Étude des propriétés fréquentistes des estimateurs bayésiens de la différence de deux proportions, du risque relatif et du rapport de cotes

François Lefebvre [1] & Nicolas Meyer [2]

[1] *Hôpital civil, Service de santé publique, Groupe méthode en recherche clinique, 1 place de l'hôpital, BP 426, 67091 Strasbourg, francois.lefebvre@chru-strasbourg.fr*
[2] *Hôpital civil, Service de santé publique, Groupe méthode en recherche clinique, 1 place de l'hôpital, BP 426, 67091 Strasbourg, nicolas.meyer@chru-strasbourg.fr*

Résumé. Les intervalles de confiance d'un estimateur ont des taux de couverture parfois éloignés de leur valeur nominale. Ils restent cependant très utilisés, notamment dans les essais thérapeutiques. D'un autre côté, alors que les intervalles de crédibilité semblent avoir de bonnes propriétés fréquentistes, les analyses statistiques des essais thérapeutiques ne se font pas souvent sous inférence bayésienne. Une étude de simulations a été réalisée afin d'étudier les propriétés fréquentistes des intervalles de confiance de la différence de deux proportions, du risque relatif et du rapport de cotes estimés selon différentes méthodes pour les comparer aux intervalles de crédibilité de ces mêmes paramètres estimés avec trois lois a priori différentes peu ou très peu informatives sur les proportions. Pour cela, les taux de couverture exacts des intervalles de confiance ont été comparés aux taux de couverture exacts des intervalles de crédibilité estimés directement d'après les lois a posteriori des paramètres. Les taux de couverture des intervalles de crédibilité pour la différence de deux proportions, le risque relatif et le rapport de cotes sont très proches de la valeur nominale, même pour des petits échantillons, notamment quand la loi a priori est uniforme alors que les taux de couverture des intervalles de confiance peuvent s'en éloigner fortement. Il est donc recommandé d'utiliser l'inférence bayésienne dans l'estimation de ces intervalles.

Mots-clés. Bayésien, intervalles de confiance, intervalles de crédibilité, taux de couverture.

Abstract. Confidence intervals may have coverage probabilities far from the nominal value but are often used, particularly in clinical trials. On the other hand, although credibility intervals seem to have good frequentist properties, statistical analyses are rarely done under Bayesian inference. A simulation study was done to analyse the frequentist properties of the confidence intervals for the difference between two proportions, the relative risk and the odds ratio estimated with different methods and the credibility intervals for this parameters estimated with three different low or very low informative priors on the proportions. For that purpose, exact coverage probabilities of the confidence intervals were compared to exact coverage probabilities of the credibility intervals directly estimated with the posteriors of the parameters. The coverage probabilities for the credibility intervals of the difference between two independent proportions, the relative risk and the odds ratio are closed to the nominal value, even when the sample sizes are small, particularly when the prior is uniform, whereas the coverage probabilities of the confidence intervals are sometimes far from the nominal value. Consequently, Bayesian inference is recommended to estimate these intervals.

Keywords. Bayesian, confidence intervals, credibility intervals, coverage probability.

Introduction

Les essais thérapeutiques requièrent souvent l'estimation d'une ou de deux proportions de succès, ou de leur rapport, assortis d'un intervalle de confiance. Plusieurs méthodes fréquentistes ont été proposées pour calculer les intervalles de confiance de ces différents estimateurs [1]. La plupart de ces méthodes ont des taux de couverture qui sont dans certains cas assez éloignés de la valeur nominale [1]. Néanmoins, ils restent communément utilisés. Par ailleurs, les intervalles de crédibilité, qui reposent sur l'inférence bayésienne, semblent avoir de bonnes propriétés fréquentistes, comme l'a montré Agresti et col [2].

Du fait du développement de l'inférence bayésienne, l'Agence américaine des produits alimentaires et médicamenteux (Food and Drug Administration, FDA) [3] accepte depuis 2012 les designs bayésiens sous réserve de démontrer que l'analyse bayésienne du critère de jugement principal respectera les risques α et β, alors qu'il n'y a pas de puissance ni de risque α sous inférence bayésienne. Aussi, pour s'assurer que les estimateurs bayésiens des paramètres évoqués ont d'aussi bonnes propriétés que les estimateurs fréquentistes, une alternative est de les comparer en utilisant leurs taux de couverture respectifs [1-2;4-5].

L'objectif de cette étude est de comparer les propriétés fréquentistes de la différence de deux proportions, du risque relatif et du rapport de cotes estimés avec une inférence fréquentiste et bayésienne.

Matériel et méthodes

Pour étudier les propriétés fréquentistes des estimateurs bayésiens, il a été décidé d'utiliser le taux de couverture calculé avec une méthode exacte à partir des intervalles de crédibilité ou de confiance.

Pour la différence de deux proportions $p_2 - p_1$, les intervalles ont été calculés de manière fréquentiste (méthode de Wald avec et sans correction de continuité), de manière bayésienne approchée (avec deux fois deux lois a priori identiques sur les proportions p_1 et p_2 : Bêta(0,5;0,5) et Bêta(1;1)) et de manière bayésienne exacte (avec trois fois deux lois a priori identiques sur les proportions p_1 et p_2 : Bêta(0,5;0,5), Bêta(1;1) et Bêta(2;2)). La méthode bayésienne approchée consiste à calculer la loi a posteriori de p_1 et de p_2

et à supposer que la différence suit une loi normale de moyenne valant la différence des moyennes a posteriori et de variance valant la somme des variances a posteriori de p_1 et de p_2. Le choix des lois a priori a porté sur la loi de Jeffreys (Bêta(0,5;0,5)), la loi uniforme (Bêta(1;1)) et une loi légèrement informative et symétrique (Bêta(2;2)).

Pour le risque relatif p_1/p_2, les intervalles ont été calculés de manière fréquentiste avec la méthode du maximum de vraisemblance, de manière bayésienne approchée (avec deux fois deux lois a priori identiques sur les proportions p_1 et p_2 : Bêta(0,5;0,5) et Bêta(1;1)) et de manière bayésienne exacte (avec trois fois deux lois a priori identiques sur les proportions p_1 et p_2 : Bêta(0,5;0,5), Bêta(1;1) et Bêta(2;2)). La méthode bayésienne approchée consiste à calculer la loi a posteriori de p_1 et de p_2 et à supposer que le logarithme du rapport suit une loi normale de moyenne valant le logarithme du rapport des moyennes a posteriori et de variance valant la somme des carrés des coefficients de variations a posteriori de p_1 et de p_2 [4].

Enfin, pour le rapport de cotes $p_1(1-p_2)/(p_2(1-p_1))$, les intervalles ont été calculés de manière fréquentiste avec la méthode du maximum de vraisemblance et celle de Lindley [4] et de manière bayésienne exacte (avec trois fois deux lois a priori identiques sur les proportions p_1 et p_2 : Bêta(0,5;0,5), Bêta(1;1) et Bêta(2;2)).

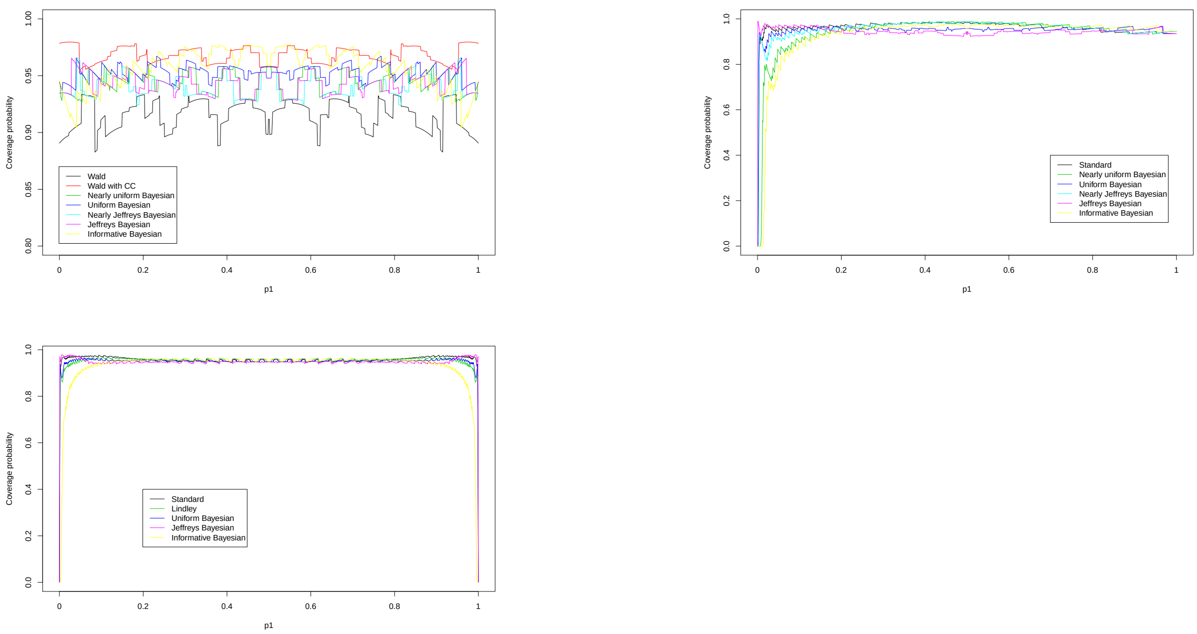
Pour toutes les estimations, les intervalles ont été calculés à 95 %. Les intervalles de crédibilité ont été calculés directement à partir des lois a posteriori sur les proportions p_1 et p_2, sans méthodes de Monte-Carlo par chaînes de Markov. Ces intervalles représentent les valeurs des 2,5 et 97,5 percentiles de la loi a posteriori. Les comparaisons des taux de couverture entre les approches fréquentiste et bayésienne ont été réalisées en utilisant des tailles d'échantillons égales dans les deux groupes allant de 10 à 100 par pas de 10, une proportion p_2 allant de 0,1 à 0,5 par pas de 0,1 et une proportion p_1 allant de 0 à 1 par pas de 0,001. Les moyennes, minimums, maximums et variances du taux de couverture ont été calculés pour chaque taille d'échantillon et proportion p_2 fixées en faisant varier p_1.

Résultats

Le taux de couverture tend vers le niveau de confiance nominal de 95 % quand les effectifs augmentent, quelles que soient les méthodes utilisées (bayésiennes et fréquentistes) et quand p_2 tend vers 0,5. La figure 1 représente les taux de couverture de la différence de deux proportions, du risque relatif et du rapport de cotes selon les différentes méthodes utilisées.

Pour la différence de deux proportions, l'intervalle qui est le plus proche de 95 % est l'intervalle de crédibilité obtenu avec deux lois a priori Bêta de paramètres Bêta(1;1). Par exemple, avec ces lois, pour des effectifs de 100 par échantillon et $p_2 = 0,5$, le taux de couverture moyen est de 0,950322. Les intervalles de crédibilité obtenus avec deux lois a priori Bêta de paramètres Bêta(0,5;0,5) et Bêta(2;2) sont respectivement légèrement anticonservateurs et conservateurs. Les intervalles bayésiens approchés sont anticonserva-

Figure 1: Taux de couverture des intervalles de la différence de deux proportions ($p_2 = 0,5$, $n_1 = n_2 = 10$), du risque relatif ($p_2 = 0,5$, $n_1 = n_2 = 10$) et du rapport de cotes ($p_2 = 0,5$, $n_1 = n_2 = 30$)



teurs. L'intervalle de Wald est anticonservateur et l'intervalle de Wald avec correction de continuité est trop conservateur (tableau 1). Avec l'intervalle confiance de Wald, le taux de couverture moyen est parfois très éloigné de la valeur nominale de 95 %, notamment en cas de faibles effectifs ou de faible proportion p_2. Par exemple, pour des effectifs de 10 sujets par groupe et $p_2 = 0,1$, le taux de couverture moyen est de 0,8720 et l'étendue est de [0,6390 - 0,9497]. Par conséquent, pour toutes les proportions p_1, le taux de couverture est inférieure à 95 %.

Différence	Wald	Wald avec CC	bayésien uniforme approché	bayésien Jeffreys approché	Bêta (1;1) bayésien	Bêta (0,5;0,5) bayésien	Bêta (2;2) bayésien
$p_2 = 0,5$							
$n = 10$	0,916 (0,013)	0,968 (0,007)	0,944 (0,009)	0,940 (0,009)	0,953 (0,007)	0,944 (0,009)	0,956 (0,017)
$n = 20$	0,934 (0,007)	0,969 (0,004)	0,946 (0,005)	0,945 (0,005)	0,952 (0,005)	0,947 (0,004)	0,953 (0,011)
$n = 30$	0,939 (0,005)	0,968 (0,003)	0,947 (0,004)	0,946 (0,004)	0,951 (0,003)	0,948 (0,003)	0,952 (0,008)
$n = 40$	0,942 (0,003)	0,967 (0,003)	0,947 (0,003)	0,947 (0,003)	0,951 (0,003)	0,948 (0,003)	0,951 (0,006)
$n = 50$	0,944 (0,003)	0,966 (0,002)	0,948 (0,003)	0,948 (0,003)	0,951 (0,003)	0,949 (0,003)	0,951 (0,005)
$n = 60$	0,945 (0,002)	0,966 (0,002)	0,948 (0,002)	0,948 (0,002)	0,951 (0,003)	0,949 (0,002)	0,951 (0,004)
$n = 100$	0,947 (0,002)	0,963 (0,002)	0,949 (0,002)	0,949 (0,002)	0,950 (0,002)	0,949 (0,002)	0,951 (0,003)

Tableau 1: Taux de couverture moyens (écart-type) des intervalles de la différence de deux proportions estimés par sept méthodes.

Pour le risque relatif, l'intervalle qui est le plus proche de 95 % est l'intervalle de crédibilité obtenu avec deux lois a priori Bêta de paramètres Bêta(1;1). Les intervalles de crédibilité obtenus avec deux lois a priori Bêta de paramètres Bêta(0,5;0,5) et Bêta(2;2) sont respectivement légèrement et trop anticonservateurs. Avec la méthode du maximum de vraisemblance, les intervalles sont trop conservateurs. Les intervalles bayésiens uniformes approchés sont légèrement anticonservateurs et les intervalles bayésiens de Jeffreys approchés sont très légèrement conservateurs (tableau 2).

Risque relatif	Standard	bayésien uniforme approché	bayésien Jeffreys approché	Bêta (1;1) bayésien	Bêta (0,5;0,5) bayésien	Bêta (2;2) bayésien
$p_2 = 0,5$						
$n = 10$	0,962 (0,058)	0,941 (0,110)	0,957 (0,059)	0,950 (0,048)	0,944 (0,032)	0,929 (0,135)
$n = 20$	0,957 (0,044)	0,943 (0,083)	0,953 (0,046)	0,950 (0,041)	0,947 (0,031)	0,934 (0,105)
$n = 30$	0,955 (0,035)	0,944 (0,071)	0,952 (0,037)	0,950 (0,031)	0,948 (0,031)	0,937 (0,089)
$n = 40$	0,954 (0,031)	0,945 (0,063)	0,952 (0,032)	0,950 (0,031)	0,948 (0,031)	0,939 (0,079)
$n = 50$	0,954 (0,031)	0,945 (0,058)	0,951 (0,032)	0,950 (0,031)	0,948 (0,030)	0,941 (0,072)
$n = 60$	0,953 (0,031)	0,946 (0,055)	0,951 (0,032)	0,948 (0,031)	0,948 (0,030)	0,942 (0,067)
$n = 100$	0,952 (0,030)	0,947 (0,046)	0,950 (0,031)	0,949 (0,030)	0,949 (0,030)	0,944 (0,056)

Tableau 2: Taux de couverture moyens (écart-type) des intervalles du risque relatif estimés par six méthodes.

Enfin, pour le rapport de cotes, l'intervalle qui est le plus proche de 95 % est l'intervalle de crédibilité obtenu avec deux lois a priori Bêta de paramètres Bêta(1;1). Les intervalles de crédibilité obtenus avec deux lois a priori Bêta de paramètres Bêta(0,5;0,5) et Bêta(2;2) sont respectivement légèrement anticonservateurs et très anticonservateurs. Avec la méthode du maximum de vraisemblance, les intervalles sont trop conservateurs. Les intervalles de Lindley sont légèrement conservateurs, notamment pour les petits effectifs (tableau 3).

Rapport de cotes	Standard	Lindley	Bêta(1;1) bayésien	Bêta(0,5;0,5) bayésien	Bêta(2;2) bayésien
$p_2 = 0,5$					
$n = 10$	0,962 (0,065)	0,958 (0,066)	0,951 (0,059)	0,947 (0,045)	0,916 (0,155)
$n = 20$	0,957 (0,055)	0,954 (0,055)	0,950 (0,050)	0,947 (0,044)	0,923 (0,129)
$n = 30$	0,956 (0,047)	0,952 (0,048)	0,950 (0,044)	0,947 (0,043)	0,928 (0,113)
$n = 40$	0,955 (0,044)	0,952 (0,044)	0,950 (0,043)	0,947 (0,043)	0,931 (0,102)
$n = 50$	0,954 (0,043)	0,951 (0,044)	0,949 (0,043)	0,947 (0,043)	0,933 (0,095)
$n = 60$	0,953 (0,043)	0,951 (0,044)	0,949 (0,043)	0,948 (0,043)	0,934 (0,090)
$n = 100$	0,952 (0,043)	0,950 (0,043)	0,949 (0,043)	0,948 (0,043)	0,939 (0,074)

Tableau 3: Taux de couverture moyens (écart-type) des intervalles du rapport de cotes estimés par cinq méthodes.

Discussion

Ces résultats montrent que les intervalles de crédibilité bayésiens ont non seulement de bonnes propriétés fréquentistes en termes de taux de couverture mais aussi de meilleures propriétés fréquentistes que les intervalles de confiance. Néanmoins, il est important de choisir certaines lois a priori comme la loi Bêta(1;1) qui donne les taux de couverture les plus proches de la valeur nominale attendue et qui est une loi aux très bonnes propriétés fréquentistes par défaut quand il n'y a pas de connaissance a priori, même pour les petits effectifs (n=10). En cas de connaissances a priori, il n'est pas possible de comparer les propriétés fréquentistes des estimateurs bayésiens par rapport aux estimateurs fréquentistes car ces derniers n'incluent que l'information contenue dans l'échantillon.

En conclusion, ces résultats montrent de meilleures performances des estimateurs bayésiens pour calculer un intervalle de confiance ou de crédibilité pour une différence de deux proportions, un risque relatif ou un rapport de cotes, ce qui justifie une utilisation systématique de l'estimation bayésienne, notamment dans les essais thérapeutiques.

Bibliographie

[1] Newcombe R. G. Interval estimation for the difference between independent proportions: comparison of eleven methods. Statist. Med. 17, 873-890 (1998).

[2] Agresti, A. Min, Y. Frequentist performance of bayesian confidence intervals for comparing proportions in 2*2 contingency tables. Biometrics 61, 515-523 (2005).

[3] Food and Drug Administration (FDA). Guidance for the use of Bayesian statistics in medical device clinical trials, 2010. disponible à :
http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/Guidance Documents/ucm071121.pdf. [accès 6 Nov 2010].

[4] Hashemi L., Nandram B., Goldberg R. Bayesian analysis for a single 2×2 table. Statistics in medicine, 16, 1311-1328 (1997).

[5] Zhang, H. Gutiérrez Rojas, HA. Cepeda Cuervo, E. Confidence and credibility intervals for the difference of two proportions. Revista Colombiana de Estadistica 33(1), 63-88 (2010).