



Forces et faiblesses de différentes approches statistiques pour l'analyse d'évènements récurrents

Jérôme Tanguy, Anissa Elfakir, Sébastien Marque

► To cite this version:

Jérôme Tanguy, Anissa Elfakir, Sébastien Marque. Forces et faiblesses de différentes approches statistiques pour l'analyse d'évènements récurrents. 42èmes Journées de Statistique, 2010, Marseille, France, France. 2010. <inria-00494822>

HAL Id: inria-00494822

<https://hal.inria.fr/inria-00494822>

Submitted on 24 Jun 2010

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.

FORCES ET FAIBLESSES DE DIFFERENTES APPROCHES STATISTIQUES POUR L'ANALYSE D'EVENEMENTS RECURRENENTS

Jérôme Tanguy, Anissa Elfakir, Sébastien Marque

Danone Research, Centre Daniel Carasso, RD128, Avenue de la Vauve, F-91767 Palaiseau Cedex

Mots clés: évènements récurrents, données de comptage, modèles modifiés en zéro, modèles à barrière, modèles à fragilité partagée

Resumé

La modélisation d'événements récurrents est une problématique rencontrée dans une large variété de disciplines telles les sciences sociales, la recherche médicale, la santé publique, l'épidémiologie ou l'économie.

Dans le cadre d'une étude clinique en nutrition, un événement de nature récurrente peut être abordé selon deux approches différentes : une approche considérant l'événement comme une donnée de comptage, et une approche considérant l'événement comme une donnée type survie pour occurrence multiple. Dans les deux cas plusieurs modélisations sont présentées et comparées d'un point de vue clinique et statistique (tests de comparaison pour des modèles emboîtés et non emboîtés [1]). Forces, faiblesses et évolutions possibles de chaque approche sont soulignées.

La structure complexe de l'événement étudié offre une occasion d'aborder une grande variété de modèles statistiques. Pour l'approche type données de comptage, des modèles classiques tels que la régression binomiale négative sont appliqués, mais aussi des modèles à mélange de distribution comme les modèles modifiés en zéro, ou les modèles dit de « barrière » utilisant des distributions tronquées [2-4]. Une autre approche utilisant un modèle logit cumulatif est également testée pour son interprétation simple [5]. Enfin, les mêmes données seront abordées sous l'angle de l'analyse de survie via un modèle à fragilités partagées, permettant entre autre de prendre en compte de la dépendance des événements entre eux [6-8].

L'exploration à travers ces multiples méthodes statistiques permet d'obtenir des résultats divers et complémentaires, et contribue à améliorer les connaissances cliniques concernant l'événement étudié.

STRENGTHS, WEAKNESSES OF DIFFERENT STATISTICAL MODELS TO ANALYSE RECURRENT EVENTS

Jérôme Tanguy, Anissa Elfakir, Sébastien Marque

Danone Research, Centre Daniel Carasso, RD128, Avenue de la Vauve, 91767 Palaiseau Cedex

Keywords: recurrent events, count data, zero-inflated regression, hurdle regression, shared frailty model

Abstract

Modelling recurrent events is a common concern in a wide range of discipline as social sciences, medical and public health research, epidemiology, economy. For a clinical study in nutrition, a recurrent event issue is considered from different perspectives using two particular frameworks: count data and recurrent survival events data. For both frameworks a range of statistical models are presented and compared from clinical and statistical point of views (goodness-of-fit tests for nested and non nested models [1]). Strengths, weaknesses and possible further development of each approach are emphasized.

The complex pattern of the data offers an opportunity to tackle a wide variety of statistical models. For count data approach, classical count models are applied i.e. Poisson regression and negative binomial regression [2]. Moreover some quite less common models are carried out like mixture distribution models where count data are assumed to be generated by different processes, i.e. zero-inflated regression and hurdle regression [3,4]. An alternative approach using a cumulative logit model is also tested for its simple and appealing interpretation [5]. The event of interest is also regarded as individual recurrent time to event data with a focus on the shared frailty model [6]. This latter model is an extension of the Cox proportional hazards model and has become increasingly popular to take into account dependency between recurrent survival times [7,8].

The exploration of these statistical methods allows to obtain diverse and complementary results, and eventually help to increase the clinical knowledge about the event of interest.

Bibliographie

- [1] Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307-333.
- [2] McCullagh P, Nelder JA (1989). *Generalized Linear Models*. 2nd edition. Chapman & Hall, London.
- [3] Lambert, D (1992), Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing, *Technometrics*, Vol. 34, No. 1, 1 – 14.
- [4] Mullahy, John, "Specification and Testing of Some Modified Count Data Models," *Journal of Econometrics*, 33 (1986), pp. 341-365
- [5] Long, JS (1997) *Regression Models for Categorical and Limited Dependent Variables*, vol 7 of *Advanced Quantitative Techniques in the Social Sciences*. Thousand Oaks, CA: Sage.
- [6] Cook, R. J. and J. F. Lawless (2007). *The statistical analysis of recurrent events*. New York, Springer.
- [7] Pickles, A. and R. Crouchley (1995). A comparison of frailty models for multivariate survival data. *Statistics in Medicine* 14(13): 1447-1461.
- [8] Duchateau, L., P. Janssen, et al. (2003). Evolution of recurrent asthma event rate over time in frailty models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 52(3): 355-363.