

**Titre:** Modélisation conjointe de trajectoire socio-professionnelle individuelle et de la survie globale ou spécifique

**Mots clés:** Statistique, Analyse de survie (biométrie), Méthode longitudinale, Modèles linéaires généralisés, Analyse de régression, Variables latentes, Sciences sociales – Méthodes statistiques, Mortalité, Modèles de hasards proportionnels, Cause de décès, Modèles conjoints, Données longitudinales, Modèles linéaires généralisés mixtes, Risques concurrents, Risque cause-spécifique, Modèle de Cox, Algorithme EM, Maximum de vraisemblance, Régression de Poisson, Effets aléatoires

**Résumé:** Appartenir à une catégorie socio-économique moins élevée est généralement associé à une mortalité plus élevée pour de nombreuses causes de décès. De précédentes études ont déjà montré l'importance de la prise en compte des différentes dimensions des trajectoires socio-économiques au cours de la vie. L'analyse des trajectoires professionnelles constitue une étape importante pour mieux comprendre ces phénomènes. L'enjeu pour mesurer l'association entre les parcours de vie des trajectoires socio-économiques et la mortalité est de décomposer la part respective de ces facteurs dans l'explication du niveau de survie des individus. La complexité de l'interprétation de cette association réside dans la causalité bidirectionnelle qui la sous-tend: Les différentiels de mortalité sont-ils dus à des différentiels d'état de santé initial influençant conjointement la situation professionnelle et la mortalité, ou l'évolution professionnelle influence-t-elle directement l'état de santé puis la mortalité?

Les méthodes usuelles ne tiennent pas compte de l'interdépendance des changements de situation professionnelle et de la bidirectionnalité de la causalité qui conduit à un biais important dans l'estimation du lien causale entre situation professionnelle et mortalité. Par conséquent, il est nécessaire de proposer des méthodes statistiques qui prennent en compte des mesures répétées (les professions) simultanément avec les variables de survie. Cette étude est motivée par la base de données Cosmop-DADS qui est un échantillon de la population salariée française.

Le premier objectif de cette thèse était d'examiner l'ensemble des trajectoires professionnelles avec une

classification professionnelle précise, au lieu d'utiliser un nombre limité d'états dans un parcours professionnel qui a été considéré précédemment. A cet effet, nous avons défini des variables dépendantes du temps afin de prendre en compte différentes dimensions des trajectoires professionnelles, à travers des modèles dits de "life-course", à savoir *critical period*, *accumulation model* et *social mobility model*, et nous avons mis en évidence l'association entre les trajectoires professionnelles et la mortalité par cause en utilisant ces variables dans un modèle de Cox.

Le deuxième objectif a consisté à intégrer les épisodes professionnel comme un sous-modèle longitudinal dans le cadre des modèles conjoints pour réduire le biais issu de l'inclusion des covariables dépendantes du temps endogènes dans le modèle de Cox. Nous avons proposé un modèle conjoint pour les données longitudinales nominales et des données de risques concurrents dans une approche basée sur la vraisemblance. En outre, nous avons proposé une approche de type méta-analyse pour résoudre les problèmes liés au temps des calculs dans les modèles conjoints appliqués à l'analyse des grandes bases de données. Cette approche consiste à combiner les résultats issus d'analyses effectuées sur les échantillons stratifiés indépendants. Dans la même perspective de l'utilisation du modèle conjoint sur les grandes bases de données, nous avons proposé une procédure basée sur l'avantage computationnel de la régression de Poisson. Cette approche consiste à trouver les trajectoires types à travers les méthodes de la classification, et d'appliquer le modèle conjoint sur ces trajectoires types.

**Title:** Joint modelling of individual socio-professional trajectory and overall or cause-specific survival

**Keywords:** Statistics, Survival analysis, Longitudinal methods, Generalized linear models, Regression analysis, Latent variables, Social science – Statistical methods, Mortality, Hazards proportional models, Causes of death, Joint models, Longitudinal data, Generalized linear mixed models, Competing risks, Cause-specific hazards, Cox model, EM algorithm, Maximum likelihood, Poisson regression, Random effects

**Abstract:** Being in low socioeconomic position is associated with increased mortality risk from various causes of death. Previous studies have already shown the importance of considering different dimensions of socioeconomic trajectories across the life-course. Analyses of professional trajectories constitute a crucial step in order to better understand the association between socio-economic position and mortality. The main challenge in measuring this association is then to decompose the respective share of these factors in explaining the survival level of individuals. The complexity lies in the bidirectional causality underlying the observed associations: Are mortality differentials due to differences in the initial health conditions that are jointly influencing employment status and mortality, or the professional trajectory influences directly health conditions and then mortality?

Standard methods do not consider the interdependence of changes in occupational status and the bidirectional causal effect underlying the observed association and that leads to substantial bias in estimating the causal link between professional trajectory and mortality. Therefore, it is necessary to propose statistical methods that consider simultaneously repeated measurements (careers) and survival variables. This study was motivated by the Cosmop-DADS database, which is a sample of the French salaried population.

The first aim of this dissertation was to consider

the whole professional trajectories and an accurate occupational classification, instead of using limited number of stages during life course and a simple occupational classification that has been considered previously. For this purpose, we defined time-dependent variables to capture different life course dimensions, namely *critical period*, *accumulation model* and *social mobility model*, and we highlighted the association between professional trajectories and cause-specific mortality using the defined variables in a Cox proportional hazards model.

The second aim was to incorporate the employment episodes in a longitudinal sub-model within the joint model framework to reduce the bias resulting from the inclusion of internal time-dependent covariates in the Cox model. We proposed a joint model for longitudinal nominal outcomes and competing risks data in a likelihood-based approach. In addition, we proposed an approach mimicking meta-analysis to address the calculation problems in joint models and large datasets, by extracting independent stratified samples from the large dataset, applying the joint model on each sample and then combining the results. In the same objective, that is fitting joint model on large-scale data, we propose a procedure based on the appeal of the Poisson regression model. This approach consist of finding representative trajectories by means of clustering methods and then applying the joint model on these representative trajectories.

