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 24, 2025

Contents

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.

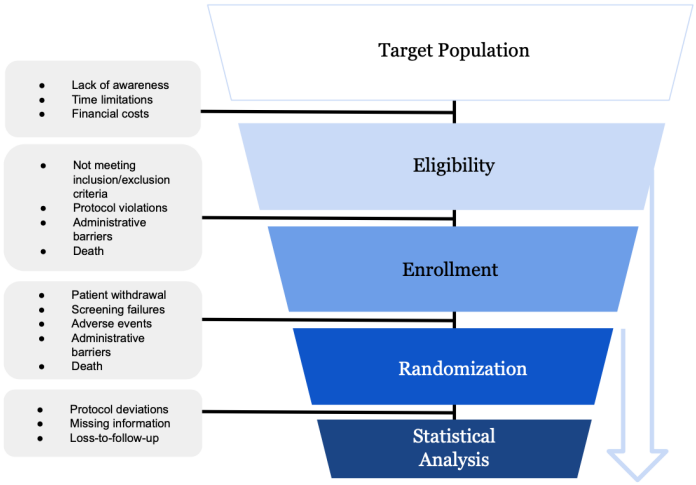

Figure 2.1: Patient Attrition at Each Stage of a Clinical Study. References: Piantadosi and Meinert (2022); Whelan *et al.* (2018); Bogin (2022)

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 define **Aleatory Uncertainty** as that which can be characterised by randomness. It is inherent, irreducible and unpredictable in nature. Whereas **Epistemic Uncertainty** arises primarily from limited or imperfect knowledge and can be reducible by obtaininng more or better information.

Let us denote

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

Methods	Counts	Expectation	Variance	Aleatory	Epistemic
Expectation	$C = \lambda T$	λT	0	No	No
Poisson	$C \sim \text{Po}(\lambda T)$	λT	λT	Yes	No
Negative Binomial	$C \sim \text{Po}(\Lambda T)$; $\Lambda \sim G(\alpha, \beta)$	$\frac{\alpha}{\beta}$	$\frac{\alpha(\beta+1)}{\beta^2}$	Yes	Yes

Table 2.1: Aleatory and epistemic uncertainty shown by different models for counts.

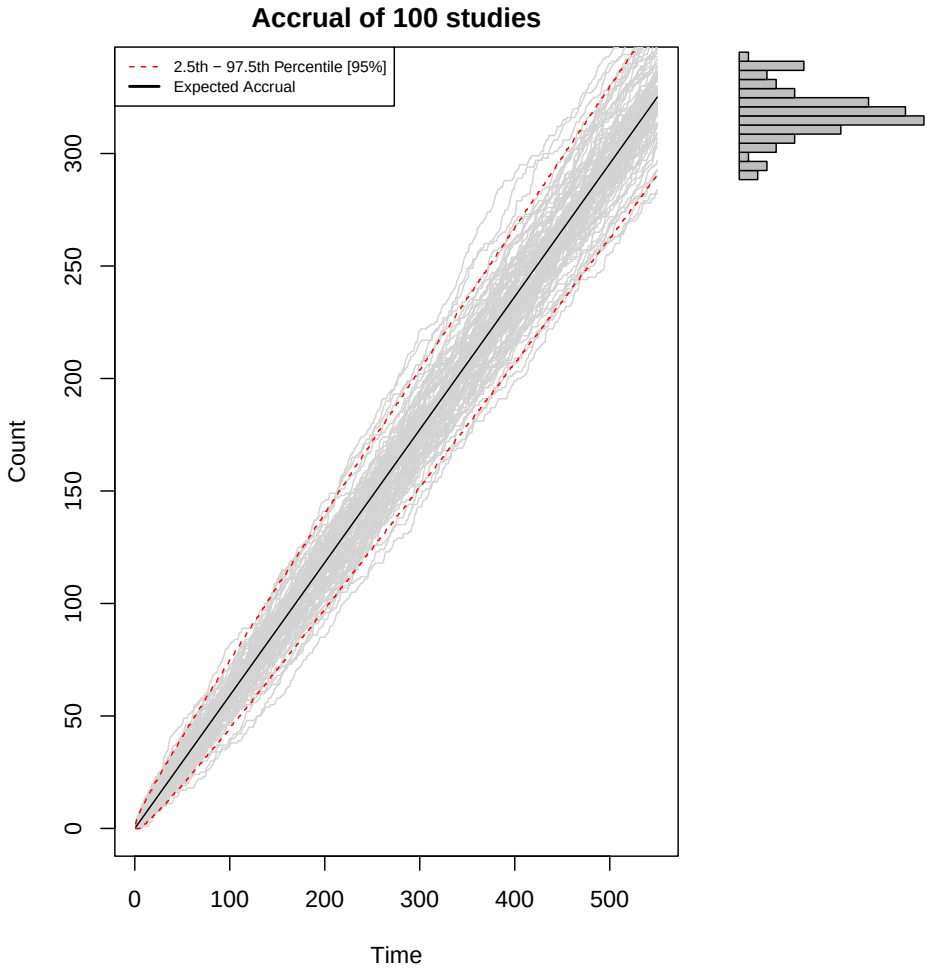

Figure 2.2: Poisson-distributed counts over time and uncertainty range. The black line represents the point estimate expectation ($\lambda = 0.591$), while the grey dashed lines indicate Poisson's 95% aleatory uncertainty. The histogram illustrates the distribution of observed counts at time $T = 550$. Ref.: Spiegelhalter *et al.* (2011).

2.3 Derivation of Negative Binomial from Poisson-Gamma model

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

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

To calculate the mean and variance:

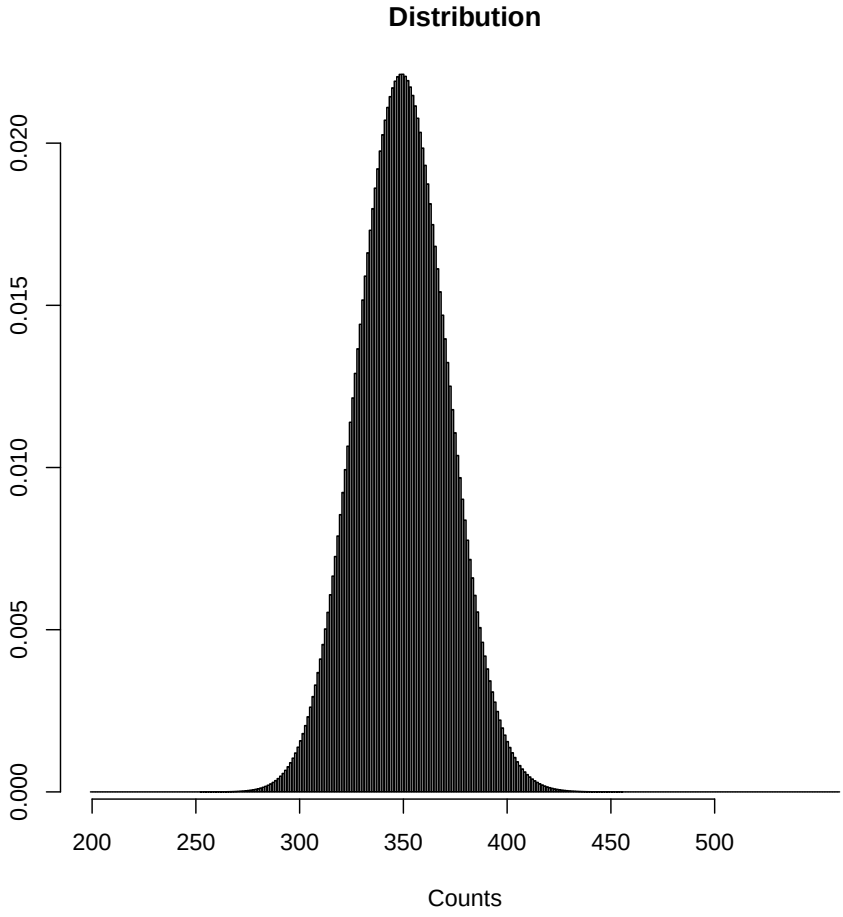

Figure 2.3: Poisson Distribution of Counts: This bar plot represents the probability mass function (PMF) of counts ranging from 200 to 500, using a Poisson distribution with a rate parameter $\lambda = 0.591$ based on 550 time periods. Ref.: .

$$
\begin{aligned}
Mean &= \frac{\alpha\left(1 - \frac{\beta}{\beta+1}\right)}{\frac{\beta}{\beta+1}} \\
&= \frac{\alpha\left(\frac{1}{\beta+1}\right)}{\frac{\beta}{\beta+1}} \\
&= \frac{\alpha(\beta+1)}{\beta(\beta+1)} \\
&= \frac{\alpha}{\beta}
\end{aligned}
$$

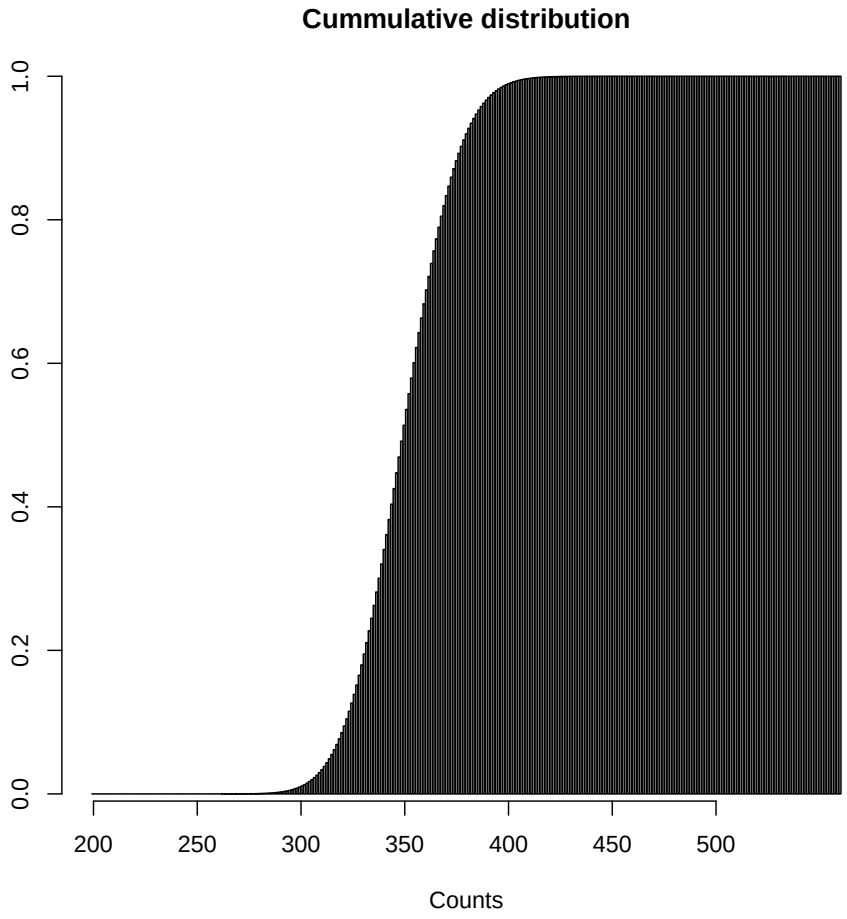

Figure 2.4: Cumulative Distribution of Poisson-Distributed Counts: The plot illustrates the cumulative probability distribution for counts within the range of 200 to 500, using a Poisson distribution with a rate parameter $\lambda = 0.591$ adjusted for 550 time periods. Ref.: .

$$
\begin{aligned}
Variance &= \frac{\alpha\left(1 - \frac{\beta}{\beta+1}\right)}{\left(\frac{\beta}{\beta+1}\right)^2} \\
&= \frac{\alpha\left(\frac{1}{\beta+1}\right)}{\left(\frac{\beta}{\beta+1}\right)^2} \\
&= \frac{\alpha(\beta+1)^2}{\beta^2(\beta+1)} \\
&= \frac{\alpha(\beta+1)}{\beta^2}
\end{aligned}
$$

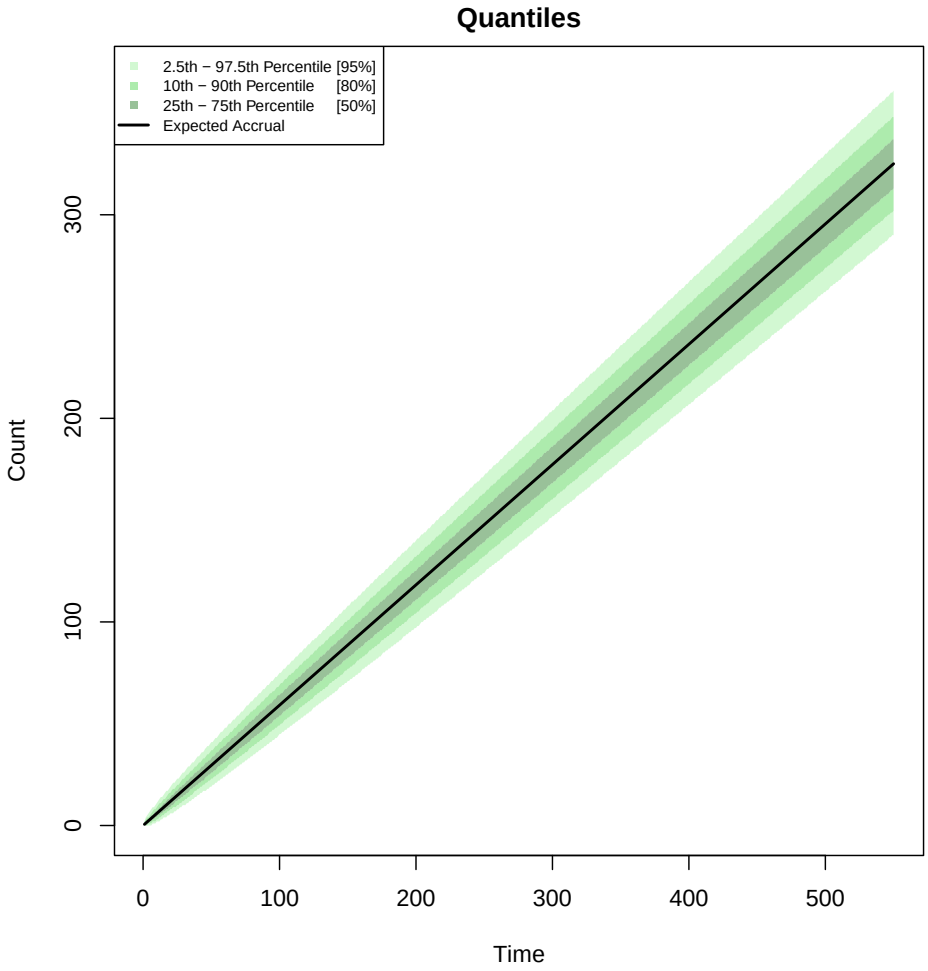

Figure 2.5: Predicted uncertainty bands for poisson process over time. The black line represents the mean trend $\lambda = 15$, while the green shaded regions indicate aleatory uncertainty: the dark green band spans the interquantile range (25th - 75th percentiles), the lighter green band cover the 10th - 90th percentile range and the light green the 2.5th - 97.5th percentile range. Ref.: Spiegelhalter *et al.* (2011).

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

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

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

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, 8

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

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