# Informative priors or noninformative priors? A Bayesian re-analysis of binary data from Macugen phase III clinical trials

Ding-Geng (Din) Chen[a,b,c], Naitee Ting[d], and Shuyen Ho[e]

[a]School of Social Work, University of North Carolina, Chapel Hill, North Carolina, USA; [b]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA; [c]Department of Statistics, University of Pretoria, Pretoria, South Africa; [d]Boehringer Ingelheim Pharmaceuticals, Inc. Ridgefield, Connecticut, USA; [e]PAREXEL InternationalE, Durham, North Carolina, USA

**ABSTRACT**

It has became more popular in the recent statistical literature to see Bayesian approaches for clinical trials, such as assurance calculations for study designs and the use of posterior probability for data analyses. When applying Bayesian analysis to clinical trial data, one common question is to use informative priors or noninformative priors. In order to explore this question, we looked for existing clinical trial data with a simple structure and a simple clinical endpoint so that the Bayesian re-analyses can be easily performed. We came across the published Phase III Macugen® data that were suitable for this exploration. In this manuscript, the Macugen Phase III development program was described, the primary data for the two Phase III pivotal studies were re-analyzed using Bayesian applications with informative priors and noninformative priors. These re-analysis results were summarized, compared, and discussed.

**ARTICLE HISTORY**

Received 3 April 2015
Accepted 13 November 2015

**KEYWORDS**

Age-related macular degeneration; Beta-binomial; Bayesian posterior distribution; Frequentist *p*-values; Hypergeometric function; Monte-Carlo simulation

## 1. Introduction

Prospective cohort studies have been widely used to help understand public health problems long before the randomized controlled clinical trials. In 1962, the Kefauver Harris Amendment required pharmaceutical manufacturers to demonstrate drug efficacy using clinical trials. This amendment revolutionized the drug development and drug evaluation processes. Drugs and biologics approved by the Food and Drug Administration (FDA) after 1962 have been mostly based on efficacy and safety evidences established from clinical trials.

The concept of randomized controlled clinical trial started with the randomization theory developed by R.A. Fisher. One of the famous applications of this concept was the example of "Lady Tasting Tea" (Salsburg, 2001). The basic decision rule is based on statistical hypothesis testing in a typical placebo controlled trial where patients are randomized to receive test drug or placebo. The null hypothesis is that the test drug is not different from placebo. However, if the treatment difference between the two groups is large enough so that the corresponding observed *p*-value is less than alpha, then the null hypothesis (that the test drug is not different from placebo) is rejected, and a statistically significant treatment difference can be claimed.

**CONTACT** Naitee Ting naitee.ting@boehringer-ingelheim.com Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, CT ,77 860USA.

This entire setting is based on the traditional Frequentist approach. For more than half of century, drug and biologics have been approved under this Frequentist frame work.

In recent years, Bayesian approaches are more widely applied in the design and analysis of clinical trials. For example, Chen and Kim (2009) presented Bayesian analyses of melanoma data from a clinical trial conducted by Eastern Cooperative Oncology Group (ECOG), and prostate cancer data from a clinical trial conducted at St. Anne's Hospital in Fall River, Massachusetts, utilizing cure rate models. Tan et al. (2002) explored a Bayesian approach to the design, analysis, and interpretation of Phase II clinical trials conducted at the National Cancer Centre in Singapore on the activity of Gemcitabine in patients with metastatic nasopharyngeal carcinoma. Connor and Berry (2005) used Bayesian methods to analyze adverse event data from medical device clinical trials. Berry et al. (2010) presented an updated summary on Bayesian adaptive clinical trials.

In clinical development of new medicinal products, the scientific learning is based on data accumulated over time. In other words, the more time it evolves, the more data are available and the more can be learned about the product under development. One key advantage of Bayesian methodology is that during the process of data accumulation, learnings from previous studies can be incorporated to the current study via the use of priors. Hence after every study, the posterior analysis reflects the learnings from the current study, in combination with the prior knowledge. In this sense, when the current study has a smaller sample size, while the prior being very informative, the posterior results can be affected more by the prior, as opposed to a larger study with a noninformative prior, in which case the prior most likely does not have a major impact.

As we follow this line of thinking, consider Phase I trials which tend to have smaller sample sizes and therefore a prior may have more influence to the posterior analysis. On the other hand, Phase III trials tend to have larger sample sizes, so the observed data could dominate the posterior results. On this basis, it seems that a non-informative prior could provide results that are more consistent with the Frequentist approach for Phase III trials, given that Bayesian analysis for Phase III is not yet popular. With this in mind, we decided to explore Bayesian analyses using some Phase III data sets to compare the performance of informative priors versus noninformative priors. In addition, we also thought it would be simpler and of interest to perform such Bayesian analyses using binary data. From our knowledge and experiences, the Phase III data from the Macugen submission fit these characteristics.

This manuscript provides detailed description of the Macugen development program in Section 2. Results of Bayesian re-analyses for Macugen data are presented in Section 3 where both informative priors and noninformative priors are used and compared. Finally, a discussion is provided in Section 4.

## 2. The Macugen clinical trial background

Age-related macular degeneration (AMD) is a devastating disease. Macula is a small area on the retina on the back of each human eye. It formulates the central vision, and transmits the image to the human brain. When macula begins to degenerate, the patient with AMD starts by seeing a black dot in the center of his/her vision. Eventually the black dot gets larger as the disease progresses, and the patient will go blind after losing all the visibility from the macula. There are two forms of AMD: dry AMD and wet AMD. The dry form of AMD results from atrophy of the retinal pigment epithelial layer below the retina, which causes vision loss through loss of photoreceptors in the central part of the eye. There is not yet any treatment for the dry AMD but research is ongoing and likely produces some treatments in the coming

years. The wet form of AMD occurs when a protein in the body known as vascular endothelial growth factor (VEGF) causes abnormal vessel growth in the macula area. These abnormal blood vessels push up through the macula, causing cracks. These vessels can begin to bleed, causing changes in vision. Bleeding vessels lead to swelling and bleeding in the macula, causing blurry, wavy, or decreased vision. This can result in the loss of central vision.

Macugen was developed as a VEGF inhibitor—in other words, it inhibits the activities of VEGF hence it would help reduce the abnormal vessel growth, and eventually Macugen would help slow down the deterioration of vision loss. On the retina, there is only one layer of photoreceptors, these cells are connected directly to the human brain so that the image can be formulated. Hence the attempt of delivering Macugen into the eye is just like the attempt of delivering the drug into the brain. It is well-known that there exists a blood-brain barrier in the human body; that is, drugs circulating in the blood may not be able to reach the brain. Hence one of the major challenges in developing Macugen would be the drug delivery. Given the devastating indication and the treatment being experimental, it was decided to deliver Macugen by direct injection into the eye with AMD.

In the clinical development of Macugen for treating patients with AMD, there were three Phase I trials – NX109-01, open-label single dose with 15 subjects; EOP1000, open-label multiple dose with 10 subjects; and EOP1001, open-label multiple dose with PDT as the concomitant treatment and 11 subjects. Based on these early results, the project team learned the treatment effect from Macugen (both benefits and risks) and proposed the range of doses as well as frequency of how to deliver the drug into patients' eyes. Subsequently a pair of Phase III trials were designed and conducted - studies EOP1003 and EOP1004 for patients with AMD treated for 2 years.

The primary endpoint is the proportion of responders based on the change in visual acuity from baseline to week 54 (approximately one year after randomization). For each patient, the eye that is suffering from AMD is identified as the study eye; the other eye is referred to as the fellow eye. When evaluating visual acuity of the study eye, the patient covers his or her fellow eye, and identifies the letters from the Early Treatment of Diabetic Retinopathy Study (ETDRS) eye chart using the study eye.

The figure below is an ETDRS eye chart where there are five letters on each row. The chart was designed such that the relative size of letters between every three consecutive rows signifies a doubling of the visual angle. Hence a loss of 15 letters (or three rows) indicates a clinically meaningful loss of visual acuity. At each clinic visit of the ophthalmologist, the number of letters identified by the patient represents the visual acuity measure of that visit. The number of letters a patient can see from the study eye at week 54 is compared with the number of letters obtained from baseline. If the reduction in visual acuity is less than 15 letters, then this patient is considered as a responder (see Figure 1). Because the visual acuity of an AMD patient is deteriorating over time, non-responders could suffer from loss of 15 or more letters by week 54.

Again the primary endpoint is the proportion of responders (losing visual acuity by less than 15 letters from baseline) in each treatment group. Macugen efficacy is assessed by comparing the primary endpoint of each dose group against the proportion of responders from the sham (placebo) group.

The frequency of dosing is one injection for every 6 weeks. For a 54-week primary endpoint, each patient will receive nine injections. There are three dose groups plus a sham. Study medication is packaged in a syringe with 0.3 mg, 1 mg, and 3 mg of Macugen, plus a syringe serves as a sham injection. Although the primary concern in developing Macugen is the fact that injection in the study eye could potentially terrify AMD patients, study results

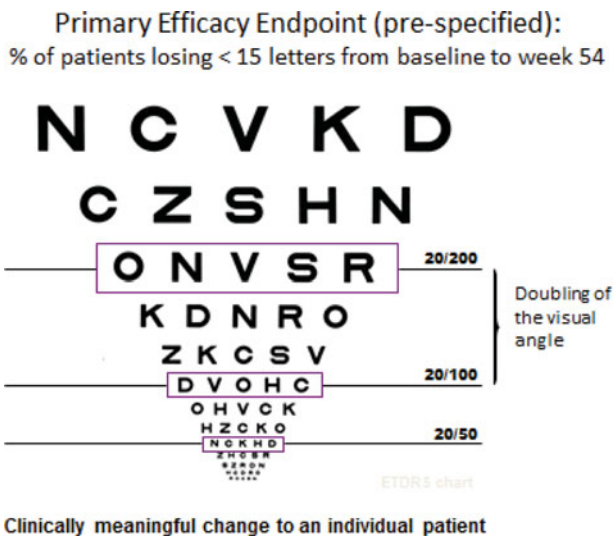


**Figure 1.** Early treatment of Diabetic Retinopathy Study (ETDRS) eye chart.

demonstrate that the compliance is very high. The mean number of injections actually received by each subject is greater than 8.5 in EOP1003, and greater than 8.4 in EOP1004, out of a protocol specified nine total injections. The high compliance rate helps with the credibility of the analytical results obtained from this pair of pivotal studies.

Based on the statistical analysis plan, the primary analysis is using the intention to treat data set. All randomized subjects are included in the analysis and hence are analyzed as randomized. Last observation carried forward was used in the two Phase III studies to handle missing data introduced by patient dropouts.

Before Macugen was approved in 2004, the only available treatment for AMD patients was photodynamic therapy (PDT). Hence in the design of EOP1003 and EOP1004, these two studies were stratified by patients' history of PDT usage. In addition, the design was further stratified by center and by lesion type. At baseline, patients could be classified into one of the three lesion subtypes—Predominantly Classic CNV, Minimally classic CNV, and Occult with no classic CNV. In the primary analysis of the proportion of responders, the pre-specified statistical approach was Cochran–Mantel–Haenszel (CMH) test, adjusted by history of PDT usage and baseline lesion type. Because there were too many centers, it was decided that the CMH test would not adjust for center.

For the three dose comparisons (each dose is compared against sham), the Hochberg procedure was selected for the multiple comparison adjustment procedure (MCP). In these two Phase III studies, each dose is compared with sham so there would be three p-values—let $p_1$, $p_2$, and $p_3$ denote the pairwise comparison between the 3 mg, 1 mg, and 0.3 mg dose group vs. sham, respectively. Hochberg procedure starts with the largest $p$-value and compares it with alpha. If max ($p_1$, $p_2$, $p_3$) $<\alpha$, then claim all three doses are statistically significantly different from sham. If max ($p_1$, $p_2$, $p_3$) $>\alpha$, then look for the second largest $p$-value, and compares it with alpha divided by two ($\alpha/2$). If the second largest $p$-value is less than $\alpha/2$, then claim the two doses with smaller p-values are significant. If the second largest p-value is greater than $\alpha/2$, then the smallest $p$-value is compared with $\alpha/3$. If the smallest $p$-value is less than $\alpha/3$ then declare the dose associated with that dose is significant, and no other dose is different from sham.

**Table 1.** Summary Statistics for Frequentist's Analysis

| | Study EOP 1003 | | | | Study EOP 1004 | | | |
|---|---|---|---|---|---|---|---|---|
| | Pegaptanib Sodium | | | | Pegaptanib Sodium | | | |
| | 0.3 mg | 1 mg | 3 mg | Sham | 0.3 mg | 1 mg | 3 mg | Sham |
| ITT population | $N = 150$ | $N = 154$ | $N = 153$ | $N = 152$ | $N = 144$ | $N = 146$ | $N = 143$ | $N = 144$ |
| | Primary Efficacy Endpoint—% of Patients losing < 15 letters VA, Baseline—Week 54 | | | | | | | |
| Vision loss < 15 letters | 73% | 75% | 69% | 59% | 67% | 66% | 61% | 52% |
| p-value | 0.0105 | 0.0035 | — | | 0.0031 | 0.0273 | 0.1294 | |

Based on the Hochberg procedure, both the 0.3 mg and the 1 mg doses of Macugen are statistically significantly different from sham for the EOP1003 results, but only the 0.3 mg dose is statistically significantly different from sham for the EOP1004 results, as seen in Table 1.

The main reason for this choice of Hochberg MCP is that it was not clear about which dose might produce the best response. If the Macugen dose–response relationship is monotonic within the range of 0.3 mg to 3 mg, then a gate-keeping procedure could have been recommended. A gate-keeping procedure is a sequential testing procedure in which only when the preceding significant test is rejected, the next test in the pre-specified sequence can be performed. Because the monotonic dose response assumption was too strong under a Phase III confirmatory setting, the project team finally decided to use Hochberg procedure. This procedure is flexible, more powerful than Bonferroni, and it is very easy to communicate with nonstatisticians.

## 3. Bayesian analysis for macugen clinical trials

### *3.1. The methods*

We now consider applying the beta-binomial posterior distribution to compare and select superior treatments in clinical trials with a binomial endpoint from the Bayesian perspective. In this case, we are interested in deciding which of the two treatments denoted by $X$ and $Y$, is superior by calculating the probability $P(X > Y)$. If $P(X > Y)$ is greater than some threshold, say 0.99, then we conclude treatment $X$ is superior to treatment $Y$. The Macugen data used in this manuscript were obtained from the public domain, which do not include individual patient level data. Therefore the two covariates, history of PDT usage and baseline lesion type, used in the analyses are not available for the Bayesian re-analyses. However, as can be seen in subsequent Sections, the Bayesian re-analyses results are strong enough without these two covariates, so we feel that it still serves our exploration purpose without incorporating the covariates into the Bayesian re-analyses.

In clinical trials with binary/binomial endpoint, say $W \sim \text{Binomial}(w; n, p)$ where $n$ is the total sample size and $p$ is the probability parameter, then the likelihood function can be written as:

$$L\left(w|n, p\right) \propto p^{w}\left(1-p\right)^{n-w}$$

Prior information can then be incorporated into this likelihood to derive the Bayesian posterior distribution. Two options are usually employed to formulate the prior distribution with one being "noninformative" when there is no prior information available and another being "informative" whenever there is prior information that can be incorporated.

When there is no prior information available, in this case, we use the Jeffery's prior $\pi\ (p) = B(0.5, 0.5)$ as a "noninformative" prior. As pointed out in Zhu and Lu (2004), this Jeffery's prior is less informative than the Uniform (0, 1) = Beta(1, 1) prior, in terms of the Bayesian posterior estimator being less different from the MLE estimator of $p$. This conjugate prior leads to the posterior distribution in the same beta-binomial family as follows:

$$\pi(p|w) \propto L(w|n, p) \times \pi\ (p) = p^{w}(1-p)^{n-w} p^{0.5-1}(1-p)^{0.5-1}$$

which is a beta distribution with parameters w+0.5 and $n$-$w$+0.5 and thereafter denoted by B(w+0.5, $n$-$w$+0.5) where B(a, b)=$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

On the other hand, when indeed there exists prior information, the commonly used prior for this probability parameter $p$ is the beta distribution of B($a$, $b$) as a conjugate prior where $a$ and $b$ are the two parameters associated with this beta distribution. With this beta conjugate prior, the Bayesian posterior distribution can be derived as follows:

$$\pi(p|w) \propto L(w|n, p) \times \pi(p) \propto p^{w}(1-p)^{n-w} p^{a-1}(1-p)^{b-1} = p^{a+w-1}(1-p)^{b+n-w-1} \tag{1}$$

which is again B($a + w$, $b + n - w$). Note that this informative posterior distribution includes the "noninformative" posterior distribution when $a = b = 0.5$.

Therefore in the comparison of two treatments, $X$ and $Y$, with posterior distributions of X$\sim$ B($s_X$, $t_X$) and Y$\sim$ B($s_Y$, $t_Y$) where ($s_X$, $t_X$) and ($s_Y$, $t_Y$) are the posterior Beta distribution parameters for $X$ and $Y$ corresponding to $s_X = a_X + w_X$, $t_X = b_X + n_X - w_X$ and $s_Y = a_Y + w_Y$, $t_Y = b_Y + n_Y - w_Y$ in equation (1), respectively. From basic probability theory, the probability that $X > Y$ is given by

$$P\ (X > Y) = \int_{-\infty}^{+\infty} f_X\ (t)\, F_Y\ (t)\, dt$$

for two continuous random variable $X$ and $Y$, where $f_X$ and $F_X$ are the PDF and CDF for random variable $X$, respectively. Typically, there is no analytical closed form for this integration; therefore, we demonstrate how to implement this integration by two different approaches. One is to employ the hypergeometric function for this integration and the other is to use the common Bayesian Monte-Carlo simulations.

For the first approach, we make use of the fact that both $X$ and $Y$ are distributed by Beta-distributions. It can be shown with some mathematical manipulations [see, for example, Cook (2005)] that

$$\begin{aligned} P(X > Y) &= \int_0^1 \frac{x^{s_x-1}(1-x)^{t_x-1}}{B(s_x, t_x)} \left( \int_0^x \frac{y^{s_y-1}(1-y)^{t_y-1}}{B(s_y, t_y)} dy \right) dx \\ &= \frac{B(sx+sy, tx+ty)}{B(sx, sy)B(sy, ty)sy}\, {}_3F_2 \left( \begin{matrix} sx+sy,\ sy+ty,\ 1 \\ sy+1,\ sx+ty+sy+ty \end{matrix} \middle|\ 1 \right) \end{aligned} \tag{2}$$

Where 3 $F_2$ is the hypergeometric function with upper parameters ($s_X + s_Y$, $s_Y + t_Y$, 1) and lower parameters ($s_Y + 1$, $s_X + t_X + s_Y + t_Y$). Therefore the calculation of the probability P($X > Y$) can be easily done with the hypergeometric function, which can be implemented in R software by calling the *hypergeo* library.

For the Bayesian Monte-Carlo simulation approach, the general algorithm is to sample the data from both Beta distributions and evaluate the total number of times where treatment $X$ is superior to treatment $Y$. It is implemented by the following steps:

**Table 2** Probabilities of "Dose Treatment" is superior than "Sham"

| $P(Dose_i > Sham)$ | Study 1003 | Study 1004 |
|---|---|---|
| 0.3 mg | 0.995 | 0.994 |
| 1 mg | 0.998 | 0.991 |
| 3 mg | 0.971 | 0.932 |

1) Define counter $I$ for iteration and the total number of simulations $N$, (say $N =$ 1,000,000). Set $I = 0$ and start looping;
2) Sample $X$ from $B(s_X, t_X)$ and $Y$ from $B(s_Y, t_Y)$. Then evaluate whether $X > Y$;
3) If it is true, increment I by 1, otherwise do not increment $I$;
4) Loop steps 2 and 3 N times. The $P(X > Y)$ is calculated by $I/N$.

Both approaches together with the data analyses are implemented in *R* and the *R* code can be obtained from the authors.

### 3.2. Analysis with noninformative prior

The percentages of responders of Macugen can be seen from Table 1 which are 73%, 75%, 69%, and 59% for study 1003 and 67%, 66%, 61%, and 52% for study 1004 corresponding to the doses of 0.3 mg, 1 mg, 3 mg, and sham, respectively. The total number of patients (ITT, i.e. $n$ in equation 1) are 150, 154, 153, and 153 for study 1003 and 144, 146, 143, and 144 for study 1004. The total number of responders (i.e. $w$ in equation 1) can then be calculated as 110, 116, 106, and 90 for study 1003 and 96, 96, 87, and 75 for study 1004.

With the "non-informative" prior where $a = b = 0.5$ in equation (1), the computation and implementation of equation (2) can be easily done using *R* package *hypergeo* (developed by Robin K. S. Hankin), which implemented the hypergeometric function for complex numbers. The step-by-step implementation using this R library can be found in Chen and Peace (2011, Section 9.4.4), where an *R* function of calculating equation (2) is created with some examples to illustrate how to use this *R* function. With this *R* function, the probabilities can be calculated, which is summarized in Table 2 for the case of "non-informative" prior. In addition, as expected, the Bayesian Monte-Carlo simulations produced almost identical results (to the 5th decimal digit) with $N = 1{,}000{,}000$. As such, those results are not duplicated in this table.

It can be seen from Table 2 that (1) the probabilities of 0.3 mg dose treatment being better than sham, are 0.995 and 0.994 for trials 1003 and 1004, respectively; (2) the probabilities of 1 mg dose treatment being better than sham are 0.998 and 0.991 for trials 1003 and 1004, respectively; (3) the probabilities of 3 mg dose treatment being better than sham are 0.972 and 0.933 for trials 1003 and 1004, respectively. This indicates that the probabilities for each of the three dose treatments being superior to sham are all greater than 0.9. Furthermore, for the 0.3 mg and 1 mg doses, these probabilities are greater than 0.99.

A graphical illustration for the posterior distribution of these four treatments and the "non-informative" prior distribution is in Figure 2. From this figure, we can observe that both 0.3 mg and 1 mg treatments overlap heavily and farther away from sham than the 3 mg treatment. This is consistent with the probability calculations and further supports that the all three treatments are better than the sham treatment, but the 0.3 mg and 1 mg treatments are numerically better than the 3 mg treatment.

If a cutoff at 0.99 is selected for this probability as the criterion for claiming superiority treatment, then both 0.3 mg and 1 mg treatments are superior to sham for both studies. However, this conclusion would be different from the Frequentist results reported in Section 2. It is a common knowledge that regulators typically require statistically significant

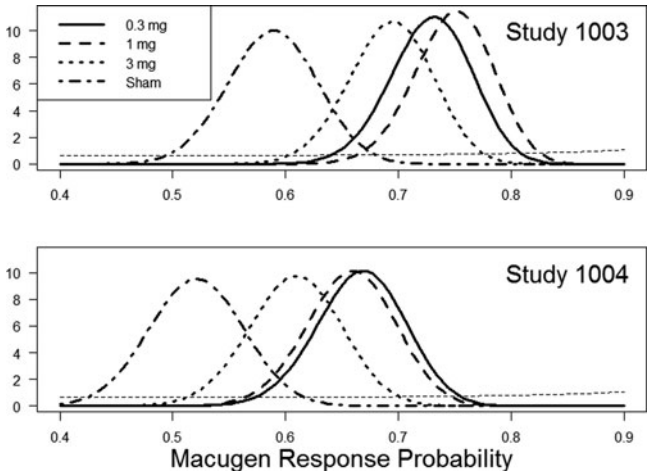


**Figure 2.** Plot for the noninformative prior of Beta (0.5,0.5) (the almost horizontal dashed line) and the four posterior distributions (beta)for four treatment groups identified in the legend.

results (multiplicity adjusted two-sided 5% significance level) from two pivotal phase III trials for approval. In this case, only 0.3 mg dose was approved for treatment by FDA and other international regulatory authorities.

Actually it can be expected, with a true non-informative prior, that the Bayesian posterior inference to be consistent with the Frequentist MLE inference. In this particular case, the trend of inference is consistent but the conclusion is different due to regulatory criteria. If the Bayesian posterior probability cutoff for superiority claim is 0.992 (instead of 0.99), then the conclusion would be consistent with the Frequentist's. This brings up a couple of issues. First the Bayesian criterion has not been well established and accepted by the statistical community, primarily because the posterior probabilities depend on the choice of priors. Second, the Bayesian approach typically does not consider multiplicity adjustments as it is a Frequentist concept and, as a result, even with identical evidence (presumably in the case of using a true non-informative prior), the inference will be different if multiplicity is considered in the Frequentist approach but not in the Bayesian approach.

### 3.3. Analysis with informative priors: Bayesian sensitivity analysis

To perform Bayesian analysis with informative prior using equation (1), the informative priors can be incorporated into the posterior from the prior parameters $a$ and $b$ in equation (1). For Macugen, we do not have prior information for the probability parameter $p$. However, we can perform a Bayesian sensitivity analysis assuming that we would have prior information for this $p$ as a Bayesian robustness analysis. To conduct this sensitivity analysis, we reparametrize the Beta distribution for ease of calculation and explanation.

As seen in Section 3.1, for any prior Beta distribution of $p$, B($a$,$b$) with parameters $a$ and $b$, its mean and variance can be shown as $\mathrm{E(p)} = \frac{a}{a+b}$ and $\mathrm{V(p)} = \frac{ab}{(a+b)^2(a+b+1)}$. We reparametrize this Beta distribution using $\mathrm{E(p)} = \frac{a}{a+b} = p_0$ as the prior mean and $m = a + b$ which lead to $a = mp_0$ and $b = m(1 - p_0)$. Then the prior mean and prior variance become $\mathrm{E}(p) = p_0$ and $\mathrm{V}(p) = \frac{p_0(1-p_0)}{m+1}$ where $m$ serves as the prior "precision" or prior "sample size" parameter with the fact that increasing $m$ decreases V(p).

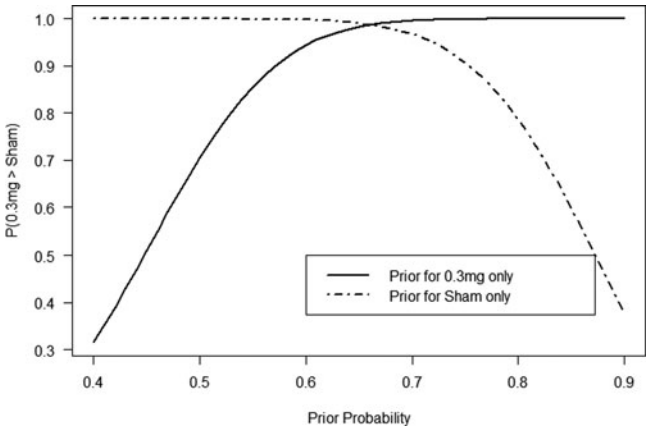


**Figure 3.** The probability of 0.3 mg is superior to sham under different prior specifications.

With this reparameterization, we can calculate probability of treatment (denoted by $X$) is superior to the control (denoted by $Y$) under different specifications of the prior distributions for $X$ and $Y$, which is then used to investigate the sensitivity of $p_0$ for $X$ and $Y$.

Although there were two Macugen trials and each involved three doses, we will perform this sensitivity analysis only for Macugen 0.3 mg (X) to sham (Y) for Study 1003. The same can be done for other doses and for Study 1004.

In Study 1003, the 0.3 mg (i.e., $X$) has 73% response rate and the sham (i.e., $Y$) has 59% response rate under the total sample size of $n_X = 150$ and $n_Y = 152$, respectively. The corresponding number of responses are then $w_X = 110$ and $w_Y = 90$, respectively. We set the prior "sample size" parameter $m$ close to this observed sample size as $m = 150$ and vary the prior probability $p_0$ from 40% to 90% by 5% for both $X$ and $Y$, to calculate the probability of 0.3 mg ($X$) being superior to sham (i.e. $Y$) using the analytical equation (1).

There are three scenarios to be investigated. The first two scenarios are to vary one treatment with another treatment fixed. Scenario one fixes sham ($Y$) and incorporates prior for 0.3 mg ($X$), which is to investigate the prior sensitivity for 0.3 mg only and scenario two does the opposite. The third scenario is to vary the prior probability for both 0.3 mg and sham.

In scenario one, prior probabilities of $p$ from 0.4 to 0.9 are incorporated into 0.3 mg using equation (1). The posterior probabilities of $P(X > Y)$ are (0.315, 0.510, 0.705, 0.856, 0.944, 0.983, 0.996, 0.999, 1, 1, 1) corresponding to the prior probabilities (0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90), respectively. The solid line in Figure 3 illustrates these calculations. It is obvious that the posterior probabilities are high when the prior is close to or above the observed response probability of 0.73 (73%). Specifically when the prior probability is greater than 0.6, the posterior probabilities of $P(X > Y)$ is greater than 0.95, indicating that the 0.3 mg is superior to the sham.

From this observation, we see that a consistent (similar to the observed data) or strong prior favoring active treatment, can strengthen and enhance treatment inference, although in this case, the data already provide strong information for treatment inference. As the prior probabilities decrease the posterior probabilities also decrease, and when the prior probabilities fall below 0.45, the posterior probabilities fall below 0.5. We do see, from this observation, that when the prior probabilities drip down below certain level, say, 0.5, the posterior probabilities (below 0.7) start to be weakened, which then also weakened the treatment inferences. In light of this, together with the observation from strong priors above, we believe the prior information needs to reflect the true knowledge and information about the new treatment.

**Table 3.** The probability that 0.3 mg (X) is superior to sham (Y) under different prior probabilities $p_0$ for X and Y.

| Prior Probability for 0.3 mg (X) | Prior Probability for Sham (Y) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.40 | 0.45 | 0.50 | 0.55 | 0.60 | 0.65 | 0.70 | 0.75 | 0.80 | 0.85 | 0.90 |
| 0.40 | 0.957 | 0.867 | 0.692 | 0.455 | 0.233 | 0.088 | 0.024 | 0.004 | 0.001 | 0.000 | 0.000 |
| 0.45 | 0.990 | 0.958 | 0.869 | 0.694 | 0.457 | 0.232 | 0.086 | 0.022 | 0.004 | 0.000 | 0.000 |
| 0.50 | 0.998 | 0.991 | 0.960 | 0.872 | 0.698 | 0.458 | 0.230 | 0.084 | 0.021 | 0.003 | 0.000 |
| 0.55 | 1.000 | 0.999 | 0.991 | 0.962 | 0.875 | 0.701 | 0.459 | 0.228 | 0.081 | 0.019 | 0.003 |
| 0.60 | 1.000 | 1.000 | 0.999 | 0.992 | 0.964 | 0.879 | 0.706 | 0.460 | 0.225 | 0.077 | 0.017 |
| 0.65 | 1.000 | 1.000 | 1.000 | 0.999 | 0.993 | 0.966 | 0.884 | 0.711 | 0.461 | 0.221 | 0.073 |
| 0.70 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.994 | 0.969 | 0.890 | 0.716 | 0.462 | 0.217 |
| 0.75 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.995 | 0.972 | 0.896 | 0.723 | 0.462 |
| 0.80 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.976 | 0.904 | 0.731 |
| 0.85 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.979 | 0.912 |
| 0.90 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.983 |

That is, it needs to be credible, in order to avoid misleading or unreliable Bayesian posterior inferences.

Similarly, in scenario two prior probabilities of $p$ from 0.4 to 0.9 are incorporated into sham using equation (1). The posterior probabilities of $P(X > Y)$ are (1, 1, 1, 1, 0.998, 0.991, 0.968, 0.908, 0.787, 0.599, 0.378) corresponding to the prior probabilities (0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90), which is illustrated in the dashed line in Figure 3. Again, the posterior probabilities are strong whenever the prior is not far from the observed probability of 0.59 (59%). Specifically when the prior probability is less than 0.75, the posterior probabilities of $P(X > Y)$ is greater than 0.90, indicating that the 0.3 mg is superior to the sham. Similar to scenario one, a weak prior for the sham, can strengthen active treatment inference. When the prior probabilities are above 0.8, the posterior probabilities are getting below 70%. One can argue that this is unrealistic because, if the sham is believed to have such a high prior probability, then the sham itself could be an effective treatment already. However, this also endorses that the prior needs to be credible; otherwise the Bayesian posterior inferences can be misleading or unreliable.

In scenario three, we incorporate prior information for both the 0.3 mg and sham with the prior probabilities from 0.4 to 0.9 by 0.05. Table 3 gives the probabilities of $P(X > Y)$ under different combinations of prior probabilities for both 0.3 mg and sham. It can be seen from this table that the posterior probabilities of $P(X > Y)$ for the combined prior information are quite strong whenever the prior information is close to the observed responses of 73% for 0.3 mg and 59% for sham, higher than 73% for 0.3 mg, or lower then 59% for sham.

Overall these three scenarios indicate that, with a reasonable informative prior, the Bayesian inferences will be consistent with those of the Frequentists. We believe this is likely due to the fact that these two pivotal trials were powered to show significant Frequentist results using significant $p$-values.

Although we do see that Bayesian informative priors can be used to strengthen and enhance statistical inferences above the Frequentist approach, we believe that the prior information needs to be credible, otherwise the Bayesian posterior inferences can be misleading or unreliable. Generally speaking, prior elicitation from experts or derived from most relevant historical data are two useful ways to obtain credible Bayesian priors.

## 4. Discussion

Because of the nature of clinical development process of new medicinal products, most of Bayesian applications in clinical trials are performed on early phase studies. As Bayesian

methods have gained popularity and regulatory acceptance, we believe there will be more interest in applying Bayesian methods on late phase clinical trials. Intuitively speaking, the same prior used for a large-scale Phase III trial tends to be less impactful than for a small early Phase trial because more observed data may dominate the posterior analytical results. Hence it seems that noninformative priors could provide results that are more consistent with the Frequentist approach for Phase III trials, especially when there is no compelling prior to use or when the regulatory agency is skeptical about informative priors.

Our non-informative prior Bayesian re-analysis of the Macugen Phase III clinical data resulted, in our view, in a consistent trend but slightly stronger inference compared with the Frequentist approach. Although one can argue that since the two approaches are based on different philosophy, direct comparisons between the two approaches should not be considered like with like.

Our informative prior Bayesian re-analyses of the Macugen Phase III clinical data indicated that the prior would have to be somewhat far from the observed response probabilities, in order to change the trend obtained from the Frequentist analyses. One key reason for this observation is that the observed sample sizes are large enough to offset minor influence from informative priors. Only when the prior information is substantially different from the observed data can the priors be influential enough to drive a different trend. We believe that in the case of smaller sample sizes, when Frequentist approach fails to show significant p-values, Bayesian approach with informative priors can be influential to show stronger results. However, we believe that the prior information needs to be credible, otherwise the Bayesian posterior inferences can be misleading or unreliable. Generally speaking, prior elicitation from experts or derived from most relevant historical data are two useful ways to obtain credible Bayesian priors.

Bayesian re-analysis of Phase III data from this manuscript implied that use of non-informative priors can produce consistent trend of inference with Frequentist approach. Furthermore, informative priors when chosen appropriately can be used to increase power and efficiency over the frequentist approach. We strongly recommend using prior elicitation or most relevant historical data to obtain priors for future new clinical trials adopting Bayesian approach. These examples also demonstrated that Bayesian analyses can be conveniently applied to binary response variables.

The methodology presented in the paper can be easily extended to incorporate covariates from the study through a formulation of $p = \frac{exp(X'\beta)}{1+exp(X'\beta)}$. However, such covariates are not available for this analysis.

## Acknowledgment

The authors would like to express their appreciation to Dr. Joe Cappelleri for his useful editorial suggestions and comments.

## References

Barry, S. M., Carlin, B. P., Lee, J.J., Muller, P. (2010). Bayesian Adaptive Methods for Clinical Trials. Boca Raton, FL: Chapman & Hall/CRC Biostatistics Series.

Chen, D.G.,Peace, K. E. (2011). Clinical Trial Data Analysis Using R. Boca Raton, FL: Chapman & Hall/CRC Biostatistics Series.

Chen, M-H.,Kim, S. (2009). Cure rate models with application to melanoma and prostate cancer data: In *Design and Analysis of Clinical Trials with Time-to-Event Endpoints* (Peace, KE Editor); Boca Raton, FL: Chapman & Hall/CRC Biostatistics Series, 349–370.

Connor, J. T.,Berry, S. M. (2005). Bayesian analysis for medical device trials. *The American Journal of Gastroenterology*, 100:1732–1735.

Cook, J. D. (2005). Exact calculation of beta inequalities. *Technical Report UTMDABTR*-005–05.

Salsburg, D. (2001). The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century. Henny Hold and Company, New York, NY

Tan, S-B, Machin, D., Tai, B-C., Foo, K-F.,Tan, E-H.(2002). A Bayesian re-assessment of two phase II trials of gemcitabine in metastatic nasopharyngeal cancer. *British Journal of Cancer*.86:843–850.

Zhu, M.,Lu, A.Y. (2004). The counter-intuitive non-informative prior for the Bernoulli family. *Journal of Statistical Education*, 12(2):1–10.