be used to visually compare the distributions. If deemed sufficiently similar, using the distribution for generating predictions for yet-unopened centers is appropriate. If the distributions are noticeably different, particularly if the true distribution is multimodal, any interim predictions for yet-unopened centers could (but need not; see robustness study in Section 6) be inaccurate.

According to the model, the counts in any initial period $[0, t']$ (such as the first month) of each center's recruitment period, follow a negative binomial distribution with shape parameter α and success probability $\phi G(t'; \theta) / (\alpha + \phi G(t'; \theta))$, similar to that given in (5.11) but using α and ϕ in place of α_c^* and ϕ_c^* . As the true parameters are unknown, we compare it to the distribution fixed at point-estimates $(\hat{\alpha}, \hat{\phi}, \hat{\theta})$. The

diagnostic indicates if the combination of the gamma random effects and the modal decay model captures the behavior over the initial period after center initiation. Again, a QQ-plot can be used for comparing the theoretical distribution to the observation, giving an indication if the fitted model under- or overestimates initial recruitment. The initial period, $[0, t']$, should be long enough that the true recruitment decay should be apparent. However, since only centers that have been recruiting for a period of at least t' can be used for the diagnostic, to ensure a reasonable power, t' should be short enough that a large number of sites have been recruiting for this duration. In this work, we set $t' = 60$ (2 months).

6. SIMULATION RESULTS

We demonstrate our flexible framework through a simulation study, using simulated data sets to illustrate model fit and prediction and to highlight the effect model misspecification can have on predictions. In practice, patterns in center initiation times can vary greatly between trials. For presenting the methodology, we consider an initiation schedule similar to that observed in a typical trial. We test the robustness of the method using a uniform initiation schedule, with another type of schedule examined in [Appendix G](#) of the [Supplementary material](#) available at *Biostatistics* online.

Our historical data set do not include the initiation times of the centers, so instead, to accurately reflect the historical data used in the meta-analysis and what is often available to researchers, we take the first recruitment time of a center as its initiation time and adjust the models to include a single deterministic recruitment at the initiation time of each center followed by stochastic recruitment as described in [Section 4](#).

We simulate a study over a course of 600 days, with 200 centers. The parameters used for simulations were $\alpha = 1.4$, $\phi = 0.01$, $\kappa = 2.7$, and $\theta = 0.02$. The inference is carried out on data observed in the first 360 days. As motivated in [Section 1](#), we condition the inference on a set of known initiation times, chosen by the practitioner; these could subsequently be varied to investigate the impact of different schedules or initiation models. We consider a set of models with flexible tails ([Section 4.1](#)) allowing $\kappa \in \{0, 0.5, 1, 2, \infty\}$, thus including the null model ([Anisimov and Fedorov, 2007](#)). The “normalization” of the curve-shapes was imposed at $\bar{\tau} = \frac{1}{C} \sum_{c=1}^C \tau_c$. We purposely simulated using a κ value outside of those considered in our models to illustrate the flexibility of the framework. For Bayesian inference, we used parameter and model priors outlined in [Sections 5.1](#) and [5.3](#), respectively. Based on the model fitted to the data at the census day 360, we wish to predict the daily accrual until day 600.

Performing the LRT and BST from [Section 3](#), we find the p -values of both tests to be < 0.001 . [Table 3](#) provides the fits for the five models. The effective samples sizes are high, which means that each of the model posteriors is represented well by its respective sample and that the marginal likelihood estimates are accurate. If the ESS values had been low, we would have retried using more samples in the importance sampler. We see that model corresponding to $\kappa = \infty$ has the highest posterior probability. A trellis plot of the posteriors for $(\tilde{\alpha}, \tilde{\phi}, \tilde{\theta})$ from the modal model (see [Appendix G](#) of the [Supplementary material](#) available at *Biostatistics* online) confirms at least approximate pairwise orthogonality between the parameters, as

Table 3. Posterior means and 95% credible intervals, posterior model probabilities and effective sample sizes, obtained using 10^4 importance samples for each model.

κ	α	ϕ	θ	$\pi(M_k \mathbf{n})$	ESS
0	1.423 (0.931, 2.239)	0.013 (0.011, 0.016)	—	6.90×10^{-37}	9000
0.5	1.632 (1.003, 2.750)	0.013 (0.011, 0.016)	0.152 (0.057, 0.495)	2.27×10^{-8}	8519
1	1.674 (1.059, 2.799)	0.013 (0.011, 0.016)	0.039 (0.025, 0.061)	3.21×10^{-3}	8471
2	1.677 (1.072, 2.693)	0.013 (0.011, 0.016)	0.021 (0.016, 0.027)	2.72×10^{-1}	8607
∞	1.690 (1.089, 2.801)	0.013 (0.011, 0.016)	0.011 (0.010, 0.013)	7.25×10^{-1}	8651

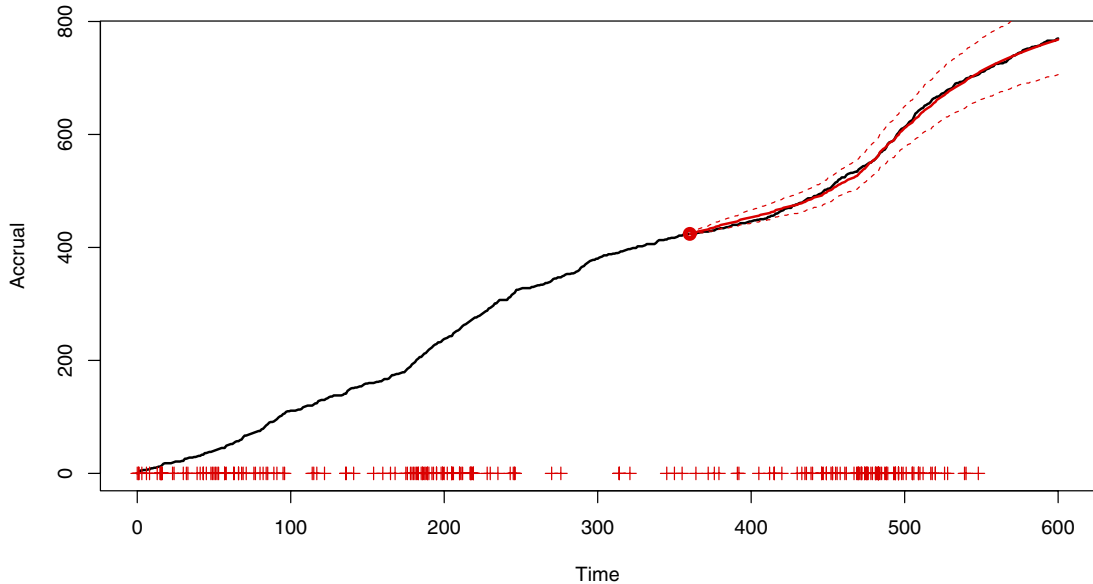


Fig. 3. Simulated accrual data from Section 6 with Bayesian model-averaged forecast predictive mean of the accrual (solid, red) and 95% prediction bands (red, dashed). Prediction bands are based on the 2.5% and 97.5% quantiles. The forecast begins from a point marked by the red dot and the “+” symbols on the abscissa indicate center initiation times.

anticipated from Sections 4 and 5.1. QQ-plots for the modal model comparing the hierarchical gamma distribution to the posterior means of the random effects and comparing the observed recruitments over the first two months of each center’s recruiting period to the model’s negative binomial distribution both show approximate straight lines with unit gradient and are provided in the [Supplementary material](#) available at *Biostatistics* online.

Figure 3 shows the accrual forecast from the census time $\tau = 360$ up to the horizon $\tau^H = 600$, superimposed onto the true accrual plot. The forecast is based on the Bayesian model-averaged posterior predictive distribution. The true accrual is contained within the 95% predictive intervals.

Figures 4 and 5 use an earlier census time ($\tau = 240$) to illustrate the issues that can arise when making predictions using maximum likelihood estimation and model selection. The inference was carried out with the same set of candidate models, and predictions were obtained by simulating from the best model ($\kappa = \infty$, chosen using AIC) with parameters fixed at the MLEs. As shown in the plots, not accounting for parameter and model uncertainty may lead to overly confident and biased predictions. Simulations with $\tau = 360$ (see [Supplementary material](#) available at *Biostatistics* online) still showed bias due to the choice of a single model, although the contrast with Figure 3 in terms of prediction interval width was less marked.

We repeated the analysis with a different distribution of initiation times, making the center initiations “clump” roughly every 2 months. The resulting forecast predictive distribution can be seen in Figure 6; performance appears to be robust to the type of initiation schedule.

To further test the robustness of the framework, we first consider the random effects λ_c^o now being generated from a mixture of two gamma distributions

$$\lambda_c^o | \alpha, \phi_1, \phi_2 \sim \frac{1}{2} \text{Gamma} \left(\alpha, \frac{\alpha}{\phi_1} \right) + \frac{1}{2} \text{Gamma} \left(\alpha, \frac{\alpha}{\phi_2} \right).$$

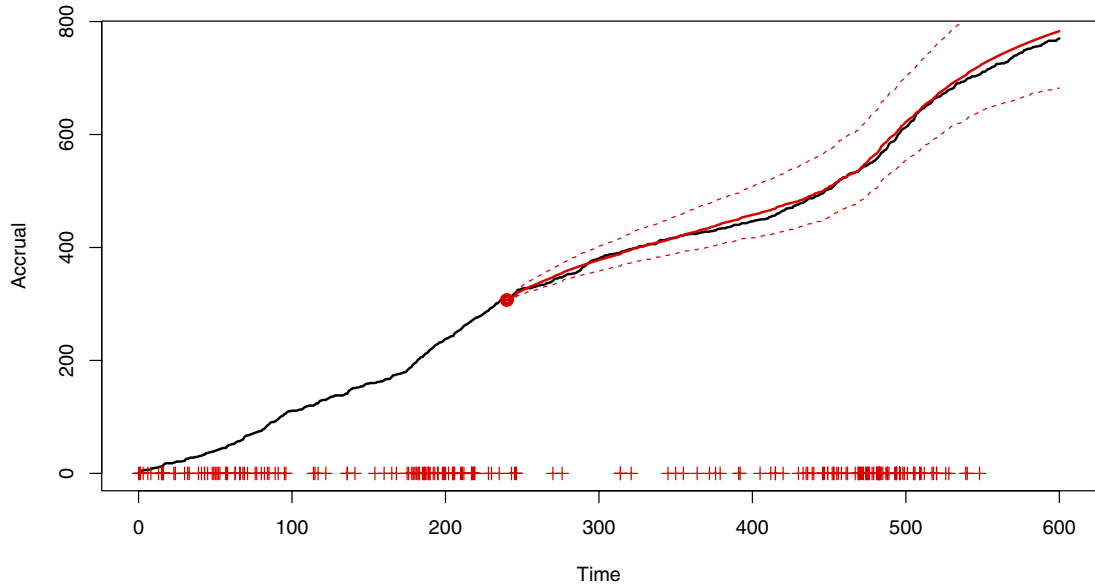


Fig. 4. Simulated accrual data from Section 6 with predictions using Bayesian model averaging.

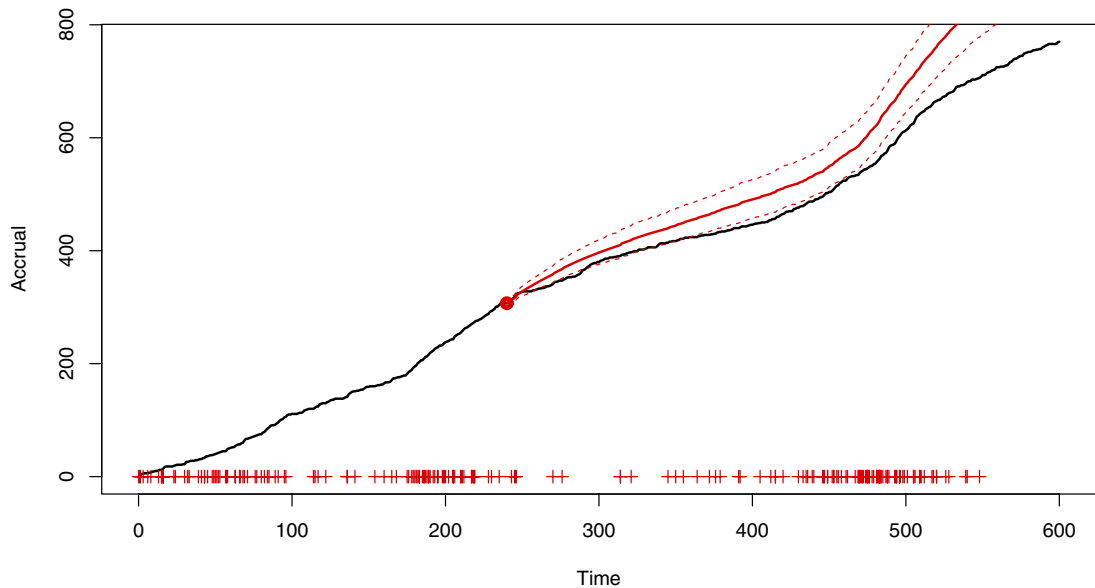


Fig. 5. Simulated accrual data from Section 6 with predictions using maximum likelihood and model selection.

We considered data generated using the same α value and curve-shape as before, but now with center initiation times uniformly sampled on the interval. The ratio of gamma expectations was fixed such that $\phi_2 = 10\phi_1$, and the random effect expectation, $E[\lambda_c^o] = (\phi_1 + \phi_2)/2$, was set to 0.01 and then 0.03. Figures 7 and 8 show example forecasts for accruals with the two different expectations. The more

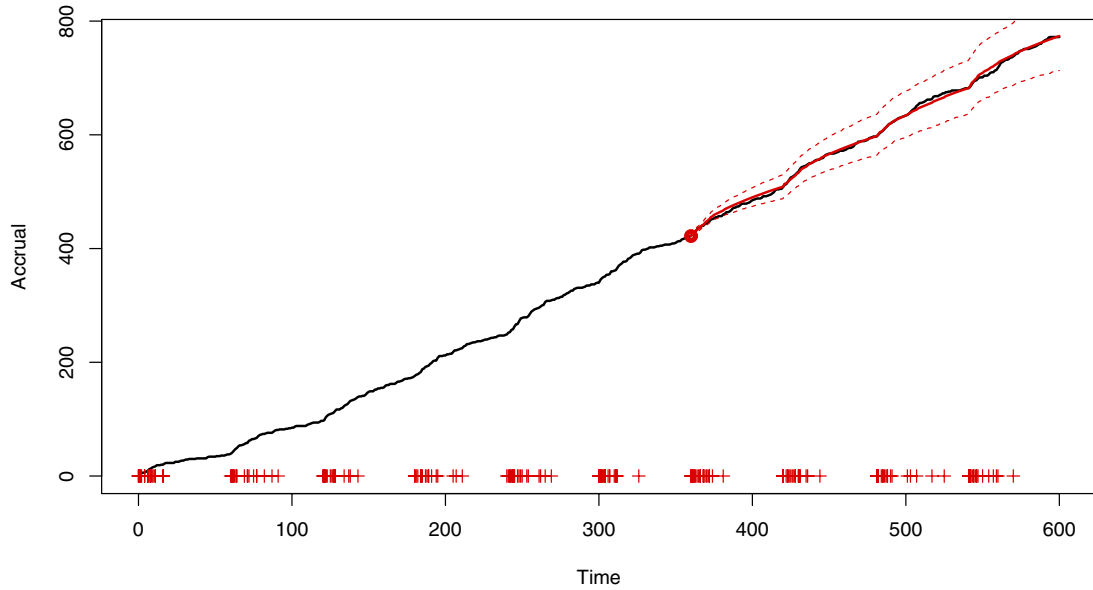


Fig. 6. Simulated accrual data from Section 6 using clumped initiations with forecast predictive accrual mean (solid, red) and 95% prediction bands (red, dashed). Prediction bands are based on the 2.5% and 97.5% quantiles. The forecast begins from a point marked by the red dot and the “+” symbols on the abscissa indicate center initiation times.

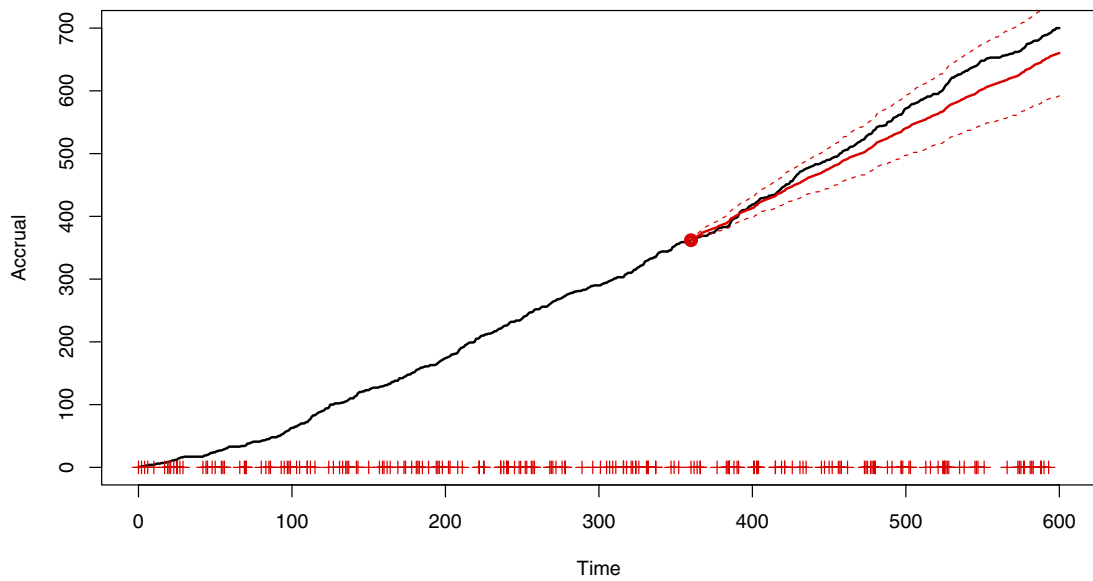


Fig. 7. Uniform initiations, $E[\lambda_c^2] = 0.01$ (mixture RE distribution)

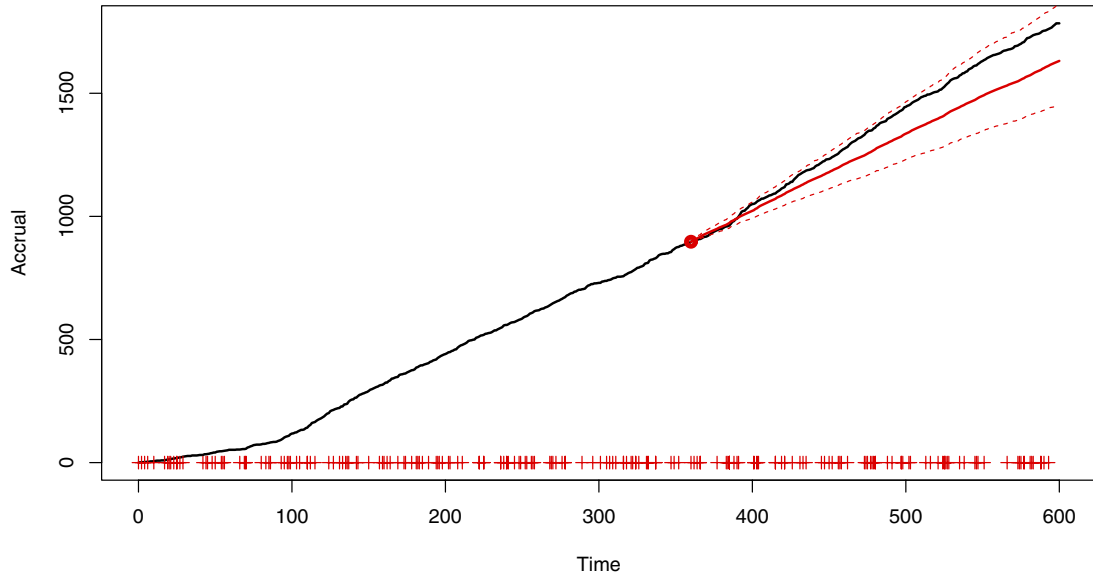


Fig. 8. Uniform initiations, $E[\lambda_c^o] = 0.03$ (mixture RE distribution)

data, that is, the larger $E[\lambda_c^o]$, the more apparent the discrepancy in the random-effect distribution, and the concomitant predictions, becomes. This is visible in the clearly non-linear diagnostic QQ-plots, and the plotted forecasts (see [Supplementary material](#) available at *Biostatistics* online). The robustness of predictions comes from the fact that the random effects for initiated centers use re-estimated data-driven distributions, reducing the importance of the random-effect prior; thus the main source of forecasting error comes from the incorrect random-effect prior for new centers. Similar plots for the “clumped” initiation schedule, provided in the [Supplementary material](#) available at *Biostatistics* online, show the same pattern. This mixture distribution of random effects represents the (extreme) scenario where roughly half of the centers recruit the vast majority of patients, with the remaining sites recruiting little to none each. When the ratio of the two means is closer to 1, the model still produces reliable predictions.

We also consider the effect of curve-shape misspecification on predictions, generating data using an intensity proportional to the Weibull density function

$$g_W(t; \theta, k) = \frac{k}{\theta} \left(\frac{t}{\theta}\right)^{k-1} \exp\left\{-\left(\frac{t}{\theta}\right)^k\right\} \tau, \quad \text{so} \quad G_W(t; \theta, k) = \frac{1 - \exp\left\{-\left(\frac{t}{\theta}\right)^k\right\}}{1 - \exp\left\{-\left(\frac{\tau}{\theta}\right)^k\right\}} \tau,$$

where $\theta, k > 0$. We simulated accrual datasets using the Weibull shape with $\theta = 30$ and $k = 1.5$, resulting in the highest recruitment rates occurring two weeks after center initiation. The random-effect distribution used $\alpha = 1.4$ and two different values ϕ were used: 0.01 and 0.03; Figures 9 and 10 show example forecasts. For lower overall recruitment levels, the model still predicts future accrual well. Forecast inaccuracies due to model misspecification become more apparent when larger recruitment rates are used. The same pattern is observed when center initiation times are clumped (see [Appendix G](#) of the [Supplementary material](#) available at *Biostatistics* online).

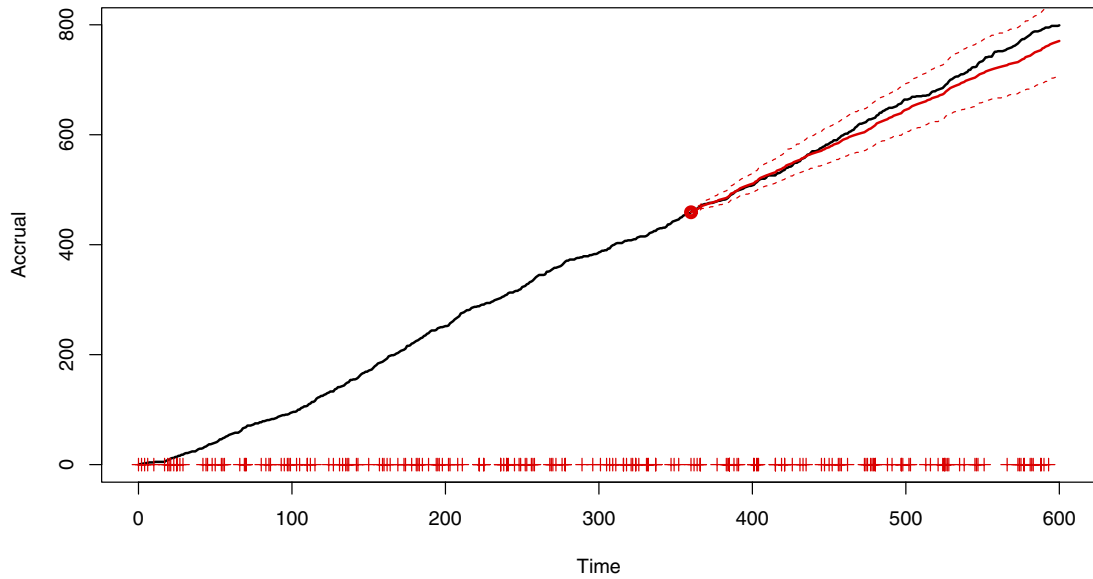


Fig. 9. Uniform initiations, $E[\lambda_c^2] = 0.01$ (Weibull-shape intensity)

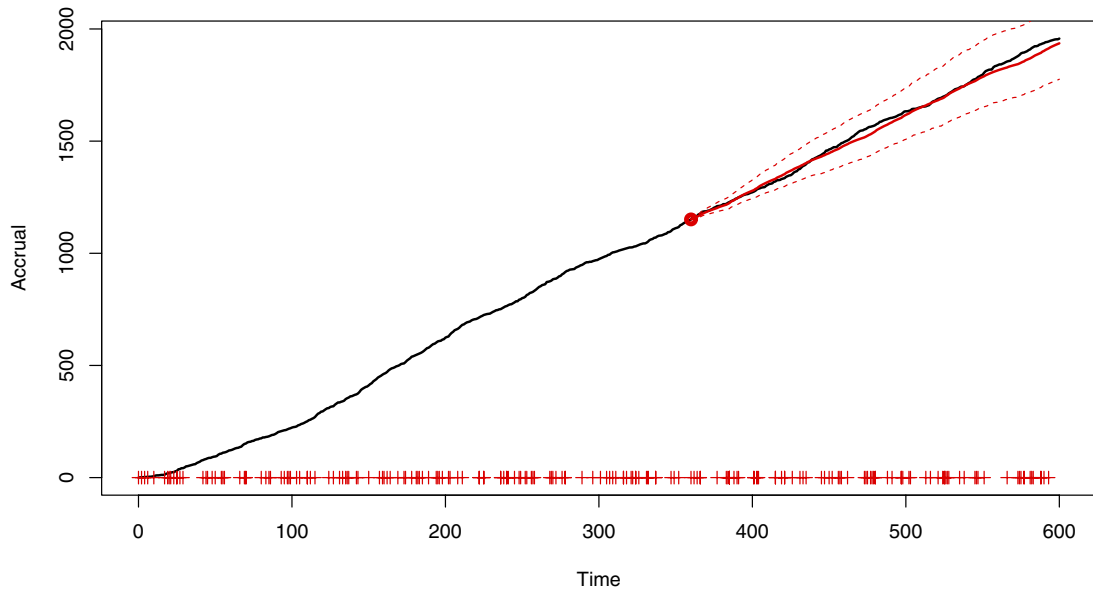


Fig. 10. Uniform initiations, $E[\lambda_c^2] = 0.03$ (Weibull-shape intensity)

7. DATA RESULTS

We fitted the same set of models to a recruitment dataset of a prostate-cancer clinical trial. The recruitment was carried out across 244 sites. The accrual is presented as the proportion of the total number enrolled. Similarly, time is given as the proportion of the total recruiting period. Figures 11 and 12 show the

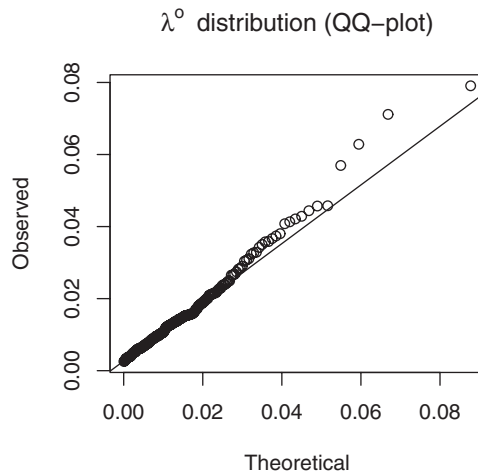


Fig. 11. Re-estimated λ_c^0 expectations compared to $\text{Gamma}(\hat{\alpha}, \hat{\alpha}/\hat{\phi})$ distribution.

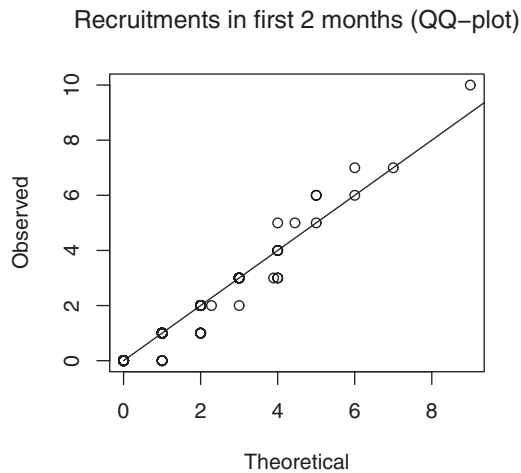


Fig. 12. Observed recruitments compared to the theoretical negative binomial distribution.

diagnostic QQ-plots for the model fitted to data available at time 0.4. They indicate that there is sufficient concordance between the assumed model and observed enrollment giving validity to potential predictions. Figure 13 shows the accrual along with forecasts from four different census times. The predictive bands become narrower and parameter uncertainty decreases at each census as more data become available for inference. After the third census, there is an unexpected jump in accrual followed by a drop around the fourth census time, suggesting a global external factor, such as a change in the protocol. Table 4 shows p -values of the LRT and BST. Initially, when the accrual is still only a small proportion of the total, it is hard to detect the time-inhomogeneity. At later census points, the test outcomes indicate that the rates are not constant.

We compare the proposed framework to the standard homogeneous PG model (2.1) as well as a homogeneous Poisson process (HPP) model fitted only to the accrual. We used the same priors as outlined

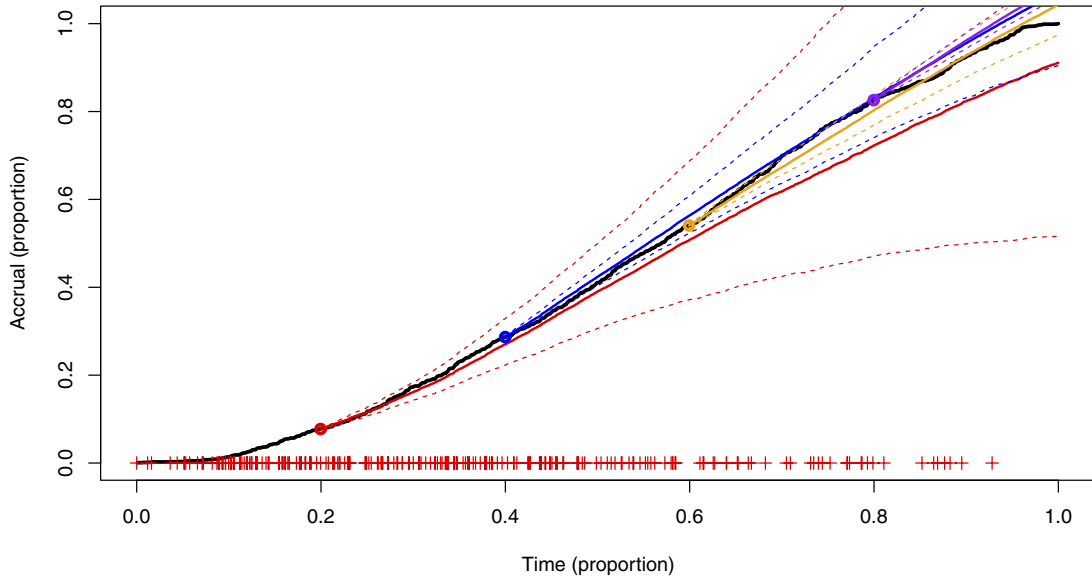


Fig. 13. Data example in Section 7 with the accrual (black, solid) for an oncology study; colored solid lines are mean predictions from census times, dashed lines are the 95% prediction bands, and the “+” symbols indicate initiation times of centers.

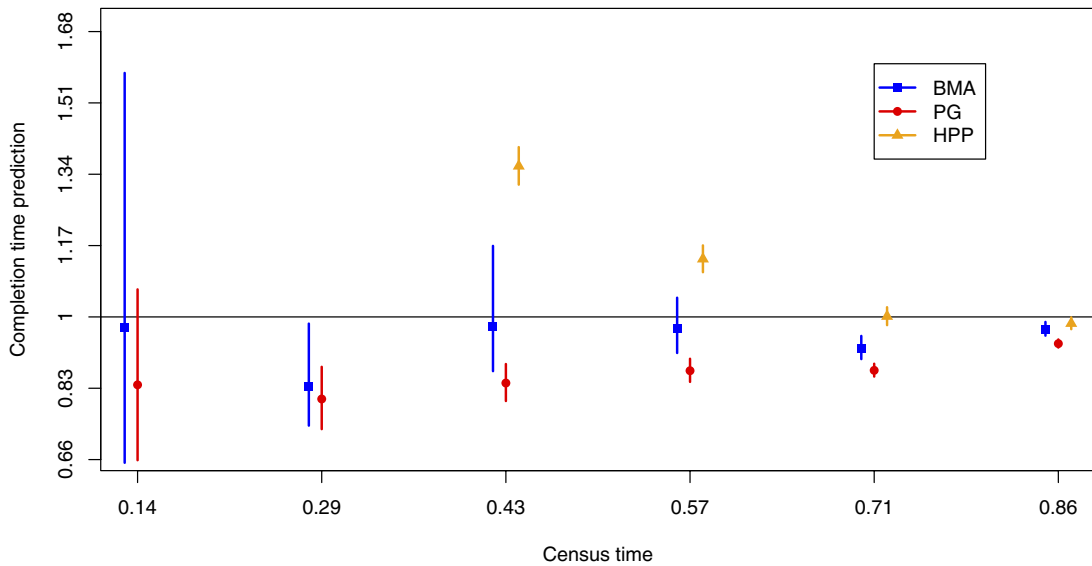


Fig. 14. Predictive distributions for time needed to make the final recruitment in the data example in Section 7, as forecast by three different modeling frameworks: Bayesian model averaging (BMA), time-homogeneous Poisson-gamma (PG), and homogeneous Poisson process fit to accrual only (HPP). The horizontal line represents the true completion time and the prediction positions of the x -axis were off-set by 0.01 for clarity.

Table 4. Decay in rate test p -values and the forecasting p -values at four census times.

Census time	BST p -value	LRT p -value
1	0.196	0.226
2	0.012	0.021
3	<0.001	<0.001
4	<0.001	<0.001

in Section 5.1 for fitting the PG model, and the HPP rate estimate was obtained using maximum likelihood. The methods were compared in terms of the predicted completion time of the recruitment for the study with the sampling details outlined in Appendix F of the [Supplementary material](#) available at *Biostatistics* online. Forecast completion time from 6 different census points and can be seen in Figure 14; the first HPP predictions were centered at 3.67 and 1.84 which were outside the plot's range. The proposed framework produces better point predictions, especially at earlier interim analyses, and more closely represents the true uncertainty. The HPP predictions near the end of the trial are very accurate. At this point, the majority of the centers having already been initiated and have been recruiting for a long period of time. As a result, the total recruitment rates are not changing by much, with the slight decreasing trend offset by the occasional initiation of a new center. This is a coincidence; if the decay rate had been sharper or shallower, or if fewer or more centers had been initiated then the naive overall Poisson process model would not have fitted as well. The underprediction of the completion time by the proposed model at the census time of $t = 0.71$ is likely a result of the unexpected surge in recruitment at around that time. The surge is examined in more detail in Appendix G of the [Supplementary material](#) available at *Biostatistics* online.

8. DISCUSSION

We have introduced a general, flexible framework for modeling and predicting recruitment to clinical trials. We suggest two tests for detecting decay in recruitment rates; comparing them both with respect to power and robustness. The particular form of the test statistic allows for a single, simple trial-level test. Alternative forms, such as splitting according to a global time, would either require a test for each center, massively reducing the power, or estimates of all of the individual center intensities which would introduces several layers of additional complexity because of the hierarchical connection between the center intensities. If it were believed *a priori* that a particular global period would be unrepresentative then this time span, and the concomitant recruitment, could simply be removed, albeit at the cost of lower power.

The parametric curve-shape forms chosen for the intensity were based on the features encountered in oncology trials. We found that the model was still robust to moderate model misspecifications in the distribution of the random effect and intensity shape. Other therapeutic areas such as pulmonary or cardiovascular diseases experience more frequent recruitments and different curve-shapes may be appropriate. As shown in Section 6, model misspecification becomes more of a problem at larger enrollment rates. However, with increased frequency, pattern changes in the early months of a center are easier to identify. Using more complex parametric forms, such as Weibull or generalized gamma shape, could lead to more accurate predictions. Alternatively, if covariate information is available, say \mathbf{x}_c for each center, the following intensity form motivated by hazard models from survival analysis could be used: $\lambda_c(t) = \lambda_c^o \exp\{\beta^\top \mathbf{x}_c\} g(t; \exp\{\eta^\top \mathbf{x}_c\})$, where λ_c^o are now random effects coming from a Gamma(α, α) distribution and β and η are vectors of unknown parameters.

As seen in the data example in Section 7, there can be external factors modulating the overall accrual. This could potentially be modeled via a short-term, constant global intensity modifier, which would maintain tractability. The framework is not constrained to parametric forms; non-parametric intensity models, such as those using B-splines (e.g., [Morgan and others, 2019](#)) or Gaussian processes (e.g., [Adams and others, 2009](#)), could be used instead. This, however, would make the intensity extrapolation problem more difficult.

For curve-shape parameter prior construction, our choice of the quantity of interest R_κ was motivated by simplicity of the form; one could just as well have used $\frac{G_\kappa(t_0/2;\theta)}{G_\kappa(t_0;\theta)}$, albeit with more algebraic manipulations. The general method was aimed at models with monotonically decreasing intensities. If curve-shapes such as Weibull are considered then constructing sensible priors will be more complicated.

In presenting the method, we condition the inference and prediction on known initiation schedules for the centers. Incorporating stochastic center initiation models, such as those in [Anisimov \(2009\)](#) and [Lan and others \(2018\)](#), into the Monte Carlo prediction framework is straight-forward, but would complicate the presentation of our methodology without adding novelty. In [Appendix H](#) of the [Supplementary material](#) available at *Biostatistics* online, we demonstrate how recruitment can be predicted using our methodology when there is uncertainty in the initiation schedule. For illustration, we imagine a Weibull-distributed delay to each center's initiation, but any other initiation model could be incorporated in a similar manner. We stress that full prediction intervals should take this uncertainty into account.

In this work, we focus on patient recruitment regardless of the numbers of dropouts observed. In practice, screening failure and patient withdrawal are both prevalent in clinical trials. Assuming the dropouts are independent of the recruitment process, existing survival analysis techniques such as Cox's proportional hazard model ([Cox, 1972](#)) or accelerated failure time frailty model ([Wei, 1992](#)) could be used in combination with the recruitment model to produce distributions of the numbers of patients in the system at a given time. Such knowledge would be useful to the practitioners and operational researchers in charge of drug-supply chains for the centers.

[Anisimov and Fedorov \(2007\)](#) introduced a method for determining the number of additional centers needed to be initiated for the study to finish on time. With minimal adaptation, the same method can also be used with our model. However, since it assumes that all new centers are initiated immediately, it may not apply in all scenarios. We would advocate a simulation-based approach, where forecasts based on different center initiation schedules are compared. As different operational costs can be associated with different schedules, this would become a resource-constrained optimization problem.

9. SOFTWARE

Software in the form of R code is available at <https://github.com/SzymonUrbas/ct-recruitment-prediction>.

SUPPLEMENTARY MATERIAL

[Supplementary material](http://biostatistics.oxfordjournals.org) is available online at <http://biostatistics.oxfordjournals.org>.

ACKNOWLEDGMENTS

Conflict of Interest: None declared.

FUNDING

The authors are grateful to UK Research and Innovation (EP/L015692/1) and AstraZeneca for sponsoring this research.

REFERENCES

- ADAMS, R. P., MURRAY, I. AND MACKAY, D. J. C. (2009). Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. pp. 9–16.
- ANISIMOV, V. (2009). Predictive modelling of recruitment and drug supply in multicenter clinical trials. In: *Proc. of Joint Statistical Meeting*. pp. 1248–1259.
- ANISIMOV, V. V. AND FEDOROV, V. V. (2007). Modelling, prediction and adaptive adjustment of recruitment in multicentre trials. *Statistics in Medicine* **26**, 4958–4975.
- COX, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* **34**, 187–202.
- GAJEWSKI, B. J., SIMON, S. D. AND CARLSON, S. E. (2008). Predicting accrual in clinical trials with Bayesian posterior predictive distributions. *Statistics in Medicine* **27**, 2328–2340.
- GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. AND RUBIN, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.
- GU, K., NG, H. K. T., TANG, M. L. AND SCHUCANY, W. R. (2008). Testing the ratio of two Poisson rates. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **50**, 283–298.
- HEITJAN, D. F., GE, Z. AND YING, G. S. (2015). Real-time prediction of clinical trial enrollment and event counts: a review. *Contemporary Clinical Trials* **45**, 26–33.
- HUZURBAZAR, V. S. (1950). Probability distributions and orthogonal parameters. Volume 46, *Mathematical Proceedings of the Cambridge Philosophical Society*. Cambridge University Press. pp. 281–284.
- JIANG, Y., SIMON, S., MAYO, M. S. AND GAJEWSKI, B. J. (2015). Modeling and validating Bayesian accrual models on clinical data and simulations using adaptive priors. *Statistics in Medicine* **34**, 613–629.
- KASENDA, B., VON ELM, E., YOU, J., BLÜMLE, A., TOMONAGA, Y., SACCILOTTO, R., AMSTUTZ, A., BENGOUGH, T., MEERPOHL, J. J., STEGERT, M. *and others.* (2014). Prevalence, characteristics, and publication of discontinued randomized trials. *JAMA* **311**, 1045–1052.
- KRISHNAMOORTHY, K. AND THOMSON, J. (2004). A more powerful test for comparing two Poisson means. *Journal of Statistical Planning and Inference* **119**, 23–35.
- LAN, Y., TANG, G. AND HEITJAN, D. F. (2018). Statistical modeling and prediction of clinical trial recruitment. *Statistics in Medicine* **38**, 945–955.
- LASAGNA, L. (1979). Problems in publication of clinical trial methodology. *Clinical Pharmacology & Therapeutics* **25**, 751–753.
- LEE, Y. J. (1983). Interim recruitment goals in clinical trials. *Journal of Chronic Diseases* **36**, 379–389.
- LINDLEY, D. V. (1957). A statistical paradox. *Biometrika* **44**, 187–192.
- MORGAN, L. E., NELSON, B. L., TITMAN, A. C. AND WORTHINGTON, D. J. (2019). A spline-based method for modelling and generating a nonhomogeneous Poisson process. In: *2019 Winter Simulation Conference (WSC)*. IEEE. pp. 356–367.
- PIANTADOSI, S. AND PATTERSON, B. (1987). A method for predicting accrual, cost, and paper flow in clinical trials. *Controlled Clinical Trials* **8**, 202–215.
- ROBERT, C. AND CASELLA, G. (2013). *Monte Carlo Statistical Methods*. Springer Science & Business Media.
- ROBERTSON, T., WRIGHT, F. T. AND DYKSTRA, R. L. (1988). *Order restricted statistical inference*. New York: Wiley.
- Doucet, A., Freitas, N. AND GORDON, N. (2001). *Sequential Monte Carlo Methods In Practice*. New York, NY: Springer New York.

- TANG, G., KONG, Y., CHANG, C.-C. H., KONG, L. AND COSTANTINO, J. P. (2012). Prediction of accrual closure date in multi-center clinical trials with discrete-time Poisson process models. *Pharmaceutical Statistics* **11**, 351–356.
- GETZ, K., AND LAMBERTI, M.J. (2013). 89% of trials meet enrollment, but timelines slip, half of sites under-enroll. *Tufts CSDD Impact Report* **15**, 1–4.
- WEI, L. J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine* **11**, 1871–1879.
- WILLIFORD, W. O., BINGHAM, S. F., WEISS, D. G., COLLINS, J. F., RAINS, K. T. AND KROL, W. F. (1987). The “constant intake rate” assumption in interim recruitment goal methodology for multicenter clinical trials. *Journal of Chronic Diseases* **40**, 297–307.
- ZHANG, X. AND LONG, Q. (2010). Stochastic modeling and prediction for accrual in clinical trials. *Statistics in Medicine* **29**, 649–658.

[Received April 1, 2020; revised August 11, 2020; accepted for publication August 15, 2020]