Thailand Statistician
April 2026; 24(2): 474-493
http://statassoc.or.th
Contributed paper

Interplay of Probability and Precision: Event Size Calculation for Hypothesis Testing, Confidence Interval and Prediction Interval of Poisson Rates

Wei-ming Luh* [a] and Jiin-Huarng Guo [b]
[a] Institute of Education, National Cheng Kung University, Tainan, Taiwan.
[b] Department of Applied Mathematics, National Pingtung University, Pingtung, Taiwan.
*Corresponding author; e-mail: luhwei@mail.ncku.edu.tw

Received: 24 June 2025
Revised: 8 October 2025
Accepted: 8 January 2026

Abstract

The Poisson distribution is a key model for describing rare events, from machinery failures in engineering to disease onset in medicine. While prior research has examined the determination of event size for testing equality between Poisson distributions, comprehensive approaches that integrate both statistical power and precision remain limited. To address this gap, our study aims to achieve two goals. First, we present methods for calculating event sizes based on incidence rate ratios, incorporating rejection, validity, and confidence interval width to ensure both power and precision. Second, we extend the discussion to prediction intervals, establishing the original event sizes needed to achieve a specified probability of coverage. Simulation studies confirm that the proposed methods yield valid results for coverage and target probabilities. To support practical use, we developed three interactive R Shiny applications, making the methodology accessible for researchers and practitioners. Together, the empirical validation and user-friendly tools demonstrate the practical value of our framework.

Keywords: Conditional probability, coverage, precision, statistical power, R Shiny app, relative width.

1. Introduction

When studying rare events—such as the number of equipment failures in a factory, the occurrence of a rare disease, or the detection of species in an ecological survey—researchers must know how many events are needed to draw reliable conclusions. If too few events are observed, the study may lack power or produce imprecise results. If too many are required, valuable resources may be wasted. Determining the correct event size is, therefore, a practical and essential step in planning scientific studies. The Poisson distribution is widely used to model rare events across diverse fields, including agriculture, ecology, biology, medicine, industry, and environmental science (Krishnamoorthy and Thomson 2004). A central challenge in applying the Poisson process is determining the appropriate event size—that is, the number of events observed or the length of the observation window. Traditionally, sample size planning has emphasized hypothesis testing, with a

focus on statistical power. More recently, confidence intervals (CIs) have gained attention as a more informative alternative, as they provide not only evidence against the null hypothesis but also estimates of precision. The 2019 special issue of The American Statistician reflects this broader shift toward interval estimation. However, designing studies based on CI width introduces challenges. Since CI width is itself a random variable, researchers may achieve the desired precision only about 50% of the time (Pan and Kupper 1999, Wang and Kupper 1997, Liu 2009). This uncertainty underscores a significant gap in current knowledge of experimental design.

To address these limitations, recent advances have incorporated probability-based criteria into sample size planning. In particular, researchers have considered three fundamental events: rejection of the null hypothesis (event rejection, denoted as R), coverage of the true parameter by the CI (event validity, denoted as V), and attainment of the desired CI width (event width, denoted as W) (Kelley and Rausch 2006, Lai and Kelley 2012, Luh 2022). Additional conditional probabilities, such as joint or intersecting events, have also been studied (Beal 1989, Cesana and Antonelli 2010, Jiroutek et al. 2003, Liu 2012). However, most of this work has focused on the normal distribution, with relatively little attention to the Poisson case. Methods for constructing CIs for ratios of two independent Poisson rates exist (e.g., Graham et al. 2003, Kharrati-Kopaei and Dorosti-Motlagh 2020, Li et al. 2014, Nelson 1970, Sahai and Khurshid 1993, Xiao et al. 2017), but systematic approaches to event size determination remain limited (Price and Bonett 2000). In the Poisson framework, event size is often more meaningful than sample size, since the number of observed rare events provides a more direct planning criterion (Luh and Guo 2022, Nelson 1987).

Beyond CIs, prediction intervals (PIs) offer another powerful but underutilized tool. PIs estimate the range in which future observations are likely to fall, thereby providing insight into study reproducibility (Anderson and Kelley 2024, Anderson and Maxwell 2015, Chiolero et al. 2012, Calin-Jageman and Cumming 2019, Cumming 2008, Hartnack and Roos 2021, IntHout et al. 2016, Preston 2000). Unlike CIs, which estimate fixed parameters, PIs directly address replication—a central concern in the ongoing reproducibility crisis (Patil et al. 2016, Spence and Stanley 2016, 2024). Yet their use in Poisson-based designs has been sparse (Faulkenberry 1973, Hahn and Meeker 1991, Kim et al. 2021, Krishnamoorthy and Peng 2011, Krishnamoorthy and Lv 2018). Extending PI-based planning to Poisson processes enables researchers to determine the number of events required in both original and replication studies, thereby enhancing predictive accuracy and reliability.

In this paper, we propose a unified framework for event size determination in Poisson models, integrating power, precision, and prediction. We also provide user-friendly R Shiny applications to facilitate adoption in practice. The remainder of this article is organized as follows. Section 2 introduces the calculation of event size for a single Poisson rate, utilizing the continuous F -statistic as described by Maguire et al. (1952) and Luh and Guo (2022). The development of corresponding R Shiny apps (Chang et al. 2023, R Development Core Team 2024) is also mentioned for ease of application. Section 3 extends the discussion to comparing two Poisson rates. These two sections first discuss testing hypotheses and constructing CIs. To be comprehensive, nine cases of probabilities are considered as follows: 1. $P(R)$ (statistical power), 2. $P(W)$ (assurance probability, or the probability of the obtained CI to achieve a desired width), 3. $P(W \cap R)$ (the probability that both event width and event rejection occur simultaneously, or the probability that the obtained confidence interval achieves the desired width and allows rejection of the null hypothesis), 4. $P(W \cap V)$ (the probability that both event width and event validity occur simultaneously), 5. $P(W \cap R \cap V)$ (the probability

that event width, event rejection, and event validity all occur), 6. $P(W \mid V)$ (the conditional probability of event width occurs, given that the event validity has occurred), 7. $P(W \cap R \mid V)$ (the conditional probability of both events width and rejection occurs, given that the event validity has occurred), 8. $P(W \mid R)$ (the conditional probability of event width occurs, given that the event rejection has occurred), and 9. $P(W \cap V \mid R)$ (the conditional probability of both events width and validity occurs, given that the event rejection has occurred). Then the event size using PIs is discussed, including two-sided and one-sided intervals. A measure of precision, relative width, is discussed. Section 4 illustrates examples and demonstrates event size planning for hypothesis testing and the construction of confidence/ prediction intervals. Section 5 presents empirical examples and simulation results. Finally, the conclusion and discussion are in Section 6.

2. The One Poisson Rate Problem

Suppose events occur randomly over time at a true rate λ; in that case, the number of observed events (n) within an observational window (time or space, denoted as sample size t) follows a Poisson distribution with a mean λt (Cox 1953). The waiting time until the n^{th} event, denoted by t, is itself random and follows a chi-squared distribution with $2n$ degrees of freedom (Cinlar 1975, Ross 2014). The maximum likelihood estimator (MLE) of λ is $\hat{\lambda} = n/t$. In this section, we describe two approaches for determining event size under the Poisson model: tests and confidence intervals (CIs), as well as prediction intervals (PIs). To make these procedures accessible for applied researchers, we provided a definition and notation table in Appendix A and developed R Shiny applications that allow users to compute the required event size directly. Researchers can also determine the corresponding sample size (i.e., the number of time intervals or observation windows, denoted as t), which must be at least equal to the ceiling function of (n/λ_1).

2.1. Tests and confidence intervals

From the maximum likelihood estimator of λ, smaller values of t lead to larger estimates of λ which increases the chance of rejecting a right-tailed test as $H_0 : \lambda \le \lambda_0$ in favor of $H_1 : \lambda > \lambda_0$; where λ_0 is provided by the researcher. The type I error rate is

$$P(2\lambda_0 t \le \chi^2_{2n;\alpha} \mid H_0 : \lambda \le \lambda_0) \le \alpha,$$

where α is the significance level and $\chi^2_{2n;\alpha}$. is the α^{th} quantile of a chi-squared distribution with $2n$ degrees of freedom. Under an alternative hypothesis $H_1 : \lambda = \lambda > \lambda_0$, the statistical power (the probability of correctly rejecting H_0) is

$$P(R) = P(2\lambda_0 t \le \chi^2_{2n;\alpha} \mid H_1 : \lambda = \lambda_1 > \lambda_0) = P\left(\chi^2 \le \frac{\lambda_1}{\lambda_0} \chi^2_{2n;\alpha} \mid H_1 : \lambda = \lambda_1 > \lambda_0 \right), \quad (1)$$

where $\chi^2 = 2\lambda_1 t$. To achieve the desired probability level $1-\beta$ for $P(R)$, the event size n can be chosen to satisfy

$$\frac{\lambda_1}{\lambda_0} \chi^2_{2n;\alpha} \ge \chi^2_{2n;1-\beta}.$$

Similarly, for the left-tailed test $H_0 : \lambda \ge \lambda_0$

$$P(R) = P\left(\chi^2 \geq \frac{\lambda_1}{\lambda_0} \chi^2_{2n;1-\alpha} \mid H_1 : \lambda = \lambda_1 < \lambda_0 \right), \tag{2}$$

and for $P(R) \geq 1-\beta$, we can find the event size, n, to meet

$$\frac{\lambda_1}{\lambda_0} \chi^2_{2n;1-\alpha} \leq \chi^2_{2n;\beta}.$$

Finally, for the two-tailed test $H_0 : \lambda = \lambda_0$,

$$P(R) = P\left(\left(\chi^2 \leq \frac{\lambda_1}{\lambda_0} \chi^2_{2n;\alpha/2} \right) \cup \left(\chi^2 \geq \frac{\lambda_1}{\lambda_0} \chi^2_{2n;1-\alpha/2} \right) \mid H_1 : \lambda = \lambda_1 \neq \lambda_0 \right). \tag{3}$$

We can find the event size, n, to achieve $P(R) \geq 1-\beta$. To construct a $(1-\alpha)100\%$ two-sided CI for λ_1 (i.e., $P(V) = 1-\alpha$), it can be shown that

$$1-\alpha = P(\chi^2_{2n;\alpha/2} \leq 2\lambda_1 t \leq \chi^2_{2n;1-\alpha/2}) = P\left(\frac{\chi^2_{2n;\alpha/2}}{2t} \leq \lambda_1 \leq \frac{\chi^2_{2n;1-\alpha/2}}{2t} \right). \tag{4}$$

Denote $L_{\lambda_1} = \chi^2_{2n;\alpha/2} / (2t)$ and $U_{\lambda_1} = \chi^2_{2n;1-\alpha/2} / (2t)$ as the CI lower limit and upper limit of λ_1, respectively, the width of CI, a random variable that can be used to determine the event size needed. Concerning the precision of absolute width, we consider the value of $W_{CI} = U_{\lambda_1} - L_{\lambda_1}$ to be less than or equal to a desired width w with a desired probability level $1-\beta$ (i.e., the event width occurs). That is,

$$\begin{aligned} P(W) &= P(U_{\lambda_1} - L_{\lambda_1} \leq w) = P((\chi^2_{2n;1-\alpha/2} - \chi^2_{2n;\alpha/2}) / (2t) \leq w) = P(2\lambda_1 t \geq \lambda_1 (\chi^2_{2n;1-\alpha/2} - \chi^2_{2n;\alpha/2}) / w), \\ &= P(\chi^2 \geq w_1) \geq 1-\beta, \end{aligned} \tag{5}$$

where $w_1 = \lambda_1 (\chi^2_{2n;1-\alpha/2} - \chi^2_{2n;\alpha/2}) / w$. For $P(W) \geq 1-\beta$, we can set $w_1 \leq \chi^2_{2n;\beta}$ to find the required event size n. The present study also considers various probabilities presented in Appendix B. The R Shiny app (I) (panel 1) (https://size-poisson.shinyapps.io/Poisson-1-rate/) is developed to find the required event size. If needed, the researcher can also determine the corresponding sample size (i.e., the number of time intervals) (t), which is equal to the ceiling (n / λ_1). To reach the target probability, the proposed R Shiny app (I) (panel 1) (see Figure 1) uses an iterative process to calculate the required event size under nine different probability scenarios. The algorithm, written in R (R Development Core Team 2024), works as follows: First, researchers select the type of probability event they are interested in (e.g., Case 1). For example: Case 1: rejection event $P(R)$, Case 2: interval width event $P(W)$, Case 3: the intersection event both width and rejection $P(W \cap R)$, and so on. Next, they enter three key values: the significance level α (e.g., 0.05), the desired probability $1-\beta$ (e.g., 0.8), and the alternative planning value λ_1 (e.g., 1.67). For cases related to event rejection (R), researchers also have to provide the null hypothesis value λ_0, which must be smaller than λ_1 for a right-tailed test (e.g., $\lambda_0 = 1$). They then indicate whether the test is one-tailed or two-tailed (e.g., "1" for a one-tailed test). For interval-width cases (event W), instead of λ_0, researchers have to specify the desired confidence interval width w (e.g., 1.1). Finally, by clicking "Submit" in the App, the required event size can be obtained.

2.2. Prediction intervals

The original sample with event size n and corresponding time t_n and the replication sample with event size m and corresponding time t_m are independent and have the same Poisson process with rate λ. Let $\hat{\lambda}_n = n/t_n$ and $\hat{\lambda}_m = m/t_m$ are the sample estimators of λ for the original sample and replication sample, respectively. We define the statistic as

$$F = \frac{t_m/m}{t_n/n} = \frac{\hat{\lambda}_n}{\hat{\lambda}_m},$$

which is distributed as an F distribution with $(2m, 2n)$ degrees of freedom (Luh and Guo 2022). For a two-sided $(1-\alpha)100\%$ prediction interval to cover $\hat{\lambda}_m$, we have

$$\begin{aligned} 1-\alpha &= P\left(F_{2m,2n;\alpha/2} < \frac{t_m/m}{t_n/n} < F_{2m,2n;1-\alpha/2} \right) = P\left(t_m\hat{\lambda}_n / F_{2m,2n;1-\alpha/2} < m < t_m\hat{\lambda}_n / F_{2m,2n;\alpha/2} \right), \\ &= P\left(\hat{\lambda}_n / F_{2m,2n;1-\alpha/2} < \hat{\lambda}_m < \hat{\lambda}_n / F_{2m,2n;\alpha/2} \right). \end{aligned} \tag{6}$$

Hence, the two-sided $(1-\alpha)100\%$ prediction interval for the replication estimated rate $\hat{\lambda}_m$ is given by $(\hat{\lambda}_n / F_{2m,2n;1-\alpha/2}, \hat{\lambda}_n / F_{2m,2n;\alpha/2})$.

2.2.1 Two-sided PIs

From (6), the width of the prediction interval of $\hat{\lambda}_m$ is $\hat{\lambda}_n(1/F_{2m,2n;\alpha/2} - 1/F_{2m,2n;1-\alpha/2})$, denoted as W_{PI}, which is a random variable. When the original event size n increases, $\hat{\lambda}_n$ approaches to λ and the statistic F converges to $\lambda t_m/m$, which follows a distribution $\chi^2_{2m}/(2m)$ with $2m$ degrees of freedom. Hence, the limiting prediction interval is $(2m\lambda/\chi^2_{2m;1-\alpha/2}, 2m\lambda/\chi^2_{2m;\alpha/2})$ with the width $2m\lambda(1/\chi^2_{2m;\alpha/2} - 1/\chi^2_{2m;1-\alpha/2})$, denoted as W_{LPI}. Then, the relative width (RW) of the prediction width to the limiting width is

$$RW = \frac{\hat{\lambda}_n(1/F_{2m,2n;\alpha/2} - 1/F_{2m,2n;1-\alpha/2})}{2m\lambda(1/\chi^2_{2m;\alpha/2} - 1/\chi^2_{2m;1-\alpha/2})}. \tag{7}$$

By applying Meeker and Hahn (1982), for a desired probability value $1-\gamma$ and a given value r (>1) for RW, we can calculate the number of original events n to satisfy $P(RW \le r) \ge 1-\gamma$, i.e.,

$$P\left(\frac{\hat{\lambda}_n(1/F_{2m,2n;\alpha/2} - 1/F_{2m,2n;1-\alpha/2})}{2m\lambda(1/\chi^2_{2m;\alpha/2} - 1/\chi^2_{2m;1-\alpha/2})} \le r \right) = P\left(\frac{n(1/F_{2m,2n;\alpha/2} - 1/F_{2m,2n;1-\alpha/2})}{m\chi^2_{2n}(1/\chi^2_{2m;\alpha/2} - 1/\chi^2_{2m;1-\alpha/2})} \le r \right) \ge 1-\gamma,$$

or

$$\frac{n(1/F_{2m,2n;\alpha/2} - 1/F_{2m,2n;1-\alpha/2})}{m\chi^2_{2n;\gamma}(1/\chi^2_{2m;\alpha/2} - 1/\chi^2_{2m;1-\alpha/2})} \le r.$$

Note that the required event size is not related to the value of λ. The proposed R Shiny app (I) (panel 2) can be used to find the needed original event size n by providing the replication event size m.

2.2.2 One-sided PIs

For a replication study with the event size m, a $(1-\alpha)100\%$ one-sided lower prediction bound of $\hat{\lambda}_m$ is $\hat{\lambda}_n / F_{2m,2n;1-\alpha}$, which converges to $2m\lambda / \chi^2_{2m;1-\alpha}$. Given a lower value λ_L and a desired probability value $1-\gamma$, we provide the R Shiny App (I) (panel 3) to find the original event size n to satisfy $P(\hat{\lambda}_n / F_{2m,2n;1-\alpha} \geq \lambda_{\mathrm{L}}) \geq 1-\gamma$ or $2n\lambda / (F_{2m,2n;1-\alpha}\chi^2_{2n;1-\gamma}) \geq \lambda_L$. Similarly, we provide the R Shiny App (I) (panel 4) for a given upper value λ_{U} to find the original event size n that satisfies either $P(\hat{\lambda}_n / F_{2m,2n;\alpha} \leq \lambda_U) \geq 1-\gamma$ or $2n\lambda / (F_{2m,2n;\alpha}\chi^2_{2n;\gamma}) \leq \lambda_U$. Note that the researcher needs to provide planning values of λ and a lower/upper bound.

3. The Two Poisson Rates Problem

Suppose we consider two independent Poisson processes with occurrence rates λ_1, λ_2 and observe the $n_1^{\text{th}}, n_2^{\text{th}}$ events occur in the number of time intervals (i.e., sample size) t_1, t_2, respectively. Then $2\lambda_1 t_1$ and $2\lambda_2 t_2$ are distributed as chi-squared distributions with $2n_1$, $2n_2$ degrees of freedom, respectively. It is noted that the maximum likelihood estimator of λ_i is $\hat{\lambda}_i = n_i / t_i$, for $i = 1, 2$. Then,

$$F = \frac{\lambda_1 t_1 / n_1}{\lambda_2 t_2 / n_2},$$

is distributed as an F distribution with $2n_1$ and $2n_2$ degrees of freedom (Cox and Lewis 1966, Maguire et al. 1952, Luh and Guo 2022).

3.1 Tests and CIs

For a right-tailed test of the ratio of two rates, the null hypothesis against the alternative hypothesis is considered as

$$H_0 : \lambda_1 / \lambda_2 \leq \theta_0 \quad \text{versus} \quad H_1 : \lambda_1 / \lambda_2 > \theta_0,$$

where θ_0 is provided by the researcher. For testing equality, setting $\theta_0 = 1$. For significance level α, a smaller value of $(t_1 / n_1) / (t_2 / n_2) = \hat{\lambda}_2 / \hat{\lambda}_1$ results in a larger estimate of λ_1 / λ_2 that tends to reject $H_0 : \lambda_1 / \lambda_2 \leq \theta_0$ in favor of $H_1 : \lambda_1 / \lambda_2 > \theta_0$. The type I error rate is

$$P(\theta_0 \frac{t_1 / n_1}{t_2 / n_2} \leq F_{2n_1,2n_2;\alpha} \mid H_0 : \lambda_1 / \lambda_2 \leq \theta_0) \leq \alpha,$$

where $F_{2n_1,2n_2;\alpha}$ is the α^{th} quantile of an F distribution with $2n_1$ and $2n_2$ degrees of freedom. When the alternative hypothesis is assumed to be $H_1 : \lambda_1 / \lambda_2 = \theta_1 > \theta_0$, we can consider the statistical power as

$$P(R) = P\left(\theta_0 \frac{t_1 / n_1}{t_2 / n_2} \leq F_{2n_1,2n_2;\alpha} \mid H_1 : \lambda_1 / \lambda_2 = \theta_1 > \theta_0 \right) = P\left(F \leq \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;\alpha} \mid H_1 : \lambda_1 / \lambda_2 = \theta_1 > \theta_0 \right), \tag{6}$$

where $F = \theta_1 (t_1 / n_1) / (t_2 / n_2)$. To achieve the desired probability level $1-\beta$ for $P(R)$, we need to find the event sizes, (n_1, n_2) to satisfy

$$\frac{\theta_1}{\theta_0} F_{2n_1,2n_2;\alpha} \geq F_{2n_1,2n_2;1-\beta}.$$

Similarly, for the left-tailed test $H_0 : \lambda_1 / \lambda_2 \geq \theta_0$

$$P(R) = P\left(F \geq \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;1-\alpha} \mid H_1 : \lambda_1 / \lambda_2 = \theta_1 < \theta_0 \right). \tag{7}$$

For $P(R) \geq 1-\beta,$ we need to find event sizes $(n_1, n_2),$ to meet

$$\frac{\theta_1}{\theta_0} F_{2n_1,2n_2;1-\alpha} \leq F_{2n_1,2n_2;\beta}.$$

For the two-tailed test $H_0 : \lambda_1 / \lambda_2 = \theta_0,$

$$P(R) = P\left(\left(F \leq \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;\alpha/2} \right) \cup \left(F \geq \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;1-\alpha/2} \right) \mid H_1 : \lambda_1 / \lambda_2 = \theta_1 \neq \theta_0 \right) \geq 1-\beta. \tag{8}$$

The required event sizes can be obtained to achieve $P(R) \geq 1-\beta.$ To construct a $(1-\alpha)100\%$ two-sided CI for $\theta_1 = \lambda_1 / \lambda_2$ (i.e., $P(V) = 1-\alpha),$ it can be shown that

$$1-\alpha = P\left(F_{2n_1,2n_2;\alpha/2} \leq \theta_1 \frac{t_1 / n_1}{t_2 / n_2} \leq F_{2n_1,2n_2;1-\alpha/2} \right) = P\left(\frac{F_{2n_1,2n_2;\alpha/2}}{(t_1 / n_1) / (t_2 / n_2)} \leq \theta_1 \leq \frac{F_{2n_1,2n_2;1-\alpha/2}}{(t_1 / n_1) / (t_2 / n_2)} \right). \tag{9}$$

Denote $L_{\theta_1} = F_{2n_1,2n_2;\alpha/2} / [(t_1 / n_1) / (t_2 / n_2)] = (\hat{\lambda}_1 / \hat{\lambda}_2) F_{2n_1,2n_2;\alpha/2}$ and $U_{\theta_1} = F_{2n_1,2n_2;1-\alpha/2} / [(t_1 / n_1) / (t_2 / n_2)] = (\hat{\lambda}_1 / \hat{\lambda}_2) F_{2n_1,2n_2;1-\alpha/2}$ as a $(1-\alpha)100\%$ lower and upper limit of $\theta_1 = \lambda_1 / \lambda_2.$ The width of CI is $U_{\theta_1} - L_{\theta_1} = (F_{2n_1,2n_2;1-\alpha/2} - F_{2n_1,2n_2;\alpha/2}) / [(t_1 / n_1) / (t_2 / n_2)],$ a random variable. Regarding the precision of absolute width, we consider the value of $W_{CI} = U_{\theta_1} - L_{\theta_1}$ to be less than or equal to a desired width w (the event *W* occurs) with a desired probability level $1-\beta,$ i.e.,

$$\begin{aligned} P(W) &= P(U_{\theta_1} - L_{\theta_1} \leq w) = P((F_{2n_1,2n_2;1-\alpha/2} - F_{2n_1,2n_2;\alpha/2}) / [(t_1 / n_1) / (t_2 / n_2)] \leq w), \\ &= P\left(\theta_1 \frac{t_1 / n_1}{t_2 / n_2} \geq \frac{\theta_1 (F_{2n_1,2n_2;1-\alpha/2} - F_{2n_1,2n_2;\alpha/2})}{w} \right) = P(F \geq w_2) \geq 1-\beta, \end{aligned} \tag{10}$$

where $w_2 = \theta_1 (F_{2n_1,2n_2;1-\alpha/2} - F_{2n_1,2n_2;\alpha/2}) / w.$ For $P(W) \geq 1-\beta,$ we need to set $w_2 \leq F_{2n_1,2n_2;\beta}$ to find the required event sizes $(n_1, n_2).$ Researchers can also determine the corresponding sample sizes (i.e., the number of time intervals) $t_i = \text{ceiling}(n_i / \hat{\lambda}_i), i = 1, 2.$ In addition to $P(R)$ and $P(W),$ other probabilities are considered and presented in Appendix C. We developed the R Shiny app (II) (panel 1) (https://size-poisson.shinyapps.io/Poisson-2-rates-Test-CI/) to obtain event sizes for various probability cases. Considering that unequal allocation can sometimes be proper (Chang et al. 2010, Luh and Guo 2022, Hsu 1994, Gu et al. 2008, Krishnamoorthy and Thomson 2004, Ng et al. 2007, Peckham et al. 2015, Shan 2016, Shiue and Bain 1982, Thode 1997), the researcher may set an allocation ratio k (i.e., $n_2 = kn_1$) to calculate the required event sizes. In addition, the App (II) (panel 2) can determine the required event size n_2 for hypothesis testing and for constructing confidence intervals of the ratio of two Poisson rates when n_1 is fixed.

3.2 Prediction intervals

For two independent Poisson processes with rates λ_1 and λ_2, respectively, considering the replication event sizes (m_1, m_2) and the corresponding times (t_{m_1}, t_{m_2}), we aim to find the prediction interval of the ratio of replication estimate rates $\hat{\lambda}_{m_1} / \hat{\lambda}_{m_2}$, where $\hat{\lambda}_{m_i} = m_i / t_{m_i}, i = 1, 2.$ Based on the same Poisson processes with the original event sizes (n_1, n_2) and the corresponding times (t_{m_1}, t_{m_2}), the statistics

$$F_1 = \frac{\lambda_1 t_{n_1} / n_1}{\lambda_2 t_{n_2} / n_2} = \frac{\lambda_1 \hat{\lambda}_{n_2}}{\lambda_2 \hat{\lambda}_{n_1}} \quad \text{and} \quad F_2 = \frac{\lambda_1 t_{m_1} / m_1}{\lambda_2 t_{m_2} / m_2} = \frac{\lambda_1 \hat{\lambda}_{m_2}}{\lambda_2 \hat{\lambda}_{m_1}},$$

which are distributed as an F-distribution with $(2n_1, 2n_2), (2m_1, 2m_2)$ degrees of freedom, respectively. We denote

$$G = \frac{F_1}{F_2} = \frac{t_{n_1} / n_1}{t_{n_2} / n_2} \frac{t_{m_2} / m_2}{t_{m_1} / m_1} = \frac{\hat{\lambda}_{n_2}}{\hat{\lambda}_{n_1}} \frac{\hat{\lambda}_{m_1}}{\hat{\lambda}_{m_2}}.$$

For the level of $1-\alpha$, we have

$$1-\alpha = P(G_{v;\alpha/2} < G < G_{v;1-\alpha/2}) = P\left(L_{PI} < \frac{\hat{\lambda}_{m_1}}{\hat{\lambda}_{m_2}} < U_{PI} \right), \tag{11}$$

where $G_{v;\alpha/2}$ denotes the $(\alpha/2)^{\text{th}}$ quantile of the G distribution with $v = (\text{df}_1; \text{df}_2)$ degrees of freedom, where $\text{df}_1 = (2n_1, 2m_2)$ and $\text{df}_2 = (2n_2, 2m_1)$ for the numerator and the denominator, respectively (Bradley and Schumann 1957, Schumann and Bradly 1957, 1959). The two-sided and one-sided PIs are demonstrated as follows.

3.2.1 Two-sided PIs

The $(1-\alpha)100\%$ prediction interval of $\hat{\lambda}_{m_1} / \hat{\lambda}_{m_2}$ is (L_{PI}, U_{PI}) with $L_{PI} = (\hat{\lambda}_{n_1} / \hat{\lambda}_{n_2}) G_{v;\alpha/2}$ and $U_{PI} = (\hat{\lambda}_{n_1} / \hat{\lambda}_{n_2}) G_{v;1-\alpha/2}$. The width of the PI is $U_{PI} - L_{PI} = (\hat{\lambda}_{n_1} / \hat{\lambda}_{n_2})(G_{v;1-\alpha/2} - G_{v;\alpha/2})$, which is a random variable. Based on Meeker and Hahn (1982), when the original event sizes increase to infinity, the random variable G converges to $(\hat{\lambda}_{m_1} / \hat{\lambda}_{m_2}) / (\lambda_1 / \lambda_2)$, distributed as an F distribution with $2m_2$ and $2m_1$ degrees of freedom. Then, the prediction interval converges to the limiting interval as $((\lambda_1 / \lambda_2) F_{2m_2, 2m_1; \alpha/2}, (\lambda_1 / \lambda_2) F_{2m_2, 2m_1; 1-\alpha/2})$, and the width of the limiting interval is $(\lambda_1 / \lambda_2)(F_{2m_2, 2m_1; 1-\alpha/2} - F_{2m_2, 2m_1; \alpha/2})$, less than $U_{PI} - L_{PI}$. Then, the relative width (RW) of the width of the prediction interval to that of the limiting interval is given by

$$RW = \frac{\hat{\lambda}_{n_1} / \hat{\lambda}_{n_2}}{\lambda_1 / \lambda_2} \frac{G_{v;1-\alpha/2} - G_{v;\alpha/2}}{(F_{2m_2, 2m_1; 1-\alpha/2} - F_{2m_2, 2m_1; \alpha/2})},$$

which is no less than 1, and the random variable $(\hat{\lambda}_{n_1} / \hat{\lambda}_{n_2}) / (\lambda_1 / \lambda_2)$ is distributed as an $F_{2n_2, 2n_1}$. Given a probability value $1-\gamma$ and a desired value r for RW, we can obtain the original event sizes (n_1, n_2) to satisfy $P(RW \le r) \ge 1-\gamma$, i.e.,

$$P\left(\frac{(\hat{\lambda}_{n_1}/\hat{\lambda}_{n_2})(G_{v;1-\alpha/2}-G_{v;\alpha/2})}{(\lambda_1/\lambda_2)(F_{2m_2,2m_1;1-\alpha/2}-F_{2m_2,2m_1;\alpha/2})}\le r\right)=P\left(\frac{\hat{\lambda}_{n_1}/\hat{\lambda}_{n_2}}{\lambda_1/\lambda_2}\le r\frac{(F_{2m_2,2m_1;1-\alpha/2}-F_{2m_2,2m_1;\alpha/2})}{(G_{v;1-\alpha/2}-G_{v;\alpha/2})}\right)\ge 1-\gamma.$$

Because $(\hat{\lambda}_{n_1}/\hat{\lambda}_{n_2})/(\lambda_1/\lambda_2)$ is distributed as an F distribution with $2n_2$ and $2n_1$ degrees of freedom, we can have

$$r\frac{(F_{2m_2,2m_1;1-\alpha/2}-F_{2m_2,2m_1;\alpha/2})}{(G_{v;1-\alpha/2}-G_{v;\alpha/2})}\ge F_{2n_2,2n_1;1-\gamma},$$

and hence

$$F_{2n_2,2n_1;1-\gamma}\frac{(G_{v;1-\alpha/2}-G_{v;\alpha/2})}{(F_{2m_2,2m_1;1-\alpha/2}-F_{2m_2,2m_1;\alpha/2})}\le r.$$

Note that the required event sizes are not related to the value of λ_1/λ_2. We developed the R Shiny App (III) (https://size-poisson.shinyapps.io/Poisson-2-rates-PI/) (panel 1) to solve for the original event sizes (n_1, n_2) provided the replication event sizes (m_1, m_2). Furthermore, if one group's event size is limited (e.g., n_1 is fixed), the researcher can find n_2 to satisfy $P(RW\le r)\ge 1-\gamma$ by applying App (III) (panel 2). Note that a caution statement will occur if n_1 is tiny and the resulting n_2 is more significant than 100,000.

3.2.2 One-sided PIs

Consider a $(1-\alpha)100\%$ one-sided lower bound of $\hat{\lambda}_{m_1}/\hat{\lambda}_{m_2}$ is $(\hat{\lambda}_{n_1}/\hat{\lambda}_{n_2})G_{v;\alpha}$, which converges to $(\lambda_1/\lambda_2)F_{2m_2,2m_1;\alpha}$. For given a lower-bound value θ_L and a desired probability of $1-\gamma$, we can determine the original event sizes (n_1 , n_2) using the R Shiny App (III) (panel 3) to satisfy $P((\hat{\lambda}_{n_1}/\hat{\lambda}_{n_2})G_{v;\alpha}\ge\theta_L)\ge 1-\gamma$ or $(\lambda_1/\lambda_2)F_{2n_2,2n_1;\gamma}G_{v;\alpha}\ge\theta_L$ with planning values of λ_1 and λ_2. On the other hand, a $(1-\alpha)100\%$ one-sided upper bound of $\hat{\lambda}_{m_1}/\hat{\lambda}_{m_2}$ is $(\hat{\lambda}_{n_1}/\hat{\lambda}_{n_2})G_{v;1-\alpha}$, which converges to $(\lambda_1/\lambda_2)F_{2m_2,2m_1;1-\alpha}$. For a given upper-bound value θ_U, we can use App (III) (panel 4) to find the original event sizes that satisfy $P((\hat{\lambda}_{n_1}/\hat{\lambda}_{n_2})G_{v;1-\alpha}\le\theta_U)\ge 1-\gamma$ or $(\lambda_1/\lambda_2)F_{2n_2,2n_1;1-\gamma}G_{v;1-\alpha}\le\theta_U$.

4. Illustrative Examples

4.1. One Poisson rate

We first revisit the surface defects data from steel plates provided by Krishnamoorthy and Lv (2018, Example 3). The counts of surface defects on 21 steel plates are: 1, 0, 4, 3, 1, 2, 0, 2, 1, 1, 0, 0, 2, 1, 3, 4, 3, 1, 0, 2, 4. In total, 35 defects were observed, leading to an estimated rate of 1.67 defects per plate. This dataset is used to illustrate different planning strategies.

4.1.1 Hypothesis testing

Suppose, based on the above dataset, we want to test the null hypothesis $H_0:\lambda\le\lambda_0=1$ against the alternative hypothesis $H_1:\lambda=\lambda_1=1.67$ and determine the required event size for $\alpha=0.05$ and $1-\beta=0.8$. Using App (I) (panel 1) and selecting case 1 for $P(R)$ with "tailed" = 1 (a right-

tailed test, $\lambda_1 > \lambda_0$), the app suggests that the event size $n = 26$ (the corresponding sample size of t approximately equals to16 (=26/1.67) plates) are sufficient, with an achieved probability of 0.8126 (see Figure 1). For "tailed" = 2 (a two-tailed test), the required event size increases to n = 33 (sample size $t = 20$ plates), with an obtained probability of 0.8098. Although the original dataset contained 35 defects, only 26 would have been sufficient to achieve 80% power in a one-sided test for $\lambda_0 = 1$ and $\lambda_1 = 1.67$. In contrast, a two-sided test demands a larger sample. These results illustrate how the choice of test direction directly affects the required event size.

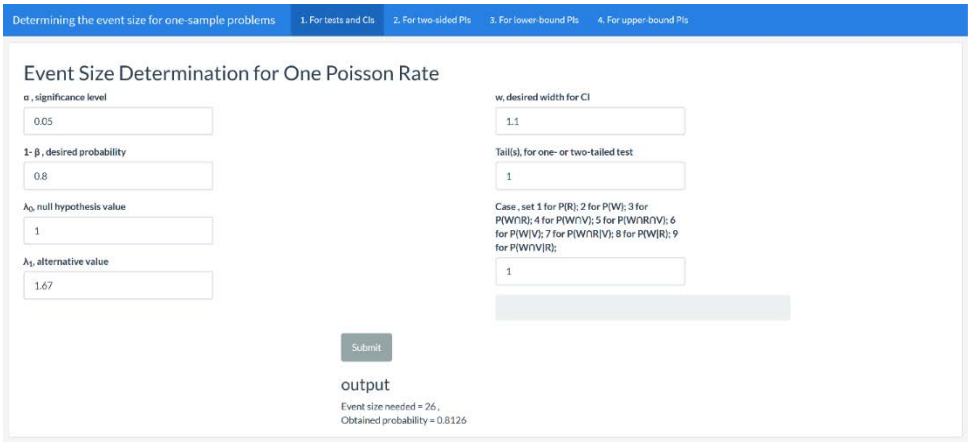


Figure 1 A screenshot of the R Shiny App (I)

4.1.2 Confidence interval construction

Using (4) with $n = 35$, we calculate a 95% confidence interval (CI) for $\lambda_1 = 1.67$ as (1.1609, 2.2625) with a CI width of 1.1016. Suppose we want a narrower CI (width ≤ 1) with at least 80% probability. That is, we aim to find the event size required to satisfy Case 2, $P(W) \geq 0.8$, with a desired width $w = 1$ and a significance level $\alpha = 0.05$. Using App (I) (panel 1) and selecting case 2 for $P(W)$ with $\lambda_1 = 1.67$, the required event size is $n = 55$ (sample size $t = 33$ plates) with an obtained probability of 0.8085. That is, to achieve a narrower width, a larger event size is required to reach the desired probability level.

4.1.3 Two-sided prediction interval

Assume in a replication study we expect the replication event size m = 25 and a desired value $r = 1.2$ to achieve $P(RW \leq r) \geq 1 - \gamma$, for $\alpha = 0.05$, a desired probability of $1 - \gamma = 0.8$, and $\lambda_1 = 1.67$. To ensure that the replication defect rate remains within 1.2 times the limiting width, with a coverage probability of 0.8, we use App (I) (panel 2) to find the required original event size $n = 111$. This corresponds to the original sample size of t_n approximately 111/1.67 = 67 plates. If m increases to 35, the required original event size n rises to 141, corresponding to 85 plates. In other words, a larger replication event size necessitates a larger original event size.

4.2. Two Poisson rates

In this example, there were 60 cases of coronary heart disease (CHD) among women who did not use postmenopausal hormones, observed over 51,477.5 person-years (i.e., t_1). In contrast, there were 30 cases of CHD among women who used postmenopausal hormones, observed over 54,308.7 person-years (t_2) (Kharrati-Kopaei and Dorosti-Motlagh 2020, Li et al. 2014). From the data, the estimates of $\hat{\lambda}_1 = 0.00117$, $\hat{\lambda}_2 = 0.00055$, and the estimated rate ratio is calculated to be 2.13 (= 0.00117/0.00055).

4.2.1 Hypothesis testing

Suppose we want to test $H_0 : \theta \le \theta_0 = 1$ vs. $H_1 : \theta = \theta_1 = 2.13$; we set a type I error rate of 0.05 and a power of 0.8. Using App (II), panel 1, with "tailed" set to 1 for a one-tailed test, and setting the allocation ratio $k = 1$ (indicating equal event sizes) and Case = 1, we obtained the required number of events as 23 for both groups. The corresponding sample sizes $t_1 =$ 23/0.00117 = 19,659 person-years and t_2 = 23/0.00055 = 41,819 person-years, respectively. The calculated probability of detecting the alternative value under these conditions is 0.8149. Despite having far larger observed person-years, the actual number of CHD events is relatively small. The app shows that even 23 events per group would have been enough to detect the observed effect size.

4.2.2 Confidence interval construction

For $\theta_1 = 2.13$, we calculate a 95% confidence interval as (1.3904, 3.3633) based on (9), resulting in a confidence interval width of 1.9729. We aim to determine the required event sizes to satisfy the desired width $w = 1.9729$ and a type I error rate α = 0.05. Using App (II), panel 1, we select case 2 and set the planning values $\theta_0 = 1$, $\theta_1 = 2.13$. This gives us the required event sizes of 53 for both groups (with $k = 1$ for equal event sizes), the corresponding sample sizes $t_1 =$ 53/0.00117 = 45,300 person-years and $t_2 =$ 53/0.00055 = 96,364 person-years, respectively, and the corresponding probability of achieving the specified condition is 0.8032.

4.2.3 Two-sided Prediction Interval

Suppose the replication event sizes are $m_1 = 40, m_2 = 40$ in a future (replication) study, and a ratio of two rates $\theta = 2.13$, we can calculate a 95% PI of $\hat{\lambda}_{m_1} / \hat{\lambda}_{m_2}$ using (11) as (1.1538, 4.0208). For $\alpha = 0.05$, to achieve the desired probability of $1-\gamma = 0.8$ with a ratio value $r = 1.5$ for $P(RW \le r) \ge 1-\gamma$, using App (III) (panel 1), we can determine the original event sizes needed as $n_1 = 66$, $n_2 = 66$ (with $k = 1$ for equal event sizes), and the corresponding sample sizes $t_1 = 66 / 0.00117 = 56,411$ person-years and $t_2 = 66 / 0.00055 = 120,000$ person-years, respectively. This example demonstrates that reproducibility requires larger event sizes than hypothesis testing alone. Even when the original effect appears strong, ensuring the replication study aligns within a reasonable range requires additional planning of the event size.

5. Event Size Tables and Simulation Results

The simulation study provides empirical validation of the proposed procedures by repeating each scenario 10,000 times. The empirical probabilities in Tables 1-4 represent the frequency with which

the desired statistical properties—power, confidence interval width, or prediction interval coverage—were achieved in practice. Table 1 presents the required event size for the nine cases of probabilities for the one-sample problem. Under the conditions of the planning values $\lambda_0 = 1,\ \lambda_1 = 1.5,$ and the desired width of $w = 1.2,$ the event sizes shown are the minimum numbers of events required to achieve approximately 80% probability for the nine cases of interest. The empirical percentages (≈80%) confirm that the proposed method produces the planned probability in repeated samples. We can find that the event size needed for $P(R) \geq 0.8$ is larger than that of $P(W) \geq 0.8.$ If the desired width becomes narrower, it can be anticipated that more event sizes might be needed in Case 2. Furthermore, more event sizes are required for the intersection of cases, reflecting the greater stringency of combined requirements. Finally, based on the 10,000 simulations, the results confirm that the proposed procedure achieves the desired probability level of 0.8.

Table 1 For the one-sample problem, the event size and simulation results for tests and CIs

Cases	Event size (n)	Empirical probability (%)
1. $P(R)$	51	80.58
2. $P(W)$	33	80.45
3. $P(W \cap R)$	52	80.72
4. $P(W \cap V)$	34	80.43
5. $P(W \cap R \cap V)$	58	82.96
6. $P(W \mid V)$	33	82.43
7. $P(W \cap R \mid V)$	52	82.74
8. $P(W \mid R)$	37	83.00
9. $P(W \cap V \mid R)$	38	85.71

Note: Setting $\alpha = 0.05,\ 1-\beta = 0.8,$ the planning values $\lambda_0 = 1,\ \lambda_1 = 1.5,$ and the desired width $w = 1.2$

Table 2 presents the required original event size given the replication event size. It is clear that the larger the given replication event size, the larger the original event size needed. The simulation results show that the coverage of the replication estimate rate $\hat{\lambda}_m$ can achieve 95% of the time by setting $\alpha = 0.05,$ and the empirical percentages of $RW \leq r$ also achieve the desired probability of 0.8, confirming that the method reliably meets the planning criteria.

Table 2 For the one-sample problem, the original event size and simulation results for PIs

Replication event size (m)	Original event size (n)	Coverage (%)	$RW \leq r$ (%)
5	44	95.16	80.65
10	62	94.72	79.74
15	79	94.99	80.32
20	95	94.94	80.35
25	111	94.86	80.52

Note: Setting $\alpha = 0.05,\ 1-\gamma = 0.8,$ and the desired relative width $r = 1.2$

For the two-sample problem, Table 3 presents the event sizes for the nine cases of probabilities and the empirical results of simulations. Considering allocations $k = 1$ (equal event size) and

$k = 2$ $(n_2 = 2n_1)$, we report the required event sizes for the conditions of planning values $\theta_0 = 1$, $\theta_1 = 1.8$, and the desired width $w = 1.94$. The empirical probabilities are again close to 80%, showing that the proposed event sizes are sufficient. Note that unequal allocation sometimes reduces the required event size in one group, but overall, the empirical performance remains stable across different allocation ratios. Similar to the one-sample case, more stringent requirements (intersections) lead to larger event sizes.

Table 3 Event sizes and simulation results for the test and CIs in the two-sample problem.

Cases	Event size (n_1, n_2)	Empirical probability (%)
$k = 1$ (equal size)		
1. $P(R)$	(46, 46)	79.82
2. $P(W)$	(42, 42)	80.91
3. $P(W \cap R)$	(53, 53)	80.44
4. $P(W \cap V)$	(43, 43)	80.80
5. $P(W \cap R \cap V)$	(55, 55)	82.33
6. $P(W \mid V)$	(41, 41)	81.29
7. $P(W \cap R \mid V)$	(52, 52)	83.40
8. $P(W \mid R)$	(44, 44)	80.04
9. $P(W \cap V \mid R)$	(46, 46)	84.65
$k = 2$ $(n_2 = 2n_1)$		
1. $P(R)$	(36, 72)	81.13
2. $P(W)$	(31, 62)	80.50
3. $P(W \cap R)$	(41, 82)	81.28
4. $P(W \cap V)$	(32, 64)	80.67
5. $P(W \cap R \cap V)$	(42, 84)	82.60
6. $P(W \mid V)$	(31, 62)	82.18
7. $P(W \cap R \mid V)$	(40, 80)	84.49
8. $P(W \mid R)$	(33, 66)	81.16
9. $P(W \cap V \mid R)$	(34, 68)	83.48

Note: Setting $\alpha = 0.05$, $1 - \beta = 0.8$, the planning values $\theta_0 = 1$, $\theta_1 = 1.8$, and the desired width $w = 1.94$

For prediction intervals of the two-sample problem, given replication event sizes, Table 4 presents the corresponding original event sizes for settings $\alpha = 0.05$, $1 - \gamma = 0.8$, and a desired relative width $r = 1.5$. It shows that larger replication event sizes require correspondingly larger original event sizes to maintain the desired precision in prediction intervals. The simulation results show that the coverage of the ratio of replication estimate rates $\hat{\lambda}_{m_1} / \hat{\lambda}_{m_2}$ can achieve 95% of the time by setting $\alpha = 0.05$, and the empirical percentages of $RW \le r$ can reach the desired probability level of 0.8. That is, the coverage percentages consistently hover around 95%, and the empirical probabilities remain at or above 80%, demonstrating that the procedure is effective in

maintaining reproducibility criteria. Both equal and unequal allocation schemes yield consistent results, highlighting the flexibility of the approach.

Across all four tables, the empirical results validate the theoretical calculations of event size. In practice, this means that researchers can rely on the proposed method to achieve the intended levels of power, confidence interval precision, and prediction interval coverage. The simulations demonstrate that the procedures are not only mathematically sound but also robust under repeated sampling, thereby supporting their use in real-world study planning. The simulations were conducted under specific parameter values (e.g., CI width, relative width, allocation ratios). Results may differ if parameters fall outside the tested ranges.

Table 4 The original event sizes and simulation results for PIs in the two-sample problem

Replication sizes (m_1, m_2)	Original event sizes (n_1, n_2)	Coverage (%)	$RW \le r$ (%)
$k = 1$ (equal sizes)			
(10, 20)	(33, 33)	95.14	80.46
(20, 20)	(43, 43)	94.76	80.89
(30, 20)	(48, 48)	95.04	80.56
(30, 30)	(54, 54)	94.80	79.96
$k = 2$ $(n_2 = 2n_1)$			
(10, 20)	(25, 50)	94.53	80.16
(20, 20)	(32, 64)	95.03	81.46
(30, 20)	(36, 72)	95.20	81.08
(30, 30)	(41, 82)	95.17	80.30

Note: Setting $\alpha = 0.05$, $1-\gamma = 0.8$, and the desired relative width $r = 1.5$

6. Conclusion and Discussion

Poisson distributions are widely used to model rare events such as manufacturing defects or disease occurrences. When event rates are low, large sample sizes are often necessary to achieve statistical significance, which complicates study planning. This paper addresses the determination of event size, rather than sample size, by integrating hypothesis testing and confidence intervals in the context of Poisson rates. We further emphasize the role of prediction intervals in evaluating study reproducibility and advocate for their broader adoption. This holistic framework strengthens statistical inference, improves precision, and contributes to mitigating the replication crisis in scientific research. The three R Shiny applications developed in this study provide researchers with practical tools for ensuring that desired probability levels are met.

Nevertheless, some limitations remain. Accurate estimation of the event rate (λ) can be challenging when prior information is scarce, potentially leading to an inaccurate estimation of the event size. However, the proposed prediction interval approach has the advantage that λ is not required to determine event size. Another limitation is that λ may vary over time, which challenges the assumption of homogeneity. Future research should investigate methods that account for time-varying or heterogeneous event rates and extend the current framework to more complex data structures, thereby enhancing the flexibility and applicability of event size determination.

When planning studies with rare events modeled by Poisson distributions, the determination of event size should reflect both the central research objective and practical considerations. To begin with, if the goal is hypothesis testing, event size should be calculated based on statistical power, so

that there is a high probability of rejecting the null hypothesis when the alternative holds. By contrast, if the focus is estimation, event size should instead be guided by the coverage and width of the confidence interval: narrower intervals yield more precise estimates, yet they also demand larger sample sizes. Moreover, in contexts where reproducibility or external validation is a priority, prediction intervals provide a more appropriate tool. Unlike confidence intervals, they explicitly capture the variability expected in replication studies and, under the proposed method, can be constructed without prior specification of the underlying event rate (λ). In summary, hypothesis testing ensures statistical significance, confidence intervals provide estimation precision, and prediction intervals safeguard reproducibility. Researchers should select the method—or combination of methods—that best fits. In experimental design, researchers must specify essential parameters before data collection begins. These specifications are necessary to calculate the required event size according to the proposed methods. Only after these parameters have been established and the required event size has been determined can the researcher proceed with the experimental work.

Acknowledgments

This research was supported by Grant MOST 110-2410-H-006-034-MY3 from the Ministry of Science and Technology, Taiwan, R.O.C. The authors thank the editor and reviewers for their insightful suggestions.

References

Anderson SF, Kelley K. Sample size planning for replication studies: The devil is in the design. Psychol Methods. 2024; 29(5): 844-867.

Anderson SF, Maxwell SE. There's more than one way to conduct a replication study: beyond statistical significance. Psychol Methods. 2016; 21(1), 1-12.

Beal SL. Sample size determination for confidence intervals on the population mean and on the difference between two population means. Biometrics. 1989; 45(3): 969-977.

Bradley RA, Schumann DEW. The comparison of the sensitivities of similar experiments: applications. Biometrics. 1957; 13(4): 496-510.

Calin-Jageman RJ, Cumming G. The new statistics for better science: Ask how much, how uncertain, and what else is known. Am Stat. 2019; 73(sup1): 271-280.

Cesana BM, Antonelli P. A new approach to sample size calculations for the power of testing and estimating population means of Gaussian distributed variables. Biomed Stat Clin Epidemiol. 2010; 4(2): 67-78.

Chang W, Cheng J, Allaire J, et al. shiny: Web application framework for R. R package version 1.7.4.9002. 2023 [cited 20 June 2025]. Available from: https://shiny.rstudio.com/

Chang C-H, Pal N, Lin J-J. A note on comparing several Poisson means. Commun Stat Simul Comput. 2010; 39(8): 1605-1627.

Chiolero A, Santschi V, Burnand B, Platt RW, Paradis G. Meta-analyses: with confidence or prediction intervals? Eur J Epidemiol. 2012; 27: 823-825.

Cinlar E. Introduction to Stochastic Processes. New Jersey: Prentice-Hall; 1975.

Cox DR. Some simple approximate tests for Poisson variates. Biometrika. 1953; 40(3-4): 354-360.

Cox DR, Lewis PA. The statistical analysis of series of events. London: Methuen; 1966.

Cumming G. Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. Perspect Psychol Sci. 2008; 3(4): 286-300.

Faulkenberry GD. A method of obtaining prediction intervals. J Am Stat Assoc. 1973; 68(342): 433-435.

Graham PL, Mengersen K, Morton AP. Confidence limits for the ratio of two rates based on likelihood scores: non-iterative method. Stat Med. 2003; 22(12): 2071-2083.

Gu K, Ng HK, Tang ML, Schucany WR. Testing the ratio of two Poisson rates. Biom J. 2008; 50(2): 283-298.

Hahn GJ, Meeker WQ. Statistical intervals: a guide for practitioners. Hoboken: John Wiley & Sons; 1991.

Hartnack S, Roos M. Teaching: confidence, prediction and tolerance intervals in scientific practice: a tutorial on binary variables. Emerg Themes Epidemiol. 2021; 18(1): 17. https://doi.org/10.1186/s12982-021-00108-1

Hsu LM. Unbalanced designs to maximize statistical power in psychotherapy efficacy studies. Psychother Res. 1994; 4(2): 95-106.

IntHout J, Ioannidis JP, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open. 2016; 6(7): e010247. https://doi.org/10.1136/bmjopen-2015-010247

Jiroutek MR, Muller KE, Kupper LL, Stewart PW. A new method for choosing sample size for confidence interval-based inferences. Biometrics. 2003; 59(3): 580-590.

Kelley K, Rausch JR. Sample size planning for the standardized mean difference: accuracy in parameter estimation via narrow confidence intervals. Psychol Methods. 2006; 11(4): 363-385.

Kharrati-Kopaei M, Dorosti-Motlagh R. Confidence intervals for the ratio of two independent Poisson rates: parametric bootstrap, modified asymptotic, and approximate-estimate approaches. Stat Methods Med Res. 2020; 29(8): 2140-2150.

Kim T, Lieberman B, Luta G, Peña EA. Prediction regions for Poisson and over-dispersed Poisson regression models with applications in forecasting the number of deaths during the COVID-19 pandemic. Open Stat. 2021; 2(1): 81-112.

Krishnamoorthy K, Lv S. Highest posterior mass prediction intervals for binomial and Poisson distributions. Metrika. 2018; 81(7): 775-796.

Krishnamoorthy K, Peng J. Improved closed-form prediction intervals for binomial and Poisson distributions. J Stat Plan Inference. 2011; 141(5): 1709-1718.

Krishnamoorthy K, Thomson J. A more powerful test for comparing two Poisson means. J Stat Plan Inference. 2004; 119(1): 23-35.

Lai K, Kelley K. Accuracy in parameter estimation for ANCOVA and ANOVA contrasts: sample size planning via narrow confidence intervals. Br J Math Stat Psychol. 2012; 65(2): 350-370.

Li HQ, Tang ML, Wong WK. Confidence intervals for ratio of two Poisson rates using the method of variance estimates recovery. Comput Stat. 2014; 29(3): 869-889.

Liu XS. Sample size and the width of the confidence interval for mean difference. Br J Math Stat Psychol. 2009; 62(Pt 2): 201-215.

Liu XS. Implications of statistical power for confidence intervals. Br J Math Stat Psychol. 2012; 65(3): 427-437.

Luh WM. Probabilistic thinking is the name of the game: integrating test and confidence intervals to plan sample sizes. Methodology. 2022; 18(2): 80-98.

Luh WM, Guo JH. Unequal allocation of sample/event sizes with considerations of sampling cost for testing equality, non-inferiority/superiority, and equivalence of two Poisson rates. Int J Biostat. 2022; 20(1): 143-156.

Maguire BA, Pearson ES, Wynn AHA. The time intervals between industrial accidents. Biometrika. 1952; 39(1-2): 168-180.

Meeker WQ, Hahn GJ. Sample sizes for prediction intervals. J Qual Tech. 1982; 14(4): 201-206.

Nelson LS. Comparison of Poisson means: the general case. J Qual Tech. 1987; 19(4): 173-179.

Nelson W. Confidence intervals for the ratio of two Poisson means and Poisson predictor intervals. IEEE Trans Reliab. 1970; R-19(2): 42-49.

Ng HKT, Gu K, Tang ML. A comparative study of tests for the difference of two Poisson means. Comput Stat Data Anal. 2007; 51(6): 3085-3099.

Pan ZY, Kupper LL. Sample size determination for multiple comparison studies treating confidence interval width as random. Stat Med. 1999; 18(12): 1475-1488.

Patil P, Peng RD, Leek JT. What should researchers expect when they replicate studies? A statistical view of replicability in psychological science. Perspect Psychol Sci. 2016; 11(4): 539-544.

Peckham E, Brabyn S, Cook L, Devlin T, Dumville J, Torgerson DJ. The use of unequal randomisation in clinical trials—An update. Contemp Clin Trials. 2015; 45(Pt A): 113-122.

Preston S. Teaching prediction intervals. J Stat Educ. 2000; 8(3). https://doi.org/10.1080/10691898.2000.12131297

Price RM, Bonett DG. Estimating the ratio of two Poisson rates. Comput Stat Data Anal. 2000; 34(3): 345-356.

R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2024 [cited 2 July 2024]. Available from: http://www.r-project.org

Ross SM. Introduction to probability models. 11th ed. San Diego: Academic Press; 2014.

Sahai H, Khurshid A. Confidence intervals for the ratio of two Poisson means. Math Sci. 1993; 18(1): 43-50.

Schumann DEW, Bradley RA. The comparison of the sensitivities of similar experiments: theory. Ann Math Stat. 1957; 28(4): 902-920.

Schumann DEW, Bradley RA. The comparison of the sensitivities of similar experiments: model II of the analysis of variance. Biometrics. 1959; 15(3): 405-416.

Shan G. Exact sample size determination for the ratio of two incident rates under the Poisson distribution. Comput Stat. 2016; 31(4): 1633-1644.

Shiue WK, Bain LJ. Experiment size and power comparisons for two-sample Poisson tests. JR Stat Soc Ser C Appl Stat. 1982; 31(2): 130-134.

Spence JR, Stanley DJ. Prediction interval: What to expect when you're expecting . . . A replication. PLoS One. 2016; 11(9): e0162874. http://doi.org/10.1371/journal.pone.0162874

Spence JR, Stanley DJ. Tempered expectations: A tutorial for calculating and interpreting prediction intervals in the context of replications. Adv Methods Pract Psychol Sci. 2024; 7(1): 25152459231217932. http://doi.org/10.1177/25152459231217932

Thode HC. Power and sample size requirements for tests of differences between two Poisson rates. Statistician. 1997; 46(2): 227-230.

Wang Y, Kupper LL. Optimal sample sizes for estimating the difference in means between two normal populations treating confidence interval length as a random variable. Commun Stat Theory Methods. 1997; 26(3): 727-741.

Xiao M, Jiang T, Zhang H, Shan G. Exact one-sided confidence limit for the ratio of two Poisson rates. Stat Biopharm Res. 2017; 9(2): 180-185.

Appendix A. Common notation used in this study

Symbol	Definition	Notes / Context
R	The event rejection	Rejection of the null hypothesis
V	The event validity	The coverage of the true parameter by the CI
W	The event width	The obtained CI is to achieve a desired width
λ	True event rate	Rate of Poisson process (events per unit time or space)
n	Event size in the original study	Number of observed events, a more meaningful planning parameter than sample size in Poisson models
t	Observation window (time or space)	Sometimes referred to as the sample size
$\hat{\lambda} = n/t$	Maximum likelihood estimator (MLE) of λ	Used for hypothesis testing, CI, and PI construction
α	Significance level	Probability of type I error (false positive)
β	Type II error probability	$1-\beta$ is the power (true positive probability), a probability of correctly rejecting H_0
CI	Confidence Interval for λ	Measures estimation precision
PI	Prediction Interval for replication $\hat{\lambda}$	Quantifies the expected range of future study results
W_{CI} (W_{PI})	Width of CI (or PI)	A random variable (A smaller value means higher precision)
RW	Relative width	Ratio of PI width to its limiting width
m	Event size in the replication study	Used for prediction interval planning

Appendix B. The various probabilities for the case of one Poisson rate

Consider the event $W \cap V$, from (4) and (5), we have

$$P(W \cap V) = P((\chi^2 \geq w_1) \cap (\chi^2_{2n;\alpha/2} \leq \chi^2 \leq \chi^2_{2n;1-\alpha/2})) = P(\max(w_1, \chi^2_{2n;\alpha/2}) \leq \chi^2 \leq \chi^2_{2n;1-\alpha/2}).$$

For right-tailed tests, from (1) and (5),

$$P(W \cap R) = P\left((\chi^2 \geq w_1) \cap \left(\chi^2 \leq \frac{\lambda_1}{\lambda_0}\chi^2_{2n;\alpha}\right)\right) = P\left(w_1 \leq \chi^2 \leq \frac{\lambda_1}{\lambda_0}\chi^2_{2n;\alpha}\right).$$

For left-tailed tests, from (2) and (5),

$$P(W \cap R) = P\left((\chi^2 \geq w_1) \cap \left(\chi^2 \geq \frac{\lambda_1}{\lambda_0}\chi^2_{2n;1-\alpha}\right)\right) = P\left(\chi^2 \geq \max\left(w_1, \frac{\lambda_1}{\lambda_0}\chi^2_{2n;1-\alpha}\right)\right).$$

For two-tailed tests and two-sided CIs, from (3) and (5),

$$P(W \cap R) = P\left((\chi^2 \geq w_1) \cap \left(\left(\chi^2 \leq \frac{\lambda_1}{\lambda_0}\chi^2_{2n;\alpha/2}\right) \cup \left(\chi^2 \geq \frac{\lambda_1}{\lambda_0}\chi^2_{2n;1-\alpha/2}\right)\right)\right),$$

$$= P\left(\left(w_1 \leq \chi^2 \leq \frac{\lambda_1}{\lambda_0}\chi^2_{2n;\alpha/2}\right) \cup \left(\chi^2 \geq \max\left(w_1, \frac{\lambda_1}{\lambda_0}\chi^2_{2n;1-\alpha/2}\right)\right)\right).$$

Consider the event $W \cap V \cap R$, for right-tailed tests and two-sided CIs, from (1), (4), and (5),

$$P(W \cap V \cap R) = P\left((\chi^2 \geq w_1) \cap (\chi^2_{2n;\alpha/2} \leq \chi^2 \leq \chi^2_{2n;1-\alpha/2}) \cap \left(\chi^2 \leq \frac{\lambda_1}{\lambda_0} \chi^2_{2n;\alpha} \right) \right),$$

$$= P\left(\max(w_1, \chi^2_{2n;\alpha/2}) \leq \chi^2 \leq \min\left(\chi^2_{2n;1-\alpha/2}, \frac{\lambda_1}{\lambda_0} \chi^2_{2n;\alpha} \right) \right).$$

For left-tailed tests and two-sided CIs, from (2), (4), and (5),

$$P(W \cap V \cap R) = P\left((\chi^2 \geq w_1) \cap (\chi^2_{2n;\alpha/2} \leq \chi^2 \leq \chi^2_{2n;1-\alpha/2}) \cap \left(\chi^2 \geq \frac{\lambda_1}{\lambda_0} \chi^2_{2n;1-\alpha} \right) \right),$$

$$= P\left(\max\left(w_1, \chi^2_{2n;\alpha/2}, \frac{\lambda_1}{\lambda_0} \chi^2_{2n;1-\alpha} \right) \leq \chi^2 \leq \chi^2_{2n;1-\alpha/2} \right).$$

For two-tailed tests and two-sided CIs, from (3), (4), and (5),

$$P(W \cap V \cap R) = P\left((\chi^2 \geq w_1) \cap (\chi^2_{2n;\alpha/2} \leq \chi^2 \leq \chi^2_{2n;1-\alpha/2}) \cap \left(\left(\chi^2 \leq \frac{\lambda_1}{\lambda_0} \chi^2_{2n;\alpha/2} \right) \cup \left(\chi^2 \geq \frac{\lambda_1}{\lambda_0} \chi^2_{2n;1-\alpha/2} \right) \right) \right),$$

$$= P\left(\begin{array}{l} \left(\max(w_1, \chi^2_{2n;\alpha/2}) \leq \chi^2 \leq \min\left(\chi^2_{2n;1-\alpha/2}, \frac{\lambda_1}{\lambda_0} \chi^2_{2n;\alpha/2} \right) \right) \\ \cup \left(\max\left(w_1, \chi^2_{2n;\alpha/2}, \frac{\lambda_1}{\lambda_0} \chi^2_{2n;1-\alpha/2} \right) \leq \chi^2 \leq \chi^2_{2n;1-\alpha/2} \right) \end{array} \right).$$

Finally, conditional probabilities are expressed as follows:

$$P(W|V) = P(W \cap V)/(1-\alpha),\ P(W \cap R|V) = P(W \cap R \cap V)/(1-\alpha),$$

$$P(W|R) = P(W \cap R)/P(R),\ \textit{and}\ P(W \cap V|R) = P(W \cap R \cap V)/P(R).$$

Appendix C. The various probabilities for the case of two Poisson rates

Consider the event $W \cap V$, from (9) and (10), we have

$$P(W \cap V) = P((F \geq w_2) \cap (F_{2n_1,2n_2;\alpha/2} \leq F \leq F_{2n_1,2n_2;1-\alpha/2})),$$

$$= P(\max(w_2, F_{2n_1,2n_2;\alpha/2}) \leq F \leq F_{2n_1,2n_2;1-\alpha/2}).$$

For right-tailed tests, from (6) and (10)

$$P(W \cap R) = P\left((F \geq w_2) \cap \left(F \leq \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;\alpha/2} \right) \right) = P\left(w_2 \leq F \leq \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;\alpha/2} \right).$$

For left-tailed tests, from (7) and (10)

$$P(W \cap R) = P\left((F \geq w_2) \cap \left(F \geq \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;1-\alpha} \right) \right) = P\left(F \geq \max\left(w_2, \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;1-\alpha} \right) \right).$$

For two-tailed tests and two-sided CIs, from (8) and (10),

$$P(W \cap R) = P\left((F \geq w_2) \cap \left(\left(F \leq \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;\alpha/2} \right) \cup \left(F \geq \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;1-\alpha/2} \right) \right) \right),$$

$$= P\left(\left(w_2 \leq F \leq \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;\alpha/2} \right) \cup \left(F \geq \max\left(w_2, \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;1-\alpha/2} \right) \right) \right).$$

To consider the event $W \cap V \cap R$, for right-tailed tests and tow-sided CIs, from (6), (9), and (10),

$$P(W \cap V \cap R) = P\left((F \geq w_2) \cap (F_{2n_1,2n_2;\alpha/2} \leq F \leq F_{2n_1,2n_2;1-\alpha/2}) \cap \left(F \leq \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;\alpha} \right) \right),$$

$$= P\left(\max(w_2, F_{2n_1,2n_2;\alpha/2}) \leq F \leq \min\left(F_{2n_1,2n_2;1-\alpha/2}, \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;\alpha} \right) \right).$$

For left-tailed tests and two-sided CIs, from (7), (9), and (10),

$$P(W \cap V \cap R) = P\left((F \geq w_2) \cap (F_{2n_1,2n_2;\alpha/2} \leq F \leq F_{2n_1,2n_2;1-\alpha/2}) \cap \left(F \geq \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;1-\alpha} \right) \right),$$

$$= P\left(\max\left(w_2, F_{2n_1,2n_2;\alpha/2}, \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;1-\alpha} \right) \leq F \leq F_{2n_1,2n_2;1-\alpha/2} \right).$$

For two-tailed tests and two-sided CIs, from (8), (9), and (10),

$$P(W \cap V \cap R) = P\left((F \geq w_2) \cap (F_{2n_1,2n_2;\alpha/2} \leq F \leq F_{2n_1,2n_2;1-\alpha/2}) \cap [(F \leq \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;\alpha/2}) \cup (F \geq \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;1-\alpha/2})] \right),$$

$$= P\left(\begin{array}{l} \left(\max(w_2, F_{2n_1,2n_2;\alpha/2}) \leq F \leq \min\left(F_{2n_1,2n_2;1-\alpha/2}, \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;\alpha/2} \right) \right) \\ \cup \left(\max\left(w_2, F_{2n_1,2n_2;\alpha/2}, \frac{\theta_1}{\theta_0} F_{2n_1,2n_2;1-\alpha/2} \right) \leq F \leq F_{2n_1,2n_2;1-\alpha/2} \right) \end{array} \right).$$

Finally, other conditional probabilities can be found as follows:

$$P(W \mid V) = P(W \cap V) / (1-\alpha),\ P(W \cap R \mid V) = P(W \cap R \cap V) / (1-\alpha),$$

$$P(W \mid R) = P(W \cap R) / P(R),\ \textit{and}\ P(W \cap V \mid R) = P(W \cap R \cap V) / P(R).$$