

Mise à jour séquentielle des prédictions d'un modèle hiérarchique pour l'aide au suivi médical

Sandrine Micallef, Billy Amzal, Frédéric Y. Bois

► **To cite this version:**

Sandrine Micallef, Billy Amzal, Frédéric Y. Bois. Mise à jour séquentielle des prédictions d'un modèle hiérarchique pour l'aide au suivi médical. 37. Journées de Statistiques, Jun 2005, Pau, France. pp.NC, 2005. <ineris-00972488>

HAL Id: ineris-00972488

<https://hal-ineris.archives-ouvertes.fr/ineris-00972488>

Submitted on 3 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TITRE (MISE À JOUR SEQUENTIELLE DES PRÉDICTIONS D'UN MODÈLE HIÉRARCHIQUE POUR L'AIDE AU SUIVI MÉDICAL)

Sandrine Micallef¹, Billy Amzal² & Frédéric Bois¹

¹ EA 3901 TOXI-DMAG, INERIS, Institut National de l'Environnement Industriel et des Risques, Parc Technologique ALATA, BP 2, 60550 Verneuil en Halatte, France

²Novartis Pharma AG. Postfach. CH-4002 Basel, Switzerland

Abstract

Caffeine treatment is widely used in nursing care to reduce the risk of apnea of premature neonates. To assure a positive effect of the treatment (reduction of apnea events) and, at the same time, reduce the occurrence of adverse effects (e.g., tachycardia), pediatricians try to keep caffeine blood concentration within a therapeutic interval. Therefore, caffeinemia is an important indicator of therapeutic efficiency against apnea in neonates. The present study aimed at building a tool as a basis for a medical decision support tool, able to predict caffeine concentration in blood after treatment taking into account the wide physiological variability of the population. We start by building a hierarchical model and use a numerical Bayesian approach to infer on its parameters values. A discussion on model building, including model checking and model adequacy test, is then presented within a Bayesian framework. The second part of this paper proposes an original way to update caffeinemia predictions as data of an incoming patient are collected. This involves the sequential updating of model parameters (at the individual and population levels) *via* a stochastic particle algorithm.

Résumé

Un traitement par la caféine est souvent administré au nouveau-né prématuré pour réduire le risque d'apnée. Pour garantir un effet positif du traitement et réduire l'occurrence des effets indésirables (tachycardie, perturbation du sommeil), les pédiatres tentent de maintenir la concentration sanguine de caféine dans un intervalle thérapeutique. Pour cette raison, la caféinémie est un indicateur important dans le traitement des apnées des enfants prématurés. La présente étude a pour objectif de construire un outil d'aide au suivi thérapeutique, capable de prédire la concentration sanguine de caféine suite au traitement d'un patient compte tenu de la grande variabilité physiologique observée dans cette population. Nous commençons par construire un modèle hiérarchique de population et utilisons l'approche bayésienne pour inférer sur les paramètres. Une discussion sur la construction, la vérification et l'adéquation de ce modèle est ensuite présentée. La seconde partie de cette étude

propose une façon originale de mettre à jour les prédictions du modèle hiérarchique lorsque des données sont collectées sur un nouveau patient. Cette mise à jour séquentielle des paramètres (individuels et de population) a été menée à partir d’un algorithme particulière.

Mots-clés *modèle hiérarchique de population, inférence Bayésienne, algorithmes MCMC, algorithme particulaire*

1 Introduction

Les nouveau-nés prématurés présentent, à la naissance, un certain nombre de déficits physiologiques en raison de leur immaturité. En particulier, des pauses respiratoires (apnées) surviennent de façon spontanée et fréquente. Pour réduire l’occurrence de ce phénomène, un traitement par la caféine est administré. Malgré des protocoles d’administration bien établis, les concentrations plasmatiques (et donc les effets résultants) sont très variables selon les nouveau-nés. La possibilité d’effectuer des prélèvements (pour mesurer la concentration plasmatique) étant limitée, les pédiatres se réfèrent généralement à des protocoles prédéfinis. Des adaptations “empiriques” de la dose sont parfois réalisées essentiellement sur la base d’observations cliniques. L’objectif de ce travail est d’apporter une aide au clinicien dans le traitement des apnées. Un outil d’aide à la décision a donc été créé. Il est basé sur des concepts statistiques bayésiens et doit permettre d’individualiser le traitement de patients sur la base de données historiques et de données issues du suivi thérapeutique des sujets considérés.

2 Matériel et Méthode

Les données pharmacocinétiques sont issues de 35 fichiers médicaux de nouveau-nés traités dans l’unité de soins intensifs du département pédiatrique de l’hôpital d’Amiens. Pour chaque patient, les données extraites regroupent de 1 à 4 caféinémie(s) mesurée(s) après l’administration d’un traitement X , et des données quotidiennes de masses, M_t . D’autres covariables, comme l’âge gestationnel (A_0) et la masse corporelle à la naissance (M_0) ont également été relevées.

2.1 Le modèle hiérarchique

Pour analyser les données de caféinémie, un modèle pharmacocinétique [1] capable de décrire sa distribution et son élimination a été utilisé. Pour être prédictif et utilisable en clinique, le modèle pharmacocinétique prend en compte l’évolution de la masse corporelle et de la clairance métabolique de la caféine. Le corps du nourrisson est modélisé par un compartiment homogène représentant l’eau du corps dont le volume est noté $V(t)$. La

caféine est éliminée avec une clairance $CL(t)$ correspondant au volume épuré par unité de temps. La variation de la concentration de caféine $c(t)$ au cours du temps est donnée par l'équation

$$\frac{dc(t)}{dt} = -\frac{CL(t)}{V(t)}c(t).$$

Le modèle décrit également la dépendance temporelle de M , V et CL à partir de fonctions linéaires déterminées à partir des paramètres $\alpha_1, \alpha_2, \tau, f_\tau, f_{25}, CL_0, \gamma$. Selon ce modèle, la réponse d'un individu au traitement par la caféine est conditionnée par le vecteur de paramètre $\theta = (\alpha_1, \alpha_2, \tau, f_\tau, f_{25}, CL_0, \gamma)$ et les covariables (M_0 et A_0).

La structure anatomo-physiologique du corps étant sensiblement la même pour tous les nouveau-nés, le modèle pharmacocinétique peut décrire l'élimination de la caféine pour tous les patients de l'étude. Les différences entre individus s'expliquent par des caractéristiques physiologiques différentes qui correspondent à des valeurs différentes de paramètres pharmacocinétiques θ . Afin de tenir compte de la variabilité inter-individuelle, le modèle pharmacocinétique a donc été inséré dans un modèle hiérarchique de population [2] (Figure 1). Au niveau de la population, les paramètres individuels du modèle sont supposés distribués normalement ou log-normalement (en fonction de leur nature).

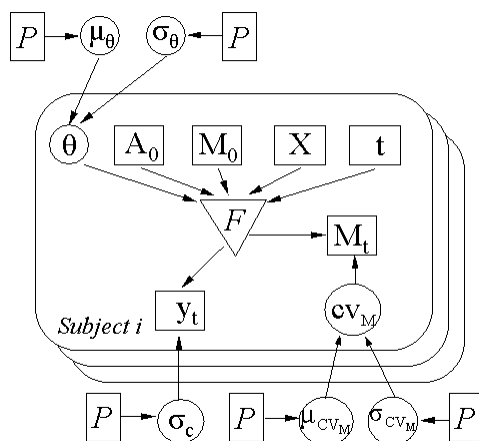


Figure 1: DAG du modèle hiérarchique. Les symboles utilisés sont : P , distribution *a priori*; μ , moyenne de population ; σ , variance de population ; t , temps d'observation ; θ , paramètres inconnus ; A_0 et M_0 , les covariables ; X , le traitement administré ; M_t , la masse corporelle mesurée au temps t ; cv_M , le coefficient de variation de la masse et Y_t , la caféinémie mesurée. F représente le modèle dynamique de bio-distribution. Les variables connues (ou supposées connues) sont contenues dans des carrés, les variables inconnues dans des cercles et le triangle correspond à un lien déterministe.

Compte tenu de la nature des observations (masse et caféinémie), un modèle d'erreur de mesure est utilisé. Nous avons supposé que les mesures de caféinémies sont entachées

d’une erreur lognormale. Concernant les masses, un modèle normal a été choisi. La variance de cette loi est supposée égale au produit de la masse à l’instant t par un coefficient de variation cv_M spécifique à l’individu.

2.2 Inférence et adéquation du modèle

Le modèle de population a été ajusté aux données observées Y_t à partir de techniques bayésiennes. Les paramètres du modèle ont été estimés par échantillonnage MCMC hors de la loi jointe *a posteriori* des paramètres individuels et de population. L’ajustement du modèle a ensuite été vérifié en étudiant successivement les distributions (*a priori* et *a posteriori*) des paramètres puis les résidus associés aux prédictions. Le test d’adéquation a consisté en différents exercices : cross-validation, étude des intervalle de confiance des prédictions, comparaison avec d’autres modèles de population. Ces tests ont montré que notre modèle était adéquat pour la prédiction de la pharmacocinétique des nouveau-nés d’Amiens.

2.3 Mise à jour séquentielle par méthode particulière

Les paramètres individuels d’un enfant venant de naître peuvent être inférés sur la base des paramètres de population. Mais lorsque des données sont collectées, les paramètres de cet individu doivent être mis à jour. Nous proposons de le faire de façon séquentielle à partir d’un algorithme particulière [3, 4, 5] puisqu’une distribution *a posteriori* a déjà été obtenue. Comme l’inférence repart de la loi *a posteriori* plutôt que de la loi *a priori*, les calculs sont plus rapides que lors d’une calibration bayésienne “classique” [6] (environ 1500 fois plus rapide, dans notre exemple). Les paramètres individuels et de population sont conjointement et rapidement mis à jour.

Soit un échantillon de paramètres $(\theta_1 \dots \theta_N)$, issu de la distribution *a posteriori* $p(\theta)$, notons y les nouvelles données. Chaque θ_i est appelé une “particule”. Nous voulons transformer notre échantillon en un autre de même taille N mais, cette fois, distribué selon $p(\theta | y)$. L’algorithme s’écrit de la façon suivante :

- Pour chaque particule $i = 1..N$, calculer les poids normalisés w_i (comme dans le schéma “Sampling Importance Resampling” de Rubin [7]) (voir aussi [8])

$$w_i \propto \frac{p(\theta_i | y)}{p(\theta_i)} \propto p(y | \theta_i)$$

Notons ici que ces poids sont facilement obtenus puisqu’ils prennent la forme d’un terme de vraisemblance.

- Rééchantillonner N particules $(\theta'_1 \dots \theta'_N)$ à partir de $(\theta_1 \dots \theta_N)$ dans une distribution multinomiale $\mathcal{M}(N, w_1, \dots, w_N)$

- Pour contourner les problèmes de dégénérescence, enrichir l'échantillon en utilisant un noyau de Markov Q dont la distribution stationnaire est $p(\theta | y)$. Pour chaque $i = 1..N$, tirer θ''_i de $Q(\cdot | \theta'_i)$. Par définition, cela transforme un échantillon approché $(\theta'_1... \theta'_N)$ de $p(\theta | y)$ en un autre échantillon $(\theta''_1... \theta''_N)$ issu de $p(\theta | y)$, mais maintenant, plus riche (plus diverse). Cet échantillon est l'échantillon final mis à jour à partir des données y .

Dans le cadre de l'exemple proposé, nous avons choisi un pas de Metropolis-Hasting avec un échantillonneur indépendant [9] dans lequel les candidats proposés pour les paramètres de population sont issus de la loi *a posteriori* originelle.

3 Résultats

Le modèle a été mis à jour séquentiellement à partir d'une mesure de caféinémie observée chez un nouveau patient. Les nouvelles prédictions (Figure 2B) sont bien meilleures que celles inférées à partir des paramètres de population avant la mise à jour (Figure 2A). Le biais a diminué et les intervalles de confiance sont également moins dispersés. Par ailleurs, puisqu'elle met à jour conjointement les paramètres de population, cette procédure rend disponible toute l'information (passée et présente) pour la prédiction d'un nouveau sujet entrant dans le service de néonatalité.

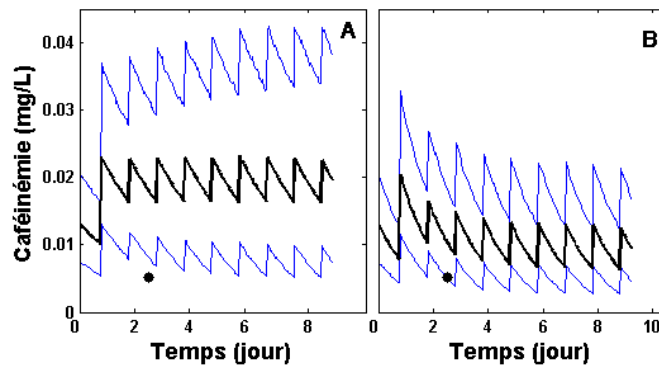


Figure 2: Prédictions de la caféinémie en fonction du temps pour le nouveau patient. Les intervalles de crédibilité [5% ; 95%] des prédictions sont tracés en gris, la moyenne en noir. La donnée est représentée à partir d'un point. Le panneau A donne les caféinémies prédites à partir de l'échantillon de paramètres issu de la loi *a posteriori* originelle, en utilisant les covariables du patient. Les prédictions du panneau B ont été calculées à partir des mêmes covariables et de l'échantillon de paramètres mis à jour *via* l'algorithme particulière.

Bibliographie

- [1] Rowland, M. Benet, L.Z. and Graham, G.G. (1973) Clearance concepts in pharmacokinetics. *Journal of Pharmacokinetics and Biopharmaceutics*, 1, 123–136.
- [2] Bois, F.Y. (2001) Applications of population approaches in toxicology. *Toxicology Letters*, 120, 385–394.
- [3] Doucet, A. and de Freitas, N. and Gordon, N. (2001) *Sequential Monte Carlo Methods in Practice*, Springer-Verlag.
- [4] Chopin, N. (2002) A sequential particle filter method for static models, *Biometrika*, 89, 539–552.
- [5] Tierney, L. and Mira, A. (1999) Some adaptive Monte Carlo methods for Bayesian inference, *Statistics in Medicine*, 18, 2507–2515.
- [6] Ridgeway, G. and Madigan, D. (2003), A sequential Monte Carlo method for Bayesian analysis of massive datasets, *Data Mining and Knowledge Discovery*, 7(3), 301–319.
- [7] Rubin, D.B. (1998), Using the SIR algorithm to simulate posterior distributions, in *Bayesian Statistics*, 3, 395–402, Oxford University Press.
- [8] Geweke, J. (1989), Bayesian inference in econometric models using Monte Carlo integration, *Econometrica*, 24, 1317–1399.
- [9] Tierney, L. (1994), Markov chains for exploring posterior distributions, *Annals of Statistics*, 22(4), 1701–1762.