

## RESEARCH ARTICLE

# Modeling time-varying recruitment rates in multicenter clinical trials

Aris Perperoglou<sup>1</sup>  | Youyi Zhang<sup>2</sup> | Dimitra-Kleio Kipourou<sup>1</sup>

<sup>1</sup>Human-Centered AI & ML, Digital Health, R&D, AstraZeneca, Cambridge, UK

<sup>2</sup>Advanced Analytics, Data Science & AI, AstraZeneca, Cambridge, UK

**Correspondence**

Dimitra-Kleio Kipourou,  
Human-Centered AI & ML, Digital Health, R&D, AstraZeneca, Cambridge, UK.

Email: [kleio.kipourou@astrazeneca.com](mailto:kleio.kipourou@astrazeneca.com)

**Abstract**

Multicenter phase II/III clinical trials are large-scale operations that often include hundreds of recruiting centers in several countries. Therefore, the operational aspects of a trial must be thoroughly planned and closely monitored to ensure better oversight and study conduct. Predicting patient recruitment plays a pivotal role in trial monitoring as it informs how many people are expected to be recruited on a given day. As such, study teams may rely on predictions to assess progress and detect any deviations from the original plan that might put the study and potentially, patients at risk. Recruitment predictions are often based on a Poisson-Gamma model that assumes that centers have a constant recruitment rate throughout the trial. The model has useful features and has extensively been used, yet its main assumption is often unrealistic. A nonhomogeneous Poisson process has been recently proposed that can incorporate time-varying recruitment rates. In this work, we predict patient recruitment using both approaches and assess the impact of said assumption in a real-world setting where studies may not necessarily have constant center-specific recruitment rates. The paper showcases experience from modeling recruitment in trials sponsored by AstraZeneca between 2005 and 2018. In these data, there is evidence of time-varying recruitment rates. The predictive performance of models that account for both constant and time-varying recruitment effects is presented. Following a descriptive analysis of data, we assess model performance and investigate the impact of model misspecification.

**KEYWORDS**

Bayesian model averaging, homogeneous Poisson-Gamma, nonhomogeneous Poisson-Gamma, patient recruitment

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.

© 2022 The Authors. *Biometrical Journal* published by Wiley-VCH GmbH.

## 1 | INTRODUCTION

A randomized multicenter clinical trial (RMCT) is a large-scale operation that involves several countries, centers, investigators, and patients. Along with the clinical aspect of the trial and what is described in the research protocol, a RMCT also entails operational planning and optimization of procedures. Some of the decisions involved in the planning phase include determining the list of countries and the number of centers that will be responsible for the recruitment, screening, treatment, and follow-up of the patients. These decisions are based on assumptions of patient recruitment capabilities for each center, based on past experience or team knowledge. For any trial to be successful, while ensuring safety of participants and minimization of their burden, recruitment prediction models are crucial to guide the planning phase. Once the trial initiates recruitment, study analytics and predictions are vital to assess whether the study is going according to plan, and indicate if action is needed to address deviations from the plan. Such deviations may jeopardize the success of the trial or even patient safety. For example, lower (than expected) recruitment rates can cause serious delays in a study, significantly increase the cost of a trial, and put patients at risk by delaying their treatment, leading some patients to seek treatment elsewhere. Equally, higher recruitment rates than expected might put patients safety at risk, if the trial supply chain is not ready, or if centers cannot provide treatment to a greater number of patients.

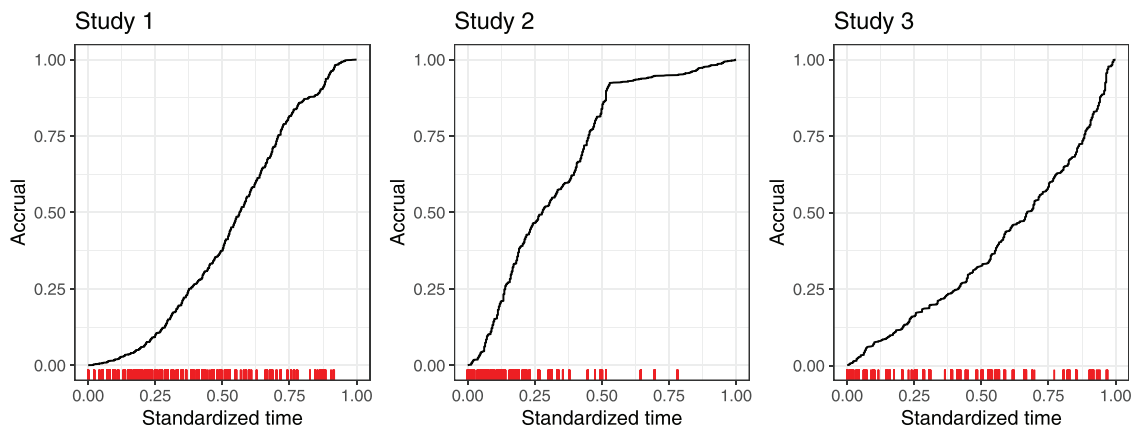
Early attempts to address the problem focused on deterministic methods in which recruitment rate was considered fixed and the recruitment period was defined by dividing the target number of recruited patients by the expected recruitment rate by unit of time, usually per month. The method—also called unconditional—was extended by Carter (2004), Comfort (2013), and Moussa (1984), with a conditional model where recruitment rates could vary with time. Although such approaches may be easy to implement, they often lack merits of stochastic approaches and heavily rely on user inputs. Lee (1983) defined a method for predicting recruitment duration based on Poisson methodology. Poisson models were also discussed in Carter (2004), Carter et al. (2005), Senn (1998), and Williford et al. (1987). Piantadosi and Patterson (1987) allowed for nonhomogeneous Poisson processes that can deal with a decaying recruitment rate over the course of the trial. Tang et al. (2012) introduced a piecewise linear intensity that could account for slow initial recruitment and a potential surge in recruitment rates as trial reaches the final stages.

Other approaches introduce Bayesian extensions of Poisson models to allow for prior knowledge to inform predictions before data are made available, or, in general, at the very early stages of the trial (Bakhshi et al., 2013; Gajewski et al., 2008; Zhang and Long, 2010), and Monte Carlo simulations (Abbas et al., 2007). For more thorough reviews, see Barnard et al. (2010), Gkioni et al. (2019), and Heitjan et al. (2015).

In this work, we focus on approaches that model recruitment at center level and also allow for centers to start and/or stop at specific times during the course of a trial. Such a model is the well-known Poisson-Gamma (PG) model, introduced by Anisimov and Fedorov (2007). The PG model is commonly used within pharmaceutical companies (also called “industry standard”) and is based on a homogeneous Poisson approach where each center is allowed to recruit with an individual  $\lambda$  rate that is coming from a Gamma distribution. This is a homogeneous approach that cannot handle recruitment rates that may vary in time. Lan et al. (2019) described a nonhomogeneous Poisson process that allows for staggered center activation and heterogeneity within centers. More importantly, the model assumes that each center will reach a plateau of recruitment capacity, from which point onward, it will start declining. Urbas et al. (2020) also extended the Poisson-Gamma approach to a nonhomogeneous approach that models recruitment rate as a decaying function of time. Urbas et al. defined a nonhomogeneous approach where recruitment rates for each center are decreasing with time by introducing the product of an initial centerwise recruitment rate with a prespecified time function.

The goal of this work is to review the performance of the homogeneous versus nonhomogeneous models in real data coming from past RMCTs. The comparison will focus on Poisson-Gamma-based models defined in Anisimov and Fedorov (2007) and Urbas et al. (2020). We draw our experience from a set of studies conducted by AstraZeneca during the period 2005–2018. The purpose is not only to identify which model performs best, but also to explore model behavior in cases where the real pattern of the data is not correctly specified in the model. In their work, Urbas et al. (2020) performed a similar comparison using simulated nonhomogeneous data. Here, we investigate the impact of model misspecification—with respect to the center-specific recruitment rate—on the predictive performance of the models, using data from a series of real RMCTs. We will investigate the ability of Bayesian model averaging (BMA, suggested as a strategy in Urbas et al., 2020) to identify an optimal model in real data and show examples where the pattern of recruitment does not seem to fit with any of the prespecified time functions.

The paper is organized as follows: in the next section, we will briefly review the PG models and nonhomogeneous PG model with BMA. In Section 3, we will describe the data in detail, and in Section 4, we will present an overall comparison of the models and a more detailed example based on three studies with various recruitment trends. The paper closes with a discussion.



**FIGURE 1** Accrual enrolment (black solid lines) with center openings marked with red vertical lines on the bottom. Time is standardized to (0–1) duration of trial and recruitment to percentage of total patients recruited.

## 2 | METHODS

### 2.1 | Poisson-Gamma model

Consider a trial involving  $c = 1, 2, \dots, C$  centers, with the aim of recruiting a total of  $n$  patients. Since centers are allowed to open and stop at any point during recruitment, each center can recruit patients during the span of  $\tau_c$  time units it remains open. Assuming that a center can recruit with a  $\lambda_c$  rate, then the number of patients  $N_c$  a center may recruit can be described by a Poisson model where  $N_c | \lambda_c \sim \text{Poisson}(\lambda_c \tau_c)$ ,  $c = 1, 2, \dots, C$ . Recruitment rate  $\lambda_c$  can be seen as a sample from a Gamma distribution such that  $\lambda_c \sim \Gamma(\alpha, \alpha/\phi)$ .

To estimate recruitment in an ongoing trial, assume that at some intermediate point  $t_1$ , the study has already recruited  $N_{t_1} = \sum_{c=1}^C m_{c,t_1}$  patients in all  $C$  centers. Then, the number of people recruited by each center up to this point,  $n_{c,t_1}$ , has a negative binomial distribution with parameters  $(\alpha, \tau_c \phi/a)$  with  $\tau_c$  the actual recruitment duration for center  $c$  up to that time. Then, with  $m = \alpha/\phi$ , we have:

$$P(n_{c,t_1} = n) = P(n, \tau_c, \alpha, m) = \frac{\Gamma(n+a)}{n! \Gamma(a)} \left( \frac{m \tau_c}{\alpha + m \tau_c} \right)^n \left( \frac{\alpha}{\alpha + m \tau_c} \right)^a, \quad n = 0, 1, \dots \quad (1)$$

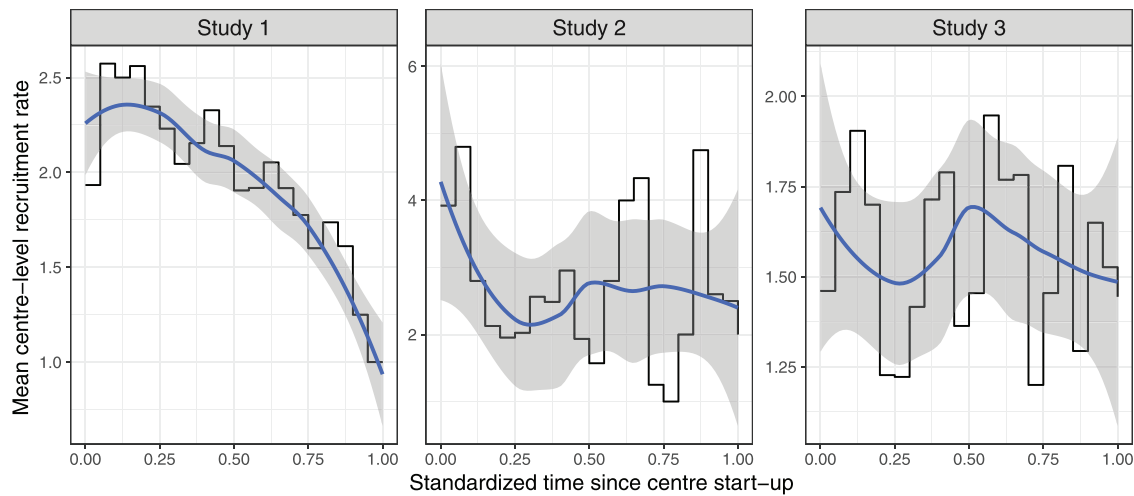
Parameters  $\alpha, \phi$  can be estimated either by maximum likelihood or empirical Bayes methods.

### 2.2 | Motivating example: Nonhomogeneous data

The homogeneous model assumes that recruitment rates remain constant throughout the duration of the study. When this assumption is met, predictions are accurate and the model seems to follow the data well. In practice, a clinical trial has to achieve the target number of patients and investigators will closely monitor recruitment. When the overall study recruitment rate shows signs of stagnation, investigators will take action to safeguard the operational success of the trial. An action that can be taken would be to activate more centers. In many cases, a constant recruitment rate at the study level can be attributed to a number of centers that are being activated in various timepoints of the study. That may lead to a constant recruitment rate at trial level but the same does not necessary hold on a center level.

Figure 1 showcases data from threephase III clinical trials.<sup>1</sup> Recruitment time (on  $x$ -axis) has been standardized throughout all data, dividing actual time by recruitment duration, computed by the time difference between the dates when the last patient was enrolled with the date the first subject enrolled in the study. Accrual on the  $y$ -axis indicates the amount of patients recruited divided by the total amount of patients who entered the study. Red points in the rug plot indicate the times at which new centers are initiated during the study. All three studies utilize more than 100 centers and aim to recruit their target population with a period of 1 year. Figure 2 presents the average recruitment rate per center for the

<sup>1</sup> All three studies presented here are part of a real dataset described in the next session. Study identifiers have been removed to safeguard data confidentiality.



**FIGURE 2** Average center-level recruitment rate for three random studies. Black lines present the average per-center recruitment rate, and smooth line is a lowess smoother.

same three studies. Rate at each center is computed as the number of patients recruited within 0.1 units of standardized time. Standardized time (presented on  $x$ -axis) is time from center activation (time 0) until the end of study. For example, for a study with recruitment period of 2 years, a 0.1 unit corresponds to approximately 2 months.

The recruitment rates for studies 1 and 2 are almost constant, with the overall recruitment curve been reasonably close to a straight line. In Study 1, overall recruitment starts at a slower pace in the first period and then speeds up at the 25th percentile of study duration. Prior to this timepoint, there is a concentration of centers being initiated (rug plot). That leads to the overall study recruitment picking up, while when looking at a center level (Figure 2) the average within center recruitment lever drops. A similar pattern can be seen for Study 2 (center graph of 2), where the overall study recruitment rate appears constant as an effect of increasing number of centers initiated, whereas center-level recruitment drops from four patients per month to about two patients per month after the first quarter of the study 2. At halfway through the study, where almost no new centers are activated, the recruitment rate drops significantly. Another interesting feature is the sharp plateau toward the end of recruitment period. There are various reasons that might explain such a plateau, for instance, it could be that recruitment pauses for safety reasons, or during the end of recruitment period, centers slow down their efforts because they do not want to overrecruit patients. Study 3 showcases an example of what seems to be a falling overall recruitment study pace (2) when centers are activated uniformly across study duration and average center recruitment rate remains around 1.25–1.75 patients throughout the duration of trial.

For the three studies above, the average recruitment rate is not monotonic, and hence, it is violating the PG's main underlying assumption. Thus, predictions produced by PG are expected to be erroneous.

### 2.3 | Nonhomogeneous Poisson model

To deal with time-varying recruitment effects, Urbas et al. (2020) extended the PG model by including an interaction of a time function with the initial recruitment rate of each center. The nonhomogeneous PG model assumes an initial per center recruitment rate  $\lambda_c^0 \sim \Gamma(\alpha, \alpha/\phi)$  that subsequently interacts with a nonnegative time function  $g(t; \theta)$  with  $\theta$  being a parameter vector associated with this function. Under this assumption, the number of people recruited at each center  $c$  depends on the initial center recruitment rate  $\lambda_c$ , the number of days the center is recruiting for  $\tau_c$ , and the time-varying recruitment intensity function  $g(t; \theta)$  such that:

$$N_c^{\tau_c} | \lambda_c^0 \sim \text{Poisson} \left( \lambda_c^0 \int_{t-1}^t g(s; \theta) ds \right), \text{ for } t = 1, \dots, \tau_c.$$

In practice, the extended approach introduces a family of models that can include a homogeneous model when  $g(t; \theta) = 1$  or several other models depending on the shape described by the time functions. In their work, Urbas et al. (2020)

suggested using a time function of the form:

$$g_k(t; \theta) \propto \left(1 + \frac{\theta t}{k}\right)^{-k}, \quad t \leq 0, \theta, k > 0.$$

Such a choice can be justified for each mathematical convenience but also for the family of options it provides. When  $k = 0$ , the model will simplify to a homogeneous PG model, while when  $k \rightarrow \infty$ , then it will have an exponential tail. The original idea was formulated around some oncology trials, where patients' arrivals are often sporadic and recruitment rates are assumed to decline with time. The authors suggested a finite choice of  $k$  values, namely, from the set  $\{0, 0.5, 1, 2, \infty\}$ , such that:

$$\begin{aligned} g_0(t) &\equiv 1, \\ g_1(t; \theta) &= \frac{\theta(1 + \theta t)^{-1}}{\log(1 + \theta \tau)^\tau}, \\ g_k(t; \theta) &= \frac{\theta(1 - k)(1 + \theta t/k)^{-k}}{k(1 + \theta \tau/k)^{1-k} - k}, \quad k \notin 0, 1, \infty, \\ g_\infty(t; \theta) &= \frac{\theta \exp(-\theta t)}{1 - \exp(-\theta \tau)^\tau}. \end{aligned}$$

In a real life example, predictions can be made using some prior knowledge on past data or historical trials and current data  $y$  that become available as the study progresses. Historical data are used to form a prior distribution  $\pi_0(\alpha, \phi, \theta)$ . Both homogeneous and nonhomogenous models require knowledge for  $\alpha$  and  $\phi$ . Urbas et al. (2020) defined a weakly informative normal prior for  $\alpha$  based on their training set on oncology data, while for  $\phi$ , which represents the mean center recruitment, they used a noninformative uniform prior. Selecting a prior for  $\theta$  is more complicated. We refer the reader to Urbas et al. (2020) for a detailed discussion. Given data  $y$  and priors, a posterior distribution is denoted as  $\pi(\alpha, \phi, \theta|y)$ .

## 2.4 | Bayesian model averaging

In order to decide if it is justified to divert from a homogeneous PG model to a nonhomogeneous PG model, a nonhomogeneity test (Fierro and Tapia, 2011; Urbas et al., 2020) does exist to help with the choice. Equally, using maximum likelihood testing approach, one may fit all models and decide on the most appropriate time function based on the best likelihood.

An alternative approach would be to use BMA (Raftery et al., 1997). For each value of  $k$  in set  $\{0, 0.5, 1, 2, \infty\}$ , a parametric model  $M_k$  is considered, with prior model probability  $\pi_0(M_k)$ . When data  $y$  are available, a posterior probability for model  $M_k$  is defined by  $\pi_p(M_k|y)$ .

Given the observed data  $y$  and  $\pi_p(M_k|y) \propto \pi(y|M_k)\pi_0(M_k)$ , then the posterior probability of  $N^+$  patients recruited is given by:

$$P(N^+ = n^+|y) = \sum_{k=1}^K P(N^+ = n^+|y, M_k)\pi_p(M_k|y).$$

BMA requires the use of marginal densities  $\pi(y|M_k)$  that, in this setting, are clearly intractable. Importance sampling is used to obtain samples from the posterior distribution of each model and an estimate of  $\pi(M_k)$ . We refer the reader to Urbas et al. (2020) for a detailed description of the approach.

One advantage of BMA is that it also considers the uncertainty of selecting an appropriate model for the data and uses a form of a weighted average of predictions under all considered models. In practice, one may use to define specific prior probabilities for each  $M_k$ . In their work, Urbas et al. (2020) used all prior model probabilities being equal and included the PG model in the set of models.

**TABLE 1** Summary statistics on study-level characteristics based on the entire sample and by therapeutic area, where CV, cardiovascular; ON, oncology; RITA, respiratory; and Other, autoimmune diseases, infection & vaccines, neuroscience, and gastroenterology

		All	CV	ON	RITA	Other
<b>Number of trials</b>		101	28	22	22	29
<b>Number of centers</b>	(median)	113	138	110	164	68
	(min)	1	13	54	20	1
	(max)	1776	1776	315	568	321
	(IQR)	(57;944)	(75;957)	(82;212)	(92;366)	(34;194)
<b>Number of countries</b>	(median)	10	8	13	12	3
	(min)	1	1	1	1	1
	(max)	43	43	28	27	30
	(IQR)	(3;17)	(3;15)	(8;20)	(4;17)	(1;14)
<b>Number of patients</b>	(median)	915	1232	841	1170	487
	(min)	57	222	88	137	57
	(max)	20,293	20,293	3075	6773	2313
	(IQR)	(486;10,604)	(727;10,762)	(464;1958)	(653;3971)	(272;1400)
<b>Recruitment duration (in years)</b>	(median)	1.2	0.8	1.7	1.2	1.2
	(min)	0.2	0.4	0.9	0.2	0.3
	(max)	3.3	2.8	3.3	2.7	3.3
	(IQR)	(0;2)	(0;1)	(1;2)	(0;1)	(0;2)

### 3 | DESCRIPTION OF DATA

Data consist of 101 clinical trials, of phase III (95%), II (3%), or II–III (2%) conducted by AstraZeneca with a recruitment period spanning between 2005 and 2018. The set contains information on a total of 3271 centers, recruiting in 56 countries with a mean recruitment period of 1.4 (min: 0.2, max: 3.3) years.

Studies were evenly distributed among three major therapeutic areas (TAs): 27% cardiovascular, 22% oncology, 22% respiratory, and 28% other (such as autoimmune diseases, infection & vaccines, neuroscience, and gastroenterology). Each study aimed to recruit and then randomize a number of people as per protocol; the smallest trial had a recruitment target of 57 patients in 1 center, while the largest study aimed to enroll 20,293 people in 1315 centers operating in 42 countries. Across studies, the median amount of recruiting centers was 91 [IQR: (45;176)]. More details can be found in Table 1.

The mean center-level recruitment rate varied in each study. Many studies show evidence of a time-varying trend that could increase/decrease over time. In practice, such patterns are rarely monotonic and hard to categorize. In a real application, trial monitoring and recruitment prediction has to be done with limited data and knowledge of how a time-dependent effect might develop. On the one hand, a PG model will only be appropriate on studies with constant recruitment rate, and, on the other hand, the extension to a nonhomogeneous model as implemented in Urbas et al. (2020) can only deal with decreasing recruitment rates. In a real-world model, assessment recruitment rates can vary nonmonotonically within center over time.

In the next sections, both homogeneous and nonhomogeneous models are fitted to obtain predictions at two distinct times points during the trial, when 25% and 50% of targeted recruitment is achieved. For each TA that we consider in this work, we have computed specific priors for parameters  $\alpha$ ,  $\phi$ , and  $\theta$  in similar manner to that described in Urbas et al. (2020).

## 4 | RESULTS

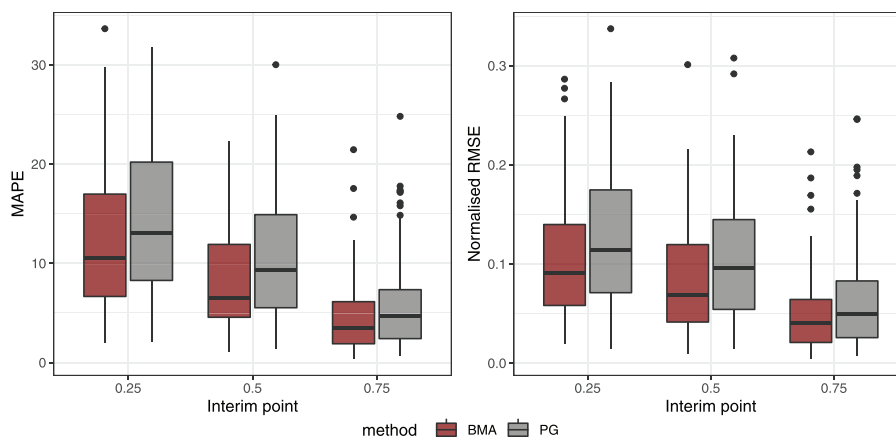
### 4.1 | Results on all studies

Predictive performance of a PG model versus the nonhomogeneous alternative with BMA was assessed at study level. The comparison focused on errors between predicted number of patients enrolled at a study level against reality, at three different timepoints  $t_i$  when 25%, 50%, and 75% of individuals have been recruited  $m = \{0.25, 0.5, 0.75\}$ . Two metrics were



**TABLE 2** Poisson-Gamma (PG) and BMA results on MAPE and RMSE with median and interquantile range (IQR) at three timepoints

Metric	Time	BMA		PG	
		Median	IQR	Median	IQR
MAPE	0.25	10.5	(6.7; 17)	13.1	(8.3; 20.2)
	0.5	6.5	(4.6; 11.9)	9.4	(5.5; 14.9)
	0.75	3.5	(1.9; 6.1)	4.7	(2.4 ;7.3)
RMSE	0.25	9.07	(5.81; 13.98)	11.39	(7.1; 17.47)
	0.5	6.83	(4.14; 11.95)	9.59	(5.41; 14.47)
	0.75	4.08	(2.08; 6.41)	4.99	(2.56; 8.28)

**FIGURE 3** Performance comparison via boxplots showing MAPE and (normalized) RMSE at three timepoints for Poisson-Gamma and BMA

used, namely, the mean absolute percentage error (MAPE) and the root mean squared error (RMSE) that are expressed as follows:

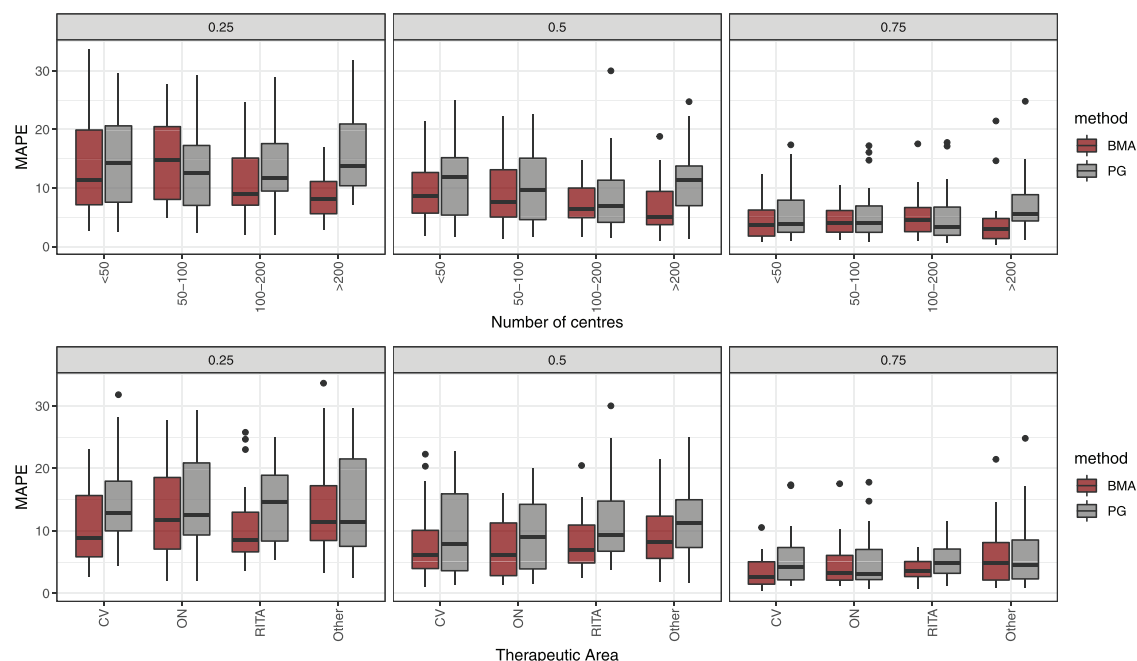
$$\text{MAPE} = \frac{1}{D} \sum_{d=1}^n \frac{|R_d - \hat{R}_d|}{R_d}, \text{RMSE} = \sqrt{\frac{1}{D} \sum_{d=1}^D \left( \frac{R_d - \hat{R}_d}{N} \right)^2}, \quad (2)$$

where  $R_d$  is the observed number of patients at a given day  $d$ , and  $\hat{R}_d$  is the predicted number of patients at a given day  $d$  when  $t_i\%$  of the target patient sample  $N$  has been recruited in the study.  $d = 1$  indicates the start of the recruitment and  $d = D$  the end of it. We calculated the metrics at different timepoints for different values of  $t_i$  to assess how MAPE and RMSE change over time. Smaller RMSE values indicate a better model fit and MAPE values that are less than 20% show good forecasting performance, while less than 10% indicate high accuracy. RMSE is also scaled to account for studies with different sample size; hence,  $R_d$  and  $\hat{R}_d$  are divided by  $N$ . All calculations have been implemented for each study separately, and figures and tables summarizing the results can be found below.

As expected, model accuracy increases when more data become available. Table 2 and Figure 3 illustrate errors for both approaches at the three timepoints of evaluation. PG had consistently higher MAPE and RMSE when compared to BMA across the board.

The same pattern repeats when looking at performance by study size—as defined by number of sites—or TA (see Figure 4 and Tables 3 and 4). BMA has consistently lower MAPE irrespective of time, number of centers, and TA with a few exceptions. Notably, the errors produced in PG models are consistently higher in studies with a large number of centers (>200) even at later stage of predictions where more data are available.

When looking at specific TAs (see Figure 4), BMA outperforms PG for all different disease indications at all timepoints where a model has been fitted.



**FIGURE 4** Performance comparison via boxplots showing MAPE by number of centres and therapeutic area (TA) at three timepoints for Poisson-Gamma (PG) and BMA

**TABLE 3** Poisson-Gamma (PG) and BMA results on MAPE with median and interquartile range (IQR) at three timepoints by therapeutic area. (CV, cardiovascular; ON, oncology; RITA, respiratory; and Other, autoimmunity, infection & vaccines, neuroscience and gastroenterology)

Therapeutic area	Time	BMA		PG	
		Median	IQR	Median	IQR
CV	0.25	8.9	(5.8;15.6)	12.8	(10;17.9)
	0.5	6	(4;10.1)	7.9	(3.6;15.9)
	0.75	2.6	(1.5;5)	4.3	(2.1;7.3)
ON	0.25	11.7	(7.1;18.5)	12.6	(9.3;20.9)
	0.5	6.1	(2.8;11.3)	9	(3.9;14.2)
	0.75	3.2	(2.1;6.1)	3.1	(2.2;7)
Other	0.25	11.5	(8.4;17.2)	11.4	(7.5;21.5)
	0.5	8.3	(5.6;12.3)	11.2	(7.3;15)
	0.75	4.8	(2.1;8.1)	4.6	(2.3;8.5)
RITA	0.25	8.6	(6.6;13)	14.7	(8.4;18.9)
	0.5	6.9	(4.8;10.9)	9.3	(6.7;14.8)
	0.75	3.6	(2.7;5.1)	4.9	(3.2;7.1)

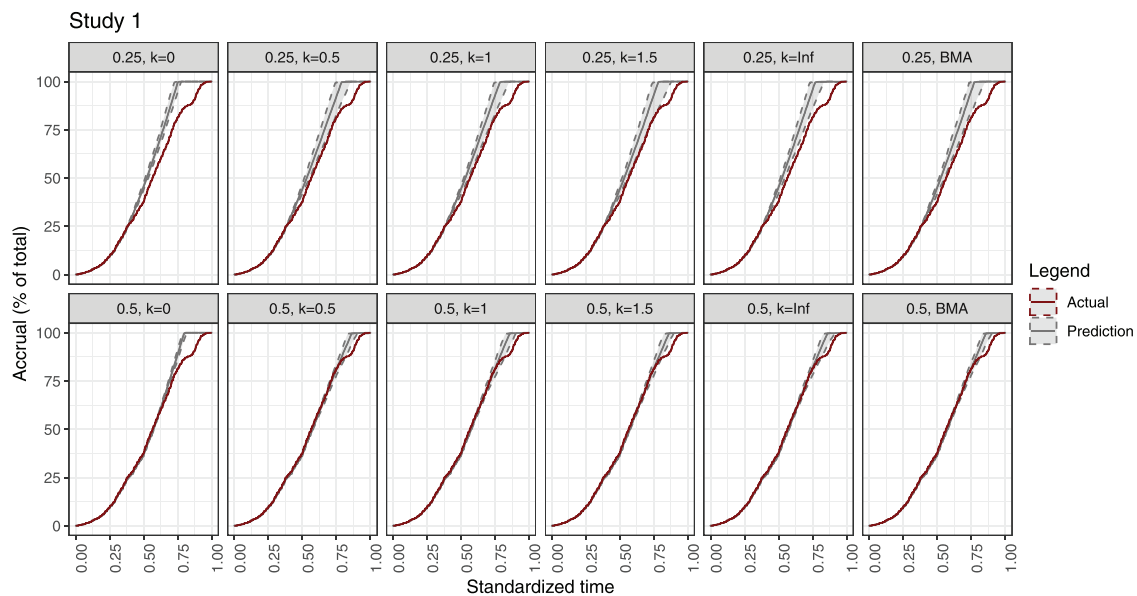
## 4.2 | Detailed example on three studies

Studies 1–3 discussed in the previous section all represent an average example of a phase III multicenter clinical trial, each with around 100 centers aiming to recruit patients with the period of around 1 year. None of these studies are on the edges of the data, neither aiming to recruit a small number of patients nor very large studies aiming to recruit more than 3000 patients. Studies were chosen for closer inspection due to their individual patterns: Study 1 represents a study in which recruitment rate seems to decrease following the first quarter of the trial duration; Study 2 has a recruitment rate that decreases after the first quarter followed by smaller fluctuations, while Study 3 seems to have a more stable recruitment rate.



**TABLE 4** Poisson-Gamma (PG) and BMA results on MAPE with median and interquantile range (IQR) at three timepoints by number of centers

Number of centers	Time	BMA		PG	
		Median	IQR	Median	IQR
<50	0.25	11.4	(7.1;19.9)	14.3	(7.6;20.6)
	0.5	8.7	(5.7;12.7)	11.8	(5.4;15.2)
	0.75	3.7	(1.8;6.2)	3.9	(2.4;7.9)
50–100	0.25	14.8	(8;20.5)	12.6	(7;17.3)
	0.5	7.6	(5;13.1)	9.7	(4.6;15.1)
	0.75	4.1	(2.5;6.2)	4	(2.4;6.9)
100–200	0.25	8.9	(7.1;15.1)	11.8	(9.5;17.6)
	0.5	6.4	(4.9;10)	7	(4.1;11.3)
	0.75	4.6	(2.5;6.7)	3.4	(1.9;6.7)
>200	0.25	8.1	(5.6;11.1)	13.8	(10.4;20.9)
	0.5	5	(3.7;9.4)	11.4	(7;13.8)
	0.75	2.9	(1.4;4.8)	5.6	(4.4;8.9)

**FIGURE 5** Predictions from several models, namely, homogeneous ( $\kappa = 0$ ), nonhomogeneous ( $\kappa=1, 1.5, \infty$ ), and BMA for Study 1

On each study, a total of  $M_k$  models have been fitted, with  $k \in \{0.5, 1, 1.5, \infty\}$ , at two different timepoints, an early point where 25% of recruitment had already been completed and midpoint at 50% of patient recruitment. At both points, future information has been censored and models predicted recruitment given observed data until that point.

Figures 5–7 show the predicted accruals (median and prediction intervals) against the actual accrual under each model and the overall Bayesian model average (BMA) at two timepoints. In Study 1, early predictions are driven by initial high recruitment rates. The optimistic prediction is dominant in  $M_{k=0}$  model with a constant recruitment rate, which overestimates recruitment accrual. The remaining four models that assume decaying rates are able to follow the data closer, although they do—at this early point—seem optimistic as well. BMA is generally predicting better than most individual models. Midway through the study, when 50% of recruitment data are available models, predictions are closer to the true curve. Again,  $M_{k=0}$  shows the larger deviation from reality (MAPE = 10) and BMA provides an overall better prediction than most individual models (MAPE = 5).

A similar pattern can be seen for Study 2, with early prediction being influenced by early recruitment rates that tend to decrease as time passes. The errors are more striking for this study because the reduction in recruitment rate is more apparent. Out of the five models,  $M_{k=1.5}$  and BMA have the smallest errors with MAPE equal to 3.9 and 4.7, respectively,

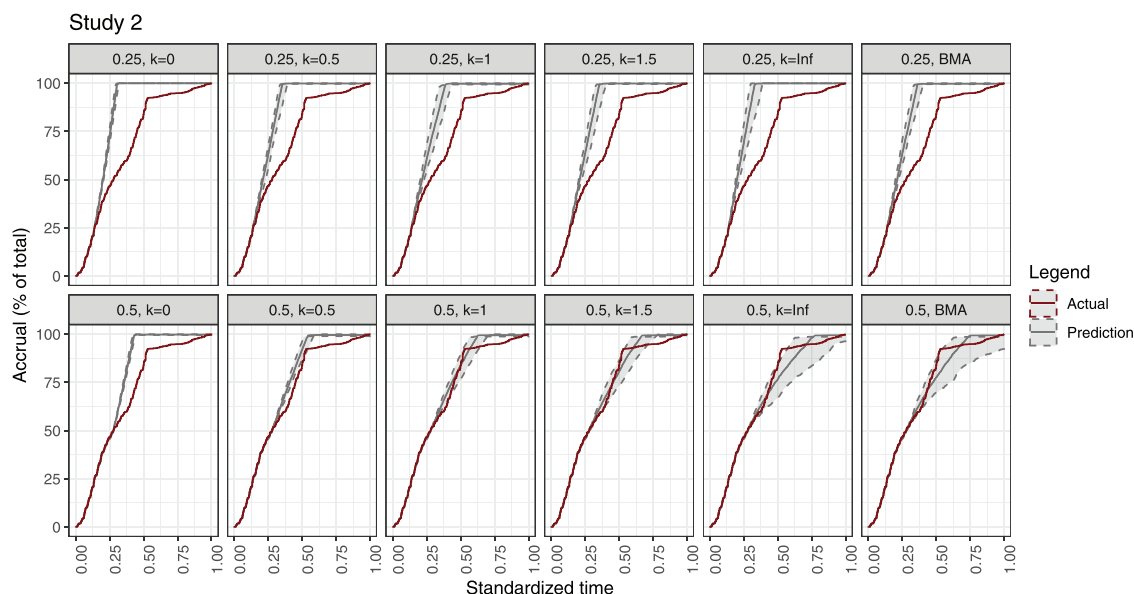


FIGURE 6 Predictions from several models, namely, homogeneous ( $\kappa = 0$ ), nonhomogeneous ( $\kappa=1, 1.5, \infty$ ), and BMA for Study 2

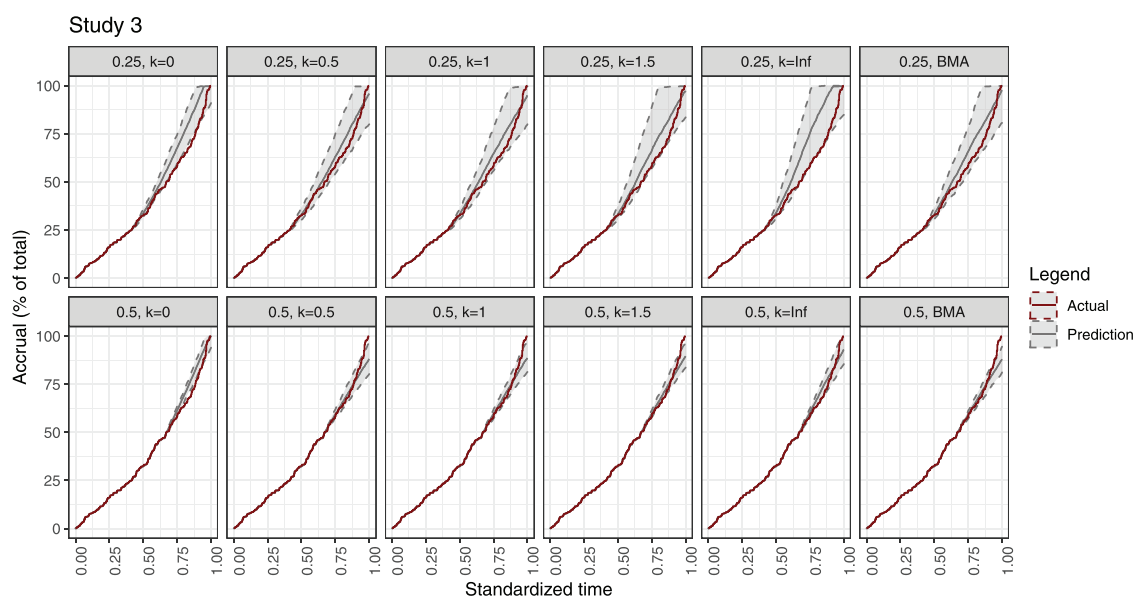


FIGURE 7 Predictions from several models, namely, homogeneous ( $\kappa = 0$ ), nonhomogeneous ( $\kappa=1, 1.5, \infty$ ), and BMA for Study 3

at the 50% of the study. In fact, all models but  $M_{k=0}$  (MAPE = 22) can follow the real data more closely showcasing that the time-varying feature of the model can identify the slowdown in recruitment rates.

In the final example, where recruitment rates appear more stable throughout the study, a constant model is comparable to the rest of the models and seems to outperform  $M_{k=\infty}$  that is more optimistic than the rest. More specifically, at 25% of the study, BMA has an MAPE tantamount to 8,  $M_{k=0}$  to 11 and  $M_{k=\infty}$  to 18. Predictions are more accurate at midpoint through the study with all predictions following closely the real data, with all MAPE for all models ranging between 4 and 6, while  $M_{k=1.5}$  was the most accurate.

As expected, prediction errors become smaller as we acquire more data. What is important to recognize though, is that the more flexible nonhomogeneous models adjust well to changes in recruitment patterns. Nonhomogeneous models ( $\kappa \in \{0.5, 1, 1.5, \infty\}$ ) with BMA capture the complicated accrual curve of study 2 at 0.5 (i.e., when 50% of the recruitment) (see Figure 6). Yet, all models fail to do so at earlier times. In contrast, in study 3, both homogeneous ( $\kappa = 0$ ) and BMA models perform well in both timepoints probably due to the roughly constant recruitment rate observed in Figure 2. BMA

assigns more weight to the homogeneous model yet it takes into account other models; hence, results are comparable to the homogeneous model but prediction intervals are wider.

## 5 | DISCUSSION

Predicting recruitment in RMCTs should be in theory a straightforward task. However, it is a very complex problem with many factors to consider. Despite the significance of accurate predictive models in such a framework, very little has been done to understand model complexities. Patient recruitment depends on a series of factors, namely, the number and location of the centers, center activation times, patient availability, disease indication, recruitment rates, randomization rates, and dropout rates. The list is not exhaustive. That underlines how complex such a problem can be. In this work, the focus was on the recruitment rate itself. Although several models have been recommended, the work by Anisimov and Fedorov on the PG model has been very popular and has been widely used in the pharmaceutical industry. The main advantage of their approach is that it models recruitment at a center level. That means that deviations from a trial plan and/or problems can be tracked down to their source. Nevertheless, we have shown that at center level, rates of patient recruitment may not be stable, or even follow a monotonic pattern. Changes in recruitment rates violate the basic assumption of the model and have a significant impact on its predictive ability.

In a set of real clinical trial data, we observed clear evidence of time-dependent recruitment rates. In many cases, patient availability wanes over time and recruitment rate slows down as a result of different causes (e.g., there might be another study team from a different trial competing for the same pool of patients). That effect of decaying recruitment rates is well established and study teams are tackling this issue by opening new study centers to keep overall study recruitment levels stable. Urbas et al. (2020) explored a different approach that can account for such an effect.

This work examined past clinical trial data to understand how this phenomenon can influence trial predictions. In a set of different studies, it was clear that there are cases where recruitment rate varies significantly and ignoring this can seriously affect the model's ability to predict a trial outcome. We have shown a specific example where even a mild recruitment decay (Study 1) will lead to large errors, especially at the beginning of a trial. Moreover, we showed that errors tend to accumulate as the number of centers increases. It is expected the accuracy of the model when predicting recruitment is limited at the start of a trial and increases as more data become available. For models that do not have the ability to capture a time-varying effect, the errors accumulate as the number of centers increases. Studies with a large number of centers may be generically fall into one of two categories: either studies that require a large number of patients and thus the need to open more sites, or studies on less common diseases. For larger studies, even though more data would be available, the large number of centers may cause predictions to derail if their corresponding recruitment rate is not constant. On the other hand, for rare disease studies, trials often run on more countries trying to recruit people from several places.

An overall comparison of a PG with the nonhomogeneous PG model has shown that the latter is a better model, especially when looking at the averaging of several models including the homogeneous. The set of functions originally defined by Urbas et al. (2020) seems to cover a wide range of cases. Yet, it is still not clear how such a model will behave when the time function is completely misspecified by the model. In our data, we have not seen a clear case of a recruiting pattern that would be completely misspecified, such as a monotonically increasing recruitment rate. Although this seems to be rare, we aim to explore this in a future work using simulation studies. Code for simulating data similar to those analyzed in this paper and fitting prediction models are presented in the [Supporting Information](#) of this paper.

## CONFLICT OF INTEREST

The authors have declared no conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from AstraZeneca. Restrictions apply to the availability of these data, which were used under license for this study.

## ORCID

Aris Perperoglou  <https://orcid.org/0000-0003-1355-2800>

## REFERENCES

- Abbas, I., Rovira, J., & Casanovas, J. (2007). Clinical trial optimization: Monte Carlo simulation Markov model for planning clinical trials recruitment. *Contemporary Clinical Trials*, 28(3), 220–231.
- Anisimov, V. V., & Fedorov, V. V. (2007). Modelling, prediction and adaptive adjustment of recruitment in multicentre trials. *Statistics in Medicine*, 26(27), 4958–4975.
- Bakhshi, A., Senn, S., & Phillips, A. (2013). Some issues in predicting patient recruitment in multi-centre clinical trials. *Statistics in Medicine*, 32(30), 5458–5468.
- Barnard, K. D., Dent, L., & Cook, A. (2010). A systematic review of models to predict recruitment to multicentre clinical trials. *BMC Medical Research Methodology*, 10(1), 1–8.
- Carter, R. E. (2004). Application of stochastic processes to participant recruitment in clinical trials. *Controlled Clinical Trials*, 25(5), 429–436.
- Carter, R. E., Sonne, S. C., & Brady, K. T. (2005). Practical considerations for estimating clinical trial accrual periods: Application to a multi-center effectiveness study. *BMC Medical Research Methodology*, 5(1), 1–5.
- Comfort, S. (2013). Improving clinical trial enrollment forecasts using SORM. *Applied Clinical Trials*, 22(5), 32.
- Fierro, R., & Tapia, A. (2011). Testing homogeneity for Poisson processes. *Revista Colombiana de Estadística*, 34(3), 421–432.
- Gajewski, B. J., Simon, S. D., & Carlson, S. E. (2008). Predicting accrual in clinical trials with Bayesian posterior predictive distributions. *Statistics in Medicine*, 27(13), 2328–2340.
- Gkioni, E., Rius, R., Dodd, S., & Gamble, C. (2019). A systematic review describes models for recruitment prediction at the design stage of a clinical trial. *Journal of Clinical Epidemiology*, 115, 141–149.
- Heitjan, D. F., Ge, Z., & Ying, G.-S. (2015). Real-time prediction of clinical trial enrollment and event counts: A review. *Contemporary Clinical Trials*, 45, 26–33.
- Lan, Y., Tang, G., & Heitjan, D. F. (2019). Statistical modeling and prediction of clinical trial recruitment. *Statistics in Medicine*, 38(6), 945–955.
- Lee, Y. J. (1983). Interim recruitment goals in clinical trials. *Journal of Chronic Diseases*, 36(5), 379–389.
- Moussa, M. (1984). Planning a clinical trial with allowance for cost and patient recruitment rate. *Computer Programs in Biomedicine*, 18(3), 173–179.
- Piantadosi, S., & Patterson, B. (1987). A method for predicting accrual, cost, and paper flow in clinical trials. *Controlled Clinical Trials*, 8(3), 202–215.
- Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437), 179–191.
- Senn, S. (1998). Some controversies in planning and analysing multi-centre trials. *Statistics in Medicine*, 17(15-16), 1753–1765.
- Tang, G., Kong, Y., Chang, C.-C. H., Kong, L., & Costantino, J. P. (2012). Prediction of accrual closure date in multi-center clinical trials with discrete-time Poisson process models. *Pharmaceutical Statistics*, 11(5), 351–356.
- Urbas, S., Sherlock, C., & Metcalfe, P. (2020). Interim recruitment prediction for multi-center clinical trials. *Biostatistics*, Kxaa036. <https://doi.org/10.1093/biostatistics/kxaa036>
- Williford, W. O., Bingham, S. F., Weiss, D. G., Collins, J. F., Rains, K. T., & Krol, W. F. (1987). The ‘constant intake rate’ assumption in interim recruitment goal methodology for multicenter clinical trials. *Journal of Chronic Diseases*, 40(4), 297–307.
- Zhang, X., & Long, Q. (2010). Stochastic modeling and prediction for accrual in clinical trials. *Statistics in Medicine*, 29(6), 649–658.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Perperoglou, A., Zhang, Y., & Kipourou, K. (2023). Modeling time-varying recruitment rates in multicenter clinical trials. *Biometrical Journal*, 65, 2100377. <https://doi.org/10.1002/bimj.202100377>