

REVIEW

# A systematic review describes models for recruitment prediction at the design stage of a clinical trial

Efstathia Gkioni<sup>a,b,c,\*</sup>, Roser Rius<sup>d</sup>, Susanna Dodd<sup>a</sup>, Carrol Gamble<sup>a</sup>

<sup>a</sup>Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom

<sup>b</sup>INSERM, U1153 Epidemiology and Biostatistics Sorbonne Paris Cité Research Center (CRESS), Methods of Therapeutic Evaluation of Chronic Diseases Team (METHODS), Paris, France

<sup>c</sup>Paris Descartes University, Paris Cité, France

<sup>d</sup>Department of Statistics and Operations Research, School of Mathematics and Statistics, BarcelonaTech (UPC), Barcelona, Spain

Accepted 4 July 2019; Published online 9 July 2019

## Abstract

**Objective:** Patient recruitment in clinical trials is challenging with failure to recruit to time and target sample size common. This may be caused by unanticipated problems or by overestimation of the recruitment rate. This study is a systematic review of statistical models to predict recruitment at the design stage of clinical trials.

**Study Design and Setting:** The Online Resource for Recruitment research in Clinical triAls database was searched to identify articles published between 2008 and 2016. Articles published before 2008 were identified from a relevant systematic review. Google search was used to find potential methods in gray literature.

**Results:** Thirteen eligible articles were identified of which, 11 focused on stochastic approaches, one on deterministic models, and one included both stochastic and deterministic methods. Models varied considerably in the factors included and in their complexity. Key aspects included their ability to condition on time; whether they used average or center-specific recruitment rates; and assumptions around center initiation rates. Lack of flexibility of some models restricts their implementation.

**Conclusion:** Deterministic models require specification of few parameters but are likely unrealistic although easy to implement. Increasingly, stochastic models require greater parameter specification, which, along with greater complexity may be a barrier to their implementation. © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Clinical trials; Recruitment prediction; Statistical models; Design stage

## 1. Introduction

Successfully recruiting the prespecified number of trial participants remains a difficult challenge negatively impacting all stakeholders [1]. Twenty-six percent of trials funded by the Swiss National Science Foundation were prematurely discontinued because of slow recruitment [2]. Despite investment targeting recruitment difficulties [3,4], there has been no improvement over time and 45% of trials

supported by two prestigious UK funding bodies, fail to meet their original recruitment targets [5].

Recruitment of patients in trials is complex and participation rates that are lower than expected may be due to several factors [6]. Common reasons for failure to recruit are delays with contracting, center initiation delays, inadequate planning, insufficient staff, and overoptimistic expectations. A cohort examination of trials that were discontinued early found overestimation of the prevalence of eligible participants was among the most frequently reported reasons [7]. A key problem is therefore within recruitment prediction at the design stage.

Many interventions targeting recruitment and retention difficulties have been suggested and implemented, despite the absence of evidence to support their use [8]. An identification and prioritization study was undertaken around recruitment in clinical trials with the aim to identify and prioritize unanswered questions around recruitment. The

Conflict of interest: The authors wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

\* Corresponding author. Department of Biostatistics, Institute of Translational Medicine, University of Liverpool, Block F/ Waterhouse Building, 1-5 Brownlow Street, Liverpool L69 3GL, United Kingdom. Tel.: +44 151 794 9743.

E-mail address: [e.gkioni@liverpool.ac.uk](mailto:e.gkioni@liverpool.ac.uk) (E. Gkioni).

<https://doi.org/10.1016/j.jclinepi.2019.07.002>

0895-4356/© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### What is new?

#### Key findings

- Decisions to fund clinical trials are a balance of factors including the importance of the clinical question and the time and costs required. Recruitment should be viewed as a stochastic rather than a deterministic process because the costs are highly driven by the length of time to recruit to the target sample size.
- Increasingly complex statistical models are being proposed as a potential solution, but they require parameter estimates for the distributions used, which may be difficult for researchers to provide.

#### What this adds to what was known?

- Thirteen articles were identified of which four were included in a previous related systematic review. The previous related review did not differentiate models that could be used during the planning stage and this article adds to that systematic review by isolating these models and updating the review.
- Models were assessed and categorized according to their nature and ability to incorporate information for recruitment prediction, including time-dependent factors such as staggered center initiations and seasonal variations, and ability to specify rates per center or average rates across centers.

#### What is the implication and what should change now?

- This systematic review and comparison of available methods will help researchers to identify models meeting the requirements of their study and to determine the information required for their implementation. Modeling recruitment as a stochastic process at the design stage allows for the uncertainty to be included in the prediction figures.

question “What are the best ways to predict recruitment rates to a randomized trial and what impact do such predictions have on recruitment?” was identified as a top ten priority question to be investigated and answered [9].

The purpose of our study is to systematically review methods to predict recruitment in the design stage of clinical trials.

## 2. Methods

### 2.1. Identification of potentially eligible studies

The search strategy was split by articles published before or after July 2008. The systematic review of Barnard

et al. [10] was used to identify eligible articles published up to July 2008 (see Fig. 1). This review includes models for recruitment prediction at the design stage or during ongoing recruitment. Articles published after July 2008 to 2016 were identified by searching the “Recruitment prediction” domain in the Online Resource for Recruitment research in Clinical trials (ORRCA) [11], a searchable database of research relevant to clinical trial recruitment. The development of ORRCA is described elsewhere [12]. At the time of the ORRCA search (June 2018), it included publications up to 2016.

References of eligible articles were also reviewed for additional articles including book chapters.

A Google search was conducted (June 2018) to identify gray literature (see [Supplementary material](#)).

### 2.2. Eligibility criteria/study selection

Research articles describing statistical methods for recruitment prediction at the design stage of a clinical trial were eligible for inclusion.

Exclusion criteria were as follows:

- 1) Articles discussing recruitment that did not include statistical methods for recruitment prediction;
- 2) Review articles that did not propose a new model, although the reference list of these articles was searched for potentially relevant articles;
- 3) Articles describing statistical models that could not be implemented before the start of participant enrollment.

The titles and abstracts of all retrieved articles were reviewed by E.G. The full texts of these articles were obtained and the final inclusion was determined after discussion by E.G., S.D., and C.G. Disagreements between reviewers were resolved by consensus.

### 2.3. Data extraction

Data extraction was determined after discussion by all the authors. The information extracted from each article included the statistical model, model parameters required for its implementation, consideration of seasonal effects, factors relating to multicentre trials such as recruitment rates variation, centers initiation rates, and whether the models had been implemented and evaluated using real data and/or simulated data. Google Form was used for data extraction, with data extracted by E.G. and then checked and discussed with C.G.

This is a systematic review of statistical models. Methods to assess risk of bias are not applicable. We did not publish a protocol for this review. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist [13] is provided as [supplementary material](#).

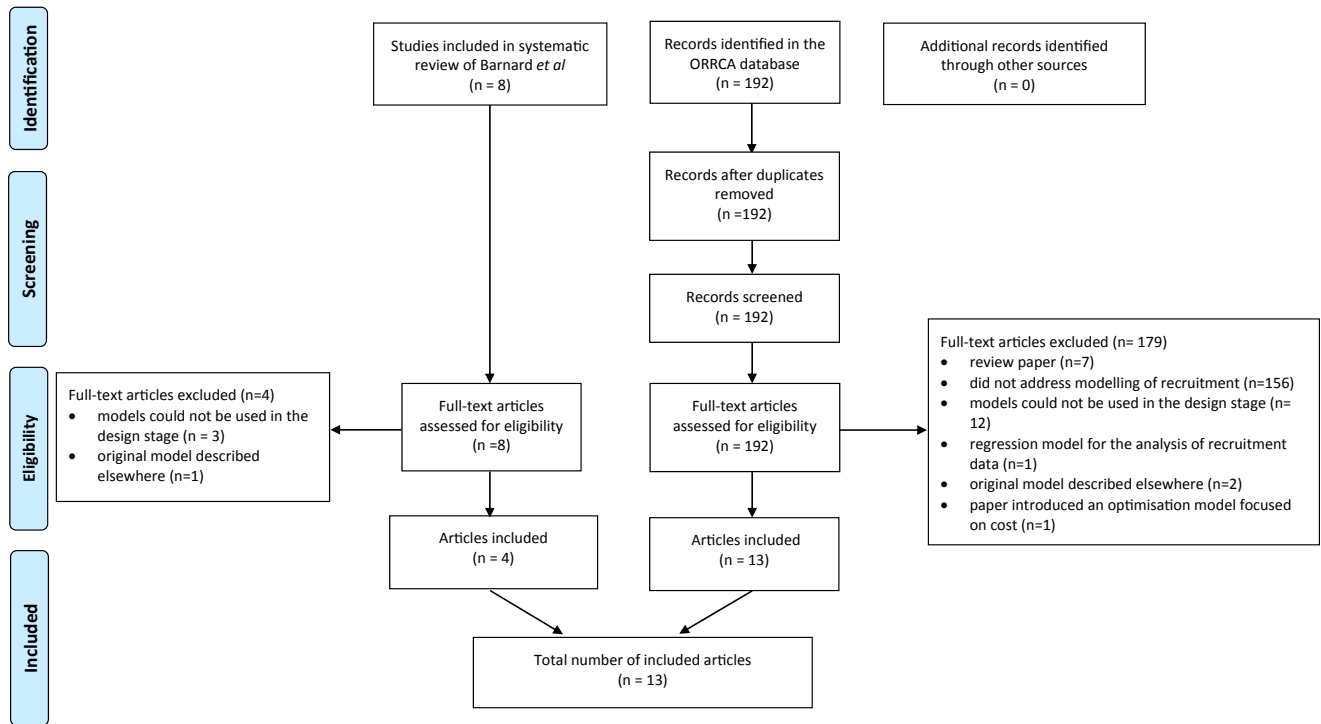


Fig. 1. PRISMA flow diagram.

### 3. Results

#### 3.1. Search results

The search results are presented in Fig. 1. A total of thirteen articles were eligible for inclusion. One article focused only on deterministic models [14], another article described both deterministic and stochastic approaches [15] and eleven articles described only stochastic methods.

#### 3.2. Classification of models

Key features of the models were identified and used to group them accordingly. Models are either deterministic in that they assume a fixed recruitment rate or stochastic in that there is random variation around an average recruitment rate. In an unconditional model, the specified recruitment rate stays the same, whereas in a conditional model, the rate varies according to time.

Recruitment rates may be homogenous within a period or nonhomogeneous incorporating variation in the expected recruitment rate.

The model classifications and data extracted are summarized in Table 1 and 2.

##### 3.2.1. Deterministic models

**3.2.1.1. Unconditional model.** The unconditional model assumes a constant linear recruitment rate. The trial's recruitment period is simply calculated as the sample size target

divided by the total number of recruited patients expected each month. The method can incorporate variability in center recruitment rates by averaging the recruitment rate across all centers, providing the overall rate is then constant over time. Carter et al. [15] and Comfort [14] described the limitations of this approach with Comfort providing a simple equation to support implementation for multicentre trials.

**3.2.1.2. Conditional model.** The conditional model allows for time-dependent changes in the overall recruitment rate [15]. It splits the overall recruitment period into successive intervals, with each interval having its own recruitment rate. The conditional model can incorporate recruitment rate variation between centers and allow for reduced capacity in recruitment. There is no closed form for this flexible but deterministic approach which is easiest to implement within a spreadsheet package [15].

As an alternative to requiring a spreadsheet implementation but at a cost of reduced flexibility, Comfort described a conditional deterministic model termed a second-order recruitment model (SORM) [14]. In this piecewise model, the recruitment rate is initially conditional on the cumulative number of centers open to recruitment with centers starting with an average initiation rate until all centers are open to recruitment. At this point, enrollment becomes linear and unconditional. Comfort provided closed form equations (mathematical processes that can be completed in a finite number of operations) to implement the model

**Table 1.** Model categorization and description of parameters required for their implementation

Model class	Description	Parameters required to implement the model at the design stage	Model output
Deterministic unconditional models Comfort [14], Carter [15]	The model assumes a constant rate of recruitment throughout the study.	<ul style="list-style-type: none"> <li>• Sample size target;</li> <li>• expected number of patients to be enrolled per month.</li> </ul>	<p>Comfort</p> <ul style="list-style-type: none"> <li>• time required to recruit to target sample size;</li> <li>• number of participants that can be achieved within a set recruitment period;</li> </ul> <p>Carter</p> <ul style="list-style-type: none"> <li>• time required to recruit to target sample size.</li> </ul>
Deterministic Conditional models Second-order recruitment model (SORM) Comfort [14], Carter [15]	<p>Piecewise model: equations are provided for recruitment, conditional on centers opening rates followed by a constant unconditional rate of recruitment once all centers are open.</p> <p>The model allows for time dependent variation in recruitment rates because of different center openings and seasonal effects. This approach does not use statistical distributions or equations and is best implemented within a spreadsheet.</p>	<ul style="list-style-type: none"> <li>• Number of centers</li> <li>• average center initiation rate</li> <li>• total recruitment time</li> <li>• recruitment rate (patient/center/month)</li> <li>• sample size target</li> <li>• sample size target</li> <li>• number of centers</li> <li>• expected number of patients per center</li> <li>• the length of time in which the center is open to recruitment</li> <li>• seasonality and other time-dependent factors impacting recruitment</li> </ul>	<ul style="list-style-type: none"> <li>• Time required to recruit to target sample size.</li> <li>• Time required to recruit the target sample size, adjusting for time-dependent changes in the overall accrual rate.</li> </ul>
Stochastic Poisson models Senn [16], Carter [17], Carter [15], Lee [18]	<p>Unconditional Poisson: the average rate at which patients arrive is constant over time [15–18];</p> <p>or</p> <p>Conditional Poisson: the recruitment rate of the Poisson process is uniformly distributed on [0,y] where y is the expected number of patients per month/per center [15,17].</p>	<ul style="list-style-type: none"> <li>• sample size target;</li> <li>• the rate at which events occur in the Poisson process;</li> <li>• start and end time;</li> </ul> <p>or</p> <ul style="list-style-type: none"> <li>• all the above and use a uniformly distributed number (0,1) to multiply the expected accrual rate.</li> </ul>	<p>Unconditional Poisson</p> <ul style="list-style-type: none"> <li>• probability of achieving the target within time specified (Senn, Carter [15,17]);</li> <li>• number of patients to be recruited by each interim time point, given the expected rate of recruitment (Lee);</li> </ul> <p>Conditional Poisson</p> <ul style="list-style-type: none"> <li>• time needed to obtain the desired sample size with a high probability allowing for a time-dependent variation in the rate.</li> </ul>
Stochastic Poisson-Gamma models Anisimov [19], Anisimov [20], Anisimov [21], Anisimov [22]	Patients arrive at clinical centers independently, based on the Poisson process and the rates among centers vary as samples from a gamma distribution.	<ul style="list-style-type: none"> <li>• sample size target;</li> <li>• total recruitment time;</li> <li>• number of centers;</li> <li>• the time when each center is open to recruitment; or with centers initiation delays described as random variables in specified intervals.</li> </ul>	<ul style="list-style-type: none"> <li>• Time required to recruit the predefined sample size accounting for the planned dates of centers initiation and the different recruitment rate per center;</li> <li>• the minimal number of centers needed to complete recruitment by a certain date with a given confidence at any stage of the study.</li> </ul>
Bayesian models Gajewski [23], Zhang [24], Bakhshi [25]	<ul style="list-style-type: none"> <li>• NB. Focus restricted to prior specifications.</li> </ul>	<p>Gajewski</p> <ul style="list-style-type: none"> <li>• sample size target;</li> <li>• recruitment duration;</li> </ul>	<p>Gajewski</p> <ul style="list-style-type: none"> <li>• patient accrual across a fixed time period;</li> </ul>

(Continued)

Table 1. Continued

Model class	Description	Parameters required to implement the model at the design stage	Model output
	<ul style="list-style-type: none"> <li>Gajewski models average waiting time between successive participants while Zhang models the average recruitment rate using the normal distribution to specify the distribution for the average. Both models assume homogeneity.</li> <li>Bakhshi requires the existence of a set of previous trials utilizing a meta-analysis approach, which incorporates variability between trials. The approach is described as not being fully Bayesian.</li> </ul>	<ul style="list-style-type: none"> <li>Investigators confidence of finishing the trial on time as a single value between 0 and 1.</li> </ul> <p>Zhang</p> <ul style="list-style-type: none"> <li>sample size target;</li> <li>the average recruitment rate;</li> <li>the variance for the average recruitment rate considered based on the certainty of investigators around the average.</li> </ul> <p>Bakhshi</p> <ul style="list-style-type: none"> <li>requires a set of previous trials with data available on the number of patients recruited per center per month.</li> </ul>	<ul style="list-style-type: none"> <li>accrual time to a target sample size.</li> </ul> <p>Zhang</p> <ul style="list-style-type: none"> <li>time to completion of patient recruitment under the assumption of the constant accrual.</li> </ul> <p>Bakhshi</p> <ul style="list-style-type: none"> <li>the time it will take to recruit the total number of subjects required.</li> </ul>
Simulation models Abbas [26]	Investigate several recruitment patterns in order to quantify the patient arrivals and calculate the length of time required to recruit the prespecified number of participants. Markov models have been used to explore the maximization of patient recruitment in a specific amount of time.	<ul style="list-style-type: none"> <li>Number of patients;</li> <li>recruitment duration;</li> <li>intervals of time T;</li> <li>duration of each interval;</li> <li>probability that determines the number of patients that arrive within the proposed duration (fixed or random probability).</li> </ul>	<ul style="list-style-type: none"> <li>Length of time required to recruit the number of patients necessary for the trial.</li> </ul>

and the model takes its name from the “quadratic” term in the equation for the conditional piece of the model. This method assumes that recruitment rate is constant once all centers are open and that each center recruits at the same average rate.

### 3.2.2. Stochastic models

The use of a stochastic process allows for fluctuations observed around an average recruitment rate and importantly the associated uncertainty and variation to be expressed.

#### 3.2.2.1. Stochastic Poisson

**3.2.2.1.1. Unconditional Poisson model.** The unconditional Poisson model is the simplest of the stochastic approaches and only requires specification of the average recruitment rate across centers in the trial, as described by Senn [16]. Carter et al. [15] described how the probability density function can be obtained for the time needed to recruit the target sample size by simulating data from the Poisson distribution.

Both Carter [17] and Lee [18] used a normal approximation to the Poisson distribution.

Carter [17] showed how, if the average recruitment rate is constant over the recruitment period, the normal approximation to the Poisson distribution can be used to formulate the recruitment period required to achieve the sample size target with a specified level of confidence. Alternatively, the formula can be rearranged to determine the required recruitment rate when the recruitment period is fixed.

Lee [18] also used a normal approximation and proposed setting interim target recruitment points that need to be attained to achieve the sample size target with a specified probability. This probability is used to determine the minimum required recruitment goal at interim points. This again uses the Poisson distribution with a constant average recruitment rate and a normal approximation with closed-form formulas provided. See Moussa [27] for code to implement Lee’s method and an extension to include cost implications.

**3.2.2.1.2. Conditional Poisson model.** Given the often unrealistic assumption of a homogenous recruitment rate, by similarly conditioning on time as per the deterministic conditional model, Carter [17] allowed for a time-varying

**Table 2.** Model classification and factors defining their flexibility and assumptions

Author	Site recruitment rates				Allows adjustments to recruitment rates, for example, seasonal variation, Yes (Y)/ No (N)	Implementation			
	Constant recruitment rate required, Yes (Y)/ No (N)	Site initiation rate: all open at TO (TO), fixed average initiation rates (D) stochastic initiation rates (S)	Variation in site recruitment rates, Yes (Y)/ No (N)	No of sites $\geq 10$ , Yes (Y)/ No (N)		Average recruitment rate $> 10$ required, Yes (Y)/ No (N)	Formulas, Yes (Y)/ No (N)	Programming Code provided, Yes (Y)/ No (N)	Model validation using real data (R)/ simulated data (S)
Carter [15] Comfort [14]	Y	TO	Y <sup>b</sup>	N	N	N	Y	N	Carter (R) Comfort (S)
Carter [15]	N	D	Y	N	Y	N	N	N	R
Carter [15] Comfort [14]	Y <sup>a</sup>	D	N	N	N	N	Y	N	Carter (R) Comfort (S)
Senn [16] Carter [15]	Y	TO	Y <sup>b</sup>	N	N	N	N	N	Senn (theoretical example) Carter (R & S)
Carter [17]	Y	TO	Y <sup>b</sup>	N	N	Y	Y	N	
Lee [18]	Y	TO	Y <sup>b</sup>	N	N	Y	Y	Y <sup>c</sup>	Theoretical example
Carter [17]	N	D	Y <sup>b</sup>	N	Y	N	N	Y	
Anisimov <sup>d</sup> [19–22]	N, rates assumed to follow a gamma distribution	S	Y	Y	N	N	Y	N	R
Gajewski [23]	N, waiting time assumed to follow an inverse gamma distribution	D	N	N	N	N	Y	Y	R
Zhang [24]	Y	D	N	N	N	N	Y	N	R & S
Bakhshi [25]	N, rates assumed to follow a gamma distribution	S	Y	N	N	N	Y	N	Real data from 18 trials used to define the prior information
Abbas [26]	Recruitment probability constant or randomly distributed	TO	Y <sup>b</sup>	N	N	Y	Y	N	S

<sup>a</sup> Average recruitment rate is assumed to increase with site initiations.

<sup>b</sup> Uses the recruitment rate across sites. Recruitment rates may vary by site so long as overall rate across sites is constant.

<sup>c</sup> Code available in Moussa [27].

<sup>d</sup> The model is Bayesian once accrual data are available. Here the focus is on the prior specification only.

accrual rate. The average recruitment rate needs to be specified for each time interval of the recruitment period and the number of observations then simulated from the Poisson distribution according to the number of units of time contained in that interval. Carter et al. [15] later adapted this approach further to allow the average recruitment rate itself from each interval to be simulated from a uniform distribution.

**3.2.2.2. Poisson-gamma model (P-G).** The Poisson-gamma model for recruitment prediction was introduced by Anisimov and Fedorov [28,29] and further described

by Anisimov [19–21]. The model uses the Poisson process to describe variation in recruitment over time, and models variation in the recruitment rates between centers using the gamma distribution. The model allows for variation in center opening with each center initiation time being uniformly distributed within a given interval [20]. This model was applied to real trial recruitment data and found to fit well with sufficient number of centers, ( $> 10$  [20] or  $> 20$  [19,29]), but advised to estimate the rates individually with fewer centers.

When applied to recruitment prediction of an ongoing study, the initial recruitment model is updated with the



observed recruitment data, becoming a Bayesian approach. Here we are focused on prediction at the design stage and therefore stop at the specification of the prior. Implementation of the model is nontrivial and code to support doing so is not available. In his later articles, Anisimov also considered the impact of randomization schemes and prediction of drug supply [20,22].

**3.2.2.3. Bayesian methodology.** The benefit of a Bayesian approach is in monitoring and prediction of ongoing recruitment. In a Bayesian method, investigators' prior beliefs of recruitment rates are combined with observed recruitment rates to predict participant accrual for the remaining recruitment period. However, the prior specification can be viewed as the predicted or expected recruitment at the design stage and therefore we restrict our consideration to the formulations of the priors. As stated by Gajewski et al. [23], thoughtful elicitation of a prior distribution for accrual rates will force issues for future expectations about accrual patterns to be faced. In this publication, the authors provided a model to estimate the average waiting time between participants, defined as the difference in dates when patients join the study. The waiting times are assumed to be exponentially distributed with mean  $\theta$ . The prior distribution for  $\theta$  is specified by the inverse gamma  $\sim \text{IG}(nP, TP)$ , where  $T$  is how long the investigator believes it will take to accrue the sample size,  $n$  is the sample size target, and  $P$ , a value between 0 and 1, is how confident the investigator is that  $n$  will be achieved by time  $T$ . If the investigator is fully confident that  $n$  can be achieved by time  $T$ , such that  $P = 1$ , then the prior sample size is the target sample size, and is otherwise reduced.  $TP$  scales the response such that the expected waiting time is approximately  $T/n$ . This model assumes a constant accrual rate averaged across all centers. R code to implement the model is available on request from the authors of the article [23].

Zhang et al. [24] also used a Bayesian framework with a nonhomogenous Poisson process that allows recruitment rates to vary over time. Although they described how their approach could be amended to predict accrual in the planning stage of a clinical trial utilizing prior information only, they do so by assuming a constant accrual rate. The prior specification requires the anticipated average accrual rate after it stabilizes to be specified, along with the level of confidence in that prior belief. Therefore, the key difference between Gajewski et al. and Zhang et al. is in the questions asked to elicit the priors. Jiang et al. [30] provided a user-friendly interface program developed in R, based on the method of Gajewski et al. [23] for the total number of patients that can be recruited within a fixed time in a clinical trial, with an updated version of the R package available [31].

Bakhshi et al. [25] used data from a meta-analysis of previous trials to estimate the parameters in the Poisson-gamma model. This method is described as not being fully

Bayesian as there are no hyperdistributions set on the parameters and the modeling of parameters has been conducted independently. They provided guidance on how to produce initial estimates at the design stage, with one parameter accounting for the trial-specific and another one reflecting the different center-specific recruitment rate. The results of the meta-analysis could be used under the empirical Bayesian framework once accrual data from the trial are available.

**3.2.2.4. Simulation models.** Abbas et al. [26] used the Monte Carlo simulation Markov models to design different recruitment patterns using time as a discrete or continuous variable. For each simulation, the time to achieve target sample size is recorded with mean and standard deviation estimated across simulation. Models allow for discrete time with and without replacements and continuous time with a constant probability or a probability distribution applied to the conversion between states. In this method, the states can be viewed as patient identification or approach to patient enrollment.

## 4. Discussion

The limited resources for funding and conducting clinical trials means that the decisions on which research to fund must balance the importance of the clinical question against the time and cost required to answer it. Factors affecting recruitment may be complex and many; thus, the accuracy of recruitment prediction at the design stage is crucial.

This article systematically reviews models implemented at the design stage of a trial to predict patient recruitment. The models' spectrum extends from simple unconditional and conditional deterministic approaches [14,15] to stochastic models that allow for variation around an average recruitment rate [17,19,22,24] and Bayesian approaches where the expectations of the investigators are translated into prior information [23–25].

Whether the recruitment rates being specified represent an average number of participants or whether they are believed to be a guaranteed minimum number is an important consideration. If it is a guaranteed minimum specified, then a deterministic model is appropriate, but to achieve this, the average recruitment rate would have to be markedly higher. If it is an average rate, then the recruitment prediction should use a stochastic approach to allow for the variation and express the uncertainty accordingly. There may be divergence between funder and researcher perspectives and the interpretation of the figures presented and this potential should be explored further.

It is understandable that researchers may be reluctant to acknowledge the impact of considering their recruitment rates as an average. At best, an average recruitment rate will reach its target with a probability of 0.5 and this alone may explain existing observations around the

percentage of trials recruiting to target as prespecified at the design stage [5]. In addition, the use of these models will not prevent overoptimistic specification of parameters but may increase awareness of factors that should be considered and how they may be incorporated. Before greater accuracy in recruitment prediction is observed, funders may also need to demonstrate willingness to fund longer recruitment periods than they may be currently perceived to do.

Although some of the authors provided the code of the method they introduced [15,23,27] (see [Supplementary material](#)), it is unclear whether this is sufficient for potential users to be able to implement them. The complexity of some models may be a barrier to their implementation. However, the simplest stochastic models use the Poisson process, which is familiar to many statisticians and yet anecdotal experience would suggest they are not widely implemented. This may be because statisticians are not inherently involved in recruitment prediction.

The approaches used to predict recruitment, including sources of information utilized, remains frequently hidden and unreported. Information may be available from other relevant clinical trials, from national disease registries and databases, from audit data from select centers that may be extrapolated across centers or a “best guess” approach based on clinician experience albeit without supporting data being readily available. Reliability of each needs to be assessed; however, not all factors that can impact recruitment may be known at the design stage.

In multicentre clinical trials, variation in center recruitment targets needs consideration. Combinations of factors, such as center size, center initiation dates, and eligible patient population, need to be considered alongside factors that may be less amenable to prediction from historic data, such as the size of existing center research portfolios and resources to support new research, the extent of support for the clinical question itself, and patient willingness to participate. Approaches such as the Poisson-gamma recruitment model attempt to incorporate this level of complexity. However, this requires a distribution to be applied to center recruitment rates and may be conceptually challenging for many in terms of specification of the parameters and how to obtain supporting data.

Although it is clear that modeling recruitment represents a statistical challenge, until these models are implemented within the mainstream, the solutions proposed will be unattainable for many and any resulting improvements will not be observed. This requires consideration before developing further academic models that may never be implemented either because of their complexity or because of the wide variation depicted for the required recruitment period.

This review is the first to our knowledge that focuses on methods to predict recruitment at the design stage of a trial. The search strategy for the review utilized two existing

resources: the review of Barnard et al. [10] and the ORRCA database [11], meaning that there are differences in the search strategies implemented. However, no additional methods were identified via Google search or screening reference lists of eligible articles. The use of ORRCA meant that article inclusion was restricted to those published in 2016 at the latest. We will continue to monitor future ORRCA updates to determine when this review should be updated.

## 5. Conclusion

Predicting recruitment is an important part of the design and planning of a clinical trial.

Modeling recruitment as a stochastic process at the design stage may lead to improvements in the prediction and in understanding deviations from the “expected average”. However, benefits may be limited if the approach taken leads to excessive variation.

This systematic review and comparison of available methods will help researchers to identify models meeting the requirements of their study and in determining the information required for their implementation. Until researchers implement these methods, they are limited in their potential to provide improved predictions. To achieve this, we believe that statisticians should routinely be involved in recruitment prediction.

## CRediT authorship contribution statement

**Efstathia Gkioni:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing - original draft, Writing - review & editing. **Roser Rius:** Formal analysis, Methodology, Writing - review & editing, Supervision. **Susanna Dodd:** Formal analysis, Methodology, Writing - review & editing, Supervision. **Carrol Gamble:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing - review & editing, Supervision.

## Acknowledgments

The authors would like to thank Ms Van Nguyen Thu, PhD candidate, on the MiRoR project, and Professor Isabelle Boutron, Paris Descartes University, for their help in reviewing the manuscript before submission.

This project is part of the MiRoR (Methods in Research on Research) Joint Doctoral Training Programme. The authors would like to acknowledge the MiRoR network for their support and guidance. The European Union’s Horizon 2020 Marie Skłodowska Curie Innovative Training Networks—European Joint Doctorate (ITN-EJD) under grant agreement No. 676207, fund MiRoR.



## Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2019.07.002>.

## References

- [1] Walters SJ, Bonacho dos Anjos Henriques-Cadby I, Bortolami O, Flight L, Hind D, Jacques RM, et al. Recruitment and retention of participants in randomised controlled trials: a review of trials funded and published by the United Kingdom Health Technology Assessment Programme. *BMJ Open* 2017;7:e015276.
- [2] Amstutz A, Schandelmaier S, Frei R, Surina J, Agarwal A, Olu KK, et al. Discontinuation and non-publication of randomised clinical trials supported by the main public funding body in Switzerland: a retrospective cohort study. *BMJ Open* 2017;7:e016216.
- [3] Treweek S, Lockhart P, Pitkethly M, Cook JA, Kjeldstrøm M, Johansen M, et al. Methods to improve recruitment to randomised controlled trials: cochrane systematic review and meta-analysis. *BMJ Open* 2013;3:e002360.
- [4] Ross S, Grant A, Counsell C, Gillespie W, Russell I, Prescott R. Barriers to participation in randomised controlled trials: a systematic review. *J Clin Epidemiol* 1999;52:1143–56.
- [5] Sully BGO, Julious SA, Nicholl J. A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies. *Trials* 2013;14:166.
- [6] McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. *Trials* 2006;7:9.
- [7] Briel M, Olu KK, von Elm E, Kasenda B, Alturki R, Agarwal A, et al. A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. *J Clin Epidemiol* 2016;80:8–15.
- [8] Bower P, Brueton V, Gamble C, Treweek S, Smith CT, Young B, et al. Interventions to improve recruitment and retention in clinical trials: a survey and workshop to assess current practice and future priorities. *Trials* 2014;15:399.
- [9] Healy P, Galvin S, Williamson PR, Treweek S, Whiting C, Maeso B, et al. Identifying trial recruitment uncertainties using a James Lind Alliance priority setting partnership – the PRioRiT (prioritising recruitment in randomised trials) study. *Trials* 2018;19:147.
- [10] Barnard KD, Dent L, Cook A. A systematic review of models to predict recruitment to multicentre clinical trials. *BMC Med Res Methodol* 2010;10:63.
- [11] ORRCA: online resource for recruitment research in clinical trials. 2018. Available at <http://www.orrca.org.uk>. Accessed June 11, 2018.
- [12] Kearney A, Harman NL, Rosala-Hallas A, Beecher C, Blazeby JM, Bower P, et al. Development of an online resource for recruitment research in clinical trials to organise and map current literature. *Clin Trials* 2018;15:533–42.
- [13] Moher D, Liberati A, Tetzlaff J, Altman DG, The PG. Preferred reporting Items for systematic reviews and meta-Analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
- [14] Comfort S. Improving clinical trial enrollment forecasts using SORM. *Appl Clin Trials* 2013;22:32–8.
- [15] Carter RE, Sonne SC, Brady KT. Practical considerations for estimating clinical trial accrual periods: application to a multi-center effectiveness study. *BMC Med Res Methodol* 2005;5:11.
- [16] Senn S. Some controversies in planning and analysing multi-centre trials. *Stat Med* 1998;17:1753–65.
- [17] Carter RE. Application of stochastic processes to participant recruitment in clinical trials. *Control Clin Trials* 2004;25:429–36.
- [18] Lee YJ. Interim recruitment goals in clinical trials. *J Clin Epidemiol* 1983;36:379–89.
- [19] Anisimov VV. Using mixed Poisson models in patient recruitment in multicentre clinical trials. In: *Proceedings of the World Congress on Engineering*; 1046–9.
- [20] Anisimov V. Predictive modelling of recruitment and drug supply in multicenter clinical trials. In: *Proc Joint Stat Meeting*; 1248–59.
- [21] Anisimov VV. Recruitment modeling and predicting in clinical trials. *Pharm Outsourcing* 2009;10:44–8.
- [22] Anisimov VV. Statistical modeling of clinical trials (recruitment and randomization). *Commun Stat Theory Methods* 2011;40:3684–99.
- [23] Gajewski BJ, Simon SD, Carlson SE. Predicting accrual in clinical trials with Bayesian posterior predictive distributions. *Stat Med* 2008;27:2328–40.
- [24] Zhang X, Long Q. Stochastic modeling and prediction for accrual in clinical trials. *Stat Med* 2010;29:649–58.
- [25] Bakhshi A, Senn S, Phillips A. Some issues in predicting patient recruitment in multi-centre clinical trials. *Stat Med* 2013;32:5458–68.
- [26] Abbas I, Rovira J, Casanovas J. Clinical trial optimization: Monte Carlo simulation Markov model for planning clinical trials recruitment. *Contemp Clin Trials* 2007;28:220–31.
- [27] Moussa MAA. Planning a clinical trial with allowance for cost and patient recruitment rate. *Computer Programs Biomed* 1984;18:173–9.
- [28] Anisimov VV, Fedorov VV. Design of multicentre clinical trials with random enrolment. In: *Advances in statistical methods for the health sciences*. Boston, MA: Birkhäuser Boston, Springer; 2007:387–400.
- [29] Anisimov VV, Fedorov VV. Modelling, prediction and adaptive adjustment of recruitment in multicentre trials. *Stat Med* 2007;26:4958–75.
- [30] Jiang Y, Guarino P, Ma S, Simon S, Mayo MS, Raghavan R, et al. Bayesian accrual prediction for interim review of clinical studies: open source R package and smartphone application. *Trials* 2016;17:336.
- [31] Liu JJY, Wu C, Simon S, Mayo MS, Raghavan R, Gajewski BJ. Package "accrual": bayesian accrual prediction. 2017. Accessed 15 December 2017. Available at <https://cran.r-project.org/web/packages/accrual/index.html>. Accessed December 15, 2017.