

# EVALUATION DE METHODES BAYESIENNES DE CONSTRUCTION DE PRIOR DANS LE CADRE DE L'ÉVALUATION DE LA SUPÉRIORITÉ D'UN EFFET TRAITEMENT PAR RAPPORT A UN GROUPE CONTROLE DANS LES ETUDES "PROOF-OF-CONCEPT" (POC) DES ESSAIS CLINIQUES EN PHASE PRECOCE.

Karine Florin<sup>1</sup> & Dorian Bankovic<sup>2</sup> & Nicolas Bonnet<sup>3</sup>  
<sup>1&3</sup> Sanofi av Prof Blayac, Montpellier, France.  
karine.florin@sanofi.com

<sup>2</sup> Arlenda, chaussée Verte, 93, 4470 Saint-George-Sur-Meuse, Belgique.  
dorian.bankovic@arlenda.com

## Résumé

Afin de diminuer les coûts d'échec des phases avancées de développement clinique mais aussi pour des raisons éthiques, une étude en phase précoce nommée « preuve de concept (PoC) » est mise en place dont l'objectif est de démontrer la supériorité d'un traitement par rapport à un groupe contrôle sur un échantillon de patient limité. En alternative aux méthodes fréquentistes actuellement utilisées, l'objectif de ce travail est de mettre en œuvre et d'évaluer des méthodes bayésiennes d'intégration des données historiques à disposition (que ce soit des publications et/ou des études antérieures) dans le but d'augmenter les probabilités de succès de ces études qui permettent la prise de décision de passer ou non aux phases suivantes du développement.

Deux types de méthodes sont explorées en particulier : celles de type « Power prior » (classique et normalisé) basées sur la pondération et celles basées sur les modèles hiérarchiques de type « MAP » (meta analytic predictive, classique et robuste). Ces méthodes sont comparées entre elles, mais aussi avec les méthodes actuellement utilisées, sur des données simulées et sur un jeu de données réels disponibles avec un critère de jugement binaire. Comme le recommande la FDA, la puissance ou le contrôle de l'erreur de type I sont évaluées, en plus d'autres critères de comparaison statistique comme le biais ou le RMSE (« root mean square error »). Des recommandations sont proposées en conclusion sur les méthodes à utiliser et comment les utiliser.

**Mots-clés.** Méthode Bayésienne , Probabilité de succès, Distribution postérieure, Power prior, Power Prior normalisé, MAP prior, MAP prior robuste, modèle hierarchique

## Abstract.

In order to reduce the costs related to failure in late phase development, and for ethical reasons, it is necessary to conduct an early phase study called "proof of concept (PoC)." The objective of such study is to evaluate the superiority of one treatment compared to a control group (placebo or Reference) in a parallel design with limited patient sample size.

Alternatively to the currently frequentist methods used, the purpose of this work is to put in place and evaluate historical data (publications and / or previous studies) borrowing Bayesian methods in order to increase the probability of success of this PoC study and to take a good (Go / No Go) decision before possibly moving to the next phases of development.

Two types of historical bayesian data integration methods are explored in particular: those as Power prior methods based on weighting and those based on hierarchical models as MAP (meta analytic predictive) methods. These methods are compared with each other, but also with frequentist methods currently used, on simulated and real available binary data. According to the FDA recommendation, control of the type 1 error rate and evaluation of power have been realized, in addition to other comparison criteria as bias or root mean square error. In conclusion, recommendations of methods are proposed.

**Keywords.** Bayesian methods, probability of success, Power prior, Normalized power prior, MAP prior, Robust MAP prior, hierarchical model, predictive distribution

## Texte long

### Contexte

Les coûts de R&D de l'industrie pharmaceutique sont dominés par les coûts d'échec en phases avancées (IIb-III) de développement. Dans ce contexte mais aussi pour des raisons éthiques, il est impératif d'établir tôt une « preuve de concept (PoC) » du fonctionnement d'un composé pour décider de poursuivre ou non le développement d'un composé. Cette preuve est obtenue lors d'une étude clinique de phase précoce, usuellement de phase IIa, dont l'objectif est de démontrer la supériorité d'un traitement par rapport à un groupe contrôle (Placebo ou Référence) en design parallèle sur un échantillon de patient limité. La décision se base alors sur les résultats de l'étude en prenant en compte le contexte environnant (études précédentes, publications, concurrence...) mais sans intégration quantitative formelle. Or l'étude « preuve de concept » intervient souvent dans un cadre où des connaissances a priori peuvent être exploitées afin d'augmenter les probabilités de succès. Les méthodes d'analyses bayésiennes sont reconnues comme étant en particulier adaptées à ce cas de figure.

En alternative aux méthodes fréquentistes actuellement utilisées l'objectif de ce travail est donc de mettre en œuvre et d'évaluer des méthodes bayésiennes d'intégration de données historiques (que ce soit des publications et/ou des études antérieures), pour l'analyse de nos études POC.

### Règle de décision

Considérons une étude PoC en développement clinique précoce dans laquelle on souhaite administrer à certains patients un nouveau traitement T et à d'autres un groupe contrôle C.

Actuellement, pour une supériorité, les méthodes fréquentistes utilisées se basent sur les tests d'hypothèses suivants :  $H_0 : \theta_T - \theta_C = 0$  versus  $H_1 : \theta_T - \theta_C > 0$ .

La décision est prise en comparant la p-value obtenue à un risque alpha de première espèce fixé par exemple à 5% pour un test unilatéral.

Avec les méthodes bayésiennes étudiées, l'objectif est calculer la probabilité de succès  $P(\theta_T - \theta_C > 0)$  à partir de la distribution à posteriori et de baser la décision sur une règle à partir d'un seuil de décision défini a priori, ici fixé à 0.95.

### Les méthodes étudiées :

- Les méthodes de type MAP prior développées par Neuenschwander *et al.* [3], se basent sur une structure des données historiques qui suggère un modèle hiérarchique avec l'incorporation d'une variance inter-étude  $\tau^2$ . La distribution prédictive a posteriori du paramètre d'intérêt  $\theta$  est générée et utilisée comme prior avec les données du futur essai. Cette méthode se décline en une version dite "robuste", où une partie de distribution non informative est ajoutée à la distribution prédictive précédente et en fonction d'une pondération  $w$ , représentant la probabilité a priori que le nouvel essai soit similaire avec les données historiques.

- Les méthodes de type Power prior, développées par Ibrahim et Chen [9], se basent sur une pondération de la vraisemblance associée aux données historiques avec un paramètre  $a_0$ . La distribution a posteriori du paramètre d'intérêt  $\theta$  est générée en prenant en compte la pondération par  $a_0$ , et cette distribution alors utilisée comme prior avec les données de la future étude. Cette méthode se décline en une version dite "Normalized Power prior", avec la prise en compte d'un prior non informatif sur le paramètre  $a_0$ .

## Comparaison des méthodes/ Etat de l'art

D'un point de vue théorique, l'intérêt des méthodes MAP prior provient de la possibilité d'intégrer la variance inter-étude dans la construction du prior, ce qui peut être avantageux en cas de fortes variances entre les études historiques. Le choix du prior sur la variance inter-études  $\tau^2$ , qui peut être vu comme un inconvénient de la méthode, peut se porter sur des distributions de type inverse-gamma  $IG(e,e)$  avec un  $e$  faible, comme le recommande Viele [7]. La robustesse apportée avec la méthode "MAP prior robuste" peut être intéressante dans le cas d'une importante hétérogénéité entre les études, le paramètre de pondération  $w$  est souvent fixé de façon arbitraire à 0.9 ou 0.5 [6]. Dans le cas où on a une idée précise du poids  $a_0$  que l'on veut attribuer aux données historiques par rapport aux données actuelles, utiliser le power prior peut être plus judicieux par rapport aux autres méthodes. Enfin, le normalized power prior constitue une réelle amélioration du power prior, car il permet d'utiliser le principe du power prior, sans avoir à fixer le poids  $a_0$  grâce à la prise en compte d'un prior peu ou non informatif sur ce paramètre. Cependant, en cas de fortes contradictions entre les données historiques et actuelles, l'erreur quadratique obtenue avec la méthode normalized power prior peut être importante [10]. Dans ce cas, le MAP prior et en particulier le MAP prior Robuste peuvent être une meilleure alternative.

### Mise en œuvre des comparaisons

Les méthodes bayésiennes ont été implémentées sous SAS 9.4 sur des données d'un exemple réel mais aussi sur des données simulées suivant une loi binomiale dans le même contexte en faisant varier certains paramètres comme le taux de réponse dans le groupe placebo. Dans l'exemple étudié, les données historiques sont considérées issues d'une publication (sur le groupe placebo) et d'une étude interne (sur les deux bras placebo et traité), elles sont prises en compte dans la construction du prior informatif de l'analyse bayésienne de l'étude POC dont l'objectif correspond à ce qui a été décrit plus haut.

Les méthodes MAP prior, MAP prior robuste, Power Prior, Power Prior normalisé sont comparées entre elles, mais aussi avec les méthodes bayésiennes classiquement utilisées, avec l'utilisation directe d'un prior informatif ou non informatif (se rapprochant des résultats des analyses fréquentistes). Comme le recommande la FDA [8], la puissance ou le contrôle de l'erreur de type I sont évaluées, en plus d'autres critères de comparaison comme le biais ou le RMSE (« root mean square error »).

### Principaux résultats sur les jeux de données étudiés et simulés

Sur les cas étudiés, l'apport des données historiques est évident, la probabilité de succès ou la puissance pour les simulations est plus forte que sans utilisation de ces données. Sur les données simulées, le risque de première espèce est bien contrôlé par les méthodes de type MAP ou power prior, ce qui n'est pas le cas avec la méthode bayésienne classique d'utilisation d'un prior informatif. Le power prior, et en particulier le normalized power prior, présente des résultats plus intéressants sur la puissance et le contrôle du risque  $\alpha$ . Les méthodes de type MAP prior sont quant à elles les plus intéressantes d'un point de vue du biais, du taux de couverture, et de la valeur du RMSE. Ces résultats restent bien sûr à nuancer du fait du cadre particulier des exemples et des simulations qui demeurent à étendre.

La limitation de ce travail réside d'une part dans le fait que l'impact du nombre d'étude historique ou de la variabilité inter-études n'a pas été encore évalué et d'autre part que l'extension au cas d'une variable réponse continue reste à être examiné.

## Bibliographie

- [1] Pocock S.J. The combination of randomized and historical controls in clinical trials, *Journal of Chronic Diseases* (1976) ; 29 :175188.
- [2] Ibrahim J.G., Chen M.H., Power prior distributions for regression models. *Statistical Science* (2000) ; 15 :4660.
- [3] Neuenschwander B., Capkun-Niggli G., Branson M., Spiegelhalter D.J., Summarizing historical information on controls in clinical trials, *Clinical Trials* (2010) ; 7 ; 5-18.
- [4] Neuenschwander B., From historical data to priors *Biometrics section* (2011)
- [5] Hobbs B.P., Carlin B.P., Mandrekar S.J., Sargent D.J., Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* (2011) ; 67(3) : 10471056.
- [6] Schmidli H., Gsteiger S., Roychoudhury S., OHagan A., Spiegelhalter D., and Beat Neuenschwander B., Robust Meta-Analytic-Predictive Priors in Clinical Trials with Historical Control Information. *Biometrics* (2014) ; DOI : 10.1111/biom.12242
- [7] Viele K., Berry S., Neuenschwander B., Amzal B., Chen F., Enas N., Hobbs B., Ibrahim J. G., Kinnersley N., Lindborg S., Micallef S., Roychoudhury S. and Thompson L., Use of historical control data for assessing treatment effects in clinical trials, *Pharmaceutical Statistics* (2014), 13 : 4154. doi : 10.1002/pst.1589
- [8] Food and Drug Administration (FDA), Guidance for the use of bayesian statistics in medical device clinical trials, *Guidance for industry and FDA staff* (2010)
- [9] Chen M.H, Ibrahim J.G, The relationship between the power prior and hierarchical models, *Bayesian Analysis* (2006)
- [10] Duan Y., Ye K., Smith P.E., Evaluating water quality using power priors to incorporate historical information, *Wiley InterScience* (2005)
- [11] Gelman A., Prior distributions for variance parameters in hierarchical models, *Bayesian Analysis* (2006)
- [12] The network meta-analytic predictive approach to non-inferiority trials, *Statistical Methods in Medical Research* (2011)
- [13] Di Scala L., Kerman J., Collection, synthesis, and interpretation of evidence : a proof-of-concept study in COPD
- [14] Neelon B., O'malley A.J., Bayesian Analysis Using Power Priors with Application to Pediatric Quality of Care *Journal of Biometrics and Biostatistics* (2010)
- [15] Ibrahim G., Chen M.H, The power prior : theory and applications *Statistics in Medicine* (2015)