

# Application of Joint Models in Genetic Association Studies

Ghislain Rocheleau<sup>1,2,3</sup>, Mickaël Canouil<sup>1,2,3</sup>, Loïc Yengo<sup>2,3</sup>, Philippe Froguel<sup>1,2,3,4</sup>

<sup>1</sup>Lille 2 University, Lille, France  
<sup>2</sup>UMR 8199 - Pasteur Institute, Lille, France  
<sup>3</sup>FR 3508 - European Genomic Institute for Diabetes, Lille, France  
<sup>4</sup>Department of Genomics of Common Disease, Imperial College London, London, United Kingdom

**Contact informations:**  
CNRS UMR8199 - Institut de Biologie de Lille  
1 Rue du Professeur Calmette  
BP 245  
F-59019 LILLE CEDEX  
  
+33(0)3-20-87-11-07  
ghislain.rocheleau@cnrs.fr

## Introduction

To optimize use of existing phenotypic data, we propose a joint model [Tsiatis and Davidian, 2004] approach aimed at identifying genetic markers simultaneously associated to temporal trajectories of a trait and an event outcome. Standard formulation of the joint model involves two components: a longitudinal component and a time-to-event component. We illustrate the application of the joint model approach in genetic epidemiology by exploiting the strong link between temporal variation of blood glucose levels (FG) and onset of type 2 diabetes (T2D). Using genotypes assayed with the Metabochip DNA arrays (Illumina) from 4,500 subjects recruited in the French cohort D.E.S.I.R. (Données Épidémiologiques sur le Syndrome d'Insulino-Résistance), we reexamine previous GWAS findings for some confirmed glycaemia and T2D loci.

## Methods

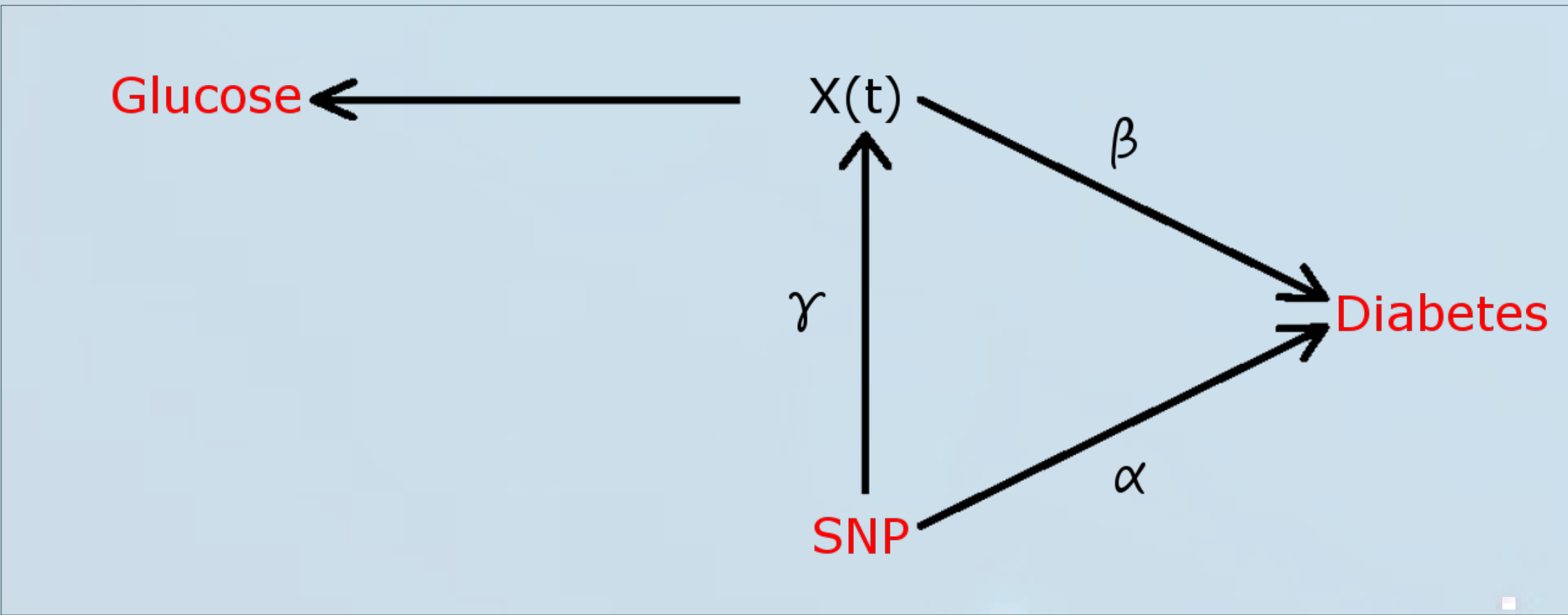


Figure 1: Causal diagram for joint modeling (adapted from Ibrahim et al. [2010]).

The longitudinal component typically consists of a linear mixed model:

$$Y_{ij} = X_{ij} + \epsilon_{ij}$$
$$X_{ij} = \theta_{0i} + \theta_{1i} \times t_{ij} + \gamma \times SNP_i$$
$$\theta \sim \mathcal{N}_2(\mu, \Sigma)$$
$$\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

$Y_{ij}$ : observed value;

$X_{ij}$ : true (unobserved) value of the longitudinal variable;

$\gamma$ : SNP effect on the trajectory function (Figure 1).

Using the Cox model for the time-to-event (survival) component, we define:

$$\lambda_i(t) = \lambda_0(t) \exp(\beta X_i(t) + \alpha SNP_i)$$

$\lambda_i(t)$ : hazard function at time  $t$

$\lambda_0(t)$ : (unspecified) baseline hazard function.

$\alpha$ : SNP effect on time-to-event (Figure 1).

$\beta$ : association between the trajectory function and time-to-event (Figure 1).

## Results

As expected, we confirm some findings from GWAS, especially the strong association between FG and some SNP ( $\gamma \neq 0$ ) in G6PC2, GCKR and MTNR1B genes (Table 1).

## References

Chen, L. M., Ibrahim, J. G., and Chu, H. (2011). Sample size and power determination in joint modeling of longitudinal and survival data. *Statistics in Medicine*, 30(18):2295–2309.

Ibrahim, J. G., Chu, H., and Chen, L. M. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *J. Clin. Oncol.*, 28(16):2796–2801.

Rizopoulos, D. (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1–33.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14:809–834.

	$\gamma$	$\alpha$	$\beta$
rs7903146 (TCF7L2)	0.02465	0.2204	3.477
rs3802177 (SLC30A8)	0.038	0.01066	3.542
rs10278336 (GCK)	0.0383	0.09214	3.527
rs560887 (G6PC2)	0.09504	-0.3237	3.568
rs780094 (GCKR)	0.06271	-0.09694	3.568
rs10830963 (MTNR1B)	0.0959	-0.3868	3.611
rs11717195 (ADCY5)	0.02581	-0.1202	3.554

Table 1: Parameter estimates from R package JM [Rizopoulos, 2010] applied to a selection of SNPs. In blue: p-value < 0.05, in red: p-value <  $5 \times 10^{-8}$ .

The association between FG and T2D risk is highly significant ( $p < 10^{-50}$ ) as a result of the T2D diagnostic ( $FG > 7mM/L$ ).

## Power estimation

$f$	$\alpha$	$\gamma$	Power
0.10	0.20	0.05	0.41
		0.06	0.49
		0.07	0.58
0.30	0.20	0.05	0.65
		0.06	0.73
		0.07	0.80
0.30	0.30	0.05	0.85
		0.06	0.89
		0.07	0.93

Table 2: Power estimation from formula derived in Chen et al. [2011].

To detect global SNP effect  $\beta\gamma + \alpha$ , we used equation:

$$z_{\beta} = \pm \sqrt{Df(1-f)(\beta\gamma + \alpha)^2} + z_{1-\tilde{\alpha}/2}$$

$D$ : number of incident T2D cases;  
 $f$ : allele frequency.

Retrospective power estimations provided by the formula in Chen et al. [2011] show consistent results with our application for an estimated effect  $\beta = 3.5$  and  $\tilde{\alpha} = 0.05$ .

## Future research

To the best of our knowledge, joint models have never been used in genetic association studies. In particular, power estimation and effect of missing data remain to be studied in further details.

