Combining probabilistic graphical models and multi-state survival analysis to compute risks of genetic predisposition. Application to the Lynch syndrome.

Alexandra Lefebvre[1], Patrick Benusiglio[2,3], Alex Duval[3] and Grégory Nuel[1].

[1] Laboratoire de Probabilités, Statistique et Modélisation, CNRS 8001, Sorbonne Université, Paris, France.
[2] UF d'Oncogénétique,et Angiogénétique Moléculaire, Département de Génétique, Hôpital Universitaire Pitié-Salpêtrière, APHP Sorbonne Université, Paris, France.
[3] Centre de Recherche hôpital Saint-Antoine (CRSA, INSERM UMR 938), Equipe "Instabilité des Microsatellites et Cancer", Paris, France.

Estimating genetic predisposition conditional on the family history of cancer of a patient is of high importance in genetic counseling to adapt surveillance and/or treatments. Mathematical models helping clinicians compute these risks are divided into two categories : linear regression (such as IBIS) and Mendelian models (such as BODICEA for breast and ovarian cancer). The main advantages of Mendelian models are the exhaustive modelization of the family history of cancer (all parental links, status towards the disease, ages at diagnosis and/or last news) and the computation of several other quantities of interest such as individual risks for all members of the family (versus proband only), the posterior probability of the number of syndrome carriers, a time to event of cancer in the whole family, etc. Their main limitation is their computational cost but, based on algorithms taken from the probabilistic graphical models theory, one can drop the computational complexity from exponential to linear in the number of individuals. In the context of the Lynch syndrome (or Hereditary Non-Polyposis Colorectal Cancer - HNPCC), MMpro is the only existing Mendelian model. It has not been updated since 2006 and suffers some technical limitations and issues regarding survival and linkage analysis. In this context, we present a new Mendelian model for risks computations in the framework of the Lynch syndrome based on probabilistic graphical models and multi-state survival with a mixture of familial data, survival data and biological data. We explain the interest of the belief propagation algorithm in this context and present results and predictive performances for chosen examples and a set of families from Saint-Antoine Hospital.