Soutenance

Mickaël Canouil

29 Septembre 2017

1 Contexte

1.1 Le Diabète dans le Monde

1.2 Les Formes de Diabète

Il existe principalement, 2 formes de diabète, le diabète de type 1, le diabète de type 2, le diabète gestationnel et les diabètes monogéniques :

- Le diabète de type 1 est le diabète dit insulinodépendant et nécessite des injections régulières d'insuline.
  Ce diabète se développe généralement chez un individu jeune qui perd rapidement sa capacité à réguler sa glycémie, suite à une réaction auto-immune contre les cellules β du pancréas (cellules sécrétrices de l'insuline).

- Le diabète de type 2 est parfois appelé diabète de l'adulte ou diabète non-insulinodépendant, par opposition au diabète de type 1.
  Il se caractérise principalement par un défaut du métabolisme de l'insuline d'un ou plusieurs organes.
  Le diabète de type 2 représente plus de 90% des diabètes dans le monde.
  Parce que les symptômes du diabète de type 2 sont moins marqués que ceux du diabète de type 1, le diabète de type 2 est souvent diagnostiqué tardivement, et notamment suite aux complications résultantes de celui-ci.

La forme de diabète étudié ici est le diabète de type 2.

1.3 Le Diabète de Type 2

1.4 La Génétique du Diabète de Type 2

1.5 Un Grand Nombre de Découverte, mais ...

2 Glycémie et Diabète de Type 2 : un Effet Génétique Conjoint ?

2.1 Contexte & Objectifs

2.2 Modèle Joint

2.2.1 Présentation

2.2.2 Motivation

The initial interest stemmed from research investigating predictiveness of repeatedly measured biomarkers for clinicalmeaningfull events.
There is also interest in these studies to determine surrogacy of the biomarker; that is, treatment effect manifests through the biomarker so that the event is independent of treatment given

biomarker trajectories

In an ideal situation, this research questions can be routinely addressed by an event time model in which the biomarker is included as a time-dependent covariate.
For example, it is straightforward to incroporate the time-dependent biomarker into a Cox proportional hazards model and study its effect on the event.
Treatment can be included as a time-fixed covariate, and its effect with and without adjustment for the biomarker is then examined to assess surrogacy.
However, application of such models in practice is often problematix.
First of all, the biomarker is not observed in a continuum interval of time; instead, measurements are available only on a set of specific time points.
Thus the biomarker may be missing at some observed event times.
Naive imputation methods, such as last observation carried forward, are likely to introduce bias into model estimation.
Secondly, the biomarker is often subject to measurement error and high biological variation, so that the observed values may not accurately reflect the true underlying trajectories.
Thus, using observed values to predict event could lead to further bias.
The third issue is regarding possible noningnorable missing data in longitudinal measurements ("endogeneity").
Statistical analysis to characterise intra-subject changes and beteen-subject heterogeneity in the biomarker should take into account the missing data mechanism.

Joint modelling of time-dependent biomarkers and event times has been developped to address the above issues.

2.3   Formulation

2.3.1   Composante Longitudinale

For subject i, let $X_i(t_{ij})$ deote the underlying, smooth trajectory of biomarker, T_i the event time, and $W_i(t_{ij})$ a set of possibly time-varying covariates.
Because the biomarker is intermittenly measured at time points t_{ij} and there are intra-subject measurements errors, $X_i(t_{ij})$ is not directly observable; instead, measurements $Y_i(t_{ij}$ are available.
Where $\epsilon_i(t_{ij})$ are measurements errors.
The longitudinal model focuses on charactering change on $X_i(t)$ over time.
If the change can be described by a polynomial function of time t, then $X_i(t)$ is specified as :

2.3.2   Composante de Survie

A Cox proportional hazards model is used to specify the interrelatinship between $X_i(t)$, T_i, Z_i, W_i.

This specification implies that given covariates and past history $\bar{Y}_i(t)$, the biomarkeris associated with the event rish through its current value $Y_i(t)$.

2.4   Inférence Statistique

2.4.1   EM

2.5   Test d'Hypothèse

$\hat{\Psi}_0$ and $\hat{\Psi}$ denotes the maximum likelihood estimates under the null and alternative hypthesis, respectively.

$S(\cdot)$ and $\mathcal{I}(\cdot)$ denotes the score function and the observed information matrix of the model under the alternative hypothesis.

under the nullhypothesis, the asymptotic distribution of each of these tests is a chi-squared distribution with p degrees of freedom, with p denoting the number of parameters being tested. For a single Ψ_j the Wald test is equivalent to $\hat{\Psi}_j - \Psi_{0j}/\hat{s.e.}(\hat{\Psi}_j)$, which under the null follows an asymptotic standard normal distribution.

2.6 Étude de Simulation

2.6.1 Objectifs

2.6.2 Génération des Données

Les jeux de données de simulation, qui ont été réalisés avec R, avec la fonction de risque de base fixée ($h_0(t) = \lambda$) de façon à obtenir une incidence équivalente à celle de la cohorte D.E.S.I.R., durant la période de suivi de neuf ans. Les temps d'événements ont été générés selon une distribution exponentielle dans le cadre du modèle de Cox à risque proportionnel.

2.6.3 Paramètres de Simulation

Due to the complexity of the estimating algorithm within JM, convergence could not be obtained ($4.53 \pm 5.81\%$ of convergence issues on average per scenario) for the whole set of 500 simulations (i.e. algorithm "piecewise-PH-aGH" for a time-dependent relative risk model with a piecewise constant baseline risk function, using the adaptive Gauss-Hermite quadrature rule to approximate integrals within the Expectation-Maximisation (EM) step)

2.6.4 Résultats

RMSE for parameter γ showed similar performance for JM and TS, which was expected given the formulation of the joint model within the "Shared Parameter Models" framework, in which Y_i (mean of Y_{ij} modelled within LME according to Equation X links the longitudinal data to the time of event.

RMSE for parameter β and for parameter α were smaller within the joint modelling framework (either JM or TS) than in the more classical CoxPH model with time varying fasting plasma glucose. While RMSE for β was the same in the CoxPH model across all scenarios, under JM or TS it decreased with increases in the sample size, incidence rate or allele frequency. Differences in RMSE for parameter α were less than for parameter β, and TS and CoxPH with time-dependent covariate, performed equally probably because partial likelihood inferences were used in both approaches. JM estimations were less biased in almost all scenarios when the sample size was greater than 2,500.

2.7 Application sur Données Réelles : D.E.S.I.R.

MTNR1B 28,4% ; TCF7L2 24,4%

3 Au delà des Études Pangénomiques

4 Apports & Perspectives