Power and sample size: a Bayesian approach

P. Fina ***(Neurosciences Discovery Biometrics, GSK Verona)***

1

Agenda

- A brief overwiev of Bayesian statistics
- From the “conditional” to the “predictive” power
- (predictive) power after an interim
- Hybrid and “full Bayesian” approaches
- Classical and bayesian power: 2 different methods addressing the same concept?

2

Which side are you on?

Bayesian
One who, vaguely expecting a horse and catching a glimpse of a donkey, strongly concludes …
he has seen a mule.

Statistician.
Someone who thinks that the rest of the world gives a damn whether he bores as a Bayesian or a frequentist.
Guernsey Mc Pearson (S.Senn)

3

Bayesian Statistics

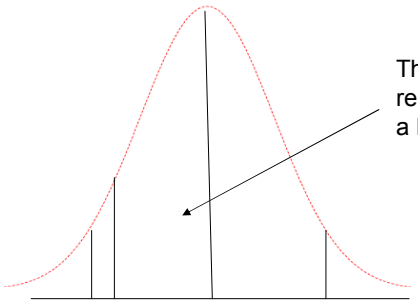
- **Before the analysis, the current belief about the treatment effect is expressed by means of a prior distribution p (θ)**

- the contribution provided by the experimental result is summarized in a distribution of probability p (y /θ) (likelihood based)

- **Bayes' rule is utilized for obtaining a posterior distribution: p (θ / y)**

4

The outcome of a Bayesian analysis: the posterior distribution



The area for values > 0 represents the Probability of a Positive Effect (PPE)

Credible interval

5

Bayesian Statistics (2)

Objective	Classical approach	Bayesian approach
Does the drug work?	p-value	Probability that the drug works = “PPE”
Quantification of the effect	Confidence interval	Credible interval

6

Bayesian Statistics (3)

- **What if... there isn't any accurate prior information? Non informative prior**
- **The posterior is entirely data-driven (likelihood-based). Specifically it is based on the estimated effect and its level of uncertainty**

Bayesian Approach	prior	vs classical approach
posterior	Non inf.	= t/normal with mean= estimated effect, Sd = standard error
PPE		= 1 – p-value (1 sided)
Credible interval		= Confidence interval

7

From the conditional to the predictive power

- the classical power is a conditional power based on a specific value of θ
- Often a "sensitivity analysis" is included considering some sparse alternative values

- Why do not consider any plausible value of θ ?
  taking into account its probability p (θ)?
- Given a prior N (μ, σ^2/n_o) with σ = sd of the diff. = $\sqrt{2}$* σ_{data}

- the predictive power is given by (Spiegelhalter) :

$$P(S_\epsilon^C) = \Phi\left[\sqrt{\frac{n_0}{n_0+n}}\left(\frac{\mu\sqrt{n}}{\sigma} + z_\epsilon\right)\right].$$

$\varepsilon = \alpha/2$, → $Z\varepsilon$= -1.96,
σ known

8

Predictive power : ex.1

$$P(S_\epsilon^C) = \Phi\left[\sqrt{\frac{n_0}{n_0+n}}\left(\frac{\mu\sqrt{n}}{\sigma} + z_\epsilon\right)\right].$$

Parallel group, 2-arm (act vs pla) study, alpha=5% (2-tails), Beta= 20%; endpoint: Hamilton (anxiety), μ =2, SD =6
→142 evaluable pts. per arm (conditional power)

Given a prior N (2, 72/50), the predictive power with 142 pts per arm is: Probnorm (0.51 * 0.85) = Probnorm (0.43) = 66.8%

Given a prior N (2,72/100000), the predictive power with 142 pts is: Probnorm (1* 0.85) = 80.2%

9

Predictive power after an interim

- after an interim ...

1. *Uncertainty is restricted to the second part of the study, one can derive the minimum difference (d2) to be observed in the second part providing an overall positive result*
2. *The conditional/predictive power of observing at least d2 in the second part, can be derived*
3. *If the predictive approach is chosen, the prior for the second part of the study may be based on the interim results*

$$p(S_\epsilon^C|y_m) = \Phi\left[\frac{z + \sqrt{f}z_\epsilon}{\sqrt{1-f}}\right].$$

f = interim / total sample size, z= z score (interim),
Assumption : same variability before and after the interim

10

Predictive power after an interim: ex. 2

$$p(S_\epsilon^C|y_m) = \Phi\left[\frac{z+\sqrt{f}z_\epsilon}{\sqrt{1-f}}\right].$$

Same study with an interim in the middle of the trial (71 pts)
Before the interim the power is = Pr. of observing >= 1.4
Effect at the interim =2.1 (sd=6) -> an effect of 0.7 in the 2nd part is enough. Therefore, the power is = Pr of observing >= 0.7 in the next 71 pts per arm
Conditional or predictive?
Predictive power = Probnorm ((2.09-0.71*1.96)/0.71) = Probnorm (0.99) = 83.9%

Effect at the interim = 0 -> Probnorm(-1.96) = 2.5%

11

Hybrids and "full bayesian" methods

- 'Till now we have analyzed hybrids methods: the analysis is still a classical one
- When a Bayesian analysis is performed, the prior information is incorporated and the definition of "positive study" may be expressed as PPE > 1-ε% ...

$$P(S_\epsilon^B) = \Phi\left[\frac{\mu\sqrt{n_0+n}\sqrt{n_0}}{\sigma\sqrt{n}} + \sqrt{\frac{n_0}{n}}z_\epsilon\right].$$

with f = no/(no+n) and z = mean/sd of the prior
you will obtain the previous formula

An even more Bayesian analysis may be performed, including utility / cost functions and investigating the corresponding optimal sample size

12

"Full bayesian" method: ex. 3

$$p(S_\epsilon^C|y_m) = \Phi\left[\frac{z+\sqrt{f}z_\epsilon}{\sqrt{1-f}}\right].$$

New definition of a positive study : PPE> 97.5 %

Coming back to the prior of ex.1 … N (2, 72/50)

Using previous slide's formula:

Probnorm (1.94 – 1.17) = Probnorm(0.77) = 78.1%

Or z = (2* √ 50)/ (6 √ 2) = 1.67, f = 50/192= 0.26

Probnorm ((1.67 – 1.96*0.51)/0.86) =

Probnorm (0.77) = 78.1 %

13

prior + interim

Hybrid approach

$$p(S_\epsilon^C|y_m,\ \text{prior}) = \Phi\left(\sqrt{\frac{n_0 n}{(n_0+m)(n_0+m+n)}}\frac{\sqrt{n_0}\mu}{\sigma} + \sqrt{\frac{m(n_0+m+n)}{n(n_0+m)}}\frac{\sqrt{m}y_m}{\sigma} + \sqrt{\frac{(m+n)(n_0+m)}{n(n_0+m+n)}}z_\epsilon\right).$$

"full bayesian"

$$p(S_\epsilon^B|y_m,\ \text{prior}) = \Phi\left[\frac{\sqrt{n_0+m+n}\,(n_0\mu+my_m)}{\sqrt{(n_0+m)n}\,\sigma} + \sqrt{\frac{n_0+m}{n}}z_\epsilon\right].$$

With no = "sample size" of the prior, m = sample size at the interim, n = post interim sample size

$$p(S_\epsilon^C|y_m) = \Phi\left[\frac{z+\sqrt{f}z_\epsilon}{\sqrt{1-f}}\right].$$

14

Conditional / predictive power: do they measure the same concept?

- It's the probability of observing a positive trial assuming that the drug works
- It's the probability of observing a positive trial taking into account the current belief about the efficacy of the drug

- you don't need to believe that the drug works (it's a kind of desired effect)
- the prior mean represents the expected effect (the most likely one)

Switching from an approach to the other could be dangerous !!

15

Last example ...

Same study, no prior at the beginning, an interim after 50 pts (f= 0.35) showing a treatment effect of 2 and a sd of 6. What's the outcome if the predictive power is utilized for a SSR ?

Final sample size per arm	Predictive power
142	73.4%
200	78.6%
225	80.0%
500	86.5 %
∞	1- p-value at interim (95.2%)

What's the role of the predictive power?

16

Conclusions

- The predictive power provides the probability of obtaining a positive trial not anchored to a specific assumption regarding the true treatment effect
- Interim results can be taken into account in a very natural way
- the definition of positive study can be the classical one (entirely based on data, p-value < ...) or the bayesian one (including external info, PPE > ...). In the first case the method is labelled as “hybrid”
- there are important logical differences between the concepts of conditional/predictive power

17

References

- D. Spiegelhalter et al “Bayesian approaches to Clinical Trials and Health-care evaluation” *John Wiley & sons (2004)*
- S. Senn “Letter to the editor, a comment on replications, p-values and evidence” *Statistics in Medicine 21, 2437, 2444 (2002)*
- J.M. Grouin et al. “Bayesian sample size determination in non-sequential trials: statistical aspects and some regulatory considerations” *Statistics in Medicine 26, 4914, 4924 (2007)*

18