Accounting for highly correlated environmental exposures and a censored disease outcome with a Bayesian profile regression mixture model. First application in ionizing radiation epidemiology

Sophie Ancelet (sophie.ancelet@irsn.fr)[1], Marion Belloni[1], Olivier Laurent[1], Chantal Guihenneuc[2]

[1]PSE-SANTE/SESANE/LEPID, Institut de Radioprotection et de Sûreté Nucléaire, Fontenay-Aux-Roses, France; [2]BioSTM UR 7537, Faculté de Pharmacie de Paris, Université de Paris, Paris, France.

Cancers are among these pathologies for which the exposome concept - that encompasses the totality of human environmental exposures throughout life - is essential. Indeed, they result from the combined influence of many genetic, environmental and behavioural stressors that may occur simultaneously and interact. In epidemiological studies, it is thus important to properly account for multifactorial exposure patterns when estimating specific cancer risks at individual or population level. However, historically, the majority of environmental health research and risk assessment approaches have focused on assessing the influence on heath of one environmental risk factor of "main interest" for investigation, considered independently of the influence of other risk factors. While some research has documented examples of synergies or antagonisms following joint exposures to different environmental agents, the health effects of mixed exposures remain poorly characterized. Moreover, one major statistical difficulty, called multicollinearity, may occur when estimating health risks from exposure-based risk factors that are highly correlated. In this context, it is well recognized that applying standard multiple regression models - in which at least two highly correlated predictors are assessed simultaneously- may lead to unstable risk coefficient estimates with high variance. In this work, we focused on the specific problem of estimating a disease risk from highly correlated environmental exposure covariates and a censored survival outcome. We extended Bayesian profile regression mixture (PRM) models to this context by assuming an instantaneous excess hazard ratio disease submodel. The proposed hierarchical model incorporates an underlying truncated Dirichlet process mixture as an attribution sub-model. A specific adaptive

Metropolis-Within-Gibbs algorithm - including label switching moves - was implemented to infer the model. This allows simultaneously clustering individuals with similar risks and similar exposure characteristics and estimating the associated risk for each group. Our Bayesian PRM model was applied to the estimation of the risk of death by lung cancer in a cohort of French uranium miners who were chronically and occupationally exposed to multiple and correlated sources of ionizing radiation, that are naturally present in uranium mines (i.e., radon gas, uranium dust and gamma rays). Our work shows that Bayesian PRM models are promising tools for exposome research. It also opens new avenues for methodological research in this class of probabilistic models.