Abstract Translations

DISCUSSION PAPER

R. L. Prentice, M. Pettinger, and
G. L. Anderson 899

Statistical Issues Arising in the Women’s Health Initiative

Nous proposons un bref panorama du plan de l’essai clinique et de l’étude observationnelle de la WHI, ainsi qu’un résumé des résultats obtenus à partir de la partie de l’essai clinique dévolue à la thérapeutique hormonale post-ménopause. Depuis son démarrage en 1992, la WHI a été confrontée à plusieurs questions de nature statistique pour lesquelles de plus amples développements méthodologiques étaient nécessaires. Parmi ceux-ci, la modélisation des erreurs de mesure, et des méthodes d’analyse pour l’évaluation de la diététique et de l’activité physique; des méthodes de surveillance d’essai clinique lorsque les traitements peuvent concerner plusieurs champs cliniques aussi bien en bénéfice qu’en réaction adverse; le plan d’étude et les procédures d’analyse pour des données génomiques et protéiques de dimension élevée; et aussi des procédures d’analyse de données de durée lorsque les rapports de risque dépendent du temps. Ce dernier point apparaît important pour comprendre les différences entre les résultats de l’essai clinique et ceux de l’étude observationnelle en ce qui concerne le traitement hormonal postménopause et la pathologie cardio-vasculaire.

REGULAR ARTICLES

E. R. DeLong, L. P. Coombs,
T. B. Ferguson, Jr., and E. D. Peterson 942

The Evaluation of Treatment When Center-Specific Selection Criteria Vary with Respect to Patient Risk

En matière de soins médicaux, beaucoup de standards reposent sur les effets démontrés de divers procédés ou stratégies thérapeutiques. A la différence des traitements pharmacologiques, ces procédés ou stratégies ne sont pas nécessairement évalués par des essais cliniques rigoureux, mais, bien souvent, à partir de données observationnelles. Lorsqu’il s’agit d’étudier l’efficacité de ces procédés, il faut non seulement se poser le problème de l’ajustement par les facteurs de risque des patients, mais aussi prendre en considération l’« effet centre », une donnée importante et souvent négligée. En effet, la qualité des soins et la tendance à employer certains traitements plutôt que d’autres varient selon les centres. Il s’ensuit une corrélation naturelle entre les résultats d’un même centre, ainsi que la possible apparition de facteurs de confusion liés à cet effet centre ; ces problèmes doivent tous deux être traités dans le cadre de l’ajustement par les facteurs de risque. De surcroît, ceux des critères de sélection de la stratégie thérapeutique qui sont spécifiques à chaque centre peuvent aussi varier en fonction du risque présenté par le patient. Pour toutes ces raisons, il est important de bien séparer, d’une part, les effets intra-centres de la stratégie thérapeutique et, d’autre part, les effets inter-centres, qui relèvent davantage des performances respectives des centres. Le premier objectif de cet article est l’exploration et l’extension, dans le cas de données binaires, des méthodes usuelles de gestion des biais de confusion d’effets liés aux centres, notamment lorsque la prise en compte du risque patient, dans la sélection de la stratégie thérapeutique, varie selon les centres. Une étude de simulation compare les résultats obtenus par différentes méthodes analytiques et met en évidence la nécessité, lorsqu’on étudie l’efficacité d’un procédé, de considérer simultanément les facteurs de risque des patients et l’effet centre. L’article présente un exemple où l’on examine l’effet d’une extubation précoce, suite à un pontage coronarien.

X. Jin, B. P. Carlin, and S. Banerjee 950

Generalized Hierarchical Multivariate CAR Models for Areal Data

Une des applications les plus communes des modèles à variables régionalisées dans les domaines de la médecine et de la santé publique est l’étude de la distribution spatiale des maladies. Si on dispose de plusieurs mesures en chaque point de l’espace (par exemple, une information sur p>=2 maladies pour des populations ou des régions identiques), on doit utiliser des modèles à variables régionalisées multivariés, ceci afin de prendre en compte la dépendance entre les variables et entre les mesures réalisées en différents points de l’espace. Nous proposons dans cet article une nouvelle famille de modèles multivariés conditionnels autorégressifs généralisés (GMCAR) pour traiter des variables régionalisés et montrons en quoi ils apportent aux MCAR. Notre approche diffère des précédentes en ce qu’elle exprime directement la loi de distribution conjointe pour un champ aléatoire de Markov multivarié (MRF) par la spécification de modèles marginaux et conditionnels plus simples. Ceci permet une réduction significative de la quantité de calcul concernant le modèle spatial hiérarchique à effets aléatoires, où des distributions a posteriori sont calculées par simulation Monte-Carlo de chaînes de Markov (MCMC). Nous comparons notre approche avec des
modèles MCAR de la littérature par simulation au moyen de l’erreur quadratique moyenne (AMSE) et d’un critère de sélection utile pour les modèles hiérarchiques, le critère d’information de déviance (DIC; Spiegelhalter et al., 2002). Enfin, nous présentons une application à des données réelles de notre approche GMCAR: la modélisation des taux de mortalité par cancer du poumon et de l’oesophage de 1991 à 1998 dans les comtés du Minnesota.

**H. Bang and J. M. Robins** 962

*Doubly Robust Estimation in Missing Data and Causal Inference Models*

Notre but est la construction d’un estimateur doublement robuste (DR) pour un modèle causal avec données manquantes. Dans un modèle avec données manquantes, un estimateur est DR si il demeure convergent quand soit le défaut des données, soit la distribution des données complètes est correctement modélisé mais pas nécessairement les deux. Comme avec des données d’observation on n’est jamais sûr que les deux soient corrects, le mieux qu’on puisse espérer est de trouver un estimateur DR. Contrairement aux estimateurs basés sur le maximum de vraisemblance standard ou une pondération par probabilité inverse (sans augmentation), un estimateur DR donne à l’analyste deux chances au lieu d’une de faire une inférence valable. Dans un modèle d’inférence causale, un estimateur DR reste convergent quand on spécifie correctement soit le modèle de l’affectation des traitements, soit celui de la distribution des effets, mais pas les deux. Comme on n’est jamais certain de ces choix avec des données d’observation, un estimateur DR devrait se montrer supérieur aux approches antérieures. Nous présentons en effet des résultats de simulations qui démontrent que les propriétés à distance finie des estimateurs DR sont aussi impressionnantes que ce que prédit la théorie. La méthode proposée est appliquée au traitement des données d’un essai clinique dans le domaine cardiovasculaire.

**M. Wang and J. M. Williamson** 973

*Generalization of the Mantel-Haenszel Estimating Function for Sparse Clustered Binary Data*

Nous étendons la fonction d’estimation de Mantel-Haenzel à l’estimation à la fois des corrélations intra-groupes et des effets principaux pour des données binaires groupées et creuses. Nous proposons à la fois une approche vraisemblance composite et une approche fonction d’estimation pour l’analyse de telles données. Les estimateurs proposés sont convergents et asymptotiquement normaux. Des résultats de simulation montrent que les deux approches sont comparables en terme de biais et d’efficacité; cependant l’approche équation d’estimation est plus simple à programmer. Une analyse des données de l’enquête «Georgia High Blood Pressure» est utilisée pour illustration.

**Y. Shen and X. Huang** 992

*Nonparametric Estimation of Asymptomatic Duration from a Randomized Prospective Cancer Screening Trial*

Nous proposons une estimation non paramétrique de la distribution de la phase pré-clinique d’un cancer dans un essai randomisé de dépistage précoce. Dans les études de dépistage en cancérologie, la durée de la phase pré-clinique est de grand intérêt pour mieux comprendre l’histoire naturelle de la maladie et pour développer des stratégies optimales de dépistage. Pour estimer non paramétriquement la distribution du temps de séjour, nous estimons d’abord non paramétriquement la distribution de l’âge au début de la phase pré-clinique à partir des données du bras expérimental d’un essai randomisé de dépistage et la distribution de l’âge au début de la phase clinique à partir des données du bras témoin de l’essai randomisé de dépistage. Finalement, par déconvolution, les deux distributions estimées conduisent à un estimateur non paramétrique du délai écoulé entre le début de la phase pré-clinique et celui de la phase clinique. Nous illustrons cette
métodologie à partir des données d’un essai randomisé de dépistage du cancer du sein.

Z. Pan and D. Y. Lin

Goodness-of-Fit Methods for Generalized Linear Mixed Models

Nous développons des méthodes graphiques et numériques pour vérifier la validité des modèles linéaires mixtes généralisés. Ces méthodes sont basées sur les sommes cumulées des résidus en fonction de covariables ou des valeurs prédites de la variable-réponse. Sous le modèle posé, les distributions asymptotiques de ces processus stochastiques peuvent être approximées par certains processus gaussiens centrés dont les structures observées sur les résidus sont le reflet d’un mauvais simulé suivant la distribution sous l’hypothèse nulle. Ces méthodes sont particulièremment utiles pour vérifier la forme des réalisations cumulées en fonction de covariables ou des valeurs prédites. Les méthodes proposées sont particulièremment utiles pour vérifier la forme fonctionnelle d’une liaison avec une covariable ou la fonction de lien. Des études de simulation extensives montrent que les tests d’adéquation au modèle ont des puissances convenables et sont sensibles à une mauvaise spécification du modèle. Des applications à deux études médicales conduisent à améliorer les modèles.

H.-C. Yang and A. Chao

Modeling Animals’ Behavioral Response by Markov Chain Models for Capture–Recapture Experiments

Nous proposons une approche de chaîne de Markov bivariée incluant à la fois des effets comportementaux durables (à long terme) et des effets éphémères (à court terme), pour l’analyse des données de capture-recapture. L’histoire de capture de chaque animal est modélisée comme une chaîne de Markov dans l’espace bivarié des états, ces états étant déterminés par le statut de capture (capturé ou non capturé) et par le statut vis-à-vis du marquage individuel (animal marqué ou non marqué). Dans ce cadre, nous estimons l’effectif de la population et les probabilités de transition par une méthode de vraisemblance conditionnelle. Le modèle comportemental classique qui suppose seulement un effet comportemental durable est vu comme un cas particulier du modèle markovien bivarié. Un autre cas particulier, qui suppose seulement un effet à court terme, se réduit à une chaîne de Markov univariée basée sur le statut de capture. En ajoutant à ce modèle des effets temporels, nous obtenons un modèle dans lequel, contrairement aux modèles dérivés du modèle comportemental classique, tous les paramètres sont identifiables. Nous analysons des données réelles pour illustrer l’utilité des modèles markoviens dans la compréhension des réponses comportementales des animaux. Nous présentons aussi les résultats de simulations permettant d’évaluer la performance des estimateurs.

C. Hans and D. B. Dunson

Bayesian Inferences on Umbrella Orderings

Dans les applications de la régression avec des prédicteurs qualitatifs, l’intérêt se focalise souvent sur la comparaison de l’hypothèse nulle d’homogénéité à une alternative ordonnée. Cet article propose une approche bayésienne pour traiter ce problème dans le cas d’ajustement à des modèles de régression gaussiens non linéaires et probit. On prend une densité conditionnelle a priori conjuguée consistant en un mélange de masses ponctuelles en zéro et de densités normales tronquées, avec un paramètre de point de rupture (qui peut être inconnu) inclus à l’ordre en \( \ll parapluie \rr \) adapté. On considère deux stratégies de choix de la loi a priori: (1) une approche bayésienne à la Bonferroni dans laquelle la probabilité de l’hypothèse nulle globale est spécifiée et les hypothèses locales sont considérées indépendantes; et (2) une approche qui traite ces probabilités comme au hasard. Une seule passe d’échantillonneur de Gibbs peut être utilisée pour obtenir les probabilités a posteriori pour les différentes hypothèses et pour estimer les coefficients de régression ainsi que les valeurs prédites, soit avec le modèle moyen soit avec le modèle préféré. Ces méthodes sont appliquées aux données d’une étude de carcinogénèse.

P. D. Hoff

Subset Clustering of Binary Sequences, with an Application to Genomic Abnormality Data

Cet article propose une méthode pour modéliser des données binaire multivariées au sein de groupe dans le cas où les caractéristiques qui identifient un groupe du reste de la population peuvent dépendre du groupe lui-même. Cette approche repose sur un modèle multivarié de mélange de processus de Dirichlet qui permet simultanément d’estimer le nombre de groupes, l’appartenance aux groupes et les paramètres spécifiques de chaque groupe. Une telle méthode pour données groupées trouvent des applications dans l’analyse de données sur des anomalies génomiques pour lesquelles le développement de différents types de tumeurs peut dépendre de la présence de certaines anomalies à différents ensembles de sites tout au long du génome. De plus, ce type de modèle de mélange fournit un procédé général d’estimation non paramétrique en présence de séries de données binaires corrélées.

J. Harezlak, L. M. Ryan, J. N. Giedd, and N. Lange

Individual and Population Penalized Regression Splines for Accelerated Longitudinal Designs

Dans le cadre de schémas longitudinaux accélérés (ALD), les individus entrent dans l’étude à différents points de leur trajectoire de croissance et sont observés sur une courte durée par rapport à la durée totale d’intérêt. Les données de type...
ALD sont combinées dans des unités indépendantes pour fournir une estimation d’une courbe de population globale et les prédictions des schémas individuels de changement. Comme une extension modeste du travail de Ruppert et al. (2003), nous développons une procédure d’estimations efficace de méthodes semi-paramétriques longitudinales dans le cadre d’échantillons de type ALD. Nous comparons les schémas longitudinaux ALD équilibrés et complets à partir des données de l’étude de croissance de Berkeley et appliquons notre méthode aux mesures longitudinales du volume de la structure cérébrale MRI longitudinales d’une étude en cours liée au développement. Des applications potentielles s’étendent au-delà des études de croissance à beaucoup d’autres domaines dans lesquels le coût et les contraintes de faisabilité imposent des restrictions sur la taille de l’échantillon et sur les nombres et le rythme des mesures répétées entre sujets.

K. Larsen

The Cox Proportional Hazards Model with a Continuous Latent Variable Measured by Multiple Binary Indicators

Cet article s’appuie sur l’étude “Santé et Vieillissement des femmes” où l’on a collecté l’information sur l’état physique d’un groupe de femmes âgées, ainsi que celle sur le décès. La question centrale était de déterminer si l’existence de difficultés à l’accomplissement des tâches quotidiennes s’accompagne d’un taux de mortalité plus élevé. Pour y répondre, on utilise un modèle de régression logistique à deux paramètres pour la modélisation des données binaires d’un questionnaire reposant sur une variable latente continue, la difficulté dans la vie quotidienne.

Le modèle de Cox est utilisé pour l’information de survie, et la variable latente y est incluse comme variable explicative ajoutée aux autres variables observées. Les paramètres sont estimés en maximisant la vraisemblance de la distribution jointe des items et de la survie. Outre la présentation d’un nouveau modèle statistique, ce papier illustre l’utilisation de ce modèle dans un cadre de données réelles, et évoque les aspects pratiques de la construction du modèle, des diagnostics, et de l’interprétation des paramètres.

D. Wu, G. L. Rosner, and L. Broemeling

MLE and Bayesian Inference of Age-Dependent Sensitivity and Transition Probability in Periodic Screening

Cet article généralise les modèles probabilistes concernant le dépistage des cancers du sein. L’objectif est ici d’effectuer des inférences statistiques concernant l’influence de l’âge sur la probabilité de transition de l’état sain à l’état pré-clinique et sur la sensibilité. La motivation vient d’un programme de dépistage du cancer du sein au cours duquel une cohorte de femmes initialement asymptomatiques doit subir une série d’exams répétés. L’âge est utilisé comme covariable pour l’estimation simultanée de la probabilité de transition et de la sensibilité du dépistage, d’abord d’un point de vue fréquentiste puis dans une perspective bayésienne. Nous appliquons notre méthode aux données de l’étude HIP.

La méthodologie développée concerne plus généralement le dépistage précoce des maladies chroniques évolutives.

H.-G. Müller and Y. Zhang

Time-Varying Functional Regression for Predicting Remaining Lifetime Distributions from Longitudinal Trajectories

Un objectif récurrent dans les études longitudinales sur le vieillissement et la longévité a été l’investigation de la relation entre l’âge au décès et les valeurs actuelles de la trajectoire d’une covariable longitudinal qui quantifie l’activité réproductive ou tout autre activité biologique. On propose une nouvelle technique pour prédire les distributions de l’âge au décès pour des situations où tout l’historique de la covariable est inclus dans la prédiction. Les trajectoires prédites jusqu’au temps courant sont représentés comme des scores en composants principaux fonctionnels qui varient dans le temps et qui sont continuellement mises à jour en fonction de la progression du temps. Ces trajectoires sont considérées comme des variables de prédiction qui varient dans le temps qui rentrent dans une classe de modèles de régression fonctionnels que nous proposons. Pour des données biodémographiques, on montre comment ces méthodes peuvent être appliquées pour obtenir des prédictions pour l’âge au décès, ainsi que les estimations de la distribution de vie restant, des estimations des quantiles et des intervalles de prédiction pour la vie restante. Des estimations et des prédictions sont obtenues pour les sujets individuellement, basés sur les trajectoires observées de leur comportement. Pour obtenir un seul index, cette technique comporte également une étape de réduction de la dimension. Les techniques proposées sont appliquées aux données longitudinales de ponte des œufs des femelles mouches, pour obtenir des prédictions de la vie restante et les distributions de l’âge au décès à partir de l’historique des événements observés jusqu’au temps en cours.

Z. Pang and A. Y. C. Kuk

A Shared Response Model for Clustered Binary Data in Developmental Toxicity Studies

Les lois de distribution actuellement disponibles pour modéliser la réponse fœtale en toxicologie du développement, telles que la loi bêta-binomiale, ont tendance à surévaluer la probabilité pour le fœtus de ne pas présenter de malformation, et ainsi à sous-estimer le risque qu’une portée comporte au moins un fœtus malformé. Par opposition au modèle extrabinomial à probabilité partagée, c’est ici un modèle à réponse partagée que l’on recommande, modèle du nombre aléatoire de fœtus d’une même portée présentant une même réponse; la formule explicite de la fonction de probabilité est fournie dans cet article. Des représentations graphiques suggèrent que ce modèle est exempt de toute surestimation de la probabilité qu’un fœtus ne présente pas de malformation. L’algorithme EM peut être utilisé pour estimer les paramètres du modèle. Les résultats d’une étude de simulation montrent que les estimateurs obtenus grâce à cet algorithme sont presque sans
biais et que le niveau de confiance réel des intervalles de confiance associés (basés sur les estimateurs usuels de l’écart-type) est proche du niveau nominal. Les résultats de la simulation suggèrent également que les estimateurs des probabilités marginales de malformation, issus du modèle à réponse partagée, sont robustes aux erreurs de spécification de la forme de la distribution, mais qu’il n’en va pas de même pour les estimateurs de la corrélation intra-portée et de la probabilité pour qu’une portée présente au moins un fœtus malformé. Le modèle proposé est appliqué à un jeu de données du United States National Toxicology Program. Pour la même relation dose-réponse, la qualité de l’ajustement du modèle à réponse partagée aux données est meilleure que celle du modèle basé sur une distribution bêta-binomiale, et comparable à celle du modèle basé sur la distribution puissance de paramètre q récemment proposée (Kuk, 2004, *Applied Statistics* 53, 369–386). Avantage du modèle à réponse partagée sur la distribution puissance de paramètre q: il est plus aisé d’interprétation et peut être généralisé plus facilement au cas multivarié. Pour illustration de cette dernière propriété, un modèle bivarié à réponse partagée est ajusté à des données de réponse fœtale faisant état de malformations des viscères et du squelette.

P. S. F. Yip, H.-Z. Lin, and L. Xi

*A Semiparametric Method for Estimating Population Size for Capture–Recapture Experiments with Random Covariates in Continuous Time*

Une procédure d’estimation semi-paramétrique est proposée pour les modèles de capture-recapture dans le but d’estimer la taille de la population dans le cas fermé. Les covariables individuelles sont possiblement dépendantes du temps et manquantes au temps de non capture et peuvent être mesurées avec erreur. Un système d’équations d’estimation basé sur un processus sur les covariables et les données est défini pour estimer les paramètres d’intérêt et la taille de la population. Ces équations d’estimation sont résolues par un algorithme similaire à l’algorithme EM. Les résultats des simulations montrent que les procédures proposées fonctionnent mieux que l’estimation naïve. Dans certains cas, elles sont même meilleures que les estimations idéales, pour lesquelles les vraies valeurs des covariables sont disponibles pour l’ensemble des sujets capturés sur l’ensemble de la période expérimentale. Nous appliquons la méthode à une expérimentation de capture-recapture sur les espèces d’oiseau *Prinia aviventris* à Hong Kong.

**CONSULTANT’S FORUM**

R. M. Dorazio, H. L. Jelks, and F. Jordan

*Improving Removal-Based Estimates of Abundance by Sampling a Population of Spatially Distinct Subpopulations*

Un cadre de modélisation statistique pour l’estimation de l’abondance de sous-populations animales subdivisées spatialement et étudiées paréchantillonnages sans remises (prélèvements) est décrit. Pour illustrer ce cadre, des modèles hiérarchiques qui utilisent les distributions de Poisson et binomiale négative pour modéliser la variation d’abondance entre sous-populations et la distribution bêta pour modéliser les variations de probabilités de capture sont développés. Ces modèles sont ajustés aux comptages de prélèvement obtenus dans le suivi d’une espèce de poisson cataloguée comme en danger par les autorités fédérales. Les estimations d’abondance obtenues ont une précision semblable ou meilleure que celles calculées avec l’approche conventionnelle où les comptages de prélèvement sont analysés séparément pour chaque sous-population. L’extension des modèles hiérarchiques à la prise en compte de covariables d’abondance spatiales est naturelle et peut être utilisée pour identifier des caractéristiques importantes de l’habitat d’un animal ou pour prédire l’abondance des animaux dans des lieux non échantillonnés.

J. M. G. Taylor, Y. Wang, and R. Thiébaut

*Counterfactual Links to the Proportion of Treatment Effect Explained by a Surrogate Marker*

Dans un essai clinique randomisé, une statistique mesurant la proportion de l’effet du traitement sur l’événement clinique d’intérêt qui est expliquée par l’effet du traitement sur un marqueur de substitution est un concept utile. Nous étudions dans quelle mesure une statistique proposée pour estimer cette proportion peut avoir une interprétation causale selon les modèles à variables contre-factuelles. Dans la situation d’un marqueur de substitution et d’un événement d’intérêt binaires, deux modèles contre-factuels sont considérés; Les deux modèles incluent le concept de proportion de l’effet du traitement passant par le marqueur de substitution. En général, la statistique n’est égale à aucune des deux proportions issues des modèles contre-factuels et peut même être franchement différente. Des conditions sont indiquées pour lesquelles la statistique est égale aux proportions des modèles contre-factuels.

**READER REACTION**

M. J. Faddy and D. M. Smith

*Modeling the Dependence between the Number of Trials and the Success Probability in Binary Trials*

Un modèle pour essais binaires reposant sur une généralisation bivariée du processus de Poisson pour le nombre de succès et le nombre d’essais avec des probabilités de transition dépendantes des nombres cumulés de succès et d’essais est utilisé pour ré-analyser des données récemment publiées de Zhu, Eickhoff et Kaiser (2003, *Biometrics* 59, 955–961). Cette modélisation admet des distributions différentes pour les nombres d’essais et les nombres de succès conditionnellement au nombre d’essais qui généralisent les distributions binomiales et de Poisson, sans certaines des restrictions apparentes dans le modèle mixte beta-binomial-Poisson.
de Zhu et coll. (2003). Quelques différences très marquées apparaissent entre les résultats de cette analyse et ceux décrits dans Zhu et coll. (2003).

P. S. Albert and S. Hunsberger

*On Analyzing Circadian Rhythms Data Using Nonlinear Mixed Models with Harmonic Terms*

Wang et al. (2003, *Biometrics* 59, 804–812) ont développé une approche de lissage pour modéliser les rythmes circadiens avec effets aléatoires. Leur approche est flexible dans le sens où les covariables fixes et aléatoires peuvent influer à la fois l’amplitude et le décalage de phase d’une fonction périodique lissée non-paramétrique. Dans la motivation de leur approche, Wang et al. partent de l’insuffisance d’une simple fonction sinusoïdale. De plus, ils considèrent que « bien que l’addition d’harmoniques puisse améliorer l’ajustement, il est difficile de choisir le nombre d’harmoniques à introduire dans le modèle, et les résultats sont délicats à interpréter ». Nous contestons l'idée que les modèles harmoniques ne soient pas un outil profitable pour la modélisation de données circadiennes. Dans cette note, nous montrons comment les modèles mixtes non-linéaires permettent une alternative simple et flexible à l’approche de Wang et al. Nous montrons comment choisir le nombre d’harmoniques en utilisant la vraisemblance pénalisée pour obtenir un modèle flexible de rythme circadien et pour estimer l’effet des covariables sur les rythmes. Nous ajustons des modèles harmoniques aux données de rythme circadien du cortisol présentées par Wang et al. pour illustrer notre approche. Par ailleurs, nous évaluons les propriétés de notre procédure avec une petite étude de simulation. L’approche paramétrique proposée est une alternative à l’approche semi-paramétrique de Wang et al., et présente l’avantage supplémentaire de pouvoir être aisément implémentée sur la plupart des progiciels statistiques.

I. Novikov, B. Oberman, and L. Freedman

*Modification of the Computational Procedure in Parker and Bregman’s Method of Calculating Sample Size from Matched Case–Control Studies with a Dichotomous Exposure*

Plusieurs auteurs se sont intéressés au problème du calcul de la taille d’échantillon nécessaire pour une étude cas-témoin apparié dans le cas d’un facteur d’exposition binaire. L’approche de Parker et Bregman (1986) est à notre sens, l’une des plus satisfaisantes car elle s’appuie sur la spécification de grandeurs aisément accessibles à l’investigateur. Cependant l’implémentation qui en a été recommandée requiert une approximation de calcul. Nous montrons ici que cette approximation conduit à de mauvaises performances dans certaines situations extrêmes, et qu’elle peut être facilement remplacée par un calcul exact.