

# Recruitment prediction for multicenter clinical trials based on a hierarchical Poisson–gamma model: Asymptotic analysis and improved intervals

Rachael Mountain  | Chris Sherlock 

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

## Correspondence

Rachael Mountain, Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK.  
Email: [r.gallop@lancaster.ac.uk](mailto:r.gallop@lancaster.ac.uk)

## Abstract

We analyze predictions of future recruitment to a multicenter clinical trial based on a maximum-likelihood fitting of a commonly used hierarchical Poisson–gamma model for recruitments at individual centers. We consider the asymptotic accuracy of quantile predictions in the limit as the number of recruitment centers grows large and find that, in an important sense, the accuracy of the quantiles does not improve as the number of centers increases. When predicting the number of further recruits in an additional time period, the accuracy degrades as the ratio of the additional time to the census time increases, whereas when predicting the amount of additional time to recruit a further  $n_+^+$  patients, the accuracy degrades as the ratio of  $n_+^+$  to the number recruited up to the census period increases. Our analysis suggests an improved quantile predictor. Simulation studies verify that the predicted pattern holds for typical recruitment scenarios in clinical trials and verify the much improved coverage properties of prediction intervals obtained from our quantile predictor. In the process of extending the applicability of our methodology, we show that in terms of the accuracy of all integer moments it is always better to approximate the sum of independent gamma random variables by a single gamma random variable matched on the first two moments than by the moment-matched Gaussian available from the central limit theorem.

## KEYWORDS

asymptotic analysis, asymptotic correction, clinical trial recruitment, multicenter clinical trial, Poisson process, recruitment prediction interval

## 1 | INTRODUCTION

Randomized controlled trials represent the gold standard for evaluating the safety and efficacy of a new health care intervention or treatment (Akobeng, 2005). Such trials can require thousands of patients, and so will typically recruit from tens or hundreds of centers. The timely recruitment

of patients is widely recognized as a key determinant of the success of a clinical trial (Carter, 2004). Nonetheless, sources suggest as many as 86% of all clinical trials fail to reach their required recruitment goals (Lamberti *et al.*, 2012; Carlisle *et al.*, 2015; Huang *et al.*, 2018). Failure to meet recruitment targets can have numerous negative implications, yet arguably the most critical is inadequate

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Biometrics* published by Wiley Periodicals LLC on behalf of International Biometric Society.

statistical power. In such a scenario, there is an increased risk of Type II error, thus potentially preventing or delaying an effective treatment from being approved (Treweek *et al.*, 2013).

Recruitment of a patient to a clinical trial can be thought of as a three-stage process. First, some recruitment centers are initiated; more centers can be initiated as the trial progresses. Second, a potential recruit is enrolled at a given center; after a lag, the potential recruit is screened for suitability, and if suitable that patient is randomized onto a particular treatment. Methods for predicting future recruitment usually model the probability of screening success separately, so we focus on the second stage of the process.

Future recruitment is often predicted using deterministic methods, based on the number already recruited up to that time, or historical data (Carter *et al.*, 2005). Such an approach is inadequate due to the stochastic nature of the recruitment process, and a number of stochastic models have been proposed.

Senn (1997) considers a Poisson-based model for a multicenter clinical trial where recruitment follows a Poisson process with a fixed study-wide rate,  $\lambda \geq 0$ . The time to recruit a given number of patients then follows a gamma distribution. The underlying assumption that recruitment follows a Poisson process is well accepted in the literature, with many articles exploring an inhomogeneous model with a time-dependent rate (Carter, 2004; Carter *et al.*, 2005; Tang *et al.*, 2012; Lan *et al.*, 2019).

The basic Poisson model outlined above fails to incorporate variation in recruitment rate across centers, as well as the uncertainty in the rate estimate. Anisimov and Fedorov (2007) propose a random effects model in which recruitment follows a homogeneous Poisson process within each center, with the center-specific rates viewed as a sample from a gamma distribution. The time to recruit a given number of patients then follows a Pearson Type VI distribution, while the number recruited in a given time is negative binomial. This model accounts for staggered center initiation times and provides a method for predicting recruitment for new centers entering the trial. Citations of Anisimov and Fedorov (2007) on Google Scholar show that it has also been used by major pharmaceutical companies and in statistical software to plan drug production and distribution across centers during clinical trials. Further details of the model will be given in Section 2.

The Anisimov and Fedorov model (henceforth AF) has been developed and extended in numerous directions. For example, Bakhshi *et al.* (2013) suggest an extra level of hierarchy to incorporate variation from trial to trial in the gamma distribution parameters, with an aim to forecast recruitment for trials yet to begin. Mijoule *et al.* (2012) propose a Pareto mixture distribution for the center rates in

place of the gamma. Further, Lan *et al.* (2019) and Urbas *et al.* (2020) both incorporate time-varying rates into the AF model, while also incorporating parameter uncertainty using the Bayesian paradigm.

Alternative methods have been suggested for modeling patient recruitment outside the Poisson approach, including Monte Carlo simulation (Abbas *et al.*, 2007), time series analysis (Haidich and Ioannidis, 2001), Brownian motions (Lai *et al.*, 2001; Zhang and Lai, 2011), and a nonparametric approach (Ying *et al.*, 2004).

We investigate future predictions based on a maximum likelihood fit of the AF model to multicenter recruitment data, where a total of  $N_+$  patients has been recruited over  $C$  centers by a census time,  $t$ . We then consider two prediction objectives, where prediction intervals are required for either (1) the total number  $N_+^+$  recruited over some additional time  $t^+$  or (2) the total time  $T^+$  to obtain  $n_+^+$  additional recruits. In this section, for brevity, we focus on objective (1); similar methods and results are obtained for objective (2).

Within the AF model, the distribution of the predicted number of recruits,  $\tilde{N}_+^+$ , has a negative binomial distribution, which depends on the observed data via the maximum likelihood estimates (MLEs) of the model parameters; in contrast, the true number recruited,  $N_+^+ \sim \text{Poisson}(\lambda, t^+)$ , where  $\lambda$  is the sum of the recruitment rates of the individual centers. Let  $\hat{q}_p$  be the  $p$ th quantile of  $\tilde{N}_+^+$ ; that is, the predicted quantile. We first investigate  $P_p := \mathbb{P}(N_+^+ \leq \hat{q}_p)$  in the limit as  $C \rightarrow \infty$ , and empirically for finite  $C$ , and show that the key determinant of the behavior is the ratio  $t^+/t$ . The desirable result of  $P_p = p$  is only recovered in the limit as  $t^+/t \rightarrow 0$ , whereas in more typical scenarios  $P_p$  can be very different from  $p$ . The underlying reason for this is that the uncertainty in the MLEs is not being accounted for. Our asymptotic approximation to  $P_p$  feeds in to a new methodology, which allows us to produce tractable prediction intervals, which have a coverage that is very close to that intended, and with a fraction of the computational cost of any bootstrap-based scheme.

Our theory, and hence our adjusted interval, is derived under the assumption that all centers opened at the same time; however, sometimes this is not the case. For example, given a predicted shortfall, perhaps based on our theory, it may be decided to open a new set of centers as well as keeping the existing centers going. Alternatively, or in addition, the existing centers may have been opened at different times. Guided by our theory, we provide an intuitive, tractable methodology for creating a prediction interval in such cases and demonstrate its accuracy in practice via extensive simulation studies.

Section 2 describes the AF model in detail, and Section 3.1 provides the asymptotic analysis in the case where all centers opened at the same time and details the

**TABLE 1** Common notations used in this article. Objectives 1 and 2 are abbreviated to O1 and O2, respectively

$C$	Number of centers
$t$	(Global) census time
$t_c$	Time center $c$ is open before census
$N_c$	Number of recruits at center $c$ at census time
$N_\bullet$	$\sum_{c=1}^C n_c$
$n_c$	Realization of $N_c$
$n_\bullet$	Realization of $N_\bullet$
$T^+$	Time from census until $n_\bullet^+$ new recruits
$t^+$	Realization of $T^+$ (O2) or specified additional recruitment time after census (O1)
$N_c^+$	Number of recruits at center $c$ during specified time $t^+$
$N_\bullet^+$	$\sum_{c=1}^C N_c^+$
$n_\bullet^+$	Realization of $N_\bullet^+$ (O1) or specified total number of additional recruits required (O2)
$\hat{q}_p$	Estimated $p^{\text{th}}$ quantile for $N_\bullet^+$
$\hat{r}_p$	Estimated $p^{\text{th}}$ quantile for $T^+$

methodology for creating prediction intervals with almost perfect coverage. Section 3.2 describes an empirical extension to this methodology for situations where the centers opened at different times. Our results and methods are verified via a detailed simulation study in Section 4, and then applied to a clinical-trial recruitment data set in Section 4.3. We conclude in Section 5 with a discussion. First, however, we define the notations that will be used throughout.

## 1.1 | Notations

Let  $C$  be the number of centers, and for  $c = 1, \dots, C$ , let  $t_c$  and  $N_c$  represent the time for which center  $c$  was open before the census time and number recruited in center  $c$  during the time  $t_c$ . The shorthand  $\underline{N}$  refers to the vector  $(N_1, \dots, N_C)$ , we let  $N_\bullet := \sum_{c=1}^C N_c$ , and when all centers are open for the same time we denote that time by  $t$ . For Objective 1, let  $t^+$  be the additional time ahead at which predictions will be made, and let  $N_c^+$  be the number recruited in center  $c$  in that time, with  $N_\bullet^+ = \sum_{c=1}^C N_c^+$ . For Objective 2, let  $n_\bullet^+$  be the additional number of recruits sought and let  $T^+$  be the additional time taken to recruit this number. Table 1 summarizes these notations, and others that will be introduced later.

The negative binomial distribution of the number of successes until there are  $a$  failures when the probability of success  $p$  is denoted  $\text{NB}(a, p)$ . We use the notation  $\xrightarrow{p}$  and  $\Rightarrow$  to indicate convergence in probability and in distribution, respectively, and  $\Phi$  to indicate the cumulative distribution function of an  $\text{N}(0, 1)$  random variable.

## 2 | MODEL AND PREDICTION SET UP

### 2.1 | Model, data, and likelihood

The model assumes that the recruitment rate at center  $c$ , for  $c = 1, \dots, C$ , is  $\lambda_c$ , where each  $\lambda_c$  is drawn independently from

$$\lambda_c \sim \text{Gam}(\alpha, \beta). \quad (1)$$

Data for center  $c$  are  $n_c^1, \dots, n_c^{t_c}$ ,  $n_c := \sum_{s=1}^{t_c} n_c^s$  and  $n_\bullet = \sum_{c=1}^C n_c$ . The likelihood for center  $c$  is

$$\begin{aligned} L(\alpha, \beta, \theta; n_c^{1:t_c}) &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda) \prod_{s=1}^{t_c} \frac{\lambda^{n_c^s}}{n_c^s!} \exp(-\lambda) d\lambda \\ &\propto \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \lambda^{\alpha+n_c-1} \exp[-\lambda(\beta+t_c)] d\lambda \\ &= \frac{\Gamma(\alpha+n_c)}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta+t_c)^{\alpha+n_c}}. \end{aligned}$$

Hence, up to an additive constant, the log-likelihood given data  $n_c^s$ ,  $s = 1, \dots, t$ ,  $c = 1, \dots, C$ , is

$$\begin{aligned} \ell(\alpha, \beta) &= C\alpha \log \beta - \sum_{c=1}^C (\alpha + n_c) \log(\beta + t_c) - C \log \Gamma(\alpha) \\ &\quad + \sum_{c=1}^C \log \Gamma(\alpha + n_c). \end{aligned} \quad (2)$$

Thus,  $\underline{n} = (n_1, \dots, n_C)$  is a sufficient statistic. In the special case where  $t_1 = \dots = t_C = t$ , the second term in (2) reduces to  $-(C\alpha + n_\bullet) \log(\beta + t)$  and, as we shall see in Lemma 1,  $\hat{\alpha}/\hat{\beta}$  depends on  $\underline{n}$  only through  $n_\bullet$ .

### 2.2 | Prediction

Since  $N_c | \lambda_c \sim \text{Po}(\lambda_c t_c)$ , given a prior of  $\text{Gam}(\hat{\alpha}, \hat{\beta})$  for  $\lambda_c$  and an observation of  $n_c$ , the posterior distribution for  $\lambda_c$  is  $\text{Gam}(\hat{\alpha} + n_c, \hat{\beta} + t_c)$ . The distribution of  $\lambda_\bullet := \sum_{c=1}^C \lambda_c$  is not tractable in general, but in the special case where  $t_1 = \dots = t_C = t$ ,  $\lambda_\bullet \sim \text{Gam}(C\hat{\alpha} + n_\bullet, \hat{\beta} + t)$ . In this case, since  $N_\bullet^+ | \lambda_\bullet \sim \text{Po}(\lambda_\bullet t^+)$ , marginalizing over  $\lambda_\bullet$ , the predicted total recruitment in further time  $t^+$  is

$$\tilde{N}_\bullet^+ \sim \text{NB}\left(C\hat{\alpha} + n_\bullet, \frac{t^+}{\hat{\beta} + t + t^+}\right), \quad (3)$$

which has moments of

$$\begin{aligned} \mathbb{E}[\tilde{N}_*^+] &= \frac{C\hat{\alpha} + N_*}{\hat{\beta} + t} \times t^+ \\ \text{and } \text{Var}[\tilde{N}_*^+] &= \frac{C\hat{\alpha} + N_*}{\hat{\beta} + t} \times t^+ \times \frac{\hat{\beta} + t + t^+}{\hat{\beta} + t}. \end{aligned} \quad (4)$$

Alternatively, if the number of additional recruits is fixed at  $n_*^+$  then,  $T^+ | \lambda_* \sim \text{Gam}(n_*^+, \lambda_*)$ , so in the case where  $t_1 = \dots = t_C = t$ , the predicted further time  $\tilde{T}^+$  to recruit these has a Pearson VI distribution (e.g., Johnson *et al.*, 1994) with a density of

$$f(\tilde{t}^+) = \frac{\Gamma(C\hat{\alpha} + n_* + n_*^+)}{\Gamma(C\hat{\alpha} + n_*)\Gamma(n_*^+)} \frac{\hat{\beta}^{C\hat{\alpha} + n_*} (\tilde{t}^+)^{n_*^+ - 1}}{(\hat{\beta} + t + \tilde{t}^+)^{C\hat{\alpha} + n_* + n_*^+}}. \quad (5)$$

Thus,  $\tilde{T}^+$  has moments of:

$$\begin{aligned} \mathbb{E}[\tilde{T}^+] &= \frac{(\hat{\beta} + t)n_*^+}{C\hat{\alpha} + N - 1} \\ \text{and } \text{Var}[\tilde{T}^+] &= \mathbb{E}[\tilde{T}^+] \times \frac{(\hat{\beta} + t)(C\hat{\alpha} + N_* + n_*^+ - 1)}{(C\hat{\alpha} + N_* - 1)(C\hat{\alpha} + N_* - 2)}. \end{aligned} \quad (6)$$

### 3 | ASYMPTOTIC ANALYSIS AND METHODOLOGY

We consider the properties of the quantile estimates under repeated sampling, so that  $\underline{N}$  is a random variable, and  $\hat{\alpha}$  and  $\hat{\beta}$  are, therefore, random. We examine the probability under the true data-generating mechanism that the quantity of interest,  $N_*^+$  or  $T^+$ , will be less than its predicted quantile. This leads to a tractable formula for an alternative probability,  $p^*(p)$ , such that  $\mathbb{P}(N_*^+ \leq \hat{q}_{p^*}) \approx p$  or  $\mathbb{P}(T^+ \leq \hat{r}_{p^*}) \approx p$ , and hence to prediction intervals with close to the intended coverage. In Section 3.1, we consider the scenario where all centers have been open for the same time; an intuitive extension for the more general scenario is given in Section 3.2.

#### 3.1 | All centers opened simultaneously

When all centers have been open for the same time,  $t$ ,  $N_* \sim \text{Po}(\lambda, t)$  is the key (random) summary of the data, instead of  $n_*$  for the specific realization; thus  $\hat{\alpha}$  and  $\hat{\beta}$  are random. Importantly, in this case  $\hat{\alpha}/\hat{\beta}$  depends on  $\underline{N}$  only through  $N_*$ .

**Lemma 1.** When  $t_1 = \dots = t_C = t$ , the MLE for the likelihood in (2) satisfies  $\hat{\alpha}/\hat{\beta} = N_*/(Ct)$ .

*Proof.* Set  $\gamma = \alpha/\beta$ ; from the invariance principle it is sufficient to show that  $\hat{\gamma} = N_*/(Ct)$ . Substituting for  $\beta$  and ignoring terms only in  $\alpha$ , (2) becomes:

$$\begin{aligned} \ell(\alpha, \gamma) &= -C\alpha \log \gamma - (C\alpha + N_*) \log(\alpha/\gamma + t) \\ &= N_* \log \gamma - (C\alpha + N_*) \log(\alpha + \gamma t). \end{aligned}$$

Thus,

$$\partial_\gamma \ell = \frac{N_*}{\gamma} - \frac{C\alpha + N_*}{\alpha + \gamma t} \times t = \frac{\alpha}{\gamma(\alpha + \gamma t)} (N_* - \gamma Ct),$$

which is zero (and a maximum for  $\ell$ ) when  $\gamma = N_*/(Ct)$ , as required.  $\square$

We now state our main result.

**Theorem 1.** Consider an infinite sequence of recruitment scenarios indexed by the number of recruitment centers,  $C = 1, 2, \dots$ . In each scenario,  $C$ , after each center has been opened for a fixed common time  $t$ ,  $(\alpha, \beta)$  is estimated from data  $\underline{N}^{(C)}$  by maximizing (2). It is used in (3) to estimate the  $p$ th quantile,  $q_p^{(C)}$ , of the total number,  $N_*^{+(C)}$  of recruits in an additional, fixed time  $t^+$ ; it is also used in (5) to estimate the  $p$ th quantile,  $r_p^{(C)}$ , of the time,  $T^{+(C)}$  until a further  $n^{+(C)}$  recruits have been obtained. Denoting the quantile estimates as  $\hat{q}_p^{(C)}$  and  $\hat{r}_p^{(C)}$ , respectively, the following results hold with  $Z \sim N(0, 1)$ :

$$\begin{aligned} 1. \quad & \lim_{C \rightarrow \infty} \mathbb{P}(N_*^{+(C)} \leq \hat{q}_p^{(C)} | \underline{N}^{(C)}) \\ & \stackrel{D}{=} \Phi \left\{ \sqrt{\frac{t^+}{t}} Z + \Phi^{-1}(p) \sqrt{1 + \frac{t^+/\beta}{1 + t/\beta}} \right\}. \end{aligned}$$

However, for large  $C$ ,  $\hat{q}_p^{(C)}/q_p^{(C)} - 1 = \mathcal{O}(1/\sqrt{C})$ ; moreover,

$$\frac{\hat{q}_{1-p/2}^{(C)} - \hat{q}_{p/2}^{(C)}}{q_{1-p/2}^{(C)} - q_{p/2}^{(C)}} \sim \sqrt{1 + \frac{t^+/\beta}{1 + t/\beta}}.$$

2. If as  $C \rightarrow \infty$ ,  $n_*^{+(C)}/C \rightarrow a > 0$ ,

$$\lim_{C \rightarrow \infty} \mathbb{P}(T^{+(C)} \leq \hat{r}_p^{(C)} | \underline{N}^{(C)})$$

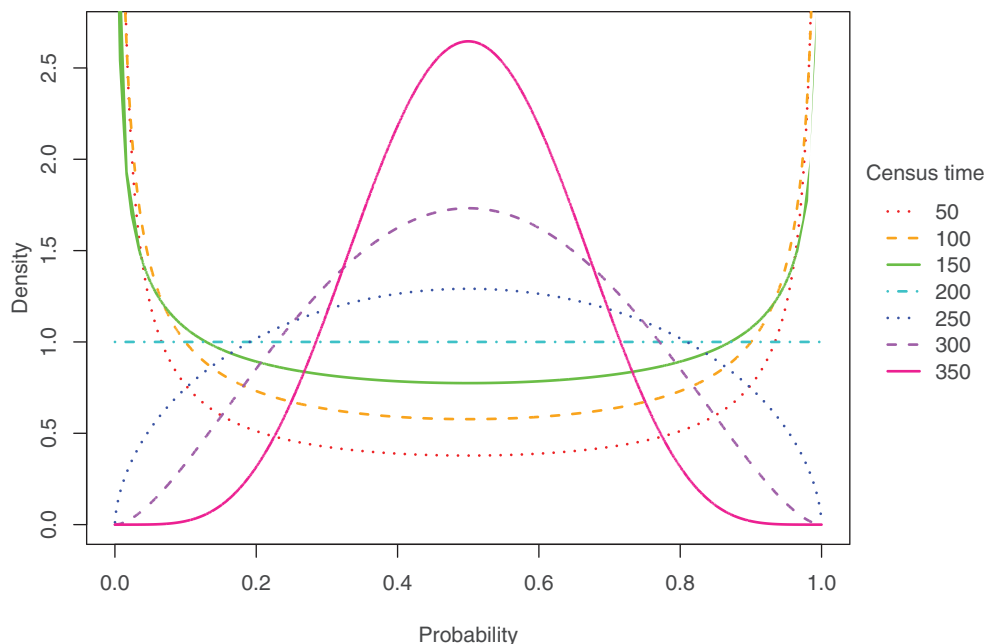


FIGURE 1 Theoretical density of  $\mathbb{P}(N^+ \leq \hat{q}_{0.5})$  as a function of census time,  $t$ , with  $t^+ = 400 - t$ ,  $\alpha = 2$ ,  $\beta = 150$ , and  $C = 150$ . This figure appears in color in the electronic version of this article

$$\stackrel{D}{=} \Phi \left\{ \sqrt{\frac{a\beta}{\alpha t}} Z + \Phi^{-1}(p) \sqrt{1 + \frac{a/\alpha}{1 + t/\beta}} \right\}.$$

However,  $\hat{r}_p^{(C)}/r_p^{(C)} - 1 = \mathcal{O}(1/\sqrt{C})$ ; moreover,

$$\frac{\hat{r}_{1-p/2}^{(C)} - \hat{r}_{p/2}^{(C)}}{r_{1-p/2}^{(C)} - r_{p/2}^{(C)}} \sim \sqrt{1 + \frac{\alpha/a}{1 + t/\beta}}.$$

Theorem 1 is proved in Section S1 of the Supporting Information. We discuss the consequences for  $N^+$  in detail; those for  $T^+$  are analogous.

Theorem 1 confirms the intuition that the width of any confidence interval estimated using  $(\hat{\alpha}, \hat{\beta})$  is wider than that which would be obtained were the total intensity,  $\lambda$ , known precisely; however, it also shows that the ratio approaches 1 as the census time increases. More importantly, for the median, Theorem 1 suggests that  $\mathbb{P}(N^+ \leq \hat{q}_{0.5}) \approx \Phi(\sqrt{t^+/t}Z)$ , so that when  $t^+ \approx t$ , this probability is approximately uniformly distributed on  $[0, 1]$ . By contrast, when  $t^+ \ll t$  the probability concentrates at  $\approx 0.5$  as is desirable, and when  $t^+ \gg t$  the probability concentrates around 0 and 1 each with a mass of 0.5, which is not desirable. The theoretical densities for  $\mathbb{P}(N^+ \leq \hat{q}_{0.5})$  as a function of  $t$  (with  $t^+ = 400 - t$ ) are given in Figure 1. For more general quantiles, with  $t$  fixed, as  $t^+ \rightarrow 0$ , the probability approaches a point mass at  $p$  as desired, but as  $t^+ \rightarrow \infty$  the same concentration around 0 and 1 happens,

however, the mass on 1 is  $\mathbb{P}\{Z \geq -\sqrt{t/(\beta + t)}\Phi^{-1}(p)\} = \Phi\{\sqrt{t/(\beta + t)}\Phi^{-1}(p)\}$ .

Despite this decidedly unintuitive behavior of the quantile probabilities, Theorem 1 also shows that the relative error in the quantile estimate decays in proportion to  $1/\sqrt{C}$  as expected. The resolution of this apparent contradiction lies in the fact that while the quantiles for  $N^+$  and  $\tilde{N}^+$  themselves are  $\mathcal{O}(C)$ , both the discrepancy between them and the widths of the distributions are  $\mathcal{O}(\sqrt{C})$ . The discrepancy between the quantiles also decreases to 0 as  $t^+/t \downarrow 0$ , so depending on this ratio, the two distributions can closely overlap or almost entirely diverge ( $t^+ \gg t$ ).

Thus, even though the point estimate of a quantile may be accurate relative to the size of the quantile ( $\mathcal{O}(\sqrt{C})$  compared with  $\mathcal{O}(C)$ ), unless  $t^+ \ll t$ , prediction intervals will not, in general, provide the intuitive and desirable coverage properties:  $\mathbb{P}(\hat{q}_{0.05} \leq N^+ \leq \hat{q}_{0.95}) \approx 0.9$ , for example. However, the (asymptotically) correct coverage can be recovered by adjusting the interval, based on Theorem 1, as we now describe.

Theorem 1 suggests that to obtain a predictive value with the true (asymptotic in  $C$ ) probability  $p$  of it not being exceeded, we must target a value  $p^*$  such that

$$p = \mathbb{E} \left[ \Phi \left\{ \sqrt{\frac{t^+}{t}} Z + \Phi^{-1}(p^*) \sqrt{\frac{\beta + t + t^+}{\beta + t}} \right\} \right].$$



Writing  $b$  for  $\Phi^{-1}(p^*)\sqrt{(\beta+t+t^+)/(\beta+t)}$  and letting  $Z' \sim N(0, 1)$  be independent of  $Z$ , the right-hand side may be rewritten as

$$\begin{aligned} \mathbb{P}\left(Z' \leq \sqrt{\frac{t^+}{t}}Z + b\right) &= \mathbb{P}\left(\sqrt{1 + \frac{t^+}{t}}N(0, 1) \leq b\right) \\ &= \Phi\left(\frac{b}{\sqrt{1 + \frac{t^+}{t}}}\right). \end{aligned}$$

Rearranging gives

$$\sqrt{\frac{t+t^+}{t}}\Phi^{-1}(p) = \sqrt{\frac{\beta+t+t^+}{\beta+t}}\Phi^{-1}(p^*),$$

so

$$p^* = \Phi\left\{\sqrt{\frac{(\beta+t)(t+t^+)}{t(\beta+t+t^+)}}\Phi^{-1}(p)\right\}. \quad (7)$$

In practice we do not know  $\beta$ , and necessarily substitute  $\hat{\beta}$  for this value. The estimator  $\hat{\beta}$  is consistent for  $\beta$ , and so we might expect this approximation to be reasonable. Section 4.2 provides empirical verification that adjustments based on this approximation lead to substantial improvements in coverage.

### 3.2 | Different center opening times

We now consider the scenario where  $t_1 = \dots = t_C$  does not hold. In this case the posterior for  $\lambda_*$  is intractable and, hence, so are the distributions for  $\tilde{N}_*^+$  and  $\tilde{T}_*^+$ . Furthermore, Lemma 1 does not hold.

Although the distribution of  $\lambda_*$  is intractable, its moments are not:

$$\mathbb{E}[\lambda_*] = \sum_{c=1}^C \frac{\hat{\alpha} + n_c}{\hat{\beta} + t_c} \quad \text{and} \quad \text{Var}[\lambda_*] = \sum_{c=1}^C \frac{\hat{\alpha} + n_c}{(\hat{\beta} + t_c)^2}.$$

**Theorem 2.** Let  $X_1, \dots, X_n$  be independent random variables with  $X_i \sim \text{Gam}(\alpha_i, \beta_i)$ . Define  $S_n = \sum_{i=1}^n X_i$ ,  $\mu_n = \mathbb{E}[S_n]$ , and  $\sigma_n^2 = \text{Var}[S_n]$ . Consider two approximations to  $S_n$  obtained by matching the first two moments:

$$Z \sim N(\mu_n, \sigma_n^2)$$

$$G \sim \text{Gam}(\alpha, \beta),$$

with  $\alpha/\beta = \mu_n$  and  $\alpha/\beta^2 = \sigma_n^2$ . Denoting the  $j$ th cumulants of  $S_n$ ,  $Z$ , and  $G$  by  $\kappa_j^{S_n}$ ,  $\kappa_j^Z$ , and  $\kappa_j^G$ , respectively,  $0 < \kappa_1^{S_n} = \kappa_1^Z = \kappa_1^G = \mu_n$  and  $0 < \kappa_2^{S_n} = \kappa_2^Z = \kappa_2^G = \sigma_n^2$  by design, and the following holds for all  $j \geq 3$ :

$$0 = \kappa_j^Z < \kappa_j^G \leq \kappa_j^{S_n}.$$

Theorem 2 is proved in Section S2 of the Supporting Information. Since the moment generating function of a random variable is  $M(t) = \exp\{K(t)\}$ , where  $K(t)$  is the cumulant generating function, the coefficient of  $t^n$  in  $M(t)$  is a linear combination of products of the cumulants  $\kappa_1, \dots, \kappa_n$  where all coefficients are positive. This immediately leads to the following:

**Corollary 1.** With  $S_n$ ,  $Z$ , and  $G$  as defined in Theorem 2,  $\mathbb{E}[Z^j] = \mathbb{E}[G^j] = \mathbb{E}[S_n^j]$  for  $j = 1, 2$ , and for all integer  $j \geq 3$ ,

$$\mathbb{E}[Z^j] < \mathbb{E}[G^j] \leq \mathbb{E}[S_n^j].$$

Theorem 2 and Corollary 1 show that a moment-matched gamma approximation to  $S_n$  is, in a sense, strictly better than the moment-matched Gaussian approximation available through the central limit theorem. We, therefore, make the approximation (see also Lemma 2.2 in Anisimov, 2011) that

$$\lambda_* \stackrel{D}{\approx} \lambda_*^* \sim \text{Gam}(C\hat{\alpha} + n^*, \hat{\beta} + t^*),$$

where  $n^*$  and  $t^*$  are chosen so that the first two moments of  $\lambda_*^*$  match those of  $\lambda_*$ . Figure S1 in the Supporting Information, and the accompanying text, demonstrate the accuracy of this approximation for two scenarios relevant to trial recruitment that we will describe in Section 4.2.

The posterior distribution for  $\lambda_*^*$  is exactly that which would arise given the  $\text{Gam}(C\hat{\alpha}, \hat{\beta})$  prior if each center had been open for the same time of  $t^*$  and a total of  $n^*$  patients had been recruited. Thus, if the MLEs from this “data”,  $\hat{\alpha}^*$  and  $\hat{\beta}^*$  were to satisfy  $\hat{\alpha}^* = \hat{\alpha}$  and  $\hat{\beta}^* = \hat{\beta}$  then the theory from Section 3.1 would follow through exactly. In reality, whatever the partitioning of  $n^*$  across centers, the data would typically lead to slightly different MLEs  $\hat{\alpha}^* \neq \hat{\alpha}$  and  $\hat{\beta}^* \neq \hat{\beta}$ ; nevertheless, in the proof of Theorem 1 the most important aspect of the MLEs is their ratio. From Lemma 1,  $\hat{\alpha}^*/\hat{\beta}^* = n^*/Ct^*$ , and empirical comparisons of  $n^*/Ct^*$  against  $\hat{\alpha}/\hat{\beta}$  (see the Supporting Information) showed a relative error of less than 0.1%.

The methodology for constructing prediction intervals for either  $\tilde{N}_*^+$  or  $\tilde{T}_*^+$  then proceeds as in Section 3.1, using  $\hat{\alpha}$  and  $\hat{\beta}$  under the assumption that  $\lambda_* \equiv \lambda_*^*$ .

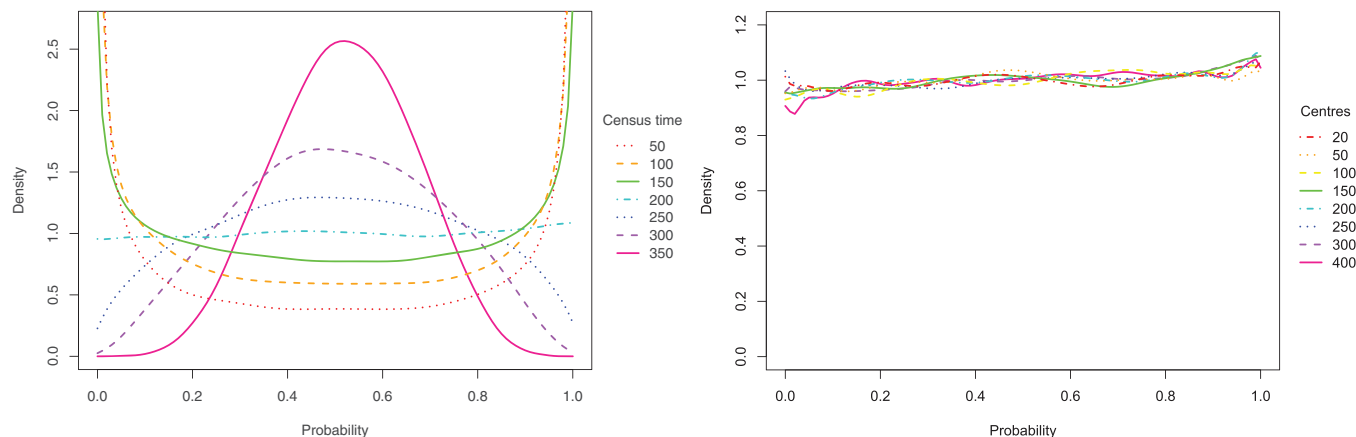


FIGURE 2 Estimated density (over repeated sampling) of  $\mathbb{P}(N^+ \leq \hat{q}_{0.5})$  for each  $t \in \mathbb{T}_1$  with  $t^+ = 400 - t$  (left) and for each  $C \in \mathbb{C}_1$  with  $t = t^+ = 200$  and  $\beta = C$  (right). This figure appears in color in the electronic version of this article

## 4 | EMPIRICAL VERIFICATION OF THEORY AND METHODOLOGY

Simulations were carried out to test the asymptotic theory and methods proposed in this paper for finite numbers of centers,  $C$ . A large number (20,000 unless otherwise stated) of realizations of the parameters  $\lambda_1, \dots, \lambda_C$ , and hence the sample  $(n_1, \dots, n_C)$  were simulated for a given set of parameter values. For each realization, the parameters  $\alpha$  and  $\beta$  were estimated using maximum likelihood and the quantile of interest,  $q_p$  or  $r_p$ , was estimated. Either  $\mathbb{P}(N^+ \leq \hat{q}_p)$  or  $\mathbb{P}(T^+ \leq \hat{r}_p)$  was then calculated exactly using the known (simulated)  $\lambda_1, \dots, \lambda_C$ . The results outlined below will primarily focus on predicting  $N^+$ .

Unless specified otherwise, the following parameter values were used:  $\alpha = 2$ ,  $\beta = 150$ ,  $C = 150$ ,  $t = 200$ . The latter two are the default values used when considering varying census times and center numbers, respectively.

When predicting  $N^+$ , the total trial length was set to  $\tau = t + t^+ = 400$ , since with the default  $C$ ,  $\mathbb{E}[N_* + N^+] = C(\alpha/\beta)(t + t^+) = 800$ , a reasonable size for a Phase III clinical trial. Furthermore, the census time  $t$  was chosen from  $\mathbb{T}_1 = \{50, 100, 150, 200, 250, 300, 350\}$  and the number of centers,  $C$ , was chosen from  $\mathbb{C}_1 = \{20, 50, 100, 150, 200, 250, 300, 400\}$ . When examining predictions of  $T^+$  we fixed  $n_*^+ = 200$  and selected  $t \in \mathbb{T}_2 = \{50, 100, 150, 200, 300, 500, 1000\}$  and  $C \in \mathbb{C}_2 = \{20, 50, 100, 150, 200, 300, 500, 1000\}$ .

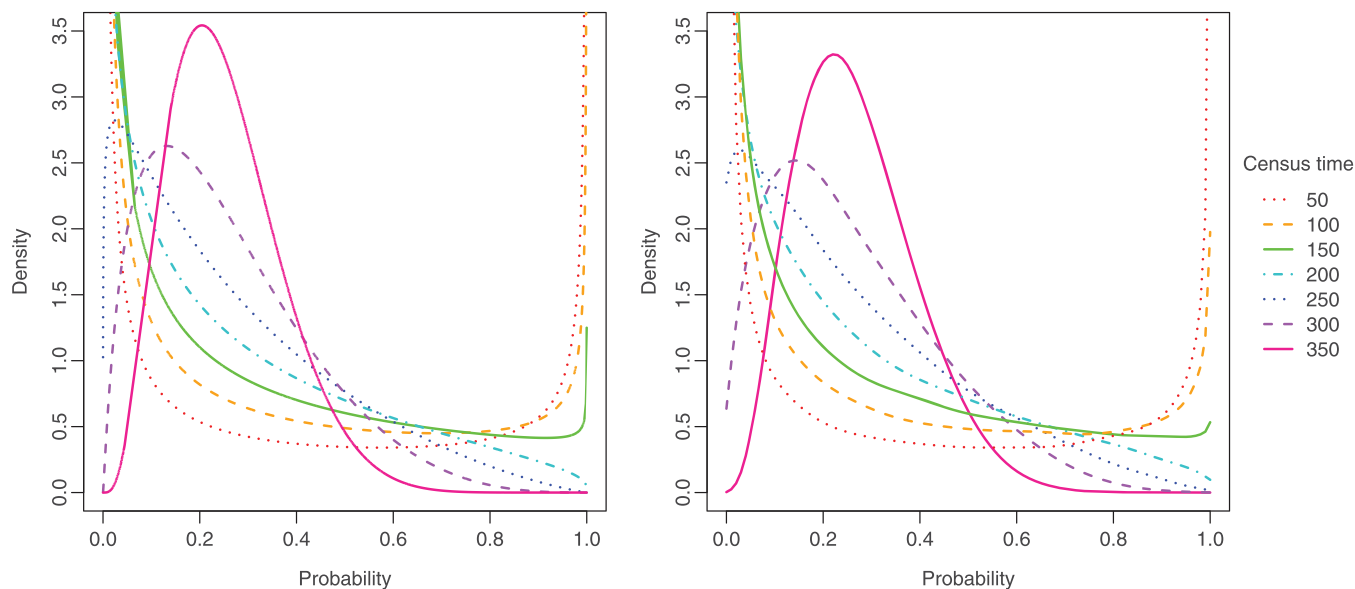
When conducting simulations with varying number of centers, we set  $\beta = C$  to maintain a fixed expected number of recruits per unit time. In the Supporting Information, we explore an alternative scenario where  $\beta = 150$  is fixed as  $C$  varies.

### 4.1 | Verification of Theorem 1

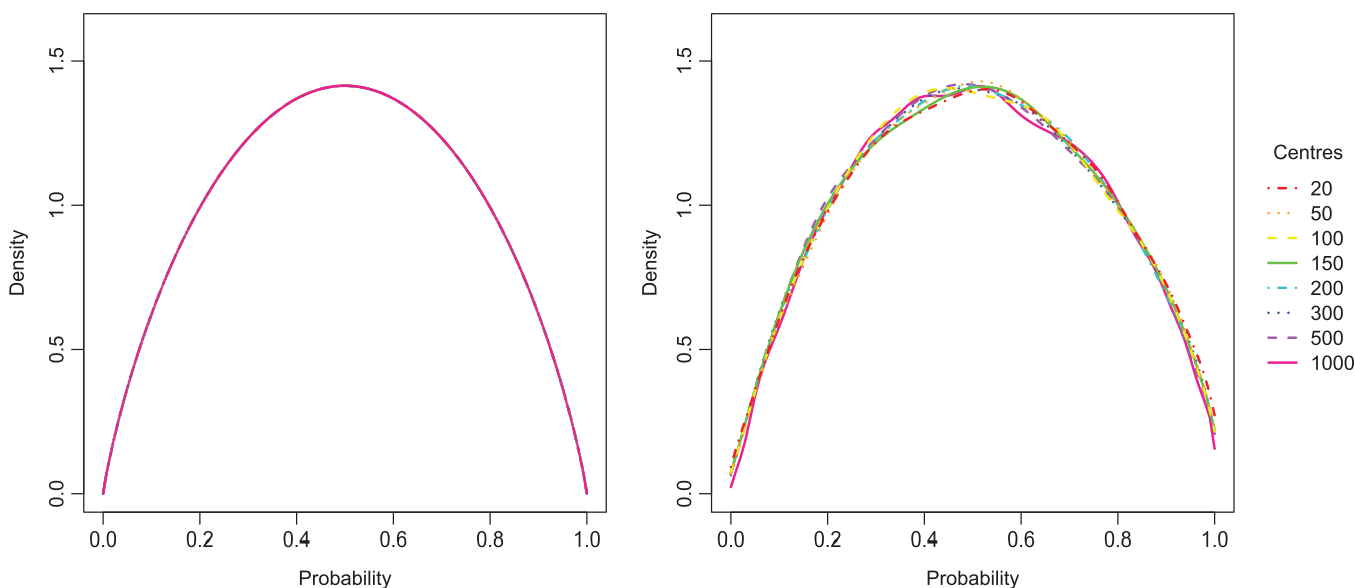
Figure 2 shows the empirical distribution of  $\mathbb{P}(N^+ \leq \hat{q}_p)$  over repeated simulation and, hence, estimates  $\hat{q}_p$ , for the median,  $p = 0.5$ . The left panel varies the census times  $t \in \mathbb{T}_1$ , while the right panel fixes  $t$  (and hence  $t^+ = \tau - t$ ) and varies the number of centers,  $C \in \mathbb{C}_1$ . The shape of the density function for  $\mathbb{P}(N^+ \leq \hat{q}_p)$  depends on the ratio of  $t^+/t$  and shows very little variation with  $C$ , just as described in Section 3.1, and matching almost perfectly the relevant theoretical curves in Figure 1. In particular, when  $t = t^+$ , as in all cases in the right panel, the distribution is very close to uniform, empirically verifying the, perhaps unintuitive, result that increasing the number of centers in the trial, thus increasing the sample size upon which the MLEs are based, does not affect the accuracy of the quantile estimates. The theory predicts that the lines in the right panel should be horizontal; however, there is a slight positive gradient. This is because the theory is based on a continuous approximation whereas  $N^+$  is a discrete random variable. The density function for  $\mathbb{P}(N^+ < \hat{q}_p)$  (not shown) exhibits a slight negative gradient, supporting this explanation.

Figure 3 repeats Figure 1 and the left panel of Figure 2 but for the  $p = 0.25$  quantile. Again, the empirical results match the theory almost perfectly. As with  $p = 0.5$ , the estimate improves with increasing census time, but as predicted in Section 3.1, when  $t \ll t^+$ , the mass is now not evenly distributed between the regions close to 0 and close to 1.

When predicting quantiles for  $T^+$ , Theorem 1 suggests that the accuracy of the quantile is primarily dependent on the ratio of  $n_*^+/n_*$ . Thus, with a fixed  $n_*^+$  and  $t$ , and with  $\beta = C$ , there is essentially no change in the prediction accuracy; Figure 4 captures the close agreement between



**FIGURE 3** Theoretical density (left) and estimated density over repeated sampling (right) of  $\mathbb{P}(N_t^+ \leq \hat{q}_{0.25})$  for each  $t \in \mathbb{T}_1$ , with  $t^+ = 400 - t$ . This figure appears in color in the electronic version of this article



**FIGURE 4** Theoretical density (left) and estimated density over repeated sampling (right) of  $\mathbb{P}(T^+ \leq \hat{r}_{0.5})$  for each  $C \in \mathbb{C}_2$  with  $n_t^+ = 200$  fixed across all simulation runs. In the left-hand figure, all eight curves coincide. This figure appears in color in the electronic version of this article

the theoretical predictions and empirical results in this case. Alternatively, with  $\beta$  fixed, as the number of centers increases and/or as the census time increases, since each of these increases  $n$ , the density curves for  $\mathbb{P}(T^+ \leq \hat{r}_p)$  transition from a concentration at 0 and 1 to a point mass at  $p$  (see Supporting Information for theoretical predictions and empirical verification). For further validation of Theorem 1, plots for  $p = 0.25$  with  $t$  varying are provided in the Supporting Information.

## 4.2 | Adjusted prediction intervals

We now study empirically the effectiveness of using quantiles based on  $p^*(p)$  to derive prediction intervals, and compare with intervals based directly on  $p$ . At each simulation, a naive, unadjusted 90% interval was estimated by calculating  $\hat{q}_p$  for  $p = 0.05$  and  $p = 0.95$ . An adjusted 90% interval was also derived by using  $p^*(p)$  from (7) instead of  $p$ , both for  $p = 0.05$  and  $p = 0.95$ . The performance of the



**TABLE 2** The mean (over repeated sampling) of the true coverage probability and width of an intended 90% prediction interval for  $N^+$  using the unadjusted and adjusted methods

	Unadjusted		Adjusted	
	Coverage (%)	$w$	Coverage (%)	$w$
$t = 50, t^+ = 350$	63.7	140.5	89.1	245.6
$t = 100, t^+ = 300$	76.3	118.2	89.5	160.9
$t = 150, t^+ = 250$	81.9	99.0	89.5	120.0
$t = 200, t^+ = 200$	84.9	82.2	89.6	92.9
$t = 250, t^+ = 150$	86.9	66.6	89.8	72.0
$t = 300, t^+ = 100$	88.2	51.3	89.8	53.6
$t = 350, t^+ = 50$	89.2	34.5	89.9	35.1

intervals was assessed for each method by calculating the mean, over 2000 simulations, of the true prediction interval coverage. The mean width of the prediction intervals was also recorded. We first consider the case where all centers opened simultaneously, then the case of different center opening times.

#### 4.2.1 | All centers opened simultaneously

Table 2 shows the results for each  $t \in \mathbb{T}_1$  and  $t^+ = \tau - t$ . The unadjusted method gives satisfactory results for  $t \ll t^+$  only, as is to be expected given Theorem 1. For all other scenarios, the quantiles are inaccurately estimated and the coverage can be far less than intended, as low as 63.7% for a census time early on in the trial. Further diagnostics showed approximately equal contributions to undercoverage from  $\hat{q}_{0.05}$  being too high and  $\hat{q}_{0.95}$  being too low. In contrast, by applying (7), the coverage is consistently improved upon and corrected to almost exactly the desired 90%. The improved coverage does come with a cost of an increased interval width, but the increase seems proportionate.

When  $\beta = 0$ , (7) gives  $p^* = p$ : no correction is needed. We, therefore, also examined the effect of our adjustment when data are simulated using a much lower true parameter value,  $\beta = 50$ . In this case, the lowest coverage was 77.8%, observed when  $(t, t^+) = (50, 350)$ , improving to 90.2% after our adjustment, while when  $t = t^+ = 200$  the coverage improved from 84.1% to 90.0%; the full tabulation is provided in the Supporting Information.

Similar improvements to those in Table 2, but for the 95% prediction interval are also provided in the Supporting Information, confirming that the  $p^*$  adjustment performs equally well when adjusting quantiles, which are further into the tails of the distribution. A further table in the Supporting Information demonstrates an even more striking improvement than in Table 2, found when creat-

ing a 90% predictive interval but with  $C = 20$ ; for example, when  $(t, t^+) = (50, 350)$  the coverage improved from 59.2% to 89.7%.

#### 4.2.2 | Different center opening times

We consider two different opening time scenarios: (1) the center opening times are drawn uniformly and independently from the interval  $[0, t]$  and (2) half of the centers are opened at time 0 and half of the centers open at time  $t$ . The former mimics a gradual coming online of new centers, while the latter scenario could occur when an initial interim analysis suggests that many new centers must be opened to achieve the required sample size.

The investigation into quantile adjustment to obtain a 90% prediction interval (Table 2) was repeated for opening time scenarios (1) and (2), and the results are provided in Tables 3 and 4, respectively. The prediction intervals for these cases were constructed according to the methodology of Section 3.2. Additional diagnostics for the moment matching were also recorded: the mean (over repeated samples) of  $t^*$ , the ratio of this to the mean (over repeated samples) of the mean (over centers) of the  $t_c$ 's, and the ratio of the mean of the  $n^*$  to the mean of the  $n_{\cdot}$ .

In both cases, the intervals obtained by combining the methodology proposed in Section 3.2 with (7) produce coverages very close to 90%, whatever the census time. By contrast the unadjusted intervals suffered from coverages as low as 48% when  $t = 50$ . Typically the values of  $t^*$  and  $n^*$  are lower than  $t$  and  $n_{\cdot}$  (although their ratio is almost unchanged; see Section 3.2), representing the increased uncertainty in parameter values because some centers have not been open for the full time interval. The especially poor coverage of the unadjusted intervals results because it is now the ratio  $t^+/t^*$  that determines the extent of the undercoverage.

Equivalent tables for  $T^+$  for opening time scenarios (1) and (2), presented in the Supporting Information, show similar dramatic improvements.

### 4.3 | Application to clinical trial recruitment data

Finally, we applied our methodology to recruitment data from an oncology clinical trial. The recruitment centers opened at different times, thus the methodology of Section 3.2 applies. For anonymization reasons, all times in the data set were jittered by up to a week, and in the plot described below both time and cumulative recruitment have been rescaled to lie in the interval  $[0, 1]$ .

**TABLE 3** The mean (over repeated sampling) of the true coverage probability and width of an intended 90% prediction interval for  $N^+$  using the unadjusted and adjusted methods for opening time scenario (1)

	$t^*$	$t^*/t_c$	$n^*/n_c$	Unadjusted Coverage (%)	$w$	Adjusted Coverage (%)	$w$
$t = 50, t^+ = 350$	24.4	0.957	0.956	49.3	143.1	89.2	341.4
$t = 100, t^+ = 300$	46.5	0.921	0.920	65.0	125.3	89.6	220.3
$t = 150, t^+ = 250$	67.2	0.891	0.890	72.7	106.7	89.6	160.0
$t = 200, t^+ = 200$	86.9	0.866	0.865	77.6	88.8	89.7	119.7
$t = 250, t^+ = 150$	105.9	0.845	0.843	81.3	71.5	89.7	88.7
$t = 300, t^+ = 100$	124.2	0.826	0.825	84.2	54.3	89.7	62.6
$t = 350, t^+ = 50$	142.1	0.810	0.809	87.1	35.5	89.8	38.2

**TABLE 4** The mean (over repeated sampling) of the true coverage probability and width of an intended 90% prediction interval for  $N^+$  using the unadjusted and adjusted method for opening time scenario (2)

	$t^*$	$t^*/t_c$	$n^*/n_c$	Unadjusted Coverage (%)	$w$	Adjusted Coverage (%)	$w$
$t = 50, t^+ = 350$	21.7	0.867	0.863	48.1	145.1	89.1	360.4
$t = 100, t^+ = 300$	38.3	0.766	0.763	60.0	126.8	89.1	240.0
$t = 150, t^+ = 250$	51.2	0.683	0.679	66.7	108.7	89.0	179.0
$t = 200, t^+ = 200$	61.7	0.612	0.614	71.1	90.9	88.9	136.0
$t = 250, t^+ = 150$	70.1	0.561	0.558	75.3	73.4	89.0	101.3
$t = 300, t^+ = 100$	77.0	0.513	0.511	79.6	55.6	89.4	70.8
$t = 350, t^+ = 50$	82.8	0.473	0.471	84.2	36.2	89.6	41.8

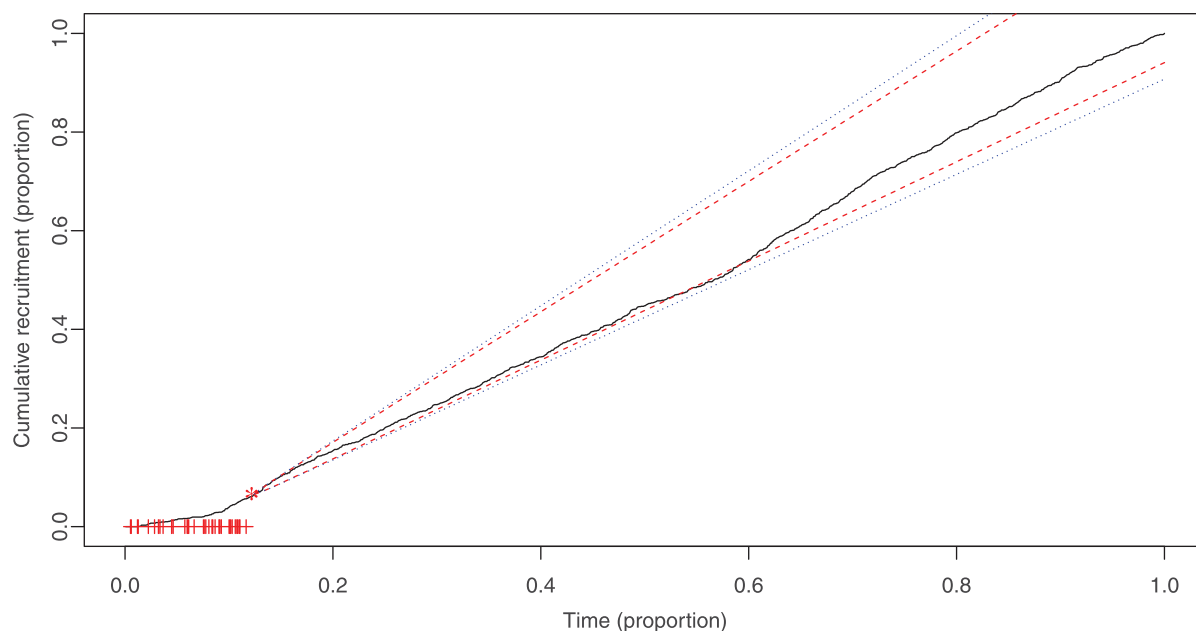
We examined the 41 centers that had opened by time 0.125 and calculated 90% prediction intervals for the total recruitment from these centers for the remainder of the recruitment period. With a single data set it is impossible to obtain true coverage probabilities, however, we can compare the predicted intervals with the true number recruited. Figure 5 shows the prediction intervals as dashed lines in red (unadjusted intervals) and dotted lines in blue (adjusted intervals) with the actual recruitment numbers shown as a solid black line (this figure appears in color in the electronic version of this article, and color refers to that version). The recruitment goes outside of the unadjusted interval just before time 0.6 yet remains entirely within the adjusted prediction interval.

A diagnostic likelihood-ratio test (see Urbas *et al.*, 2020) with a null hypothesis that the Poisson process is time-homogeneous (as assumed by the model) produced  $p$ -values of 0.39 (data up to the census time) and 0.50 (all data). Diagnostic Q-Q plots (see the Supporting Information) suggest that the assumption of a gamma distribution in (1) is reasonable.

## 5 | DISCUSSION

We must start by pointing out that the model described in Section 2 is just that: a model. The hierarchical nature allows the borrowing of information from centers that have been open for some time and enables sensible predictions for newly opened centers, the Poisson process is a reasonable first approximation for the recruitment process at an individual center, and the gamma hierarchical distribution is chosen for tractability. The model does not account for the myriad logistical issues that might occur during a trial, affecting recruitment, and even were this not the case, data do not arise from the model. However, the model has gained traction in the industry and has been developed further by a number of authors (see the introduction of this article).

Theorem 1 first provides insight into when prediction intervals obtained by simply plugging in the parameter point estimates might be adequate; for example, when the future time horizon is small compared with the time for which the trial has been running. However, often the future time horizon is at least as long as the current



**FIGURE 5** Recruitment (black, solid) for an oncology clinical trial. The estimated 90% prediction intervals are shown by red dashed lines (unadjusted intervals) and dotted blue lines (adjusted intervals), and the “+” symbols indicate center opening times. This figure appears in color in the electronic version of this article, and color refers to that version

length of the trial, and in this situation the coverage of plug-in intervals is poor. The methodology resulting from Theorem 1 essentially, takes parameter uncertainty into account to create prediction intervals with almost exactly the intended coverage.

Alternatives that allow parameter uncertainty to inform prediction intervals include Bayesian methods (e.g., Zhang and Long, 2010; Urbas *et al.*, 2020), which are typically computationally expensive, or the bootstrap, which is usually even more expensive. Our method has the same cost as the standard plug-in, frequentist approach.

The diagnostics detailed at the end of Section 4.3 suggest that the model of Anisimov and Fedorov (2007) is suitable for the oncology data set that we examine, but this might not always be the case. The simulation study in Section S5 of the Supporting Information suggests robustness to moderate departures from the hierarchical gamma distribution and robustness of improvements to the prediction intervals through our method. However, as demonstrated, for example, in Urbas *et al.* (2020), if the intensity curve for each center is strongly time-dependent, predictions based on the assumption of a homogeneous Poisson process can be wildly inaccurate, and a time-inhomogeneous Poisson process might be more appropriate (e.g., Lan *et al.*, 2019; Urbas *et al.*, 2020). If the exact form of the time-inhomogeneity is known then the standard time transformation used for the Poisson process ( $t \rightarrow \lambda_c \int_0^t a(s)ds$ , where  $\lambda_c a(s)$  is intensity at true time  $s$  for center  $c$ ), with one transformed time scale per center, permits the application of our correction to predicting the number of new recruits in a given addi-

tional (true) time. However, the time-dependency typically contains unknown parameters, and our correction as it stands cannot account for the uncertainty in these. Future research could look into extending our method to allow for this.

Theorem 1, upon which our prediction adjustment is based, describes the limit as the number of centers  $C \rightarrow \infty$ . Our simulations suggest that the approximation based on the limit result works well even when  $C$  is as low as 20; however, it is unlikely to hold for very low center numbers. Furthermore, experience has shown that for very low center numbers it is possible for the likelihood to increase monotonically as  $\alpha \rightarrow \infty$  and  $\beta \rightarrow \infty$  with  $\alpha/\beta$  fixed (this can occur when the counts for individual centers are underdispersed). Relevant historical data might then be brought in to make parameter estimation more robust; however  $C$  would still be low and the intended coverage might not be achieved.

This article has considered scenarios where centers can open at different times up until the census time, additional centers may be opened at the census time (perhaps driven by the results of the analysis) and we wish to predict the total recruitment for these centers into the future. A more general opening time scenario would also allow for centers coming online at different times after the census time. This could be incorporated into predictions of recruitment over the remainder of the recruitment period via a more general definition of  $\lambda_*$ , which would become a weighted sum of the individual intensities, with a center's weight being the fraction of the future time that it would be open for.

How to deal with the converse problem in this scenario: predicting the time to recruit a certain number of patients, is an open problem.

Prediction using the model of Anisimov and Fedorov (2007) relies on the true center opening times, which are rarely known in advance. There is often a plan and a back up plan, however, and it is straightforward (see the appendix of Urbas *et al.*, 2020) to combine the Anisimov and Fedorov (2007) model with a standard survival model for the opening time of each center conditional on the planned opening time and, potentially, other covariates. Alternatively, Lan *et al.* (2019) models center opening times as realizations from an inhomogeneous Poisson process. With either of these approaches, once the model has been fitted using the data up to the census time, it is straightforward to repeatedly simulate sets of future opening times. One would then obtain a mixture of negative-binomial distributions for the distribution of the number of additional recruits over additional time  $t^+$ . The mixture could be approximated by a single negative-binomial distribution and our method applied directly to that. This would mainly be an extension of the model of Anisimov and Fedorov (2007), and would certainly be interesting to explore; the aim of this paper, however, is to analyze the existing method of Anisimov and Fedorov (2007) and, in addition to new insights on performance, provide improved prediction intervals.

## ACKNOWLEDGMENTS

The first author acknowledges support from award: NIHR-MS-2016-03-01 Lancaster University.

## DATA AVAILABILITY STATEMENT

The results in this paper are mainly from simulations studies; the one clinical-trial data set that we use is not shared due to privacy/ethical restrictions.

## ORCID

Rachael Mountain  <https://orcid.org/0000-0002-5597-4529>

Chris Sherlock  <https://orcid.org/0000-0002-2429-3157>

## References

- Abbas, I., Rovira, J. and Casanovas, J. (2007) Clinical trial optimization: Monte carlo simulation Markov model for planning clinical trials recruitment. *Contemporary Clinical Trials*, 28, 220–231.
- Akobeng, A.K. (2005) Understanding randomised controlled trials. *Archives of Disease in Childhood*, 90, 840–844.
- Anisimov, V.V. (2011) Statistical modeling of clinical trials (recruitment and randomization). *Communications in Statistics - Theory and Methods*, 40, 3684–3699.
- Anisimov, V.V. and Fedorov, V.V. (2007) Modelling, prediction and adaptive adjustment of recruitment in multicentre trials. *Statistics in Medicine*, 26, 4958–4975.
- Bakhshi, A., Senn, S. and Phillips, A. (2013) Some issues in predicting patient recruitment in multi-centre clinical trials. *Statistics in Medicine*, 32, 5458–5468.
- Carlisle, B., Kimmelman, J., Ramsay, T. and MacKinnon, N. (2015) Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. *Clinical Trials*, 12, 77–83.
- Carter, R.E. (2004) Application of stochastic processes to participant recruitment in clinical trials. *Controlled Clinical Trials*, 25, 429–436.
- Carter, R.E., Sonne, S.C. and Brady, K.T. (2005) Practical considerations for estimating clinical trial accrual periods: application to a multi-center effectiveness study. *BMC Medical Research Methodology*, 5, Article number: 11.
- Haidich, A.-B. and Ioannidis, J.P. (2001) Determinants of patient recruitment in a multicenter clinical trials group: trends, seasonality and the effect of large studies. *BMC Medical Research Methodology*, 1, 1–11.
- Huang, G.D., Bull, J., McKee, K.J., Mahon, E., Harper, B. and Roberts, J.N. (2018) Clinical trials recruitment planning: a proposed framework from the clinical trials transformation initiative. *Contemporary Clinical Trials*, 66, 74–79.
- Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994) *Continuous Univariate Distributions*, Wiley Series in Probability and Mathematical Statistics, 2nd edition. New York: Wiley.
- Lai, D., Moyé, L.A., Davis, B.R., Brown, L.E. and Sacks, F.M. (2001) Brownian motion and long-term clinical trial recruitment. *Journal of Statistical Planning and Inference*, 93, 239–246.
- Lamberti, M.J., Mathias, A., Myles, J.E., Howe, D. and Getz, K. (2012) Evaluating the impact of patient recruitment and retention practices. *Drug Information Journal*, 46, 573–580.
- Lan, Y., Tang, G. and Heitjan, D.F. (2019) Statistical modelling and prediction of clinical trial recruitment. *Statistics in Medicine*, 38, 945–955.
- Mijoule, G., Savy, S. and Savy, N. (2012) Models for patients' recruitment in clinical trials and sensitivity analysis. *Statistics in Medicine*, 31, 1655–1674.
- Senn, S. (1997) *Statistical Issues in Drug Development*. Statistics in Practice. Chichester, New York: John Wiley.
- 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.
- Treweek, S., Lockhart, P., Pitkethly, M., Cook, J.A., Kjeldstrøm, M., Johansen, M., Taskila, T.K., Sullivan, F.M., Wilson, S., Jackson, C., Jones, R. and Mitchell, E.D. (2013) Methods to improve recruitment to randomised controlled trials: Cochrane systematic review and meta-analysis. *BMJ Open*, 3, e002360.
- Urbas, S., Sherlock, C. and Metcalfe, P. (2020) Interim recruitment prediction for multi-center clinical trials. *Biostatistics*. <https://doi.org/10.1093/biostatistics/kxaa036>
- Ying, G., Heitjan, D.F. and Chen, T. (2004) Nonparametric prediction of event times in randomized clinical trials. *Clinical Trials*, 1, 352–362.
- Zhang, Q. and Lai, D. (2011) Fractional Brownian motion and long term clinical trial recruitment. *Journal of Statistical Planning and Inference*, 141, 1783–1788.
- Zhang, X. and Long, Q. (2010) Stochastic modeling and prediction for accrual in clinical trials. *Statistics in Medicine*, 29, 649–658.

**SUPPORTING INFORMATION**

Web Appendices, Tables, and Figures referenced in Sections 3, 4, and 5 are available with this paper at the Biometrics website on Wiley Online Library. In addition, the computing code used in Sections 3 and 4 is also available in the Supporting Information.

**How to cite this article:** Mountain R, Sherlock C. Recruitment prediction for multicenter clinical trials based on a hierarchical Poisson–gamma model: Asymptotic analysis and improved intervals. *Biometrics*. 2022;78:636–648.  
<https://doi.org/10.1111/biom.13447>