

**Frequentists and Bayesian methods to incorporate  
recruitment rate stochasticity  
at the design stage of a clinical trial**

---

Master Thesis in Biostatistics (STA495)

by

Pilar Pastor  
23-733-975

supervised by

PD Dr. Malgorzata Roos

Zurich, month year



# Frequentists and Bayesian methods to incorporate recruitment rate stochasticity at the design stage of a clinical trial

Pilar Pastor

Version February 18, 2025



# Contents

<b>Preface</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Methodology</b>	<b>3</b>
2.1 Definitions . . . . .	3
2.2 Uncertainty and models for counts . . . . .	4
2.3 Derivation of Negative Binomial from Poisson-Gamma model . . . . .	5
2.4 Citations . . . . .	6
<b>3 Results</b>	<b>9</b>
<b>4 Discussion and Outlook</b>	<b>11</b>
<b>5 Conclusions</b>	<b>13</b>
<b>Bibliography</b>	<b>15</b>



# Preface

Howdy!

Pilar Pastor  
June 2025





# Chapter 1

## Introduction

Why, what and how...



## Chapter 2

# Methodology

### 2.1 Definitions

The **Target Population** is a specific group within the broader population, defined by attributes relevant to the research question. This group is selected based on criteria that match the study's goals, helping researchers focus on the most pertinent segments of the population (Willie, 2024). Defining the target population allows researchers to refine their objectives and sampling methods to align with the study's aims.

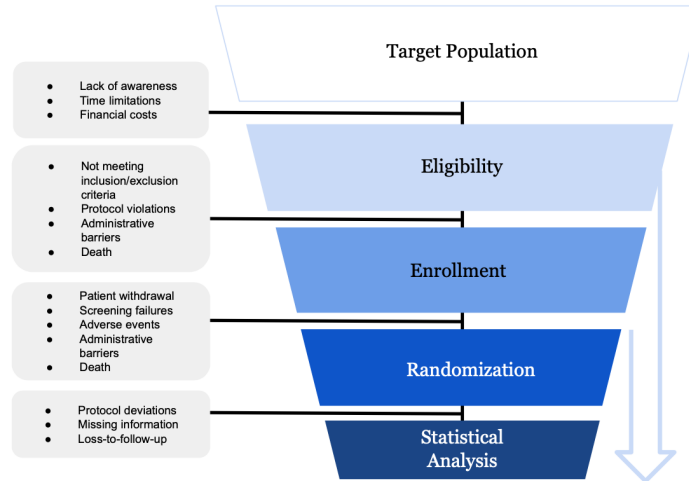
The **Eligibility** criteria are the specific requirements that individuals must meet to participate in a study. Eligible patients will be selected from the target population. It is important to note that eligibility criteria also include exclusion factors, conditions or circumstances that disqualify potential participants (Food *et al.*, 2018). Inclusion criteria specify the conditions that allow individuals to participate in the trial, particularly focusing on the medical condition of interest. Any other factors that limit eligibility are classified as exclusion criteria (Van Spall *et al.*, 2007).

In clinical trials, **Enrollment** refers to the formal process of registering participants into a study after they have met all eligibility criteria and provided informed consent. This process includes verifying that each participant satisfies the inclusion and exclusion criteria outlined in the study protocol (National Institute of Allergy and Infectious Diseases, 2021). It is important to distinguish between recruitment and enrollment. Recruitment involves identifying and inviting potential participants to join the study, whereas enrollment occurs after these individuals have been screened, consented, and officially registered into the trial (Frank, 2004).

Once enrolled, participants are assigned to specific treatment groups or interventions as defined by the study design. The most common practice been **Randomization**. In clinical research, randomization is the process of assigning participants to different treatment groups using chance methods, such as random number generators or coin flips (Lim and In, 2019). Randomized controlled trials (RCTs) are considered the most effective method for preventing bias in the evaluation of new interventions, drugs, or devices. (Van Spall *et al.*, 2007).

In clinical research, **Statistical Analysis** involves applying statistical methods to collect, summarize, interpret, and present data derived from clinical studies. This process is essential for evaluating the safety, efficacy, and overall outcomes of medical interventions, ensuring that conclusions drawn are both reliable and valid (Panos and Boeckler, 2023). Not all participants who are randomized may be included in the final statistical analysis due to protocol deviations of patients not adhering to the protocol (Rehman *et al.*, 2020), missing data (Shih, 2002) or loss-to-follow-up, some participants may become unreachable or withdraw consent during the study, resulting in missing outcome data (Nüesch *et al.*, 2009).

The number of patients decreases at each stage of a clinical study, from defining the target population to final statistical analysis, see Figure 2.1. Eligibility criteria narrow down participants, and enrollment further reduces numbers as only those meeting strict criteria are registered.



**Figure 2.1:** Patient Attrition at Each Stage of a Clinical Study. References: [Piantadosi and Meinert \(2022\)](#); [Whelan \*et al.\* \(2018\)](#); [Bogin \(2022\)](#)

Randomization assigns individuals to treatment groups, but some may later be excluded due to protocol deviations, missing data, or loss to follow-up.

## 2.2 Uncertainty and models for counts

Let us denote

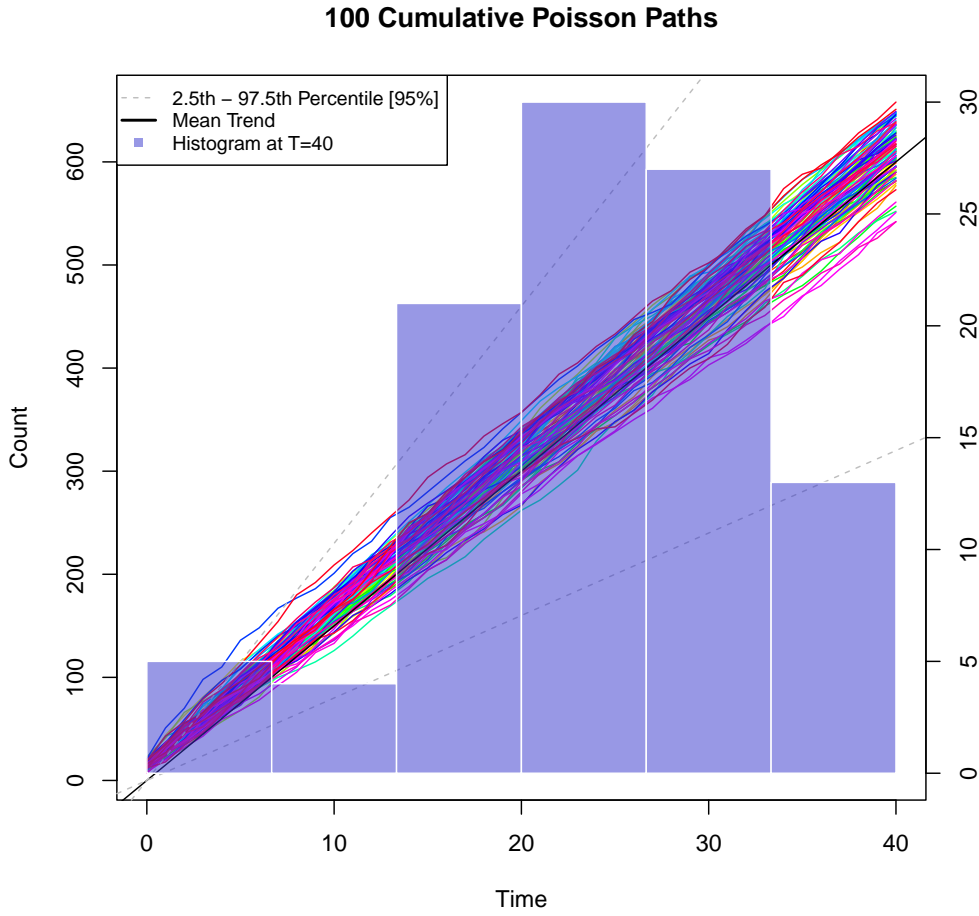
- $T = \text{time}$
- $C = \text{counts}$
- $\lambda = \frac{C}{T}$

	Counts	Expectation	Standard Deviation	Aleatory	Epistemic
Expectation	$C = \lambda \cdot T$	$\lambda \cdot T$	0	No	No
Poisson	$C \sim \text{Poi}(\lambda \cdot T)$	$\lambda \cdot T$	$\lambda \cdot T$	Yes	No
Negative Binomial	$C \sim \text{Poi}(\Lambda \cdot T); \Lambda \sim G(\alpha, \beta)$	?	?	Yes	Yes

**Table 2.1:** Modeling Count Data with Specified Levels of Aleatory and Epistemic Uncertainty.

This options are best placed in the main document at the beginning. Otherwise a `cache=FALSE` as knitr option is necessary to overrule a possible `cache=TRUE` flag.

Notice how in Figure 2.4 everything is properly scaled.

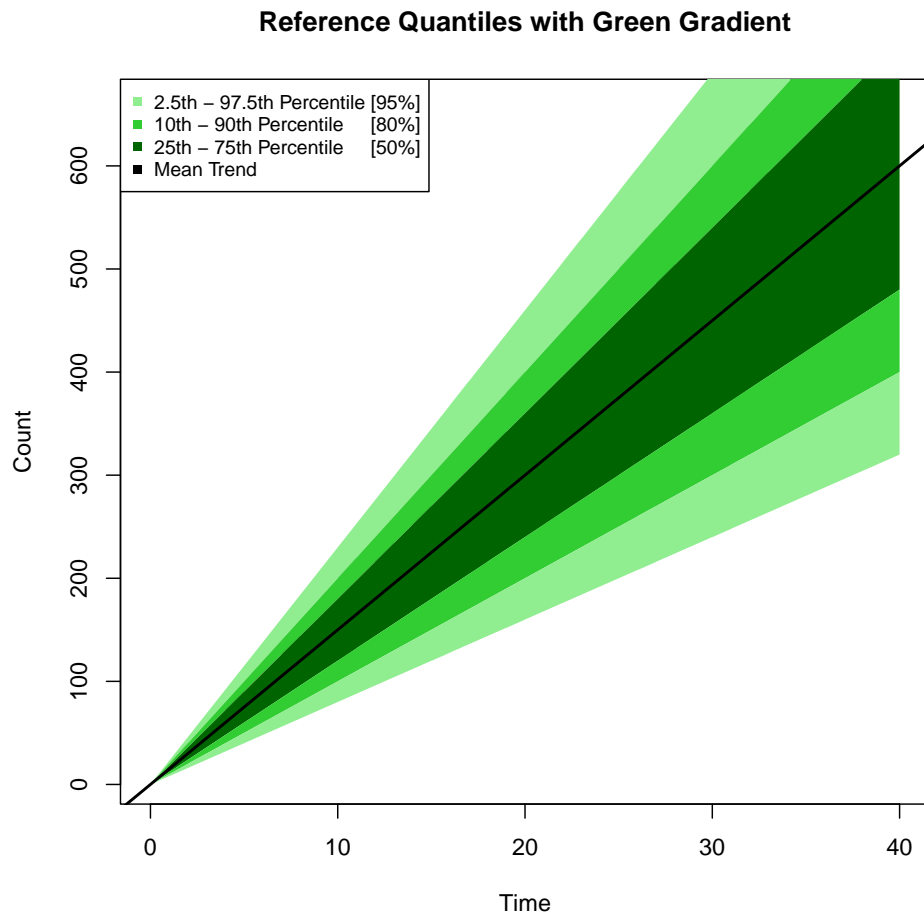


**Figure 2.2:** Poisson-distributed counts over time with estimated expectation and uncertainty. The red line represents the point estimate expectation ( $\lambda t$ ), while the blue dashed lines indicate Poisson aleatory uncertainty ( $\lambda t \pm \sqrt{\lambda t}$ ). The histogram on the right illustrates the distribution of observed counts. Ref.: [Spiegelhalter et al. \(2011\)](#)

## 2.3 Derivation of Negative Binomial from Poisson-Gamma model

Let  $Y|\lambda \sim Po(\lambda)$  and  $\lambda \sim G(\alpha, \beta)$

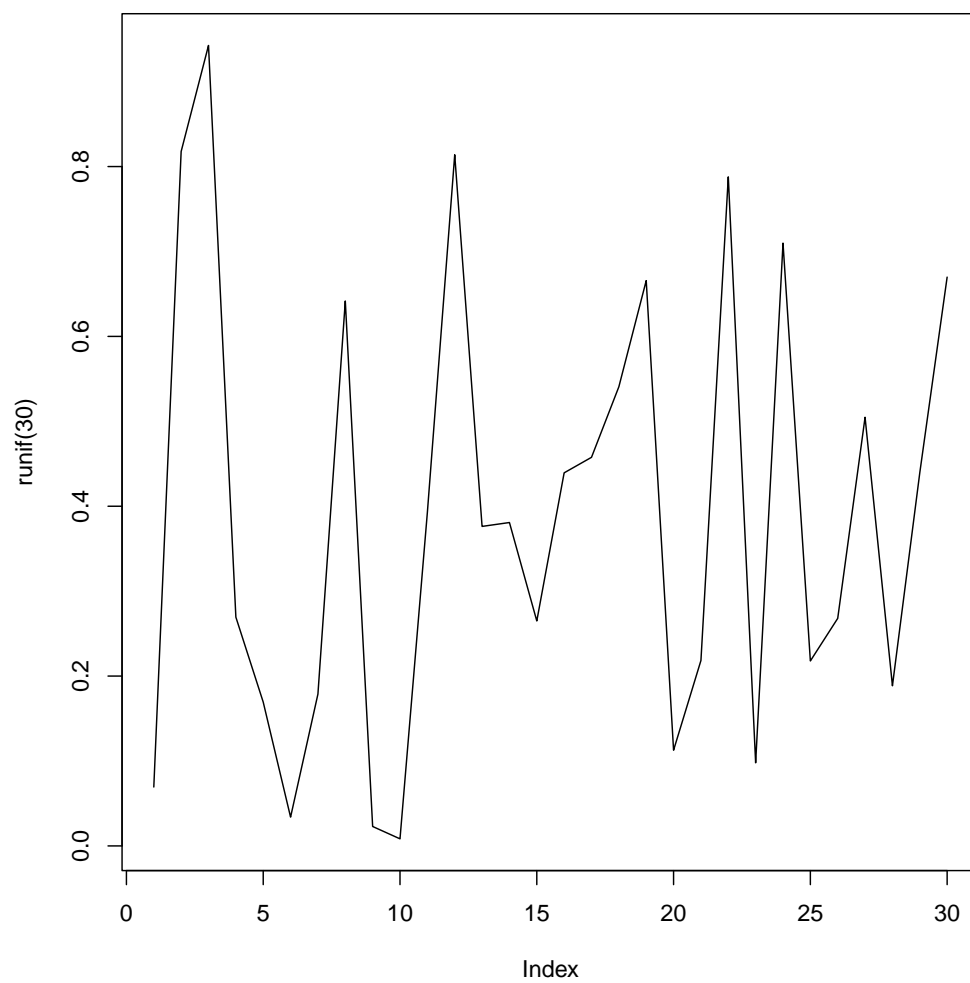
$$\begin{aligned}
 p(y) &= \int_0^\infty p(y|\lambda)p(\lambda)d\lambda \\
 &= \int_0^\infty \frac{\lambda^y \exp(-\lambda)}{y!} \left[ \lambda^{\alpha-1} \exp(-\beta\lambda) \frac{\beta^\alpha}{\Gamma(\alpha)} \right] d\lambda \\
 &= \frac{\beta^\alpha}{y! \Gamma(\alpha)} \int_0^\infty \lambda^{\alpha+y-1} \exp(-\lambda) \exp(-\lambda\beta) d\lambda \\
 &= \frac{\Gamma(\alpha+y)}{y! \Gamma(\alpha)} \int_0^\infty \beta^\alpha \exp(-\lambda\beta) d\lambda \\
 &= \frac{\Gamma(\alpha+y)}{y! \Gamma(\alpha)} \left( 1 - \frac{\beta}{\beta+1} \right)^\alpha \left( \frac{\beta}{\beta+1} \right)^y \sim NBin\left(\alpha, \frac{\beta}{\beta+1}\right)
 \end{aligned}$$



**Figure 2.3:** Poisson process with uncertainty bands over time. The black line represents the median count, while the green shaded regions indicate uncertainty: the dark green band spans the interquantile range (25th - 75th percentiles), and the lighter green band cover the 5th - 95th percentile range, based on 1,000 simulations. Ref.: Spiegelhalter *et al.* (2011)

## 2.4 Citations

Recall the difference between `\citet{}` (e.g., Chu and George (1999)), `\citep{}` (e.g., (Chu and George, 1999)) and `\citealp{}` (e.g., Chu and George, 1999). For simplicity, we include here all references in the file `biblio.bib` with the command `\nocite{*}`.



**Figure 2.4:** Test figure to illustrate figure options used by knitr.





## Chapter 3

## Results



## Chapter 4

# Discussion and Outlook



## Chapter 5

# Conclusions

### 5.0.1 Personal Statement

NOT AI: This thesis was not written by any generative AI. It was written independently and without assistance from third parties. All sources utilized in this thesis are appropriately cited in the references

AI: During the preparation of this Master Thesis, I used [NAME OF TOOLS AND SERVICES] in order to [REASON]. After using this tool/service, I reviewed and edited the content as needed and I take full responsibility for the content of the Master Thesis.



# Bibliography

- Bogin, V. (2022). Lasagna’s law: a dish best served early. *Contemporary Clinical Trials Communications*, **26**, 100900. [4](#)
- Chu, E. and George, A. (1999). *Inside the FFT Black Box: Serial and Parallel Fast Fourier Transform Algorithms*. CRC Press. [6](#)
- Collett, D. (2014). *Modelling Survival Data in Medical Research*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, third edition.
- Food, Administration, D., *et al.* (2018). Evaluating inclusion and exclusion criteria in clinical trials. In *Workshop report. The National Press Club, Washington DC*. [3](#)
- Frank, G. (2004). Current challenges in clinical trial patient recruitment and enrollment. *SoCRA Source*, **2**, 30–38. [3](#)
- Hurrell, J. W., Kushnir, Y., Ottersen, G., and Visbeck, M. (2013). *An Overview of the North Atlantic Oscillation*, 1–35. American Geophysical Union.
- Kratzer, G. and Furrer, R. (2018). *varrank: an R package for variable ranking based on mutual information with applications to observed systemic datasets*. R package version 0.3.
- Lim, C.-Y. and In, J. (2019). Randomization in clinical studies. *Korean journal of anesthesiology*, **72**, 221–232. [3](#)
- National Institute of Allergy and Infectious Diseases (2021). Screening, enrollment, and unblinding of participants. <https://www.niaid.nih.gov/sites/default/files/screening-enrollment-unblinding-of-participants.pdf>. Accessed: 2025-02-10. [3](#)
- Nüesch, E., Trelle, S., Reichenbach, S., Rutjes, A. W., Bürgi, E., Scherer, M., Altman, D. G., and Jüni, P. (2009). The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. *Bmj*, **339**, . [3](#)
- Panos, G. D. and Boeckler, F. M. (2023). Statistical analysis in clinical and experimental medical research: simplified guidance for authors and reviewers. [3](#)
- Piantadosi, S. and Meinert, C. L. (2022). *Principles and practice of clinical trials*. Springer Nature. [4](#)
- Porcu, E., Furrer, R., and Nychka, D. (2020). 30 years of space-time covariance functions. *Wiley Interdisciplinary Reviews. Computational Statistics*, **13**:e1512, 1–24.
- Rehman, A. M., Ferrand, R., Allen, E., Simms, V., McHugh, G., and Weiss, H. A. (2020). Exclusion of enrolled participants in randomised controlled trials: what to do with ineligible participants? *BMJ open*, **10**, e039546. [3](#)
- Shih, W. J. (2002). Problems in dealing with missing data and informative censoring in clinical trials. *Current controlled trials in cardiovascular medicine*, **3**, 1–7. [3](#)

- Spiegelhalter, D., Pearson, M., and Short, I. (2011). Visualizing uncertainty about the future. *science*, **333**, 1393–1400. [5](#), [6](#)
- Van Spall, H. G., Toren, A., Kiss, A., and Fowler, R. A. (2007). Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *Jama*, **297**, 1233–1240. [3](#)
- Wang, C. and Furrer, R. (2020). Monte Carlo permutation tests for assessing spatial dependence at different scales. In La Rocca, M., Liseo, B., and Salmaso, L., editors, *Nonparametric Statistics. ISNPS 2018. Springer Proceedings in Mathematics & Statistics*, volume 339, 503–511. Springer.
- Whelan, J., Le Deley, M.-C., Dirksen, U., Le Teuff, G., Brennan, B., Gaspar, N., Hawkins, D. S., Amler, S., Bauer, S., Bielack, S., *et al.* (2018). High-dose chemotherapy and blood autologous stem-cell rescue compared with standard chemotherapy in localized high-risk ewing sarcoma: results of euro-ewing 99 and ewing-2008. *Journal of Clinical Oncology*, **36**, 3110–3119. [4](#)
- Willie, M. M. (2024). Population and target population in research methodology. *Golden Ratio of Social Science and Education*, **4**, 75–79. [3](#)