

INCERTITUDE SUR LE MODÈLE DOSE-RÉPONSE EN ÉPIDÉMIOLOGIE DES RAYONNEMENTS IONISANTS: ÉTUDE COMPARATIVE DES APPROCHES MULTI-MODEL INFERENCE ET BAYESIAN MODEL AVERAGING.

Sophie Ancelet ¹ & Merlin Keller ²

¹ *Institut de Radioprotection et de Sûreté Nucléaire (IRSN), Laboratoire d'épidémiologie (LEPID), BP17, 92262 Fontenay-aux-Roses, France. Email: sophie.ancelet@irsn.fr*

² *EDF R&D, Département PRISME, 6 quai Wattier, Chatou, France. Email: merlin.keller@edf.fr*

Résumé.

En épidémiologie des rayonnements ionisants (RI), plusieurs modèles dose-réponse sont souvent proposés pour décrire un même jeu de données. Certains modèles peuvent être très difficiles à discriminer en termes de qualité d'ajustement aux données mais conduire à des estimations différentes d'un même risque sanitaire radio-induit. Afin de tenir compte de l'incertitude sur la forme du modèle dose-réponse associé à une pathologie radio-induite, plusieurs travaux de recherche se sont récemment basés sur l'approche fréquentiste proposée par Burnham et Anderson (2002), appelée Multi-Model Inference (MMI). L'estimateur d'un risque d'intérêt est défini comme l'estimateur de risque moyen dans lequel chaque modèle possible a un poids qui dépend de sa capacité d'ajustement aux données, quantifiée par le critère d'information d'Akaike (AIC). L'utilisation de tels poids AIC peut en effet permettre de réduire les biais potentiellement induits par la sélection d'un modèle unique. Malgré la simplicité de mise en oeuvre d'une telle approche, à notre connaissance, aucune théorie n'existe quant à l'optimalité de l'utilisation de tels poids dans une procédure de combinaison de modèles. Dans ce travail, nous réalisons une étude par simulations afin de comparer les performances d'ajustement de l'approche MMI par rapport à une approche alternative bayésienne, connue pour ses bonnes propriétés en termes d'estimation et de prédiction: le Bayesian Model Averaging (BMA). Les simulations réalisées sont motivées par un cas d'étude réel portant sur l'estimation de l'excès de risque de leucémie potentiel due à l'exposition aux RI chez les survivants des bombardements d'Hiroshima-Nagasaki. Pour inférer les modèles candidats et mettre en oeuvre le BMA, une approche par échantillonnage préférentiel adaptatif et multiple a été implémentée.

Mots-clés. Bayesian Model Averaging, critère d'information d'Akaike, épidémiologie, Multi-Model Inference, rayonnements ionisants, méthodes bayésiennes, modèle dose-réponse, incertitude.

Abstract.

In radiation epidemiology, several dose-response models are often proposed to describe a same dataset. Some of these models may be very difficult to discriminate in terms of fitting abilities but lead to different estimates of a same radiation-induced disease risk. In order to account for the dose-response model uncertainty associated with a radiation-induced disease, several research studies have recently focused on the frequentist approach proposed by Burnham et Anderson (2002): Multi-Model Inference (MMI). Risk estimate is defined as a model-averaged risk estimate in which each possible model has a weight depending on its fitting performances, given by the Akaike Information Criterion (AIC). Using AIC weights can allow to reduce potential biases induced when selecting a single model. Despite the simplicity of such an approach, to our knowledge, no theory exists as to the optimality of including such weights in a model-averaging procedure. In this work, we perform a simulation study in order to compare the fitting performances of the MMI approach compared to a Bayesian alternative approach, known for its good properties in terms of both estimation and prediction: Bayesian Model Averaging (BMA). Our simulation study is motivated by a real-life case study dealing with the estimation of potential leukemia excess risks due to ionizing radiations among the survivors of the atomic bombings of Hiroshima and Nagasaki. To infer the candidate models and carry out BMA, an adaptive multiple importance sampling approach has been implemented.

Keywords. Akaike Information Criterion, Bayesian methods, Bayesian Model Averaging, dose-response model, epidemiology, ionizing radiation, Multi Model Inference, uncertainty.

1 Introduction

En épidémiologie des rayonnements ionisants (RI), plusieurs modèles décrivant la relation entre une dose reçue de RI et un risque sanitaire radio-induit sont souvent considérés pour un même jeu de données. Ces modèles diffèrent de par le choix de la forme de la relation dose-réponse (e.g., linéaire, linéaire quadratique exponentielle, linéaire par morceaux, . . .), la forme du risque de base (i.e., à dose nulle), le choix d'une relation dose-risque en excès de risque relatif ou absolu et des facteurs de risque susceptibles de modifier cette relation. Certains de ces modèles peuvent être très difficiles à discriminer en termes de qualité d'ajustement aux données mais conduire à des estimations différentes d'un même risque sanitaire radio-induit.

L'incertitude sur le modèle dose-réponse est le plus souvent ignorée en pratique: un seul modèle est généralement sélectionné pour estimer les risques de cancers radio-induits. Une telle approche peut mener à des estimations biaisées des excès de risque d'intérêt, en forçant un unique modèle à s'ajuster à des données qui ne vérifient pas nécessairement toutes les caractéristiques du modèle, et à une sous-estimation de l'incertitude de ces estimations. À terme, cela peut mener à des conclusions erronées quant à l'ampleur et la significativité statistique de l'impact de l'exposition aux RI sur le risque sanitaire étudié.

L'impact de la combinaison de plusieurs modèles dose-réponse sur l'estimation de risques sanitaires radio-induits est examiné depuis peu en épidémiologie des RI. Plusieurs travaux de recherche se sont ainsi récemment concentrés sur l'approche fréquentiste proposée par Burnham et Anderson (2002) et appelée Multi-Model Inference (MMI) [1][2]. L'estimateur d'un risque d'intérêt est défini comme l'estimateur de risque moyen dans lequel chaque modèle possible a un poids qui dépend de sa capacité d'ajustement aux données, quantifiée par le critère d'information d'Akaike (AIC). Malgré la simplicité de mise en oeuvre d'une telle approche, à notre connaissance, aucune théorie n'existe quant à l'optimalité de l'utilisation de tels poids dans une procédure de combinaison de modèles.

Dans ce travail, nous réalisons une étude par simulations, motivée par un cas d'étude en épidémiologie des rayonnements ionisants, afin de comparer les performances d'ajustement de l'approche MMI par rapport à une approche alternative bayésienne, connue pour ses bonnes propriétés en termes d'estimation et de prédiction: le Bayesian Model Averaging (BMA) [3]. Placés dans la situation utopique où le vrai modèle dose-réponse est connu, nous comparons également les performances de ces approches à celles d'une procédure de sélection de modèle pour l'estimation d'un risque sanitaire radio-induit.

2 Cas d'étude

L'estimation de l'excès de risque de leucémie radio-induite chez les survivants des bombardements d'Hiroshima Nagasaki a motivé notre étude par simulations. En effet, l'incertitude sur la forme du modèle dose-réponse est importante dans ce cas d'étude. Pas moins de treize modèles de régression de Poisson distincts et non emboîtés ont été publiés dans la littérature pour estimer cet excès de risque à partir des données de mortalité de la cohorte Life Span Study (LSS). Ces données, fournies par la "Radiation Effects Research Foundation" (RERF, Japon) [4], correspondent au suivi de 86611 survivants sur la période 1950-2003. A la fin de l'année 2003, 318 survivants sont décédés d'une leucémie. 11624 strates constituent la base de données LSS. Chaque strate regroupe les personnes-années homogènes selon un certain nombre de covariables comme le sexe, la ville (Hiroshima/Nagasaki), l'âge à l'exposition, l'âge atteint, la dose pondérée à la moelle osseuse. La variable réponse est le nombre de décès par leucémie, observés dans chaque strate.

Des ajustements fréquentiste et bayésien aux données de la LSS ont été préalablement réalisés et ont permis de calculer le critère d'AIC et la vraisemblance marginale pour les 13 modèles dose-réponse publiés.

3 Méthode

3.1 L'approche par Bayesian Model Averaging (BMA)

Soient M_k ($k=1, \dots, K$) K modèles candidats de risque de leucémie radio-induite. Chaque modèle est défini par un jeu de paramètres θ_k . Soit Δ une quantité d'intérêt à estimer ou prédire, i.e., ici, l'excès de risque relatif (ERR) de leucémie radio-induite pour un âge, sexe et dose de RI donnés. L'approche BMA [3] consiste à obtenir la loi a posteriori de Δ intégrée sur les différents modèles candidats, définie par:

$$[\Delta|y] = \sum_{k=1}^K [\Delta(\theta_k)|y, M_k] \omega_k \quad (1)$$

Elle s'écrit comme le mélange des lois a posteriori de Δ sous chaque modèle M_k ($k=1, \dots, K$) pondéré par ω_k , le poids du modèle M_k . Ce dernier est défini comme la probabilité a posteriori de M_k sachant les données observées y :

$$\omega_k = [M_k|y] = \frac{[y|M_k][M_k]}{\sum_{l=1}^K [y|M_l][M_l]} \quad (2)$$

où $[M_k]$ désigne la probabilité a priori du modèle M_k et $[y|M_k]$ la vraisemblance marginale (ou évidence) du modèle M_k . Celle-ci intègre l'incertitude sur les paramètres du modèle M_k . Par la suite, une loi a priori uniforme est assignée à l'espace des modèles candidats.

3.2 Simulation de données

Parmi les treize modèles de régression de Poisson publiés dans la littérature pour décrire le risque de mortalité par leucémie radio-induite, cinq ont été utilisés pour générer des données de mortalité par leucémie. Les cinq modèles dose-réponse choisis correspondent aux modèles dotés des meilleures capacités d'ajustement aux données de la LSS d'après le critère AIC et le calcul de la vraisemblance marginale.

Pour chacun des cinq modèles considérés, 100 jeux de données ont été simulés. L'objectif était de simuler des données de comptage de décès par leucémie similaires à celles observées dans la cohorte LSS (même effectif et caractéristiques de la population, mêmes scénarios d'exposition aux RI). Ainsi, chaque jeu de données contient 11624 réalisations de lois de Poisson dont les intensités respectives (de forme fonctionnelle dépendant du modèle dose-réponse génératif) ont été fixées en considérant: 1/ les valeurs des covariables (i.e., sexe, ville, âge moyen à l'exposition, âge atteint moyen, dose moyenne pondérée à la moelle osseuse) associées aux 11624 strates de la LSS; 2/ un vecteur de paramètres tiré aléatoirement dans la loi a posteriori jointe obtenue à partir d'un ajustement bayésien préalable du modèle génératif aux données de la LSS.

Dans une optique de comparaison de résultats, nous avons suivi les travaux de Walsh et Kaiser (2011) en nous concentrant sur l'estimation de deux excès de risque relatif

spécifiques. L'un correspond à l'ERR d'un homme de 55 ans ayant reçu une dose à la moelle osseuse de 1 Sievert à l'âge de 30 ans (noté ERR1 par la suite); le second correspond à l'ERR d'un garçon de 7 ans ayant reçu une dose de 0.1 Sievert à l'âge de 2 ans (noté ERR2 par la suite). ERR1 permet de considérer le cas d'un homme fortement exposé et ayant les caractéristiques d'âge moyennes observées dans la LSS soit 55 ans à la fin du suivi et 30 ans au moment des bombardements d'Hiroshima-Nagasaki. Cette population est fortement représentée dans la LSS. Par comparaison avec ERR1, ERR2 permet de s'intéresser au cas d'un enfant exposé très jeune et ayant reçu une dose à la moelle osseuse 10 fois plus petite. Cette population est faiblement représentée dans la LSS.

3.3 Inférence statistique

Dans le contexte fréquentiste, l'estimateur du maximum de vraisemblance des paramètres inconnus, l'intervalle de confiance à 95% associé ainsi que les critères AIC et BIC ont été calculés pour chacun des cinq modèles candidats et chacun des 500 jeux de données simulés.

Concernant la mise en oeuvre de l'approche BMA, nous avons implémenté un algorithme d'échantillonnage préférentiel adaptatif [5] qui a permis d'obtenir simultanément, pour chaque jeu de données simulées:

- un estimateur sans biais de la vraisemblance marginale de chaque modèle candidat - quantité indispensable au calcul des poids des modèles dans une approche BMA (cf. équation (2))- à partir d'un échantillon de particules issues d'une fonction d'importance Student multivariée;
- un échantillon approché de la loi a posteriori des paramètres inconnus de chaque modèle candidat, moyennant une étape de ré-échantillonnage des particules. Ces échantillons sont indispensables à l'estimation des lois a posteriori de ERR1 et ERR2 (cf. section 3.2), intégrées sur les 5 modèles candidats. De ces lois a posteriori ont pu être dérivés les médianes a posteriori de ERR1 et ERR2 ainsi que les intervalles de crédibilité à 95% associés.

Afin de recycler les simulations obtenues au cours de l'échantillonnage préférentiel adaptatif et ainsi de réduire les temps de calculs, rendre la répartition des poids des particules générées la plus uniforme possible, et ainsi réduire au mieux la variance de l'estimateur de Monte-Carlo (MC) résultant pour la vraisemblance marginale, nous avons considéré la stratégie de post-traitement des poids des particules générées proposée par Owen et Zhou [5], qui considèrent l'ensemble des particules générées comme étant issues d'un mélange déterministe de toutes les lois d'importance considérées.

Les estimateurs MMI et BMA ainsi que les intervalles de confiance et de crédibilité à 95% respectivement associés ont été calculés dans le contexte où le vrai modèle ayant servi à générer les données appartient ou non à la liste des modèles candidats.

4 Résultats

Trois statistiques ont été calculées afin de comparer les performances d’ajustement des approches BMA, MMI et de sélection de modèle (fréquentiste et bayésienne): 1/ le taux de couverture à 95% des excès de risque d’intérêt ERR1 et ERR2 ainsi que la variabilité de MC associée; 2/ le biais moyen sur 500 jeux de données associé à l’estimation de ERR1 et ERR2 ainsi que la variabilité de MC associée; 3/ l’erreur quadratique moyenne sur 500 jeux de données associée à l’estimation de ERR1 et ERR2 ainsi que la variabilité de MC associée. Ces statistiques ont également été calculées sous le ”vrai” modèle génératif.

Les premiers résultats obtenus indiquent que, lorsque le ”vrai” modèle ayant servi à générer les données ne fait pas parti des modèles candidats, l’approche BMA fournit globalement de meilleures performances d’ajustement aux données que l’approche MMI et qu’une approche par sélection bayésienne de modèle basée sur le facteur de Bayes. Enfin, l’approche BMA semble fournir des résultats d’ajustement presque aussi bons que sous le ”vrai” modèle ajusté. Les performances d’ajustement sont, comme prévu, globalement meilleures pour l’estimation de ERR1 par rapport à ERR2.

5 Perspectives

A notre connaissance, il s’agit de la première étude par simulations réalisée pour comparer l’approche BMA à l’approche MMI, qui est de plus en plus utilisée en épidémiologie des rayonnements ionisants mais également dans d’autres disciplines comme l’écologie. Cette étude illustre la possible valeur ajoutée d’une approche par combinaison de modèles par rapport à une approche par sélection de modèle pour l’estimation d’un risque. Des analyses complémentaires sont actuellement en cours afin de consolider ces premiers résultats.

Bibliographie

- [1] Burnham, K.P et Anderson, D.R. (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd Edition, Springer, New-York.
- [2] Walsh, L. et Kaiser, J.C. (2011), Multi-model inference of adult and childhood leukaemia excess relative risks based on the Japanese A-bomb survivors mortality data (1950-2000), *Radiation and Environmental Biophysics*, 50(1):21-35.
- [3] Hoeting, J.A, Madigan, D., Raftery, A.E. et Volinsky, C.T. (1999) Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382-417.
- [4] Ozasa, K. et al. (2012) Studies of the Mortality of Atomic Bomb Survivors, Report 14, 1950-2003: An Overview of Cancer and Noncancer Diseases. *Radiation Research*, 177(3):229-243.
- [5] Owen, A. et Zhou, Y. (2000), Safe and effective importance sampling, *Journal of the American Statistical Association*, 95, 135-143.