

## A BAYESIAN FRAMEWORK FOR THE RATIO OF TWO POISSON RATES IN THE CONTEXT OF VACCINE EFFICACY TRIALS \*

STÉPHANE LAURENT,<sup>1</sup> AND CATHERINE LEGRAND<sup>1</sup>

**Abstract.** In many applications, we assume that two random observations  $x$  and  $y$  are generated according to independent Poisson distributions  $\mathcal{P}(\lambda S)$  and  $\mathcal{P}(\mu T)$  and we are interested in performing statistical inference on the ratio  $\phi = \lambda/\mu$  of the two incidence rates. In vaccine efficacy trials,  $x$  and  $y$  are typically the numbers of cases in the vaccine and the control groups respectively,  $\phi$  is called the relative risk and the statistical model is called ‘partial immunity model’. In this paper we start by defining a natural semi-conjugate family of prior distributions for this model, allowing straightforward computation of the posterior inference. Following theory on reference priors, we define the reference prior for the partial immunity model when  $\phi$  is the parameter of interest. We also define a family of reference priors with partial information on  $\mu$  while remaining uninformative about  $\phi$ . We notice that these priors belong to the semi-conjugate family. We then demonstrate using numerical examples that Bayesian credible intervals for  $\phi$  enjoy attractive frequentist properties when using reference priors, a typical property of reference priors.

**1991 Mathematics Subject Classification.** 62F15,62F03,62F25,62P10.

June 11, 2010.

### INTRODUCTION

Inference on the ratio of two Poisson rates has broad applications in: agriculture, ecology, medicine, quality control, physics; see [16, 19, 20, 22, 23, 27, 36] and the references given therein. The statistical model is given by two count observations  $x$  and  $y$  assumed to be realizations from two independent random variables distributed according to Poisson distributions  $\mathcal{P}(\lambda S)$  and  $\mathcal{P}(\mu T)$  respectively. The *incidence rates* are the two unknown parameters  $\lambda$  and  $\mu$ , and the *‘observation-opportunity sizes’* or *‘sample sizes’* are the fixed numbers  $S$  and  $T$ , which could represent for example a time-period or an area. The problem of interest is to perform statistical inference on the ratio  $\phi = \lambda/\mu$ , typically termed as the *relative risk*.

Throughout this paper, we will work in the context of *Phase III vaccine efficacy trials*. These trials are typically very large, often enrolling thousands of patients, due to low incidence of the disease of interest. These patients are randomly divided into two groups, the vaccine group and the control group (see figure 1), and followed for the incidence of the particular disease of interest. The statistical model defined above is then commonly used, in which context it is called the *partial immunity model*. The counts  $x$  and  $y$  are the numbers of patients who developed the disease (shorter, the numbers of ‘cases’). The observation-opportunity sizes  $S$  and  $T$  are *person times at risk*, defined in each group as the sum of the follow-up periods for all individuals belonging to this

---

*Keywords and phrases:* Poisson rates, relative risk, vaccine efficacy, partial immunity model, semi-conjugate family, reference prior, Jeffreys’ prior, frequentist coverage, Beta prime distribution, Beta-negative binomial distribution.

\* *Research supported by IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy).*

<sup>1</sup> Institut de Statistique, Biostatistique et Sciences Actuarielles (ISBA). Université catholique de Louvain, Voie du Roman Pays, 20, 1348 Louvain la Neuve, Belgium

group. The disease incidence rates are also called *disease attack rates*, or shorter, *attack rates*. The *vaccine efficacy parameter* is defined as  $VE = 1 - \phi$ .

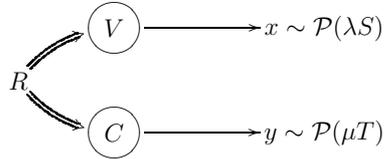


FIGURE 1. Partial immunity model

One of the main objectives of this paper is to investigate an ‘objective Bayesian’ statistical inference on  $\phi$  when considering *reference priors*, introduced by Bernardo [10] and further developed by Berger & Bernardo [3–6]. The objective approach consider an a priori law on the parameters which is said to be ‘uninformative’, aiming to yield a posterior which essentially reflects the information brought by the data. The required conditions for such a law were discovered along with emerging flaws and paradoxes raised by candidate a priori laws. The *reference a priori law*, whose popularity is increasing in theoretical and applied statistics, is nowadays recognized as the only uninformative prior satisfying all these required conditions (see [13]).

We will discuss frequentist properties of Bayesian credible intervals for  $\phi$  when using these kinds of priors. There are several known frequentist confidence intervals for the ratio or the difference of two Poisson rates; see for example [20, 22, 27, 32, 35]. A common way is to perform a conditional frequentist inference about the relative risk  $\phi$  is by conditioning on the sum  $x + y$  of the two counts. The conditional distribution of  $x$  given  $x + y$  is binomial with size parameter  $n = x + y$  and proportion parameter

$$p = \frac{\lambda S}{\lambda S + \mu T} = \frac{\phi \frac{S}{T}}{\phi \frac{S}{T} + 1}, \quad (1)$$

which does not involve the nuisance parameter  $\mu$  and is a one-to-one function of  $\phi$ . Thus, conditional statistical inference on  $\phi$  can be derived from the inference on the proportion parameter in the well-known binomial model. This conditional method is appealing as it eliminates the nuisance parameter  $\mu$ . A noteworthy discussion on conditional inference is provided in [34], giving several other arguments for claiming that conditional inference is appealing in frequentist statistics. However, under the unconditional distributional assumption, the conditional confidence intervals are too conservative and the conditional hypothesis tests achieve lower power than unconditional tests. We note this quite interesting property<sup>1</sup> for the partial immunity model: *considering either the unconditional model or the conditional model, the reference posteriors on  $\phi$  are the same*. Indeed we know (see for example [15] or [9]) that the reference posterior on the proportion parameter for the binomial model with size  $n = x + y$  after  $x$  has been observed equals Jeffreys’ posterior which is the Beta distribution  $\mathcal{B}(x + \frac{1}{2}, y + \frac{1}{2})$ , and we shall see that this is also the marginal reference posterior on  $p$  defined by (1) for the partial immunity model after  $x$  and  $y$  have been observed. This fact has a quite attractive consequence for the frequentist performance of the reference posterior inference. Indeed, it is known that credible intervals based on reference posteriors typically satisfy a ‘*frequentist-matching property*’: even for a moderate sample size, a 95% posterior credibility interval is also, roughly, a 95% confidence interval in the frequentist sense. Consequently, the credible intervals based on the reference posterior on  $\phi$  for the partial immunity model satisfy a “simultaneous” frequentist-matching property, considering either the unconditional ‘*two Poisson samples*’ model or the conditional ‘*one binomial sample*’ model. A remarkable frequentist performance of credible intervals based on the Jeffreys posterior for the binomial model is demonstrated in [17, 18]. When using the reference prior on  $\phi$  for the partial immunity model, we will see on numerical examples that posterior credible intervals for  $\phi$

<sup>1</sup>This fact is not expected as a consequence of the general properties of reference priors (we thank J. Berger for this remark).

enjoy an attractive unconditional frequentist-matching performance. Investigations of some possible theoretical frequentist-matching property (see [21]) is however not in the scope of this paper.

The main topics addressed in this paper are summarized as follows:

- *Semi-conjugate family* — Typical statistical parametric models admit a natural conjugate family of prior distributions (see for example [33]). We shall define a family of prior distributions with a conjugacy-like property, usually called *semi-conjugacy* property. These distributions are defined through standard distributions, such as Gamma or Beta, so that the Bayesian posterior inference can be handled in any standard statistical software. When we are interested in  $\phi$ , we will see (section 1.4) that this family of priors has a better interpretation as compared to the natural conjugate family, given by two independent Gamma distributions on  $\mu$  and  $\lambda$ .
- *Reference prior* — We shall derive the ‘ $\phi$ -reference prior’ for the partial immunity model, that is, the reference prior in the case of  $\phi$  is the quantity of interest. Indeed, for a parametric model with two parameters, the definition of the reference prior pertains to the choice of a particular quantity of interest. This  $\phi$ -reference prior will be short termed as reference prior because it is understood that we are interested in  $\phi$  throughout the whole paper. It happens that it belongs to the semi-conjugate family which will be defined. We will be particularly interested in the frequentist properties of the usual Bayesian posterior credible intervals for  $\phi$  when making use of the reference prior.
- *‘Semi-reference’ prior* — Instead of a ‘fully non-informative’ prior, one may be interested in a prior which represents some personal beliefs about the attack rate  $\mu$  while remaining non-informative about  $\phi$ . We will derive a family of such kind of priors contained in the semi-conjugate family. We follow a method proposed by Berger and Sun [8] which, as said by the authors, closely mirrors the underlying motivation of reference priors. We will then particularly focus on the comparison of the Bayesian inference when making use of the reference prior and the ‘semi-reference priors’.

This paper is organized as follows. In section 1, we introduce some notations and define a natural semi-conjugate family for the partial immunity model, which contains the reference prior and the family of semi-reference priors. Section 2 discusses the posterior inference using these priors. We end this paper by illustrating these methods in the setting of vaccine efficacy trials in section 3.

## 1. SEMI-CONJUGATE FAMILY FOR THE PARTIAL IMMUNITY MODEL

All notations around the partial immunity model are given in section 1.1. In subsection 1.2, we define the so-called semi-conjugate family of prior distributions. In subsection 1.3, we derive a ‘noninformative prior’ and a family of ‘semi-informative’ priors for the relative risk  $\phi$ , following original ideas of Bernardo (1979). We will see that they belong to the semi-conjugate family. In subsection 1.4, we sum up the results and discuss non-informative priors as well as the advantages of this semi-conjugate family of priors as compared to the conjugate family.

### 1.1. Notations

The numbers of cases  $x$  and  $y$  are assumed to be generated according to independent Poisson distributions  $\mathcal{P}(\mu T)$  and  $\mathcal{P}(\lambda S)$  respectively, and  $\phi = \lambda/\mu$  is the relative risk. The person times at risk  $S$  and  $T$  are fixed positive numbers, which will be omitted in the notations when it is clear from the context that they are fixed. It is understood that the model is parameterized with  $\phi$  (parameter of interest) and  $\mu$  (nuisance parameter). For given values of  $\mu$  and  $\phi$ , the notation  $p(x, y | \mu, \phi, S, T)$  or  $p(x, y | \mu, \phi)$  denotes the probability to observe  $x$  and  $y$ , as well as the probability distribution, so called *sampling distribution*, of  $x$  and  $y$ . The parametric model  $\mathcal{M} = \{p(x, y | \mu, \phi); \mu > 0, \phi \geq 0\}$  is the family of sampling distributions indexed by the unknown parameters  $\mu$  and  $\phi$ .

As usual in Bayesian statistics, the notation  $\pi(\cdot)$  is used as generic notation for probabilities on the parameters space or their densities. For example, the notation  $\pi(\mu)$  denotes the density of the prior distribution on  $\mu$ , the

notation  $\pi(\mu > 1)$  denotes the prior probability that  $\mu$  is greater than 1, the notation  $\pi(\mu | \phi)$  denotes the conditional prior distribution on  $\mu$  given  $\phi$ . In the same way, the notation  $\pi(\cdot | x, y)$  is used as generic notation for posterior probabilities or their densities.

## 1.2. Semi-conjugate family

The family of joint priors  $\pi(\mu, \phi)$  we define is called *semi-conjugate* in the sense that the conditional posterior  $\pi(\mu | \phi, x, y)$  on  $\mu$  given  $\phi$  belongs to the same family of distributions as the conditional prior  $\pi(\mu | \phi)$  on  $\mu$  given  $\phi$ , and the marginal posterior  $\pi(\phi | x, y)$  on  $\phi$  belongs to the same family as the marginal prior  $\pi(\phi)$  on  $\phi$ .

To define this family, we will need the *Beta prime* distribution  $\mathcal{B}'(c, d)$  with shape parameters  $c, d > 0$ . This distribution is defined as the distribution of the random variable  $Z := \frac{U}{1-U}$  where  $U$  is a random variable distributed according to the Beta distribution  $\mathcal{B}(c, d)$ . Note that the distribution of  $1/Z$  is then  $\mathcal{B}'(d, c)$ . More details on this distribution are provided in Appendix A.

As shown in the remainder of this section, the following result defines a semi-conjugate family for the partial immunity model:

*For any positive numbers  $a, b, c, d$ , if the joint prior on  $(\mu, \phi)$  is given by the marginal-conditional factorization  $\pi(\mu, \phi) = \pi(\mu | \phi)\pi(\phi)$  with*

$$(\mu | \phi) \sim \mathcal{G}(a, b) \quad \text{and} \quad \phi \sim \frac{T+b}{S} \times \mathcal{B}'(c, d),$$

*then the joint posterior on  $(\mu, \phi)$  is given by the marginal-conditional factorization  $\pi(\mu, \phi | x, y) = \pi(\mu | \phi, x, y)\pi(\phi | x, y)$  with*

$$(\mu | \phi, x, y) \sim \mathcal{G}(x + y + a, \phi S + T + b) \quad \text{and} \quad (\phi | x, y) \sim \frac{T+b}{S} \times \mathcal{B}'(x + c, y + a + d).$$

This result can be proved by using expression of the Gamma and Beta prime density functions. However, a more intuitive proof can be based on well known results for the Poisson model and the negative binomial model (see Appendix B and C).

From now on we assume that  $\pi(\mu | \phi) = \mathcal{G}(\mu | a, b)$  and  $\pi(\phi) = \frac{T+b}{S} \times \mathcal{B}'(\phi | c, d)$ . To derive the posterior, we introduce the likelihood  $L(\mu, \phi | x, y) = p(x, y | \mu, \phi)$ , which for our parametric model can be written

$$L(\mu, \phi | x, y) \underset{\mu, \phi}{\propto} \phi^x e^{-(\phi S + T)\mu} \mu^{x+y}, \quad (2)$$

where the symbol " $\underset{\mu, \phi}{\propto}$ " means that the two members are proportional functions of  $(\mu, \phi)$ .

Bayes' formula can be written as

$$\pi(\mu | \phi, x, y)\pi(\phi | x, y) \underset{\mu, \phi}{\propto} L(\mu, \phi | x, y)\pi(\mu | \phi)\pi(\phi), \quad (3)$$

from which follows the conditional Bayes' formula

$$\begin{aligned} \pi(\mu | \phi, x, y) &\underset{\mu}{\propto} L(\mu, \phi | x, y)\pi(\mu | \phi) \\ &\underset{\mu}{\propto} e^{-(\phi S + T)\mu} \mu^{x+y} \mathcal{G}(\mu | a, b), \end{aligned}$$

and from which it is straightforward to check that  $\pi(\mu | \phi, x, y) = \mathcal{G}(\mu | x + y + a, \phi S + T + b)$ .

By integrating (3) with respect to  $\mu$ , we have

$$\pi(\phi | x, y) \propto_{\phi} \tilde{L}(\phi | x, y) \pi(\phi) \quad (4)$$

where the integrated likelihood

$$\tilde{L}(\phi | x, y) = \int L(\mu, \phi | x, y) \pi(\mu | \phi) d\mu$$

is the likelihood on  $\phi$  for the one-parameter model  $\tilde{\mathcal{M}} = \{\tilde{p}_{a,b}(x, y | \phi); \phi \geq 0\}$  whose sampling distribution is  $\tilde{p}_{a,b}(x, y | \phi) = \int p(x, y | \mu, \phi) \pi(\mu | \phi) d\mu$ , hereafter called the *marginal sampling distribution*.

With the help of the  $T \star \mathcal{P}\mathcal{G}$  distribution defined in Appendix B, we now establish that the marginal sampling distribution is the bivariate Poisson-Gamma (or negative binomial) distribution defined as follows. One has

$$(\mu | \phi) \sim \mathcal{G}(a, b) \quad \text{and} \quad (y | \phi, \mu) \sim \mathcal{P}(\mu T), \quad \text{therefore} \quad (y | \phi) \sim T \star \mathcal{P}\mathcal{G}(a, b),$$

and in the same way

$$(\mu | \phi) \sim \mathcal{G}(a, b) \quad \text{and} \quad (x | \phi, \mu) \sim \mathcal{P}(\mu \phi S), \quad \text{therefore} \quad (x | \phi) \sim \phi S \star \mathcal{P}\mathcal{G}(a, b).$$

Thus both margins  $\tilde{p}_{a,b}(x | \phi)$  and  $\tilde{p}_{a,b}(y | \phi)$  of  $\tilde{p}_{a,b}(x, y | \phi)$  are determined. The conditional distribution  $\tilde{p}_{a,b}(x | y, \phi)$  is determined in the same way. Indeed, we know from the conjugacy relation (12) for the Poisson model that  $\pi(\mu | y, \phi) = \mathcal{G}(\mu | y + a, T + b)$ . As  $p(x | y, \phi, \mu) = \mathcal{P}(x | \mu \phi S)$ , one has therefore

$$(x | y, \phi) \sim \phi S \star \mathcal{P}\mathcal{G}(y + a, T + b).$$

Thus, the joint marginal distribution is now given by  $\tilde{p}_{a,b}(x, y | \phi) = \tilde{p}_{a,b}(x | y, \phi) \tilde{p}_{a,b}(y | \phi)$ .

We still have to determine  $\pi(\phi | x, y)$ . As  $\tilde{p}_{a,b}(y | \phi) = T \star \mathcal{P}\mathcal{G}(y | a, b)$  does not involve  $\phi$ , one has then

$$\tilde{L}(\phi | x, y) \propto_{\phi} \tilde{p}_{a,b}(x | y, \phi), \quad (5)$$

and we have seen that  $\tilde{p}_{a,b}(x | y, \phi) = \phi S \star \mathcal{P}\mathcal{G}(x | y + a, T + b)$ . From (4) and (5), we can compute the expression of  $\pi(\phi | x, y)$  using the expression of the density  $\pi(\phi)$  provided in Appendix A and the expression of  $\tilde{p}_{a,b}(x | y, \phi)$  provided in Appendix B. If we want to avoid these computations, another possibility is to proceed as follows. From (5) and (4), the posterior  $\pi(\phi | x, y)$  on  $\phi$  for the marginal model  $\tilde{\mathcal{M}}_{a,b} = \{\tilde{p}_{a,b}(x, y | \phi); \phi \geq 0\}$  is the same as the posterior on  $\phi$  for the negative binomial model  $\{\tilde{p}_{a,b}(x | y, \phi); \phi \geq 0\}$ . It is well known that the Beta distributions form a conjugate family of prior distributions on the proportion parameter of a negative binomial model, with conjugacy relation (13) provided in Appendix C. Now,  $\phi S \star \mathcal{P}\mathcal{G}(y + a, T + b)$  is a negative binomial distribution with size parameter  $y + a$  and proportion parameter  $\theta$  given by (see Appendix C):

$$\theta = \frac{\frac{T + b}{\phi S}}{1 + \frac{T + b}{\phi S}} = \frac{T + b}{\phi S + T + b}. \quad (6)$$

The  $\mathcal{B}(d, c)$  distribution on  $\theta$  induces the Beta prime distribution  $\mathcal{B}'(d, c)$  on  $\frac{T + b}{\phi S}$ , which induces the  $\mathcal{B}'(c, d)$  distribution on  $\frac{\phi S}{T + b}$ , and we know from the conjugacy relation (13) that if  $\theta \sim \mathcal{B}(d, c)$  then  $(\theta | x, y) \sim \mathcal{B}(y + a + d, x + c)$ . In the same way the  $\mathcal{B}(y + a + d, x + c)$  distribution on  $\theta$  induces the  $\mathcal{B}'(x + c, y + a + d)$

distribution on  $\frac{\phi S}{T+b}$ . This terminates to prove our result on the semi-conjugate family for the partial immunity model.

### 1.3. Semi-informative prior and non-informative prior

Following original ideas of Bernardo [10] and further developments by Berger & Bernardo [3–6], we derive the ‘reference’ prior, and a family of ‘semi-reference’ priors when considering  $\phi$  as the quantity of interest. The reference prior is described as ‘noninformative’ whereas the semi-reference priors are described as ‘semi-informative’, as it reflects some information about  $\mu$  while remaining noninformative about  $\phi$ . The semi-informative case is of particular interest for the partial immunity model; indeed, most often, some reliable information is available on the theoretical incidence rate  $\mu$  in the unvaccinated population.

The construction of the semi-reference priors involves the well-known Jeffreys prior for one-parameter models, defined as the square root of the Fisher information. Although the formal definition of reference priors is recent [7], it was clear since the earlier work of Bernardo [10] that for a one-parameter model, under appropriate regularity conditions, the reference prior coincides with the Jeffreys prior.

**Semi-informative prior** — The semi-informative approach consists in first choosing subjectively any probability distribution for the conditional prior  $\pi(\mu | \phi)$  of  $\mu$  given  $\phi$ . Then we define the one-parameter marginal model  $\widetilde{\mathcal{M}} = \{\widetilde{p}(x, y | \phi); \phi > 0\}$  with sampling distribution  $\widetilde{p}(x, y | \phi) = \int p(x, y | \mu, \phi) \pi(\mu | \phi) d\mu$ , and we take as marginal prior on  $\phi$  the reference prior  $\widetilde{\pi}^{\text{ref}}(\phi)$  on  $\phi$  for  $\widetilde{\mathcal{M}}$ . Finally the joint prior on  $\mu$  and  $\phi$  is defined by  $\pi^{1/2\text{ref}}(\mu, \phi) = \pi(\mu | \phi) \widetilde{\pi}^{\text{ref}}(\phi)$ .

We use  $\pi(\mu | \phi) \equiv \pi(\mu) = \mathcal{G}(\mu | a, b)$ , as in the semi-conjugate family defined in the preceding section. In the context of the partial immunity model, it is natural that  $\phi$  does not pertain to the information available on  $\mu$ , and thus to set  $\pi(\mu | \phi) \equiv \pi(\mu)$ . As the marginal model is then defined through  $a$  and  $b$ , we denote it  $\widetilde{\mathcal{M}}_{a,b}$ , we denote by  $\widetilde{p}_{a,b}(x, y | \phi)$  the marginal sampling distribution, and we use  $\pi_{a,b}^{1/2\text{ref}}$  as the generic notation for prior and posterior distributions.

We now derive the semi-informative prior  $\pi_{a,b}^{1/2\text{ref}}$  following the method described above. We have seen in subsection 1.2 that the marginal sampling distribution  $\widetilde{p}_{a,b}(x, y | \phi)$  is the bivariate Poisson-Gamma distribution given by the marginal-conditional decomposition

$$(y | \phi) \sim T \star \mathcal{PG}(a, b) \quad \text{and} \quad (x | y, \phi) \sim \phi S \star \mathcal{PG}(y + a, T + b). \quad (7)$$

With the help of known results about the negative binomial model (Appendix C), we will show that the reference prior on  $\phi$  for the marginal model  $\widetilde{\mathcal{M}}_{a,b} = \{\widetilde{p}_{a,b}(x, y | \phi); \phi \geq 0\}$  is the improper scaled Beta prime distribution  $\frac{T+b}{S} \times \mathcal{B}'(\frac{1}{2}, 0)$ . To do so, we first note that the marginal distribution  $\widetilde{p}_{a,b}(y | \phi)$  of  $y$  does not involve  $\phi$ . Consequently, the model  $\widetilde{\mathcal{M}}_{a,b}$  inherits the regularity conditions from the negative binomial model  $\{\widetilde{p}_{a,b}(x | y, \phi); \phi \geq 0\}$  with known size parameter  $a + y$  and unknown proportion parameter  $\theta := \frac{T+b}{\phi S + T + b}$ . Therefore the reference prior for  $\widetilde{\mathcal{M}}_{a,b}$  equals the Jeffreys prior, defined as the square root of the Fisher information. Then it appears that Fisher’s information of the model  $\widetilde{\mathcal{M}}_{a,b}$  is a weighted mean over  $y \in \mathbb{N}$  of Fisher’s informations of the models  $\{\widetilde{p}_{a,b}(x | y, \phi); \phi > 0\}$ . It is known (see Appendix C) that the Jeffreys prior on the proportion parameter  $\theta$  in the negative binomial model is the improper  $\mathcal{B}(0, \frac{1}{2})$  distribution, independently of the size parameter. Consequently, the Jeffreys prior on  $\theta$  for  $\widetilde{\mathcal{M}}_{a,b}$  is also the improper  $\mathcal{B}(0, \frac{1}{2})$  distribution. It induces the improper  $\frac{T+b}{S} \times \mathcal{B}'(\frac{1}{2}, 0)$  distribution on  $\phi$ , corresponding to  $c = \frac{1}{2}$  and  $d = 0$  in the semi-conjugate family we have defined in subsection 1.2.

**Noninformative prior** — The noninformative approach consists of deriving the  $\phi$ -reference prior, shorter termed as *reference prior* (see the introduction). The underlying idea in the definition of the  $\phi$ -reference prior is to adapt the construction of the semi-reference prior as above but when at first step, instead of selecting

a subjective conditional prior  $\pi(\mu \mid \phi)$  on  $\mu$  given  $\phi$ , we set it to be reference prior  $\pi^{\text{ref}}(\mu \mid \phi)$  for the one-parameter conditional model  $\mathcal{M}_\phi = \{p(x, y \mid \mu, \phi); \mu > 0\}$ . In case of the partial immunity model,  $\pi^{\text{ref}}(\mu \mid \phi)$  is the Jeffreys prior for the Poisson model; indeed, we see that for a fixed value of  $\phi$ , the likelihood given by (2) is a function of  $\mu$  proportional to  $e^{-(\phi S + T)\mu} \mu^{x+y}$ , thus  $x + y$  is a sufficient statistic for  $\mathcal{M}_\phi$  and one has  $x + y \sim \mathcal{P}((\phi S + T)\mu)$ . It is well known that the reference prior on the rate parameter of the Poisson model is the improper Jeffreys prior  $\mathcal{G}(\frac{1}{2}, 0)$  (see Appendix B). Thus, we could *heuristically* claim that the reference prior  $\pi^{\text{ref}}(\mu, \phi)$  is the semi-reference prior in case when  $a = \frac{1}{2}$  and  $b = 0$ , that is, the distribution of the semi-conjugate family with parameters  $a = c = \frac{1}{2}$  and  $b = d = 0$ . This approach is only heuristic as the construction for the semi-reference prior is not valid in this case, because  $\mathcal{G}(\frac{1}{2}, 0)$  is not a probability distribution and consequently the integration  $\int p(x, y \mid \mu, \phi) \pi^{\text{ref}}(\mu \mid \phi) d\mu$  does not provide a valid statistical model. Nevertheless, it yields the correct reference prior.

Further developments of the pioneering ideas of Bernardo provide a general definition of the reference prior. The reference prior for the partial immunity model has been derived by Liseo [31], who followed the algorithm given by Berger and Bernardo in [3] (see also Berger and Bernardo [5, 6]), and is given in the *catalog of noninformative priors* [9] by Berger and Yang. As a result, the reference prior is indeed the semi-reference prior in case of  $a = \frac{1}{2}$  and  $b = 0$ . Another way to derive the reference prior for the partial immunity model is to use the *theorem on reference prior under factorization* originally established [12] (see also [13]). This theorem is based on the fact that for any typically regular model with two parameters  $\mu$  and  $\phi$  and any suitable prior distribution, the joint posterior distribution on  $\mu$  and  $\phi$  is asymptotically close to a normal distribution with covariance matrix  $I^{-1}(\hat{\mu}, \hat{\phi})$  where  $(\hat{\mu}, \hat{\phi})$  is the maximum likelihood estimation and  $I$  is the Fisher information matrix of the model (see [15]: Section 5.3.). In case of the partial immunity model  $\mathcal{M} = \{p(x, y \mid \mu, \phi); \mu > 0, \phi > 0\}$ , we can check that the inverse of the Fisher information matrix  $I(\mu, \phi)$  satisfies

$$I(\mu, \phi)^{-1} \underset{\mu, \phi}{\propto} \begin{pmatrix} \mu^{-1} \phi \left( \phi + \frac{T}{S} \right) & -\phi \\ -\phi & \mu \end{pmatrix}.$$

Thus the conditions required by the theorem on reference prior under factorization (as stated in Bernardo [13]: Theorem 12) are fulfilled. This theorem then asserts that  $\pi^{\text{ref}}(\mu \mid \phi) \propto \mu^{-\frac{1}{2}}$  and  $\pi^{\text{ref}}(\phi) \propto \phi^{-\frac{1}{2}} \left( \phi + \frac{T}{S} \right)^{-\frac{1}{2}}$ , corresponding to  $a = c = \frac{1}{2}$  and  $b = d = 0$  in the semi-conjugate family defined in subsection 1.2. Thus

$$\pi^{\text{ref}}(\phi \mid x, y) = \frac{T}{S} \times \mathcal{B}' \left( \phi \mid x + \frac{1}{2}, y + \frac{1}{2} \right).$$

We note that  $\pi^{\text{ref}}(\phi \mid x, y)$  induces the  $\mathcal{B}(x + \frac{1}{2}, y + \frac{1}{2})$  reference posterior distribution on the parameter  $p$  defined by (1). This establishes what we claimed about the coincidence of the reference posteriors considering either the conditional or the unconditional model. However, the reference priors on  $p$  differ: this is the improper  $\mathcal{B}(\frac{1}{2}, 0)$  distribution for the partial immunity model, whereas the Jeffreys prior for the binomial model is the proper  $\mathcal{B}(\frac{1}{2}, \frac{1}{2})$  distribution.

**Remark.** (*Jeffreys prior for the partial immunity model*) The Jeffreys prior for the partial immunity model, defined as the square root of Fisher's information matrix, is given by two independent  $\mathcal{G}(\frac{1}{2}, 0)$  distributions on  $\mu$  and  $\lambda$ , and the Jeffreys posterior is given by two independent  $\mathcal{G}(y + \frac{1}{2}, T)$  and  $\mathcal{G}(x + \frac{1}{2}, S)$  distributions on  $\mu$  and  $\lambda$  respectively. Consequently the marginal Jeffreys posterior on  $\phi = \lambda/\mu$  equals the marginal reference posterior on  $\phi$ , as we have (Appendix A)

$$\frac{\mathcal{G}(x + \frac{1}{2}, S)}{\mathcal{G}(y + \frac{1}{2}, T)} = \frac{T}{S} \times \frac{\mathcal{G}(x + \frac{1}{2}, 1)}{\mathcal{G}(y + \frac{1}{2}, 1)} = \frac{T}{S} \times \mathcal{B}' \left( x + \frac{1}{2}, y + \frac{1}{2} \right).$$

Thus, posterior inference about  $\phi$  is the same when using either the Jeffreys posterior or the reference posterior (as well as the "fiducial" distribution; see [28]). Nevertheless, the joint posteriors on  $\mu$  and  $\phi$  differ, and thus

in particular the Jeffreys posterior predictive distribution of  $x$  and  $y$  is not the same as the reference posterior predictive distribution.

#### 1.4. Discussion

We have defined a semi-conjugate family of prior distributions for the partial immunity model, indexed by the four positive hyperparameters  $a, b, c$  and  $d$ . We have seen that particular choices of  $a, b, c, d$  yield the so-called *reference prior* or a certain *semi-reference prior*. The reference prior  $\pi^{\text{ref}}$  is the prior of the semi-conjugate family obtained when  $a = c = \frac{1}{2}$  and  $b = d = 0$ . In the case where  $c = \frac{1}{2}$  and  $d = 0$ , but  $a$  and  $b$  are set to other values, then the prior of the semi-conjugate family is the semi-reference prior  $\pi_{a,b}^{1/2\text{ref}}$ .

*Reference priors* — A noninformative Bayesian  $B^{\text{ref}}$  primarily interested in the relative risk  $\phi$  would use the reference prior  $\pi^{\text{ref}}$ . Note that strictly speaking  $B^{\text{ref}}$  does *not* assign a prior distribution on the parameters (a fortiori, there does not exist any prior predictive distribution for  $B^{\text{ref}}$ ). Indeed the reference prior is *not* a probability distribution, it is not even intended to describe personal beliefs, and should *not* be interpreted as a representation of prior ignorance; it is only a function to be formally used in Bayes theorem to obtain the reference posterior  $\pi^{\text{ref}}(\phi | x, y)$ , which is supposed to describe whatever the data “have to say” about  $\phi$  (Bernardo [11]). A semi-informative Bayesian  $B_{a,b}^{1/2\text{ref}}$ , primarily interested in  $\phi$  too, would use a semi-reference prior  $\pi_{a,b}^{1/2\text{ref}}$ : he/she chooses to describe its personal beliefs about the disease incidence rate  $\mu$  through a  $\mathcal{G}(a, b)$  distribution, no matter the value of  $\phi$ , and has no personal beliefs about the parameter of interest  $\phi$ ; the semi-reference posterior distribution  $\pi_{a,b}^{1/2\text{ref}}(\phi | x, y)$  is supposed to describe whatever the data “have to say” about  $\phi$ , by taking into account the prior information contained in  $\pi(\mu | \phi)$ .

As we have seen,  $\pi_{a,b}^{1/2\text{ref}}(\phi | x, y)$  is defined as the reference posterior on  $\phi$  for the marginal model  $\widetilde{\mathcal{M}}_{a,b}$  whose sampling distribution  $\widetilde{p}_{a,b}(x, y | \phi) = \int p(x, y | \mu, \phi) \mathcal{G}(\mu | a, b) d\mu$  is the bivariate Poisson-Gamma given by (7). Thus, we could consider that  $B^{\text{ref}}$  and  $B_{a,b}^{1/2\text{ref}}$  adopt *the same objective Bayesian methodology*, but make *different distributional assumptions* for the numbers of cases  $x$  and  $y$ :  $B^{\text{ref}}$  assumes that the sampling distribution of  $x$  and  $y$  is  $p(x, y | \mu, \phi)$ , whereas  $B_{a,b}^{1/2\text{ref}}$  assumes that it is  $\widetilde{p}_{a,b}(x, y | \phi)$  (which could be termed as the *conditional prior predictive distribution of  $x$  and  $y$  given  $\phi$* ), and each of the two Bayesians uses the reference prior corresponding to the distributional assumption of  $x$  and  $y$ .

*Semi-conjugate vs conjugate* — The prior distributions of the form  $\pi(\mu, \lambda) = \pi(\mu)\pi(\lambda)$  where each of  $\pi(\mu)$  and  $\pi(\lambda)$  is a Gamma distribution, obviously form a conjugate family for the partial immunity model. Considering a prior belonging to this family leads to independent posterior predictive distributions for  $x$  and  $y$ , which sounds strange in case we are interested in  $\phi$ . Indeed, our aim is to acquire some information about  $\phi$  after the experiment is performed, and thus we should naturally expect that a future observation of the number of cases  $y^*$  in the control group pertains to the prediction of the future observation of the number of cases  $x^*$  in the vaccine group. Another inconvenience of the conjugate family is that a prior from this family yields a rather complicated conditional prior predictive distribution of  $x$  and  $y$  given  $\phi$ , under which the marginal of  $y$  depends on  $\phi$ , whereas the conditional prior predictive distribution (7) for the semi-conjugate family is easily handled and has a clear interpretation.

The Jeffreys prior belongs to the conjugate family, and the Jeffreys posterior on  $\phi$  equals the marginal reference posterior on  $\phi$ , as we noted at the end of subsection 1.3. One could be tempted to define a semi-informative prior in the conjugate family by setting  $\pi(\mu) \sim \mathcal{G}(a, b)$  and  $\pi(\lambda) \sim \mathcal{G}(\frac{1}{2}, 0)$ . But we should be warned about such a method. This kind of semi-informative prior is not derived by a formal rule such as the semi-reference prior in the semi-conjugate family, which is formally defined as the reference prior for the marginal model. It causes no apparent problem in this case, but in general making use of improper prior distributions without precautions can lead to a deficient posterior inference. We refer to [26] for the reader interested in the existing different kinds of noninformative priors defined by formal rules and in the paradoxes which could be caused when making use of improper prior distributions.

## 2. POSTERIOR INFERENCE

In this section we study and compare the posterior inference on the relative risk  $\phi$  in case of the noninformative prior and in case of the semi-informative prior. Some preliminary points are presented in subsection 2.1. The posterior credible intervals that we consider are defined in subsection 2.2. These are the classical posterior credible intervals used in Bayesian statistics. In subsection 2.3 we define some frequentist confidence intervals that will be used in order to do comparisons with the Bayesian confidence intervals. We will focus on frequentist properties of the posterior inference in subsections 2.4 and 2.5.

As we will see, posterior credible intervals enjoy attractive frequentist performance when using the noninformative prior. But note that, although using the reference prior, it is *not* reasonable to term them as “*reference posterior credible intervals*”. Indeed, these procedures do *not* deal with *reference analysis*, as we do *not* derive any decision by minimizing a reference posterior expected loss corresponding to an information theory based loss function such as the *the intrinsic discrepancy* (see [13, 14]). Intrinsic credible regions and intrinsic hypothesis testing for the relative risk  $\phi$  in the partial immunity model will be the purpose of another paper.

### 2.1. OBayesian and Semi-OBayesian

Let us consider again a non-informative Bayesian  $B^{\text{ref}}$  and a semi-informative Bayesian  $B_{a,b}^{1/2\text{ref}}$  as in section 1.4. Throughout this section, we consider as an example that the person times at risk are  $S = T = 10000$ , and that  $B_{a,b}^{1/2\text{ref}}$  chooses the  $\mathcal{G}(a, b)$  prior distribution with  $a = 30k$  and  $b = 10000k$  for a certain value of  $k > 0$ . Thus the prior mean of  $\mu$  equals  $a/b = 30/10000$  whatever the choice of  $k$ .

To observe the difference between the posterior inference for  $B^{\text{ref}}$  and  $B_{a,b}^{1/2\text{ref}}$ , we start by considering three scenarios for the observed values  $x^{\text{obs}}$  and  $y^{\text{obs}}$  of the number of cases: we consider  $x^{\text{obs}} = 5$  for each scenario, and  $y^{\text{obs}} = 23, y^{\text{obs}} = 30, y^{\text{obs}} = 37$ . With  $k = 1$ , these values of  $y$  respectively correspond to a case where  $y^{\text{obs}}$  is a little below the range, in the range, and a little above the range of most likely values of the marginal sampling distribution  $\tilde{p}_{a,b}(y | \phi) \equiv \tilde{p}_{a,b}(y)$  of  $y$ , say between 25 and 35. The marginal posterior distributions on  $\phi$  are plotted on Figure 2.

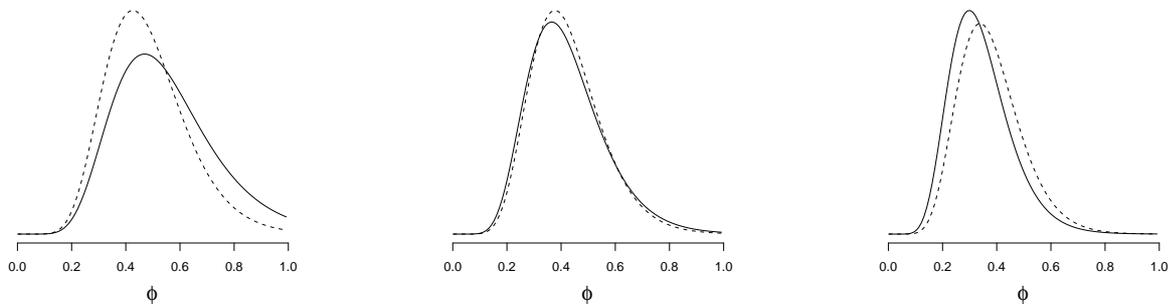


FIGURE 2. Marginal posterior on  $\phi$  with  $x = 12$  and  $y = 23$  (left),  $y = 30$  (middle),  $y = 37$  (right). Solid:  $\pi^{\text{ref}}(\phi | x, y)$ . Dashed:  $\pi_{a,b}^{1/2\text{ref}}(\phi | x, y)$

What we observe on Figure 2 is not surprising, and is made more clear by noting the following fact, which immediately results from the expressions of the posterior distributions in the semi-conjugate family:

If we denote by  $\pi(\cdot | x, y, S, T)$  the generic notation for posterior distributions (thus including the dependence on  $S$  and  $T$  in the notations), then we have the following relation:

$$\pi_{a,b}^{1/2\text{ref}}(\cdot | x, y, S, T) = \pi^{\text{ref}}(\cdot | x, a + y - 0.5, S, b + T).$$

In other words, the posterior on  $\phi$  for  $B_{a,b}^{1/2\text{ref}}$  when the observed values of  $x$  and  $y$  are  $x^{\text{obs}}$  and  $y^{\text{obs}}$  and the person-times at risk are  $S$  and  $T$ , equals the posterior on  $\phi$  for  $B^{\text{ref}}$  when the observed values of  $x$  and  $y$  are  $x^{\text{obs}}$  and  $a + y^{\text{obs}} - 0.5$  and the person-times at risk are  $S$  and  $T + b$ .

Thus, in the second case the posteriors for  $B^{\text{ref}}$  and  $B_{a,b}^{1/2\text{ref}}$  are in agreement because the value  $y^{\text{obs}} = 30$  is in agreement with  $\tilde{p}_{a,b}(y)$ , but the posterior for  $B_{a,b}^{1/2\text{ref}}$  is more concentrated; the semi-reference posterior  $\pi_{a,b}^{1/2\text{ref}}(\phi | x, y)$  would be the same as the posterior for  $B^{\text{ref}}$  if instead of  $y^{\text{obs}} = 30$  and  $T = 10000$  he/she would have  $y^{\text{obs}} = a + 30 - 0.5 = 59.5$  and  $T = b + 10000 = 20000$ . In the other cases  $y^{\text{obs}} = 23$  and  $y^{\text{obs}} = 37$ , the marginal semi-reference posterior  $\pi_{a,b}^{1/2\text{ref}}(\phi | x, y)$  would be the same as the posterior for  $B^{\text{ref}}$  with  $y^{\text{obs}} = 52.5$  and  $y^{\text{obs}} = 66.5$  respectively, and  $T = 20000$ .

Now, consider that both of  $B^{\text{ref}}$  and  $B_{a,b}^{1/2\text{ref}}$  perform an hypothesis testing for  $H_0 : \phi \geq 75\%$  vs  $H_1 : \phi < 75\%$  by adopting the following rejection rule:

$$\begin{cases} \text{if } \pi(H_1 | x, y) \leq 97.5\% & \text{then do not reject } H_0 \\ \text{if } \pi(H_1 | x, y) > 97.5\% & \text{then reject } H_0, \end{cases} \quad (8)$$

where, of course,  $\pi = \pi^{\text{ref}}$  for  $B^{\text{ref}}$  and  $\pi = \pi_{a,b}^{1/2\text{ref}}$  for  $B_{a,b}^{1/2\text{ref}}$ .

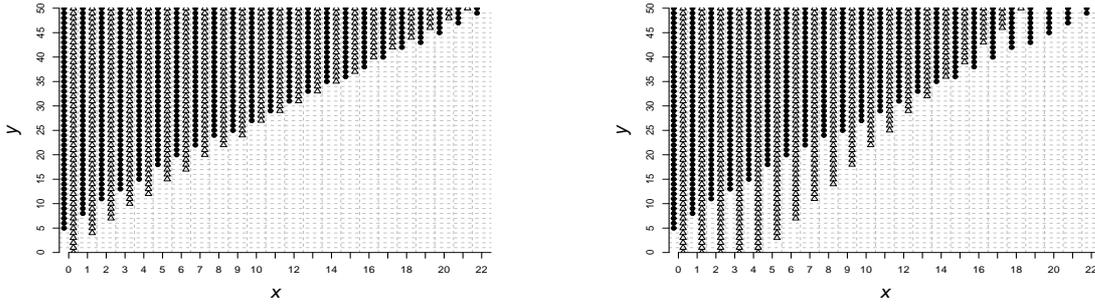


FIGURE 3. Decisions of the test defined by the rejection rule (8). A black circle  $\bullet$  means that  $H_0$  is rejected for  $B^{\text{ref}}$ , and a white triangle  $\triangle$  means that  $H_0$  is rejected for  $B_{a,b}^{1/2\text{ref}}$ . *Left:*  $k = 0.2$ . *Right:*  $k = 1$ .

Using  $k = 0.2$  and  $k = 1$ , the possible decisions of the test are displayed on Figure 3 for  $x$  varying from 0 to 22 and  $y$  varying from 0 to 50. If we consider, for example,  $x = 2$ ,  $B^{\text{ref}}$  rejects  $H_0$  whenever  $y \geq 11$ , whereas  $B_{a,b}^{1/2\text{ref}}$  rejects  $H_0$  whatever the value of  $y$  in case of  $k = 1$ . Looking at  $x = 12$ , it is easy to link how the possible decisions of the test differ between the two Bayesians with results displayed on Figure 2. Note that the test decisions do not greatly differ between the two Bayesians when the number of cases  $y$  is among the most likely values of  $\tilde{p}_{a,b}(\cdot | \phi)$ , say between 25 and 35.

We now study the frequentist characteristics of the posterior credible intervals. Frequentist properties are expressed in terms of probabilities under repeated sampling for various fixed values of the parameters. Motivated by our discussion in section 1.4, we will make two different assumptions for the sampling distribution: the first one is that  $(x, y) \sim p(x, y | \mu, \phi)$ , the second one is that  $(x, y) \sim \tilde{p}_{a,b}(x, y | \phi)$ ; the latter assumption could be considered as implicit for  $B_{a,b}^{1/2\text{ref}}$ . We have seen in section 1.2 that the marginal sampling distribution  $\tilde{p}_{a,b}(x, y | \phi)$  is given by the marginal-conditional factorization

$$(y | \phi) \sim T \star \mathcal{PG}(a, b) \quad \text{and} \quad (x | y, \phi) \sim \phi S \star \mathcal{PG}(y + a, T + b),$$

hence in particular the distribution on  $y$  does not involve any unknown parameter.

## 2.2. Posterior credible intervals

For each of the posterior credibility intervals  $I_{x,y}$  of  $\phi$  that we will study, we associate, as frequentist statisticians do for confidence intervals, the family of tests  $T_{\phi^*}$ ,  $\phi^* > 0$  for  $H_0 : \{\phi = \phi^*\}$  versus some  $H_1$ , defined by the rejection region  $R(\phi^*) = \{(x, y) \mid \phi^* \notin I_{x,y}\}$ . As we shall note, tests associated to one-sided intervals are natural in a Bayesian framework, and could also be derived from the Bayesian decision-theoretic approach, whereas tests associated with two-sided intervals do *not* have a Bayesian interpretation. Our motivation for introducing these tests is that their frequentist properties are directly related to the frequentist properties of the posterior credibility intervals.

**Left one-sided intervals.** — We focus on left one-sided credibility intervals since vaccine efficacy trials primarily aim to show a significant evidence that the relative risk  $\phi$  is small. The  $100(1 - \alpha)\%$  posterior credibility left one-sided interval  $I_{x,y}^{\text{left}}(\alpha)$  is defined as  $I_{x,y}^{\text{left}}(\alpha) = [0, z_\alpha(x, y)]$  where  $z_\alpha(x, y)$  is the  $100(1 - \alpha)\%$  upper quantile of the posterior distribution  $\pi(\phi \mid x, y)$  of  $\phi$ , defined by  $\pi(\phi > z_\alpha(x, y) \mid x, y) = \alpha$ .

They are associated to the family of tests  $T_{\phi^*}$ ,  $\phi^* > 0$ , for  $H_0 : \{\phi = \phi^*\}$  vs  $H_1 : \{\phi < \phi^*\}$  whose rejection rule could be written as:

$$\begin{cases} \text{if } \pi(H_1 \mid x, y) \leq 1 - \alpha & \text{then do not reject } H_0 \\ \text{if } \pi(H_1 \mid x, y) > 1 - \alpha & \text{then reject } H_0. \end{cases} \quad (9)$$

The same rejection rule for  $H_0 : \{\phi \geq \phi^*\}$  vs  $H_1 : \{\phi < \phi^*\}$  is quite natural in a Bayesian framework, and it also derives from the Bayesian decision theory with the help of a weighted 0–1 loss function (see [33]).

**Two-sided intervals.** — The  $100(1 - \alpha)\%$  posterior credibility *equi-tailed* interval  $I_{x,y}^{\text{equi}}(\alpha)$  is defined as  $I_{x,y}^{\text{equi}}(\alpha) = [z_{1-\frac{\alpha}{2}}(x, y), z_{\frac{\alpha}{2}}(x, y)]$ . When using the reference prior, one could prefer to use the *modified equi-tailed* interval defined as

$$I_{x,y}^{\text{equi}^*}(\alpha) = \begin{cases} I_{x,y}^{\text{equi}}(\alpha) & \text{if } x \neq 0 \\ I_{x,y}^{\text{left}}(\frac{\alpha}{2}) & \text{if } x = 0. \end{cases}$$

Indeed, the reference posterior density of  $\phi$  is the scaled Beta prime distribution  $\frac{T}{S} \times \mathcal{B}'(\frac{1}{2}, y + \frac{1}{2})$  when  $x = 0$ , which is decreasing and concentrated on the left side of its support  $(0, +\infty)$  (see Figure 16). Because the reference posterior is supposed to describe whatever the data “have to say” about  $\phi$ , it is then counter-intuitive to consider the equi-tailed interval as a “confidence” set in that case. Of course we should not accept to call  $I_{x,y}^{\text{equi}^*}(\alpha)$  a  $100(1 - \alpha)\%$  posterior credibility interval because the equality  $\pi(\phi \in I_{x,y}^{\text{equi}^*}(\alpha) \mid x, y) = 1 - \alpha$  fails to be true when  $x = 0$ . The same modification is proposed in [17] for the posterior credible intervals of a binomial proportion based on Jeffreys’ prior (see also [9]).

The  $100(1 - \alpha)\%$  posterior credibility *HPD* intervals  $I_{x,y}^{\text{HPD}}(\alpha)$  are well defined when using a prior belonging to the semi-conjugate family: the marginal posterior distribution on  $\phi$  is always unimodal or decreasing (when  $c \leq 1$  and  $x = 0$ ), hence the smallest region with posterior probability  $1 - \alpha$  is indeed an interval.

Two-sided credibility intervals are associated to tests for  $H_0 : \{\phi = \phi^*\}$  vs  $H_1 : \{\phi \neq \phi^*\}$ . Contrary to the tests associated to one-sided intervals, the rejection rule cannot be based on the posterior probability of  $H_1$  as in (9) because here we have  $\pi(H_1 \mid x, y) = 1$  whatever the values of  $x$  and  $y$ .

## 2.3. Sahai & Khurshid frequentist confidence intervals

In order to study the frequentist properties of the Bayesian posterior credible intervals, we need some frequentist confidence intervals for comparison. We will use those obtained by the method described by Sahai & Khurshid in [35] based on the approximation<sup>2</sup> of  $2(\sqrt{W + 0.5} - \sqrt{m})$  by a standard normal distribution, where  $W$  has Poisson distribution with mean  $m$ . The two-sided Sahai & Khurshid (SK) confidence interval has been compared to other confidence intervals in [32]. It enjoys a particularly good frequentist performance. We will

<sup>2</sup>This convergence result can be proved with the help of the Delta Method.

consider the left and right one-sided and the two-sided confidence intervals based on this approximation. The bounds of the SK  $100(1 - 2\alpha)\%$  centered confidence intervals  $I_{x,y}^{SK}$  are given by

$$\frac{T}{S} \left( \frac{\sqrt{(x+0.5)(y+0.5)} \pm 0.5z_\alpha \sqrt{x+y+1-0.25z_\alpha^2}}{y+0.5-0.25z_\alpha^2} \right)^2$$

where  $\mathbb{P}(\mathcal{N}(0,1) > z_\alpha) = \alpha$ , and we define the bounds for the situation  $x = y = 0$  to be 0 and  $+\infty$  respectively<sup>3</sup>. One can verify that these bounds are well defined as long as  $z_\alpha^2 < 8$  which is the case whenever  $\alpha \geq 0.5\%$ .

From now on, in addition to  $B^{\text{ref}}$  and  $B^{1/2\text{ref}}$ , we will consider a frequentist statistician, namely  $F^{SK}$ , who performs inference on  $\phi$  using the SK confidence intervals. In subsection 2.4 we will draw the frequentist properties of the posterior credible intervals based on the reference posterior distribution  $\pi^{\text{ref}}(\phi | x, y)$ . Conversely, it is interesting to compute the reference posterior probabilities  $\pi^{\text{ref}}(\phi \in I_{x,y} | x, y)$  of the SK confidence interval  $I_{x,y}$  used by  $F^{SK}$  (see the discussion in [14]). We show in Table 1 the values of  $\pi^{\text{ref}}(\phi \in I_{x,y} | x, y)$  for some values of  $x$  and  $y$  where  $I_{x,y}$  is the left one-sided SK 97.5% confidence interval.

We see in Table 1 that the realizations of the left one-sided SK 97.5% confidence interval have a reference posterior probability slightly higher than 97.5%, except for  $y = 0$ . In view of this, we could add a modification to the SK intervals when  $y = 0$ . However we will not do so as it would not have any consequence of interest in this paper.

$x \backslash y$	0	1	2	4	6	10	20	30
0	100.0	99.1	98.5	98.4	98.3	98.3	98.3	98.3
1	69.8	99.0	98.3	98.0	97.9	97.9	97.9	97.9
2	70.0	99.0	98.2	97.9	97.8	97.8	97.8	97.8
3	70.0	98.9	98.2	97.9	97.8	97.7	97.7	97.7
4	70.0	98.9	98.2	97.9	97.8	97.7	97.7	97.7
5	70.0	98.9	98.2	97.8	97.7	97.7	97.7	97.6
6	70.0	98.9	98.2	97.8	97.7	97.7	97.6	97.6

TABLE 1. Reference posterior probabilities (in %) of left one-sided 97.5% confidence intervals for  $F^{SK}$ .

#### 2.4. Frequentist properties with non-informative prior

We compare the frequentist properties of the inference performed by  $B^{\text{ref}}$  and  $F^{SK}$  under the assumption of the partial immunity model  $\mathcal{M} = \{p(x, y | \mu, \phi); \mu > 0, \phi \geq 0\}$ . It is clear from their definition that both the posterior  $100(1 - \alpha)\%$ -credibility intervals for  $B^{\text{ref}}$  (subsection 2.2) and the  $100(1 - \alpha)\%$ -confidence intervals for  $F^{SK}$  (subsection 2.3) depend only on  $\alpha$  and the ratio  $S/T$ . As a consequence, for fixed values of  $\alpha$  and  $S/T$ , their frequentist coverage depends only on the relative risk  $\phi$  and the expected number of cases  $\mu T$  in the control group. Indeed, considering a probability  $P(\cdot | \mu, \phi)$  under which  $x$  and  $y$  are distributed according to  $p(x, y | \mu, \phi)$ , the frequentist coverage of a random set  $I_{x,y}$  depending on  $x$  and  $y$  is the function  $(\mu, \phi) \mapsto P(I_{x,y} \ni \phi | \mu, \phi)$ , which is given by

$$P(I_{x,y} \ni \phi | \mu, \phi) = \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} \mathbf{1}_{I_{x,y}}(\phi) p(x, y | \mu, \phi), \quad (10)$$

<sup>3</sup>Our convention for the situation  $x = y = 0$  is slightly different from the one in [32] but this has no impact on the numerical results drawn by the authors.

with

$$p(x, y | \mu, \phi) = \frac{1}{x!y!} e^{-\mu T(\phi \frac{S}{T} + 1)} \left( \phi \frac{S}{T} \right)^x (\mu T)^{x+y}.$$

For a fixed ratio  $S/T = 1$  and various values of  $\phi$  and  $\mu T$ , Tables 2 and 3 show the frequentist coverage probability (10) of the left one-sided posterior credibility interval for  $B^{\text{ref}}$  and the left one-sided confidence interval for  $F^{SK}$  with  $\alpha = 2.5\%$ . We see on these tables that, roughly, the frequentist coverage for  $B^{\text{ref}}$  behaves in the same way as for  $F^{SK}$ .

$\phi \backslash \mu T$	2	4	10	15	20	30	40	50
0.1	100.0	100.0	100.0	99.9	99.0	96.3	97.5	97.3
0.3	100.0	99.8	97.1	97.5	97.4	97.5	97.4	97.5
0.5	99.4	97.0	97.2	97.5	97.5	97.4	97.5	97.4
0.75	98.8	97.3	97.3	97.6	97.5	97.5	97.5	97.5
1	97.9	97.7	97.3	97.4	97.5	97.4	97.5	97.5
2	97.2	97.2	97.5	97.5	97.5	97.5	97.5	97.5
4	97.6	97.4	97.5	97.5	97.5	97.5	97.5	97.5

TABLE 2. Frequentist coverage probabilities  $P(I_{x,y} \ni \phi | \mu, \phi)$  of left one-sided 97.5% posterior credibility interval for  $B^{\text{ref}}$  with  $S/T = 1$ .

$\phi \backslash \mu T$	2	4	10	15	20	30	40	50
0.1	100.0	100.0	100.0	100.0	99.7	97.4	97.9	97.7
0.3	100.0	99.9	97.7	97.5	97.5	97.6	97.6	97.6
0.5	99.8	98.5	97.6	97.5	97.5	97.5	97.6	97.6
0.75	98.8	97.3	97.6	97.8	97.6	97.6	97.5	97.5
1	97.9	97.8	97.5	97.5	97.5	97.5	97.5	97.5
2	97.2	97.2	97.5	97.5	97.5	97.5	97.5	97.5
4	97.9	97.6	97.6	97.6	97.5	97.5	97.5	97.5

TABLE 3. Frequentist coverage probabilities  $P(I_{x,y} \ni \phi | \mu, \phi)$  of left one-sided 97.5% confidence interval for  $F^{SK}$  with  $S/T = 1$ .

A column in Tables 2 and 3 corresponds to the function  $\phi \mapsto P(I_{x,y} \ni \phi | \mu, \phi)$ , termed as the frequentist coverage with respect to the one-parameter conditional model  $\mathcal{M}_\mu = \{p(x, y | \mu, \phi); \phi \geq 0\}$  given a fixed value of  $\mu$ . We display this function for  $B^{\text{ref}}$  and  $F^{SK}$  on Figure 4 for  $\mu T \in \{10, 30\}$ . We see that the curves for  $B^{\text{ref}}$  and  $F^{SK}$ , unless  $\phi$  is too small, are both close to the nominal level and behave similarly.

The curves on Figure 5 correspond to one minus the fourth row of Tables 2 and 3: they display  $P(I_{x,y} \not\ni \phi^* | \mu, \phi^*)$  in function of  $\mu T$  for  $\phi^* = 75\%$ ; considering the test for  $H_0 : \{\phi = \phi^*\}$  vs  $H_1 : \{\phi < \phi^*\}$  defined by the rejection rule (9), we have  $P(I_{x,y} \not\ni \phi^* | \mu, \phi^*) = P(\text{reject } H_0 | \mu, \phi^*)$ , hence these curves show the significance level of the test with respect to  $\mathcal{M}_\mu$  for various values of  $\mu$ . The significance level of this test with respect to the two-parameters partial immunity model  $\mathcal{M}$  is then the supremum over  $\mu$  of  $P(I_{x,y} \not\ni \phi^* | \mu, \phi^*)$ .

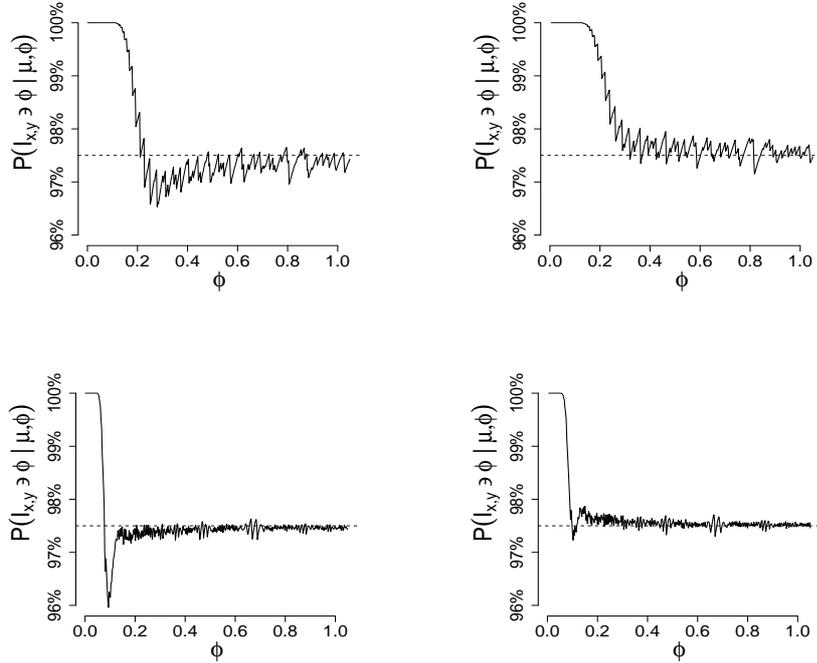


FIGURE 4. Frequentist coverage probability of left one-sided intervals with  $S/T = 1$ . *Left:* 97.5%  $B^{\text{ref}}$ . *Right:* 97.5%  $F^{\text{SK}}$ . *Top:*  $\mu T = 10$ . *Bottom:*  $\mu T = 30$ .

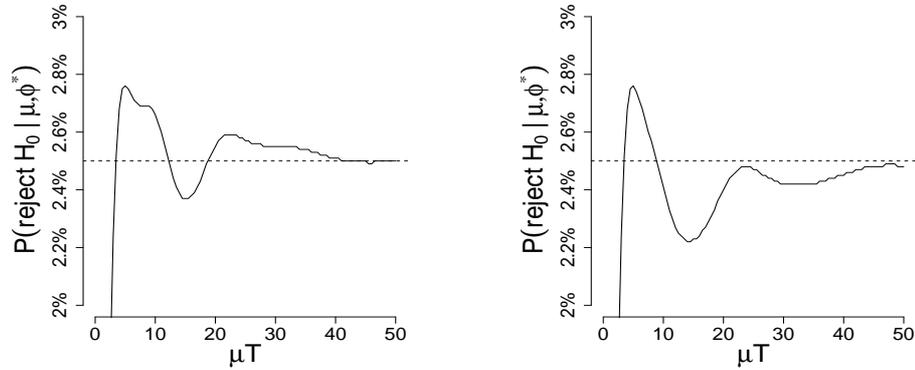


FIGURE 5. Significance level of the test for  $H_0 : \{\phi = 75\%\}$  vs  $H_1 : \{\phi < 75\%\}$  in function of the expected number of cases under control, with  $\phi^* = 75\%$  and  $S/T = 1$ . *Left:* 97.5%  $B^{\text{ref}}$ . *Right:* 97.5%  $F^{\text{SK}}$ .

The power curve  $\phi \mapsto P(I_{x,y} \not\ni \phi^* \mid \mu, \phi)$  of this test with respect to  $\mathcal{M}_\mu$  is plotted on Figure 6 for  $\mu T \in \{10, 30\}$ . Curves for  $B^{\text{ref}}$  and  $F^{\text{SK}}$  are so close that they are indistinguishable. As we see, the curves are decreasing, so that our test has the same significance level considering either  $H_0 : \{\phi = 75\%\}$  or  $H_0 : \{\phi \geq 75\%\}$ .

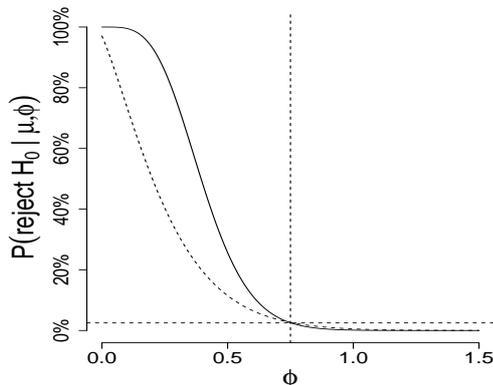


FIGURE 6. Probability for rejecting  $H_0 : \{\phi = 75\%\}$  vs  $H_1 : \{\phi < 75\%\}$  with  $S/T = 1$ . *Dashed:*  $\mu T = 10$ . *Solid:*  $\mu T = 30$ . Curves for  $B^{\text{ref}}$  and  $F^{\text{SK}}$  are confounded.

For  $\alpha = 5\%$ , Figure 7 shows the frequentist coverage  $P(I_{x,y} \ni \phi \mid \mu, \phi)$  for the two-sided confidence intervals  $I_{x,y}$  in function of  $\phi$  for  $S/T = 1$  and  $\mu T \in \{10, 30\}$ . Figure 8 shows the power of the associated tests for  $H_0 : \{\phi = \phi^*\}$  vs  $H_1 : \{\phi < \phi^*\}$  with  $\phi^* = 75\%$ . Curves for  $F^{\text{SK}}$  are confounded with curves for  $B^{\text{ref}}$  in case of equi-tailed posterior credibility intervals. We see that the test associated with the HPD interval is *biased* (see for example [30]), i.e. the power is lower than  $\alpha$  for some values of  $\phi$  (we also say that the HPD credibility interval defines a *biased* confidence interval).

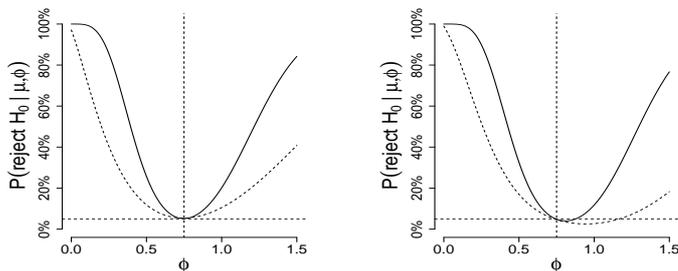


FIGURE 8. Probability for rejecting  $H_0 : \phi = 75\%$  with  $S/T = 1$ . *Solid:*  $\mu T = 30$ . *Dotted:*  $\mu T = 10$ . *Left:* 95% equi-tailed for  $B^{\text{ref}}$  or 95% two-sided for  $F^{\text{SK}}$  (curves are confounded). *Right:* 95% HPD for  $B^{\text{ref}}$ .

## 2.5. Frequentist properties with semi-informative prior

As announced in section 2.1, we now compare the frequentist properties between  $B^{\text{ref}}$  and  $B_{a,b}^{1/2\text{ref}}$  assuming either  $(x, y) \sim p(x, y \mid \mu, \phi)$  or  $(x, y) \sim \tilde{p}_{a,b}(x, y \mid \phi)$ . As in section 2.1, we consider that  $S = T = 10000$  and that  $a = 30k$  and  $b = 10000k$  for a certain value of  $k$  that we set to  $k = 0.2$  and  $k = 1$ . We will study only left one-sided intervals as the same type of reasoning can be applied to the other types of intervals.

We first consider the case where the distributional assumption is  $(x, y) \sim p(x, y \mid \mu, \phi)$ . We focus on the cases when  $\mu = 10/10000$  (above the range of the prior) and  $\mu = 50/10000$  (below the range of the prior). Figure 9 shows the power curves of the test  $H_0 : \{\phi = \phi^*\}$  vs  $H_1 : \{\phi < \phi^*\}$  defined by the rejection rule (8) with  $\phi^* = 75\%$  and  $\alpha = 2.5\%$ . These results can be better understood when considering them in light of the results displayed in Figure 3. Let's first have a look at the right side of Figure 9, that is considering that  $x$  and  $y$  are generated with  $\mu = 10/10000$ . Having in mind that the most likely values for  $y$  are around 10 in this situation, Figure 3 shows that, especially for the larger value of  $k$ , the semi-informative Bayesian  $B_{a,b}^{1/2\text{ref}}$ ,

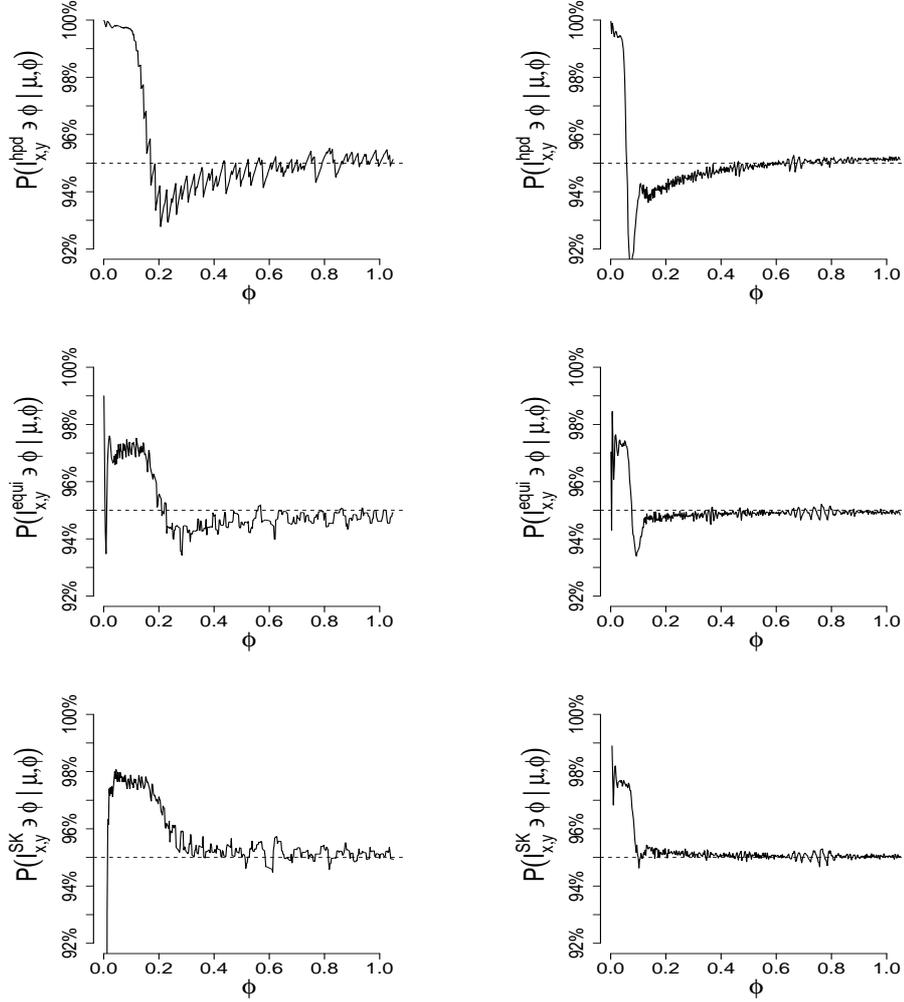


FIGURE 7. Frequentist coverage probability of two-sided intervals with  $S/T = 1$ . *Left:*  $\mu T = 10$ . *Right:*  $\mu T = 30$ . *Top:* 95%  $B^{\text{ref}}$  HPD. *Middle:* 95%  $B^{\text{ref}}$  equi-tailed. *Bottom:* 95%  $F^{\text{SK}}$ .

rejects  $H_0$  much more often than the non-informative Bayesian  $B^{\text{ref}}$ , and this is even more striking for small values of  $x$ . Therefore, as expected, we see on Figure 9 that the probability for rejecting  $H_0$  is higher for  $B_{a,b}^{1/2\text{ref}}$  than for  $B^{\text{ref}}$ , especially for the larger value of  $k$ , and the difference is more pronounced for small values of  $\phi$ , under which the generated values of  $x$  are typically small. Of course the same type of reasoning applies for  $\mu = 50/10000$  and explains the ‘reverse behavior’ that we observe on the right side of Figure 9.

Figure 10 displays the significance level  $\mu \mapsto P(I_{x,y} \not\ni \phi^* | \mu, \phi^*)$  of this test in function of  $\mu$ ; this figure is comparable with Figure 5 given for  $B^{\text{ref}}$ . Figure 11 displays the frequentist coverage  $\phi \mapsto P(I_{x,y} \ni \phi | \mu, \phi)$  for  $B_{a,b}^{1/2\text{ref}}$ ; this figure is comparable with Figure 4 where we have seen that the frequentist coverage for  $B^{\text{ref}}$  is more or less close to the credibility level whatever the value of  $\mu$ .

The link between Figure 9 and Figure 10 is clear. Figure 11 is more complicated to interpret in details, but it is clear that, roughly, the frequentist coverage curve for  $B_{a,b}^{1/2\text{ref}}$  is below the one for  $B^{\text{ref}}$  in case  $\mu = 10/10000$ , and conversely in case  $\mu = 50/10000$ .

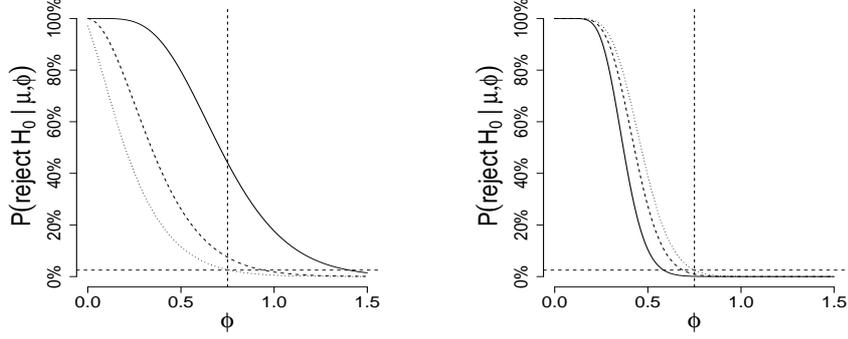


FIGURE 9. Probability for rejecting  $H_0 : \{\phi = 75\%\}$  vs  $H_1 : \{\phi < 75\%\}$  assuming  $(x, y) \sim p(x, y | \mu, \phi)$ . *Left:*  $\mu = 10/10000$ . *Right:*  $\mu = 50/10000$ . *Dotted:*  $B^{\text{ref}}$ . *Dashed:*  $B_{a,b}^{1/2\text{ref}}$  with  $k = 0.2$ . *Solid:*  $B_{a,b}^{1/2\text{ref}}$  with  $k = 1$ .

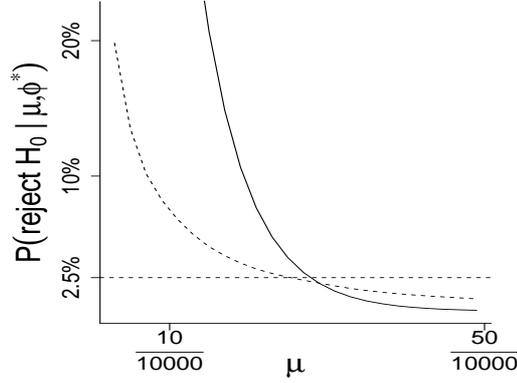


FIGURE 10. Significance level for  $B^{1/2\text{ref}}$  in function of the attack rate  $\mu$ , with  $\phi^* = 75\%$ . *Dashed:*  $k = 0.2$ . *Solid:*  $k = 1$ .

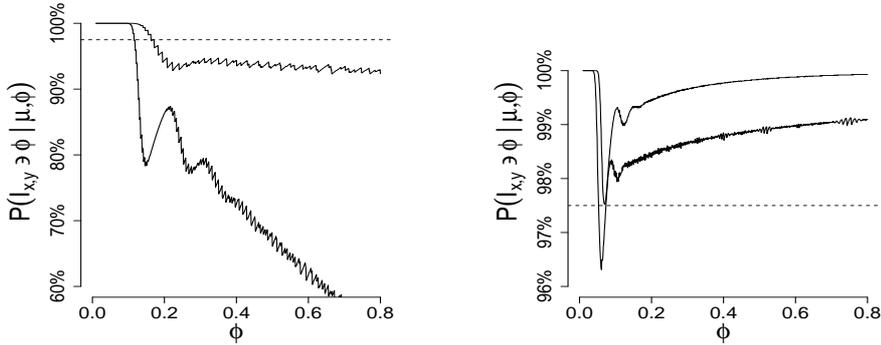


FIGURE 11. Frequentist coverage  $\phi \mapsto P(I_{x,y} \ni \phi | \mu, \phi)$  for  $B^{1/2\text{ref}}$  with  $k = 0.2$  and  $k = 1$ . The curve deviating the more from the nominal level the larger  $k$ . *Left:*  $\mu = 10/10000$ . *Right:*  $\mu = 50/10000$ .

Now, let us see what happens under the assumption that  $(x, y) \sim \tilde{p}_{a,b}(x, y | \phi)$ . Denoting by  $\tilde{P}_{a,b}(\cdot | \phi)$  a probability under which  $(x, y) \sim \tilde{p}_{a,b}(x, y | \phi)$ , then the frequentist coverage of a random set  $I_{x,y}$  depending on  $x$  and  $y$  is the function  $\phi \mapsto \tilde{P}_{a,b}(I_{x,y} \ni \phi | \phi)$ .

Of course, if a posterior credible set  $I_{x,y}$  enjoys a good *frequentist-matching property* under the distributional assumption  $(x, y) \sim p(x, y | \mu, \phi)$ , i.e. its frequentist coverage (10) is not far from the credibility level, then it has also a good frequentist-matching property under the distributional assumption  $(x, y) \sim \tilde{p}_{a,b}(x, y | \phi)$  because of

$$\tilde{P}_{a,b}(I_{x,y} \ni \phi | \phi) = \int P(I_{x,y} \ni \phi | \mu, \phi) \mathcal{G}(\mu | a, b) d\mu.$$

Figure 12 shows that the frequentist coverage curves for  $B^{\text{ref}}$  and  $B_{a,b}^{1/2\text{ref}}$  are both close to the credibility level  $1 - \alpha$  unless  $\phi$  is too small (and the peak in the beginning of the curve is deeper for  $B_{a,b}^{1/2\text{ref}}$  when  $k = 1$ ). Therefore, for both  $B^{\text{ref}}$  and  $B_{a,b}^{1/2\text{ref}}$ , the significance level  $\tilde{P}_{a,b}(I_{x,y} \not\ni \phi^* | \phi^*)$  of the associated tests for  $H_0 : \{\phi = \phi^*\}$  vs  $H_1 : \{\phi < \phi^*\}$  defined by the rejection rule (8), are close to  $\alpha$  unless  $\phi^*$  is too small. Figure 13 shows the power  $\phi \mapsto \tilde{P}_{a,b}(I_{x,y} \not\ni \phi^* | \phi)$  for  $\phi^* = 75\%$ . An expected difference occurs :  $B_{a,b}^{1/2\text{ref}}$  attains a higher power than  $B^{\text{ref}}$ .

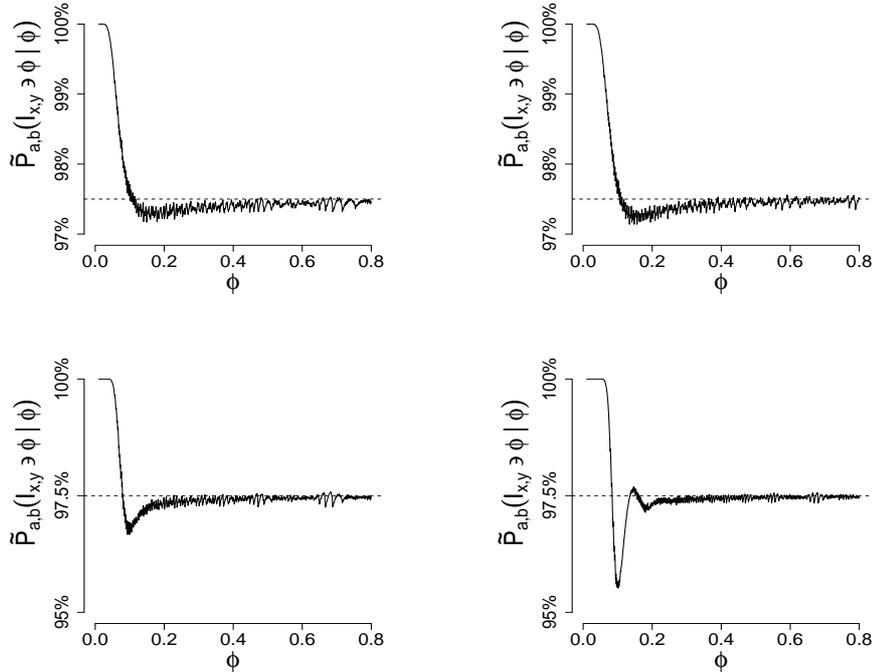


FIGURE 12. Frequentist coverage for  $B^{\text{ref}}$  and  $B_{a,b}^{1/2\text{ref}}$  assuming  $(x, y) \sim \tilde{p}_{a,b}(x, y | \phi)$  Top:  $k = 0.2$ . Bottom:  $k = 1$ . Left:  $B^{\text{ref}}$ . Right:  $B_{a,b}^{1/2\text{ref}}$ .

## 2.6. Discussion

One sometimes hear statements such as “inclusion of prior information increases power, or equivalently, reduces the number of subjects required to achieve trial objective”. However this can generate a lot of misinterpretations.

Our investigations in section 2.4 show that the inference on the relative risk  $\phi$  for the noninformative Bayesian statistician  $B^{\text{ref}}$  achieves a frequentist performance comparable to the frequentist statistician  $F^{SK}$ . From what

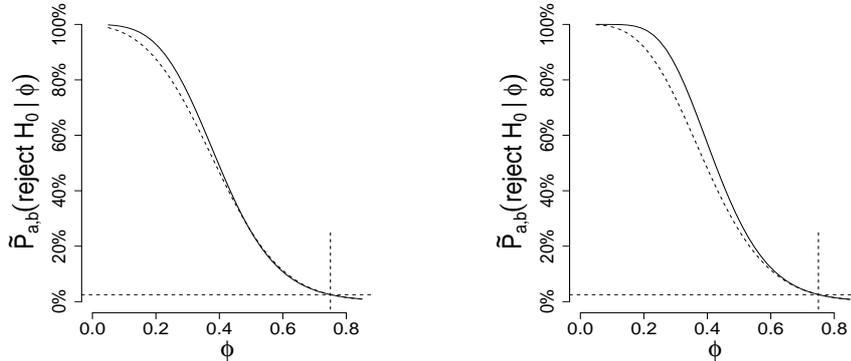


FIGURE 13. Probability for rejecting  $H_0 : \phi = 75\%$  vs  $H_1 : \phi < 75\%$  for  $B^{\text{ref}}$  and  $B_{a,b}^{1/2\text{ref}}$  assuming  $(x, y) \sim \tilde{p}_{a,b}(x, y | \phi)$ . Dashed:  $B^{\text{ref}}$ . Solid:  $B_{a,b}^{1/2\text{ref}}$ . Left:  $k = 0.2$ . Right:  $k = 1$ .

we have seen in section 2.5, we could claim without ambiguity that *the inference on  $\phi$  for the semi-informative Bayesian statistician  $B_{a,b}^{1/2\text{ref}}$  achieves higher power than  $B^{\text{ref}}$  under the hypothesis that the data is generated from the marginal sampling distribution  $\tilde{p}_{a,b}(\cdot | \phi)$* . However the claim that *inclusion of prior information increases power* can only be made in light of the following comments.

- Under repeated sampling of  $x$  and  $y$  from  $\tilde{p}_{a,b}(x, y | \phi)$ , the semi-informative Bayesian  $B_{a,b}^{1/2\text{ref}}$  achieves higher power than  $B^{\text{ref}}$  and  $F^{SK}$  not because  $B_{a,b}^{1/2\text{ref}}$  is Bayesian, but because  $B_{a,b}^{1/2\text{ref}}$  implicitly makes the assumption that  $x$  and  $y$  are precisely generated from  $\tilde{p}_{a,b}(x, y | \phi)$ , whereas  $B^{\text{ref}}$  and  $F^{SK}$  assume the original sampling distribution  $p(x, y | \mu, \phi)$  of the partial immunity model. A comparable power could a priori be achieved by any inference on the parameter  $\phi$  of the marginal model  $\tilde{\mathcal{M}}_{a,b} = \{\tilde{p}_{a,b}(\cdot | \phi); \phi > 0\}$  based on a frequentist method<sup>4</sup>. Nevertheless, considering the marginal model makes more sense in the Bayesian context as it corresponds to incorporating prior information about  $\mu$  in the original sampling distribution.
- The assumption that  $\tilde{p}_{a,b}(x, y | \phi)$  is the sampling distribution, is strong as it assumes in particular that there is no unknown parameter in the sampling distribution of the number of cases  $y$  in the control group.
- The inference on  $\phi$  performed by  $B_{a,b}^{1/2\text{ref}}$  does not provide a valid frequentist inference under the assumption of the original sampling distribution  $p(x, y | \mu, \phi)$ . On the contrary the type I error is controlled by  $B^{\text{ref}}$  and  $F^{SK}$  under the assumption of any sampling distribution obtained by eliminating the nuisance parameter  $\mu$  through integration over an arbitrary probability distribution.

Besides the good frequentist properties achieved by the reference prior, and the possibility to formally include the prior information available on the disease incidence at the time of designing the study, the main force of the Bayesian approach is the appealing interpretation of the inference results. In the context of vaccine efficacy trials, the *vaccine efficacy acceptability curve* [19], defined as  $v^* \mapsto \pi(VE > v^* | x, y)$ , allows a quantitative description of  $VE$  much more rich than a frequentist confidence interval, and is much more convenient to interpret for practitioners. The concise illustration we give in the next section shows an example of such a curve (Figure 15) when considering the reference prior. Furthermore, the semi-conjugate family allows to consider not only one prior but a variety of priors with straightforward computations; this can considerably enrich the discussion on  $VE$ .

<sup>4</sup>An easy way to perform a frequentist inference on  $\phi$  involves conditioning with respect to  $y$  and thus considering the univariate negative binomial conditional sampling distribution  $\tilde{p}_{a,b}(x | y, \phi)$  of  $x$  given  $y$ .

As we have seen, inclusion of prior information yields a posterior inference which is not acceptable in a regulatory context which requires to control the type I error with respect to the original assumed sampling distribution. Nevertheless, in either the frequentist or the Bayesian context, some prior information is obviously used in order to set up the study design, and Bayesian subjective prior predictive distributions can be helpful for this step. An example is provided in the illustration given in the next section.

### 3. ILLUSTRATION

Consider the set-up of a vaccine efficacy trial in which a new vaccine would be declared to be *effective* if we reject  $H_0$  when testing for  $H_0 : \{\phi \geq 75\%\}$  vs  $H_1 : \{\phi < 75\%\}$  by adopting the following rejection rule:

$$\begin{cases} \text{if } \pi^{\text{ref}}(H_1 | x, y) \leq 97.5\% & \text{then do not reject } H_0; \\ \text{if } \pi^{\text{ref}}(H_1 | x, y) > 97.5\% & \text{then reject } H_0. \end{cases} \quad (11)$$

Although the study statistician will, at the time of the analysis, perform inference using the reference posterior, he may want to use prior knowledge about  $\mu$  and  $\phi$  at the time of setting-up the study to determine the values of  $S$  and  $T$  (as usually done in statistics). To do so, the study statistician will first choose an a priori law in the semi-conjugate family which represents his beliefs on  $\mu$  and  $\phi$ . Then, assuming that  $x$  and  $y$  are generated following the corresponding a priori predictive law, he will choose  $S$  and  $T$  such that the predictive power of the test (11) is 80% or above, i.e.:

$$\bar{P}_{a,b,c,d}(\pi^{\text{ref}}(H_1 | x, y) > 97.5\%) \geq 80\%$$

where  $\bar{P}_{a,b,c,d}$  is a probability under which the distribution of  $x$  and  $y$  is the prior predictive law  $\bar{p}_{a,b,c,d}$  associated with values  $a, b, c, d$  in the semi-conjugate family.

We have seen that the conditional prior predictive  $\tilde{p}_{a,b}(y | \phi)$  of  $y$  given  $\phi$  is the Poisson-Gamma distribution  $T \star \mathcal{PG}(a, b)$  which does not depend on  $\phi$ , therefore the prior predictive distribution  $\bar{p}_{a,b,c,d}(y)$  of  $y$  is  $T \star \mathcal{PG}(a, b)$  too. We have also seen that conditional distribution  $\tilde{p}_{a,b}(x | y, \phi)$  of  $x$  given  $y$  is  $\phi S \star \mathcal{PG}(y + a, T + b)$ . Thus, as we can check that  $y$  and  $\phi$  are independent, it follows that the prior conditional predictive distribution  $\bar{p}_{a,b,c,d}(x | y)$  of  $x$  given  $y$  is the Beta-negative binomial distribution  $\mathcal{BNB}(y + a, d, c)$  (see Appendix C). In particular,  $\bar{p}_{a,b,c,d}(y)$  does not depend on  $S$  and  $\bar{p}_{a,b,c,d}(x | y)$  does not depend on  $S$  nor  $T$ , and as a consequence the joint prior predictive  $\bar{p}_{a,b,c,d}(x, y)$  does not depend on  $S$ .

For our illustration, imagine that our statistician decides, based for example on information obtained in previous trials, to use a  $\mathcal{G}(a, b)$  distribution with  $a = 6$  and  $b = 2000$  (corresponding to  $k = 0.2$ ). To choose the values of  $c$  and  $d$  before knowing  $S$  and  $T$ , the statistician first imposes a proportionality between  $T$  and  $S$ , for example that  $T = S$ , and then determines  $c$  and  $d$  based on the prior conditional predictive law  $\bar{p}_{a,b,c,d}(x | y)$  of  $x$  given  $y$  for a particular value of  $y$ . Here, for say  $y = 30$ , and  $S/T = 1$ , our statistician thinks that his prior beliefs on  $x$  are well represented by  $\bar{p}_{a,b,c,d}(x | y)$  when  $c = 2$  and  $d = 12$  (Figure 14 at left); the prior predictive power is displayed on the right side of Figure 14 for  $T$  ranging from 2000 to 14000. Based on this curve, our statistician will choose  $S = T = 10000$  so that  $\bar{P}_{a,b,c,d}(\text{reject } H_0) \approx 80\%$ .

The vaccine efficacy trial is then run aiming at  $S \approx T \approx 10000$ . Assume that at the time of analyzing the study results, one observes  $x = 4$  cases in the vaccine group and  $y = 28$  cases in the control group and  $S = T = 10000$ . Results of such a trial are often expressed in terms of the vaccine efficacy parameter  $VE = 1 - \phi$ . Figure 15 shows, with the terminology of [19], the *vaccine acceptability efficacy curve*, defined as the posterior probability that  $VE > v^*$  in function of  $v^*$ . Based on this curve, we can conclude to a high posterior probability of a real  $\phi$  value below 75%, or equivalently  $VE$  above 25%.

**On the reference vaccine acceptability curve** — This curve (Figure 15) illustrates the undeniable appealing interpretation of Bayesian inference in general, and how reference priors are appealing candidates to *standard* prior distributions for scientific communication.

Currently, in vaccine efficacy studies, the statistical inference on  $VE$  is performed by frequentist methods. A 97.5%-confidence interval  $[v^*, 1]$  is then given. It is well-known that many practitioners misinterpret the frequentist notion of confidence level, as if they were “Bayesian without knowing it” [29]. The reference vaccine acceptability curve would not be subject to such a misinterpretation. We could also, of course, draw such a curve when using arbitrary prior distributions. But what is remarkable in the case of the reference prior, is the ‘matching’ property of both notions of confidence level: considering the value of  $v^*$  corresponding to 97.5% on the  $y$ -axis yields an interval  $[v^*, 1]$  with a confidence level of 97.5% in the Bayesian coverage sense, and approximately 97.5% in the frequentist coverage sense.

Hence, it would be possible to abandon frequentist approaches for objective Bayesian approaches, gaining advantages of the Bayesian framework and without really breaking old conventions.

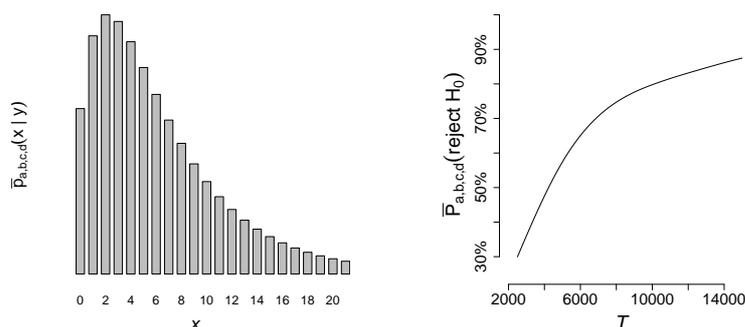


FIGURE 14. *Left*: Prior conditional predictive distribution of  $x$  given  $y = 30$  with  $a = 6$ ,  $c = 2$ ,  $d = 12$  (Beta-negative binomial distribution  $\mathcal{BNB}(36, 12, 2)$ ). *Right*: Prior predictive power of the test (11) in function of  $T$  when  $S = T$  and  $a = 6$ ,  $b = 2000$ ,  $c = 2$ ,  $d = 12$ .

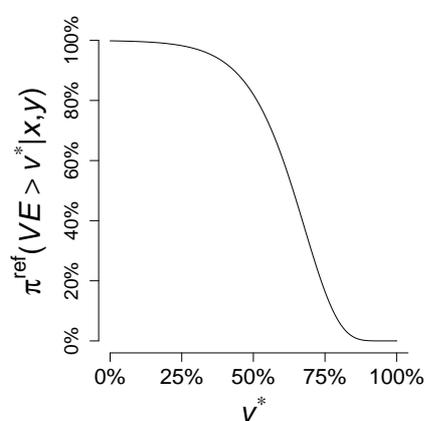


FIGURE 15. Reference vaccine acceptability curve.

## APPENDIX

## APPENDIX A. BETA PRIME DISTRIBUTION

We refer to Balakrishnan, Johnson and Kotz (1995) and to Kleiber and Kotz (2003) for this Appendix.

The *Beta prime* distribution  $\mathcal{B}'(c, d)$  with positive shape parameters  $c$  and  $d$ , is defined as the distribution of the random variable  $Z := \frac{U}{1-U}$  where the random variable  $U$  is distributed according to the Beta distribution  $\mathcal{B}(c, d)$  whose density function is  $\mathcal{B}(u | c, d) = \frac{1}{B(c, d)} u^{c-1} (1-u)^{d-1}$ ,  $0 < u < 1$ . Note that  $1-U \sim \mathcal{B}(d, c)$  and then  $Z^{-1} \sim \mathcal{B}'(d, c)$ . The *scaled Beta prime* distribution  $\rho \times \mathcal{B}'(c, d)$  is also called a *Beta distribution of the second kind* or a *Pearson type VI* distribution on  $(0, +\infty)$ .

The well-known  $F$ -distribution is a scaled Beta prime distribution, more precisely  $\mathcal{B}'(c, d) = \frac{c}{d} \times F(2c, 2d)$ . The  $\mathcal{B}'(c, d)$  distribution is also the distribution of the random variable  $X_1/X_2$  where  $X_1 \sim \mathcal{G}(c, b)$  and  $X_2 \sim \mathcal{G}(d, b)$  are two independent random variables, whatever the value of the rate parameter  $b$ . The  $\mathcal{B}'(c, d)$  distribution has density function

$$\mathcal{B}'(\psi | c, d) = \frac{1}{B(c, d)} \frac{\psi^{c-1}}{(1+\psi)^{c+d}}, \quad \psi > 0.$$

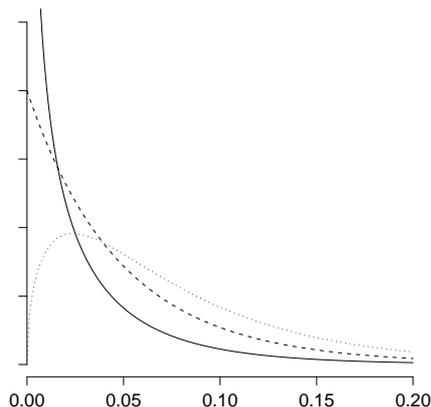


FIGURE 16.  $\mathcal{B}'(c, d)$  density function for  $d = 20$  and  $c = 0.5$  (solid),  $c = 1$  (dashed),  $c = 1.5$  (dotted).

## APPENDIX B. POISSON MODEL

We refer to Bernardo and Smith (1994) for this Appendix.

The Poisson model is the statistical model given by one observation  $x$  assumed to be realized from a Poisson distribution  $\mathcal{P}(\theta)$  with unknown mean  $\theta \geq 0$ . When the mean is written in the form  $\theta = \mu T$  where  $T > 0$  is fixed and the model is parameterized with  $\mu$ , the Gamma distributions  $\mathcal{G}(a, b)$  constitute a conjugate family of prior distributions on  $\mu$  with conjugacy relation:

$$\text{if } \mu \sim \mathcal{G}(a, b) \quad \text{then} \quad (\mu | x) \sim \mathcal{G}(a + x, b + T), \quad (12)$$

where the Gamma distribution  $\mathcal{G}(a, b)$  is parametrized in terms of its shape parameter  $a > 0$ , and its rate parameter  $b > 0$ , thus with density function  $\mathcal{G}(\mu | a, b) = \frac{1}{\Gamma(a)} b^a \mu^{a-1} \exp(-b\mu)$ . The reference prior on  $\mu$  is the improper Jeffreys' prior  $\mathcal{G}(\frac{1}{2}, 0)$ , yielding the  $\mathcal{G}(x + \frac{1}{2}, T)$  distribution as the reference posterior on  $\mu$ .

We denote by  $T \star \mathcal{PG}(a, b)$  the prior predictive distribution of  $x$  when  $\mathcal{G}(a, b)$  is the prior distribution on  $\mu$  in the context above. Since we have  $T \times \mathcal{G}(a, b) = \mathcal{G}(a, b/T)$ , then  $T \star \mathcal{PG}(a, b)$  is a Poisson-Gamma distribution:  $T \star \mathcal{PG}(a, b) = \mathcal{PG}(a, b/T)$ . Its probability masses are given by

$$T \star \mathcal{PG}(x | a, b) = \frac{(a)_x}{x!} \frac{b^a T^x}{(b + T)^{a+x}}, \quad x \in \mathbb{N},$$

where  $(a)_x = a(a + 1) \dots (a + x - 1) = \Gamma(a + x)/\Gamma(a)$  is Pochhammer's symbol for ascending factorials.

## APPENDIX C. NEGATIVE BINOMIAL MODEL

We refer to Bernardo and Smith (1994) and Johnson, Kemp and Kotz (2005) for this Appendix.

The negative binomial distribution  $\mathcal{NB}(a, \theta)$  with size parameter  $a > 0$  and proportion parameter  $\theta \in (0, 1)$  is the Poisson-Gamma distribution  $\mathcal{PG}(a, \psi)$  where  $\psi = \frac{\theta}{1-\theta}$  is the *odds parameter* of  $\mathcal{NB}(a, \theta)$ .

The negative binomial model is the statistical model given by one observation  $x$  considered to be generated from a negative binomial distribution with known size parameter  $a$  and unknown proportion parameter  $\theta$ . The Beta distributions  $\mathcal{B}(c, d)$  constitute a conjugate family of prior distributions on  $\theta$ , with conjugacy relation:

$$\text{if } \theta \sim \mathcal{B}(c, d) \quad \text{then} \quad (\theta | x) \sim \mathcal{B}(a + c, x + d). \quad (13)$$

The reference prior on  $\theta$  for the negative binomial model is the improper Jeffreys' prior  $\mathcal{B}(0, \frac{1}{2})$ , yielding  $\mathcal{B}(a, x + \frac{1}{2})$  as the reference posterior distribution on  $\theta$ .

When  $\mathcal{B}(c, d)$  is the prior on  $\theta$ , the prior predictive distribution of  $x$  is the *Beta-negative binomial* distribution  $\mathcal{BNB}(a, c, d)$  whose probability mass at  $x$  is

$$\mathcal{BNB}(x | a, c, d) = \frac{(a)_x}{x!} \frac{B(a + c, x + d)}{B(c, d)}, \quad x \in \mathbb{N}.$$

Beta-negative binomial distribution is also known as type IV general hypergeometric distribution.

This work was initiated while the first author was working at GSK Biologicals (Rixensart, Belgium) in connection with Université de Strasbourg.

## REFERENCES

- [1] Balakrishnan, N., Johnson, N.L., Kotz, S., 1995. Continuous Univariate Distributions, Vol 1, 2nd edition. John Wiley, New York.
- [2] Bayarri, M. J., Berger, J., 2004. The Interplay of Bayesian and Frequentist Analysis. *Statist. Sci.* 19, 58–80
- [3] Berger, J.O., Bernardo, J.M., 1989a. Ordered Group Reference Priors With Applications to Multinomial and Variance Component Problems. Technical report Dept. of Statistics, Purdue University.
- [4] Berger, J.O., Bernardo, J.M., 1989b. Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.*, 84, 200–207.
- [5] Berger, J.O., Bernardo, J.M., 1992a. Ordered group reference priors, with applications to multinomial problems. *Biometrika*, 79, 25–37
- [6] Berger, J.O., Bernardo, J.M., 1992b. On the development of reference priors. *Bayesian Statistics 4* (Bernardo, J.M., Berger, J.O., Dawid, A.P., and Smith, A.F.M., eds.) Oxford: University Press, 35–60 (with discussion).
- [7] Berger, J.O., Bernardo, J.M., Sun, D., 2009. The formal definition of reference priors. *Annals of Statistics*, 37, to appear.
- [8] Berger, J.O., Sun, D., 1998. Reference priors with partial information. *Biometrika*, 85, 1, 55–71.
- [9] Berger, J.O., Yang, R., 1997. A catalog of noninformative priors. ISDS Discussion Paper, 97–42, Duke Univ.
- [10] Bernardo, J.M., 1979. Reference posterior distributions for Bayesian inference (with discussion). *J. R. Statist. Soc.*, B, 41, 113–148.
- [11] Bernardo, J.M., 1997. Noninformative priors do not exist: A discussion. *J. Statistical Planning and Inference*, 65, 159–189 (with discussion).

- [12] Bernardo, J.M., Ramon, J.M., 1998. An introduction to Bayesian reference analysis: inference on the ratio of multinomial parameters. *J. Roy. Statist. Soc., D*, 47, 101–135.
- [13] Bernardo, J.M., 2005a. Reference Analysis. *Handbook of Statistics* (Dey, D.K., and Rao, C.R., eds.), 25, 17–90. Elsevier.
- [14] Bernardo, J.M., 2005b. Intrinsic Credible Regions: an Objective Bayesian approach to Interval Estimation (with discussion). *Test*, 14, 317–384.
- [15] Bernardo, J.M., Smith, A.F.M., 1994. *Bayesian Theory*. Wiley, Chichester.
- [16] Berry, D.A., Wolff, M. C., Sack, D., 1994. Decision Making during a Phase III Randomized Controlled Trial. *Controlled Clinical Trials*, 15, 360–378.
- [17] Brown, L.D., Cai, T.T., DasGupta, A., 2001. Interval estimation for a binomial proportion (with discussion). *Statistical Science*, 16, 101–133.
- [18] Brown, L.D., Cai, T.T., DasGupta, A., 2002. Confidence Intervals for a Binomial Proportion and Edgeworth Expansions. *The Annals of Statistics*, 30, 1, 160–201.
- [19] Chu, H., Halloran, M. E., 2004. Bayesian estimation of vaccine efficacy. *Clinical Trials*, 1, 306–314.
- [20] Cousins, R.D., 1998. Improved central confidence intervals for the ratio of Poisson means. *Nucl. Instrum. and Meth. in Phys. Res., A* 417, 391–399.
- [21] Datta, G. S., Mukerjee, R., 2004. *Probability Matching Priors: Higher Order Asymptotics*. Springer, New-York.
- [22] Ewell, M., 1996. Comparing Methods for Calculating Confidence Intervals for Vaccine Efficacy. *Statistics in Medicine*, 15, 2379–2392.
- [23] Halloran, M.E., Longini, I.M. Jr, Struchiner, C.J., 1999. Design and interpretation of vaccine field studies. *Epidemiological Reviews*, 21, 1, 73–88.
- [24] Johnson, N.L., Kemp, A.W., Kotz, S., 2005. *Univariate Discrete Distributions*. 3rd edition. New York: John Wiley
- [25] Kleiber C., Kotz S., 2003. *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley.
- [26] Kass, R.E., Wasserman, L., 1996. The selection of prior distributions by formal rules. *J. Am. Statist. Assoc.*, 91, 1343–1370.
- [27] Krishnamoorthy, K., Thomson, J., 2004. A more powerful test for comparing two Poisson means. *J. Statist. Plann. Inference*, v119, 23–35.
- [28] Krishnamoorthy, K., Lee, M., 2009. Inference for functions of parameters in discrete distributions based on fiducial approach: Binomial and Poisson cases. *J. Statist. Plann. Inference*, v140, 1182–1192.
- [29] Lecoutre, B., 2006. And if you were a Bayesian without knowing it? *Bayesian Inference and Maximum Entropy Methods In Science and Engineering*. AIP Conference Proceedings, Volume 872, 15–22.
- [30] Lehmann, E.L., Romano, J.P, 2005. *Testing Statistical Hypotheses*, 3rd ed., New York, NY: Springer.
- [31] Liseo, B., 1990. Elimination of Nuisance Parameters with Reference Noninformative Priors. Technical report #90-58C, Purdue University, Department of Statistics.
- [32] Price R.M., Bonett, D.G., 2000. Estimating the ratio of two Poisson rates. *Computational Statistics and Data Analysis*, 34, 345–356.
- [33] Robert C., 2001. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer Texts in Statistics. 2nd ed.
- [34] Robins, J., Wasserman, L., 2000. Conditioning, likelihood and coherence: A review of some foundational concepts. *J. Amer. Statist. Assoc.*, 95, 1340–1346.
- [35] Sahai, H., Khurshid, A., 1993. Confidence intervals for the ratio of two Poisson means. *The Mathematical Scientist*, 18, 43–50.
- [36] Stamey, J.D., Young, D.M., Bratcher, T. L., 2006. Bayesian Sample-size Determination for One and Two Poisson Rate Parameters with Applications to Quality Control. *Journal of Applied Statistics*, 33, 6, 583–594.