

Génomique Intégrative et Modélisation des Maladies Métaboliques  
UMR 8199 (CNRS / Université de Lille 2 / Institut Pasteur de Lille)

CNRS UMR8199 - Institut de Biologie de Lille  
1 Rue du Professeur Calmette  
BP 245  
F-59019 LILLE CEDEX

# Développement et Application de Méthodologies Statistiques pour Etudes Longitudinales d'Association Génétique

*Comité de suivi de thèse: première année*

Mickaël Canouil

Doctorant en Biostatistique  
([mickael.canouil@cnrs.fr](mailto:mickael.canouil@cnrs.fr))

**Directeur de thèse:**

Pr. Philippe Froguel

**Co-directeur de thèse:**

Dr. Ghislain Rocheleau (Lille 2)

**Membres du CST:**

Dr. Hélène Jacqmin-Gadda (INSERM U 897, Bordeaux)

Pr. Cristian Preda (Ecole Polytechnique, Lille)

28 septembre 2015



**Table des matières**

<b>1</b>	<b>Introduction et contexte</b>	<b>2</b>
<b>2</b>	<b>Objectif de la thèse</b>	<b>2</b>
<b>3</b>	<b>Matériels</b>	<b>3</b>
<b>4</b>	<b>Méthodes</b>	<b>4</b>
<b>5</b>	<b>Résultats préliminaires</b>	<b>5</b>
5.1	Cohorte D.E.S.I.R . . . . .	5
5.2	Simulations . . . . .	8
5.2.1	Scénario 1 : variation de la fréquence allélique . . . . .	9
5.2.2	Scénario 2 : variation du nombre de mesures longitudinales . . . . .	10
5.2.3	Conclusion . . . . .	11
<b>6</b>	<b>Perspectives</b>	<b>13</b>
6.1	En cours / À court terme . . . . .	13
6.2	À long terme . . . . .	13
<b>7</b>	<b>Formations</b>	<b>14</b>
7.1	Passées . . . . .	14
7.2	Futures . . . . .	14
	<b>Références</b>	<b>15</b>

## 1 Introduction et contexte

La prévalence mondiale du diabète de type 2 (T2D) est en constante augmentation et génère par le fait même des efforts soutenus dans la découverte de nouveaux marqueurs génétiques impliqués dans le risque de la maladie [World Health Organization, 2015]. Ces dernières années, les études d'association pangénomiques (GWAS) ont identifié plusieurs variants génétiques associés à des traits métaboliques et au T2D, dont pas moins de 65 loci associés à la susceptibilité du T2D [Morris et al., 2012]. Les GWAS ont notamment identifié 36 variants associés à la glycémie à jeun chez des individus normoglycémiques [Dupuis et al., 2010; Scott et al., 2012]. Bien que le rôle fonctionnel de la plupart de ces loci reste à déterminer, ils ont tout de même permis de mettre en lumière des liens intéressants entre le glucose sanguin et le T2D, et souligné quelques différences entre l'homéostasie du glucose dans la population générale et les niveaux physiopathologiques observés chez les diabétiques. La majorité de ces loci ont été trouvés par méta-analyse de GWAS, malgré la très grande hétérogénéité des études et des cohortes incluses dans la méta-analyse.

La grande majorité des GWAS publiées ont utilisé un design transversal plutôt qu'un design longitudinal. Cependant, une étude prospective longitudinale offre la possibilité de mesurer une même variable plusieurs fois au cours de l'étude, permettant également une description de ses fluctuations temporelles. Il est attendu que la puissance de détecter des variants génétiques associés à ces variables soit augmentée grâce aux multiples mesures prises à plusieurs moments, puisqu'elles donnent globalement plus d'information qu'une seule mesure prise à un moment donné. De plus, un design longitudinal permet d'étudier les variants associés aux changements (trajectoires) de ces mesures à travers le temps, tout en augmentant la puissance de détecter des différences entre les groupes (définis par leur génotype, dans un contexte génétique). Tout porte à croire qu'une bonne modélisation de ces trajectoires optimiserait les tests d'association et permettrait une meilleure exploitation des phénotypes disponibles.

## 2 Objectif de la thèse

Cette thèse propose de développer et implémenter des approches pangénomiques basées sur des modèles de régression, tels les modèles linéaires mixtes (LMM), les équations estimantes généralisées (GEE) [Sitlani et al., 2015], et les modèles joints (JM) liant de façon statistique la variation temporelle d'une variable (p.ex., glucose sanguin) avec la survenue d'une maladie (p.ex., T2D). Ces approches pourraient mener à l'identification de nouveaux loci. De plus, ces méthodologies permettront de confronter la vision actuelle selon laquelle les gènes

impliqués dans les niveaux de glucose chez les individus normoglycémiques diffèrent de ceux observés dans les niveaux pathophysiologiques des sujets atteints de T2D [Yaghoodeh and Frayling, 2013]. Aussi, ces approches statistiques appliquées à un contexte génétique pourraient permettre d'expliquer pourquoi certains individus en légère hyperglycémie ne développent pas le T2D. Les marqueurs ainsi identifiés renforceront du même coup l'utilité clinique de la stratification génétique du risque dans la population générale [Pal and McCarthy, 2013].

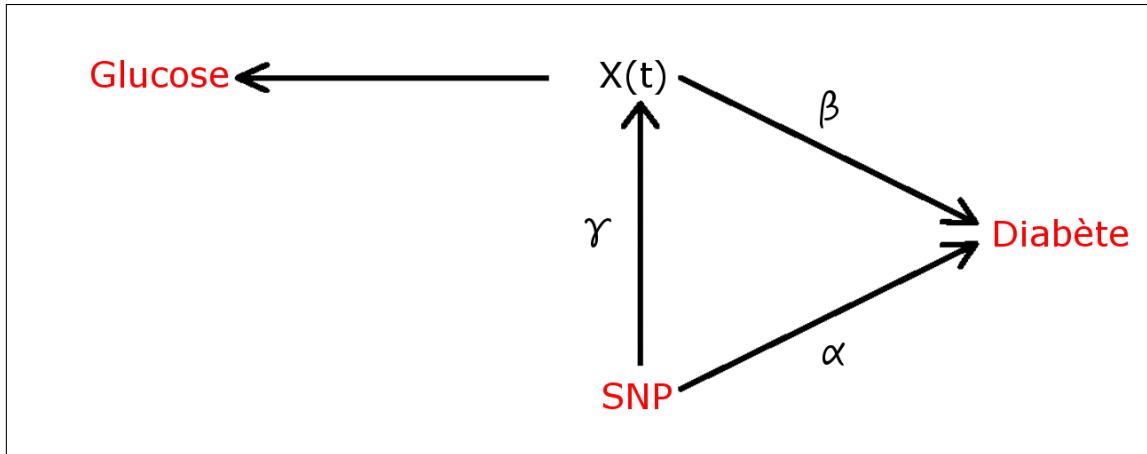
Néanmoins, les approches LMM, GEE et JM nécessitent de grandes quantités de ressources lorsqu'appliquées sur des millions de loci et des milliers d'individus. Aussi, ces approches sont le plus souvent utilisées sur un sous-ensemble de loci dans des régions d'intérêts de la pathologie étudiée [Beyene and Hamid, 2014; Wu et al., 2014]. Des approches dites "rapides" ont été développées pour pallier au problème du temps de calcul, présents notamment sur les implémentations actuelles des LMM (`nlme` et `lme4` pour R), et se basent sur des méthodes d'approximation des LMM [Sikorska et al., 2013, 2015; Sikorska, 2014]. A ce jour, deux implantations des JM existent sur la plateforme CRAN pour le logiciel R [R Core Team, 2014]. La première est développée par Dimitris Rizopoulos (JM) [Rizopoulos, 2010] et se focalise sur l'estimation des paramètres du sous modèle de survie, quand la seconde, développée par Pete Philipson (joineR) [Philipson et al., 2012], vise à estimer principalement les paramètres du sous modèle LMM.

### 3 Matériels

La cohorte prospective D.E.S.I.R (Données Épidémiologiques sur le Syndrome d'Insulino-Résistance), comprend 5 214 sujets suivis pendant neuf années au cours desquelles plusieurs mesures phénotypiques ont été prises à quatre occasions uniformément réparties durant la période de suivi (0, 3, 6 et 9 ans). En exploitant l'approche par GWAS dans la cohorte D.E.S.I.R, une percée majeure a été réalisée dans la génétique des traits métaboliques menant, d'une part, à la découverte de nouveaux loci associés au T2D, et d'autre part, à l'identification de loci impliqués dans le glucose sanguin de sujets normoglycémiques [Dupuis et al., 2010; Voight et al., 2010]. La base de données phénotypique et génotypique existante de D.E.S.I.R offre l'opportunité unique de modéliser la trajectoire temporelle de plusieurs phénotypes dont le glucose sanguin, et peut-être même permettra de découvrir de nouveaux loci simultanément associés à l'incidence du T2D et à un taux élevé de glucose sanguin.

L'étude des données a été réalisée avec le logiciel R sur un serveur de calcul de 80 coeurs et 500 Go de mémoire vive, accessible dans le laboratoire de l'unité UMR8199 dirigée par le Pr. Philippe Froguel.

## 4 Méthodes



**Figure 1:** Diagramme général d'un modèle joint pour le T2D (adapté de [Ibrahim et al. \[2010\]](#))

$X(t)$  : trajectoire de glucose sanguin inférée des données longitudinales observées ;  $\alpha$  : effet du SNP sur le diabète ;  $\gamma$  : effet du SNP sur la trajectoire de glucose sanguin ;  $\beta$  : effet de la trajectoire de glucose sanguin sur le diabète.

Le risque d'incidence de T2D a d'abord été évalué à l'aide d'un modèle de survie de Cox, avec une variable dépendante du temps [[Fisher and Lin, 1999](#)], le glucose sanguin, ainsi qu'une variable fixe non dépendante du temps, le SNP. Pour un individu  $i$  au temps  $t$ , la fonction de risque  $\lambda_i(t)$  est exprimée par

$$\lambda_i(t) = \lambda_0(t) \times \exp(\beta \times X_i(t) + \alpha \times SNP_i) \quad (1)$$

où  $\lambda_0(t)$  est la fonction de risque de base. Le paramètre  $\alpha$  représente l'effet du SNP sur le temps de survenue de l'événement, quand  $\beta$  est l'association entre la trajectoire de glucose sanguin et le temps du T2D.

Dans un second temps, les approches de type modèle joint ont été appliquées (paquet R : **JM** et **joineR**). Les modèles joints intègrent deux composantes, une composante de survie ([Equation 1](#)) et une composante longitudinale, généralement modélisée par un modèle linéaire mixte

$$Y_{ij} = X_{ij} + \epsilon_{ij} \quad (2)$$

avec  $Y_{ij}$  la valeur observée,  $X_{ij}$  la vraie valeur (non observée) de la variable longitudinale, et  $\epsilon_{ij}$  l'erreur de mesure pour l'individu  $i$  au temps  $j$ , supposée normalement distribuée. La variable  $X_{ij}$  représente la trajectoire modélisée comme une fonction du temps (linéaire, quadratique, etc), et peut inclure des covariables. Dans la suite du texte, nous supposons que

$$X_{ij} = \theta_{0i} + \theta_{1i} \times t_{ij} + \gamma \times SNP_i \quad (3)$$

Nous supposons que les paramètres  $\theta_{0i}$  et  $\theta_{1i}$  suivent une loi normale bivariée. Le paramètre  $\gamma$  représente l'effet du SNP sur la trajectoire de glucose sanguin.

## 5 Résultats préliminaires

### 5.1 Cohorte D.E.S.I.R

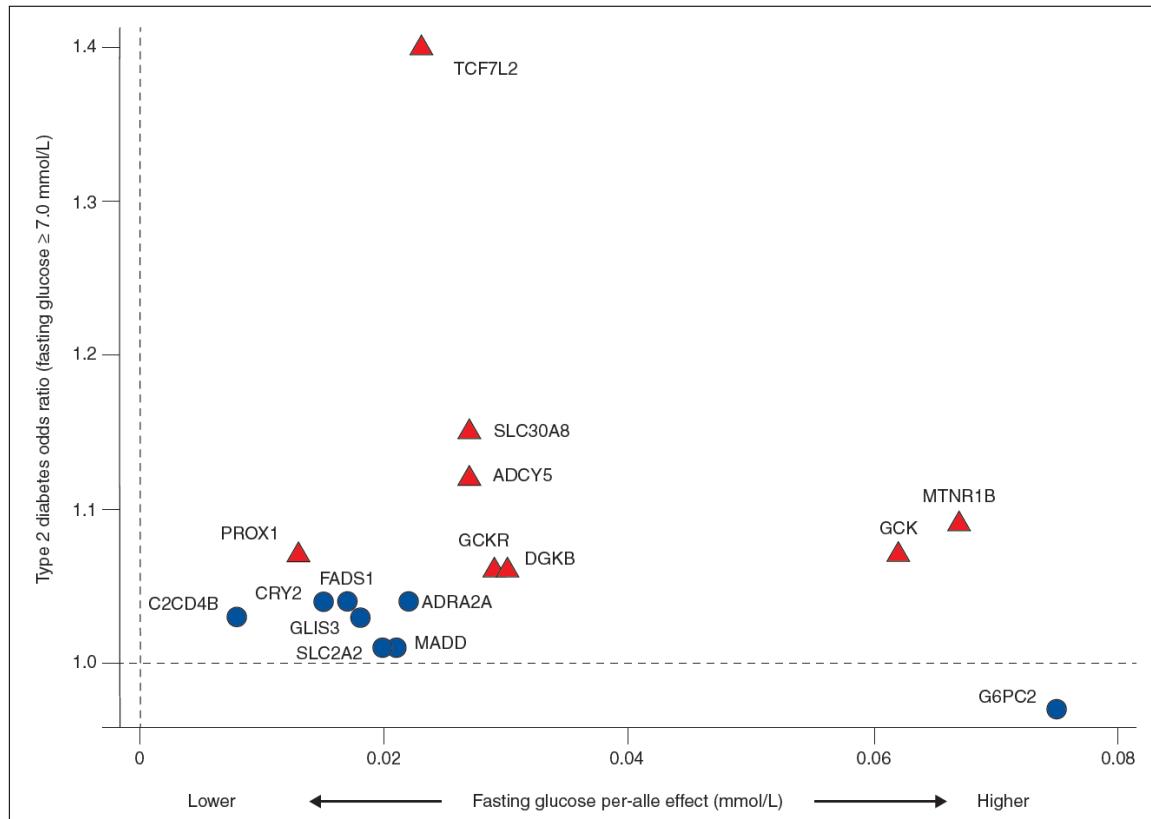
Dans le but de tester les approches LMM et JM, des SNPs, découverts par méta-analyses de GWAS, associés au glucose sanguin et au T2D (consortium [DIAGRAM](#) et [MAGIC](#)), ont été sélectionnés. Les individus présents dans la cohorte D.E.S.I.R ont pour la plupart (environ 4 500) été génotypés à l'aide d'une Metabochip DNA arrays (Illumina) [[Voight et al., 2012](#)]. Les approches précédemment décrites à la [Section 4](#) ont été appliquées. Il s'agissait ici de confirmer ces marqueurs génétiques dans la cohorte D.E.S.I.R, en incluant, dans un même modèle, l'information d'incidence de T2D et le glucose sanguin.

	$\gamma$		$\alpha$		$\beta$	
	Estimation	Valeur-p	Estimation	Valeur-p	Estimation	Valeur-p
rs7903146_A (TCF7L2)	0.02465	1.2e-02	0.2204	0.0629	3.477	1.5e-53
rs3802177_G (SLC30A8)	0.038	1.2e-04	0.01066	0.9329	3.542	3.0e-55
rs7756992_G (CDKAL1)	0.01842	6.6e-02	0.08661	0.4651	3.511	5.2e-55
rs10278336_G (GCK)	0.0383	3.5e-05	0.09214	0.4197	3.527	1.7e-54
rs1552224_A (ARAP1)	0.03644	4.5e-03	0.1077	0.5177	3.529	1.5e-55
rs560887_C (G6PC2)	0.09504	1.2e-22	-0.3237	0.0076	3.568	9.4e-56
rs780094_C (GCKR)	0.06271	4.4e-12	-0.09694	0.3938	3.568	2.8e-55
rs10830963_G (MTNR1B)	0.0959	4.1e-22	-0.3868	0.0021	3.611	8.5e-56
rs2075423_G (PROX1)	0.02666	4.0e-03	-0.03139	0.7784	3.538	1.7e-55
rs10811661_T (CDKN2A)	0.01671	1.4e-01	0.0003004	0.9983	3.636	2.4e-55
rs11717195_T (ADCY5)	0.02581	1.8e-02	-0.1202	0.3788	3.554	1.5e-55
rs4402960_A (IGF2BP2)	0.01174	2.3e-01	0.06249	0.5958	3.558	1.2e-55
rs6878122_G (ZBED3)	0.01349	1.6e-01	-0.000382	0.9974	3.444	1.0e-54

**Table 1:** Résultat d'analyse pour une sélection de SNPs connus pour être associés au T2D ou au glucose sanguin. Résultats obtenus avec le paquet **JM** et ajustés à l'âge, au sexe et à l'IMC. En bleu, valeur-p < 0.05 et en rouge, valeur-p <  $5 \times 10^{-8}$ .

Comme prévu, ces résultats confirment les résultats obtenus par GWAS ([Figure 2](#)), notamment la forte association entre le glucose sanguin et le SNP ( $\gamma \neq 0$ ). Les valeurs-p mises en

évidence dans la [Table 1](#) ne sont pas aussi petites que celles rapportées dans la littérature, en raison de l'effectif réduit de la cohorte D.E.S.I.R, en comparaison des milliers d'individus présents dans les méta-analyses de GWAS. L'association entre la trajectoire du glucose et le risque d'incidence de T2D est très significative ( $p < 10^{-50}$ ), de par la définition même de l'incidence de T2D (définie selon un seuil de glucose dans le sang supérieur à 7mM/L). Rappelons que le seuil de significativité pangénomique est communément fixé à  $5 \times 10^{-8}$ .



**Figure 2:** Résultat d'association des allèles à risques pour le T2D et le glucose sanguin

[[Yaghootkar and Frayling, 2013](#)]. Cercle bleu : l'association au glucose sanguin pour valeur-p  $< 5 \times 10^{-8}$ ; Triangle rouge : l'association au T2D et au glucose sanguin valeur-p  $< 5 \times 10^{-8}$ .

Les données de la cohorte D.E.S.I.R ont également été analysées avec des approches classiques de LMM (paquet [nlme](#)) et par une approche dite "Two-Step" (TS) [[Self and Pawitan, 1992](#)]. Cette dernière approche consiste en la succession de deux étapes :

**Etape 1** Ajustement d'un modèle linéaire mixte pour estimer la "vraie" trajectoire.

**Etape 2** Ajustement d'un modèle de survie incorporant la trajectoire estimée de l'étape 1 en tant que covariable dépendante du temps.

L'approche TS présente des résultats proches de ceux obtenus par l'approche JM. Néanmoins, l'estimation du paramètre  $\beta$  est systématiquement inférieure à celle obtenue par JM. L'estima-

	$\alpha$		$\beta$	
	Estimation	Valeur-p	Estimation	Valeur-p
rs7903146_A (TCF7L2)	0.317	1.0e-04	3.356	< 2 × 10 <sup>-16</sup>
rs3802177_G (SLC30A8)	0.1291	1.5e-01	3.375	< 2 × 10 <sup>-16</sup>
rs7756992_G (CDKAL1)	0.1829	2.4e-02	3.372	< 2 × 10 <sup>-16</sup>
rs10278336_G (GCK)	-0.1373	8.2e-02	3.376	< 2 × 10 <sup>-16</sup>
rs1552224_A (ARAP1)	0.06174	5.8e-01	3.392	< 2 × 10 <sup>-16</sup>
rs560887_C (G6PC2)	-0.2725	9.6e-04	3.432	< 2 × 10 <sup>-16</sup>
rs780094_C (GCKR)	-0.04382	5.8e-01	3.404	< 2 × 10 <sup>-16</sup>
rs10830963_G (MTNR1B)	-0.2683	1.4e-03	3.456	< 2 × 10 <sup>-16</sup>
rs2075423_G (PROX1)	-0.1184	1.2e-01	3.407	< 2 × 10 <sup>-16</sup>
rs10811661_T (CDKN2A)	0.03332	7.3e-01	3.364	< 2 × 10 <sup>-16</sup>
rs11717195_T (ADCY5)	-0.08409	3.7e-01	3.405	< 2 × 10 <sup>-16</sup>
rs4402960_A (IGF2BP2)	0.09775	2.2e-01	3.391	< 2 × 10 <sup>-16</sup>
rs6878122_G (ZBED3)	0.07204	3.7e-01	3.386	< 2 × 10 <sup>-16</sup>

**Table 2:** Résultat d'analyse pour une sélection de SNPs connus pour être associés au T2D et au glucose sanguin. Résultats obtenus avec l'approche "Two-Step", ajustés à l'âge, au sexe et à l'IMC (paquets **nlme** et **survival**).

En bleu, valeur-p < 0.05 et en rouge, valeur-p < 5 × 10<sup>-8</sup>.

tion du paramètre  $\alpha$  (effet du SNP sur l'incidence de T2D) diffère substantiellement de celle obtenue par le modèle joint.

L'approche TS permet de réduire le biais introduit par l'ajout d'une covariable dépendante du temps dans un modèle de survie. Cependant, cette approche n'est pas sans biais, puisque l'information d'incidence de T2D n'est pas intégrée dans l'estimation de la "vraie" trajectoire [McCrink et al., 2013].

## 5.2 Simulations

Paramètres	Valeurs
Effectif ( $N$ )	5000
Temps de mesures (en années)	0, 3, 6, 9
Incidence à neuf ans ( $I$ )	5%
LMM : Trajectoire $\begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$	$\mathcal{N}_2 \left( \begin{bmatrix} 4.50 \\ 0.013 \end{bmatrix}, \begin{bmatrix} 0.16 & 0 \\ 0 & 1 \times 10^{-3} \end{bmatrix} \right)$
LMM : Effet du SNP ( $\gamma$ )	0.025
Cox : Effet du SNP ( $\alpha$ )	0.2
JM : Effet de la trajectoire ( $\beta$ )	3.50

**Table 3:** Paramètres initiaux pour la simulation des données. Caractéristiques basées sur le SNP de TCF7L2 (SNP le plus fortement associé au T2D).

Dans le but d'identifier les avantages et limites des approches de type modèle joint (**JM** et **joineR**), un jeu de données a été simulé sur la base de la cohorte D.E.S.I.R. Les jeux de données de simulation, qui ont été réalisés avec R, suivent les [Equations 1 à 3](#) et les paramètres donnés dans la [Table 3](#). La fonction de risque de base  $\lambda_0$  a été fixée de façon à obtenir une incidence équivalente à celle de la cohorte D.E.S.I.R (environ 5%), durant la période de suivi de neuf ans.

Plusieurs scénarios de simulation ont été réalisés pour tester la robustesse des estimations des paramètres en présence de données manquantes, en utilisant la classification usuelle :

**MCAR (missing completely at random) :** les données sont manquantes indépendamment des données observées et non observées ;

**MAR (missing at random) :** conditionnellement aux données observées, les données manquantes sont indépendantes des données non observées ;

**MNAR (missing not at random) :** les données manquantes sont dépendantes de variables non observées.

D'autres paramètres ont également été étudiés, tels que l'effectif de la population, la fréquence du marqueur génétique et dans le cas plus général des LMM, le nombre de mesures longitudinales. Ces scénarios ont été étudiés avec les paquets **JM** et **joineR**.

Les scénarios adoptés et étudiés avec les paquets **JM** et **joineR** sont les suivants :

1. Données complètes et variation de la fréquence allélique ;
2. Données complètes et variation du nombre de mesures longitudinales ;

3. Données complètes et variation de l'effectif ;
4. Données manquantes distribuées de façon uniforme (MCAR) ;
5. Perte au suivi (MCAR) ;
6. Données manquantes conditionnellement au génotype (MAR) ;
7. Données manquantes conditionnellement à une variable non observée (MNAR).  
(Scénario non encore implémenté).

En raison de la complexité du calcul, un faible nombre de simulations (1000) a été réalisé pour produire les figures suivantes.

### 5.2.1 Scénario 1 : variation de la fréquence allélique

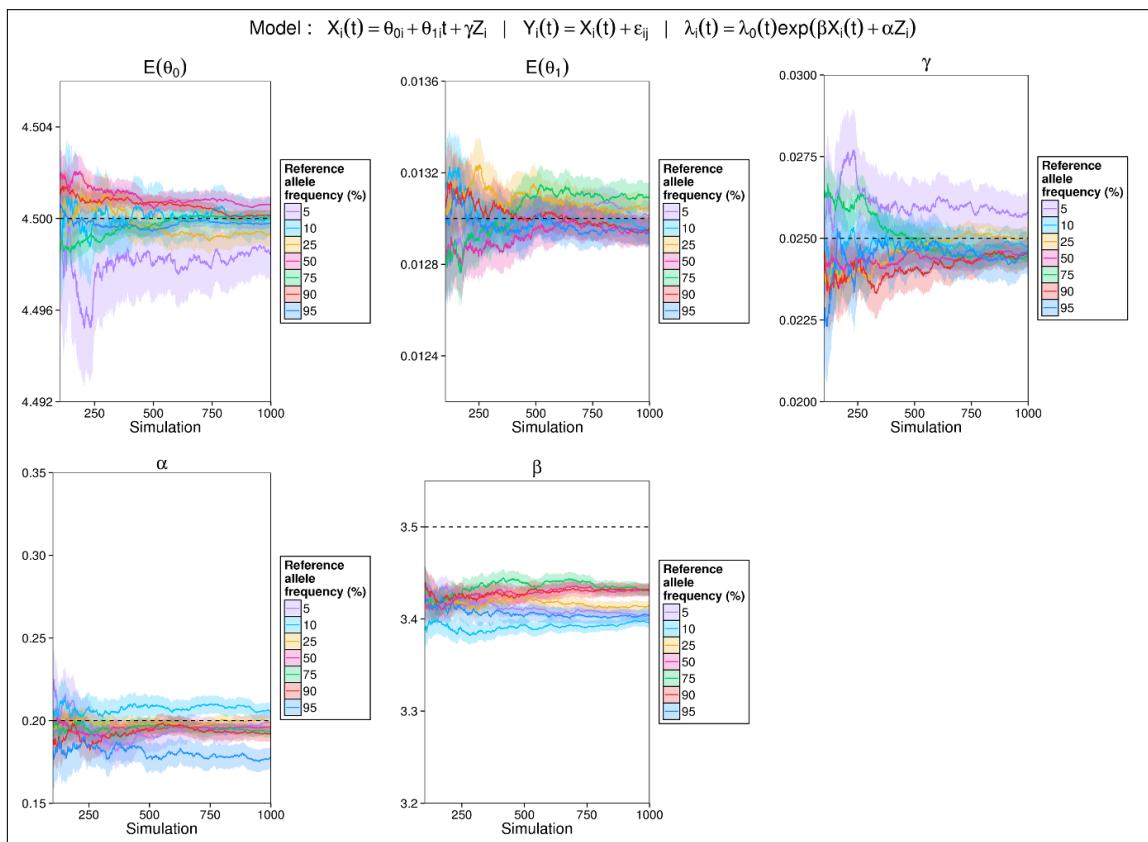


Figure 3: Estimation par JM (Scénario 1).

Ces premières simulations (Figure 3) révèlent que JM ne semble pas présenter de biais important pour l'estimation des paramètres liés au LMM. Il en va de même pour le paramètre  $\alpha$  du modèle de survie, ce qui était attendu : en effet, JM se focalise sur l'estimation des paramètres de la composante de survie du modèle joint. Néanmoins, l'estimation du  $\beta$ , issue du modèle de survie de JM, est sous-estimée de façon systématique dans ce scénario. La fréquence

allélique semble influencer, de façon plus ou moins importante, l'estimation des paramètres lorsque celle-ci est faible ( $f < 5\%$ ).

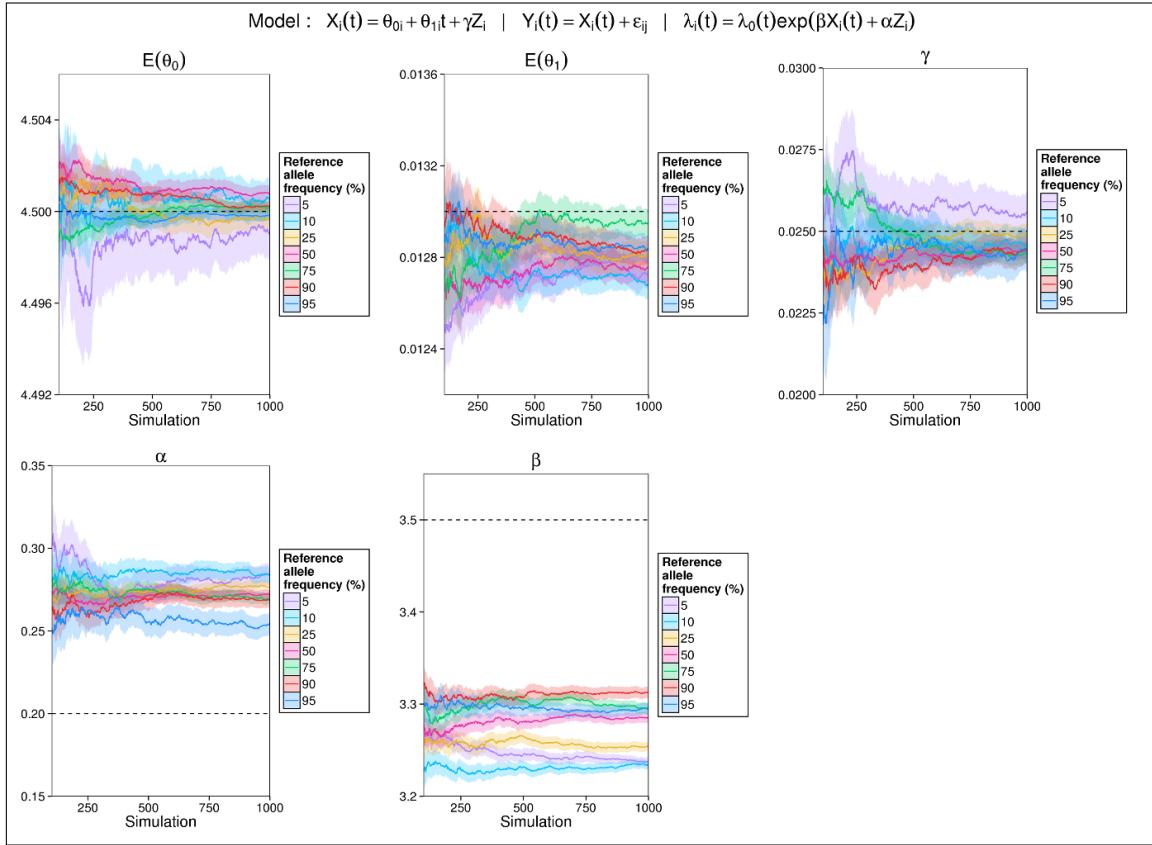


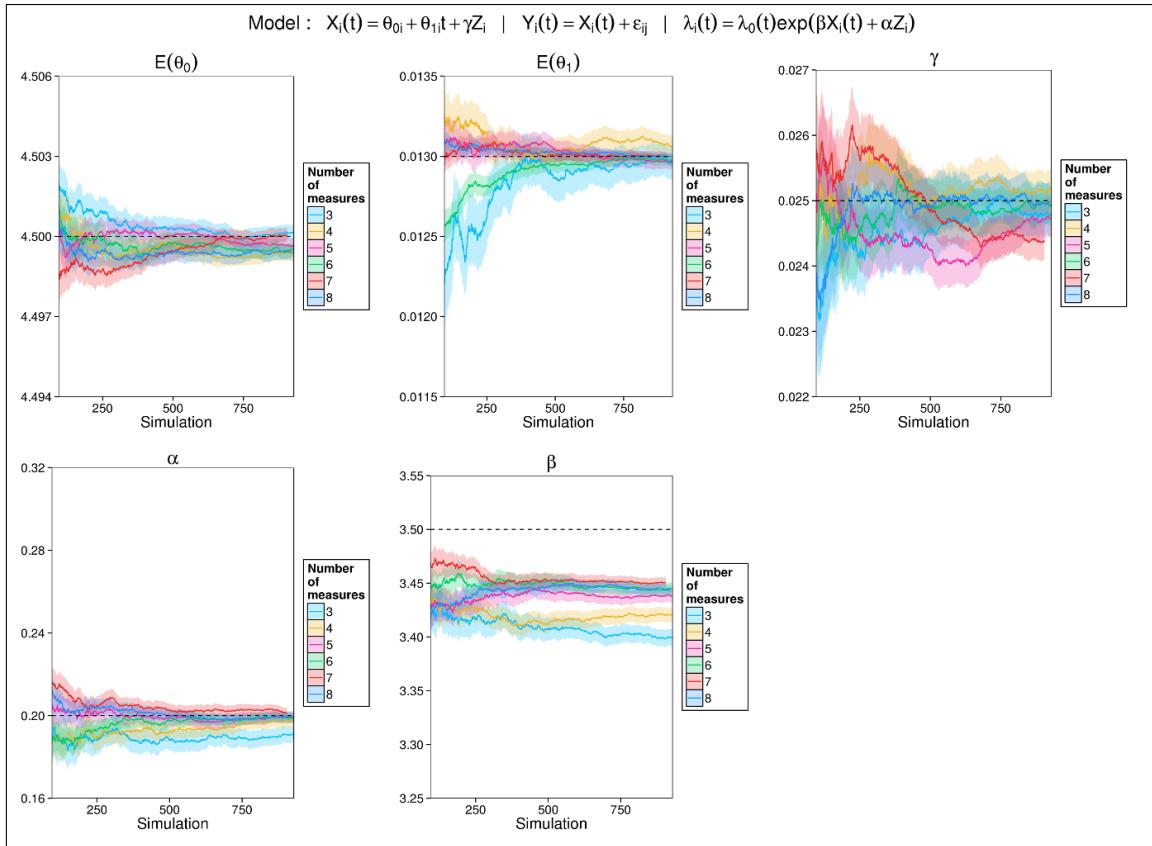
Figure 4: Estimation par **joineR** (Scénario 1).

Dans le cas des estimations obtenues à l'aide de **joineR** (Figure 4), qui se focalise principalement sur les paramètres de la composante LMM, les estimations semblent non biaisées pour  $E(\theta_0)$  et  $\gamma$ . Cependant, sur les 1000 simulations réalisées, un biais sur le paramètre  $E(\theta_1)$  est présent, en comparaison du biais de l'estimation de **JM**.

En ce qui concerne la composante de survie, les estimations de  $\alpha$  (l'effet du SNP) et de  $\beta$  (l'effet de la trajectoire) sont respectivement surestimées et sous-estimées. Encore une fois, la fréquence allélique ne doit pas être trop faible pour l'estimation des paramètres du modèle, notamment pour les paramètres de la composante de survie.

### 5.2.2 Scénario 2 : variation du nombre de mesures longitudinales

En ce qui concerne **JM**, la variation du nombre de mesures ne révèle que peu de changement concernant les paramètres de la composante LMM (Figure 5). Le changement le plus notable réside dans l'estimation du paramètre  $\beta$  de l'effet de la trajectoire sur la survenue de l'évènement. En effet, malgré le faible nombre de simulations, le biais de l'estimation se réduit avec l'augmentation du nombre de mesures.



**Figure 5:** Estimation par **JM** (Scénario 2).

Le biais des estimations de **joineR** (Figure 6) reste plus important que celui de **JM**, même si celui-ci se réduit avec l'augmentation du nombre de mesures longitudinales. Cependant, le biais d'estimation de  $\alpha$  varie très peu selon le nombre de mesures. L'estimation de  $E(\theta_1)$  demeure légèrement biaisée dans ce scénario, malgré le fait que le paquet **joineR** se concentre principalement sur l'estimation des paramètres issus de la composante longitudinale du modèle joint. Ce phénomène peut être dû à des problèmes de convergence, pouvant être causés par l'algorithme utilisé (par défaut) dans **joineR**, à savoir **nlinlmb** du paquet **stats**.

### 5.2.3 Conclusion

Ces simulations ont permis de mettre en évidence des points forts et des points faibles des implémentations actuelles des modèles joints (**JM** et **joineR**), et de se familiariser avec le processus de simulation des données. Le biais des estimateurs devra être étudié de façon plus approfondie dans la suite de la thèse.

Une étude détaillée des algorithmes sera nécessaire pour mieux comprendre ces deux implémentations et plus fondamentalement les modèles joints. En effet, le raffinement de l'algorithme de simulation pourra permettre de mieux caractériser le biais observé (Section 5.2.1 et 5.2.2). Les scénarios de simulation de données manquantes (MCAR, MAR et MNAR) ont pour la plu-

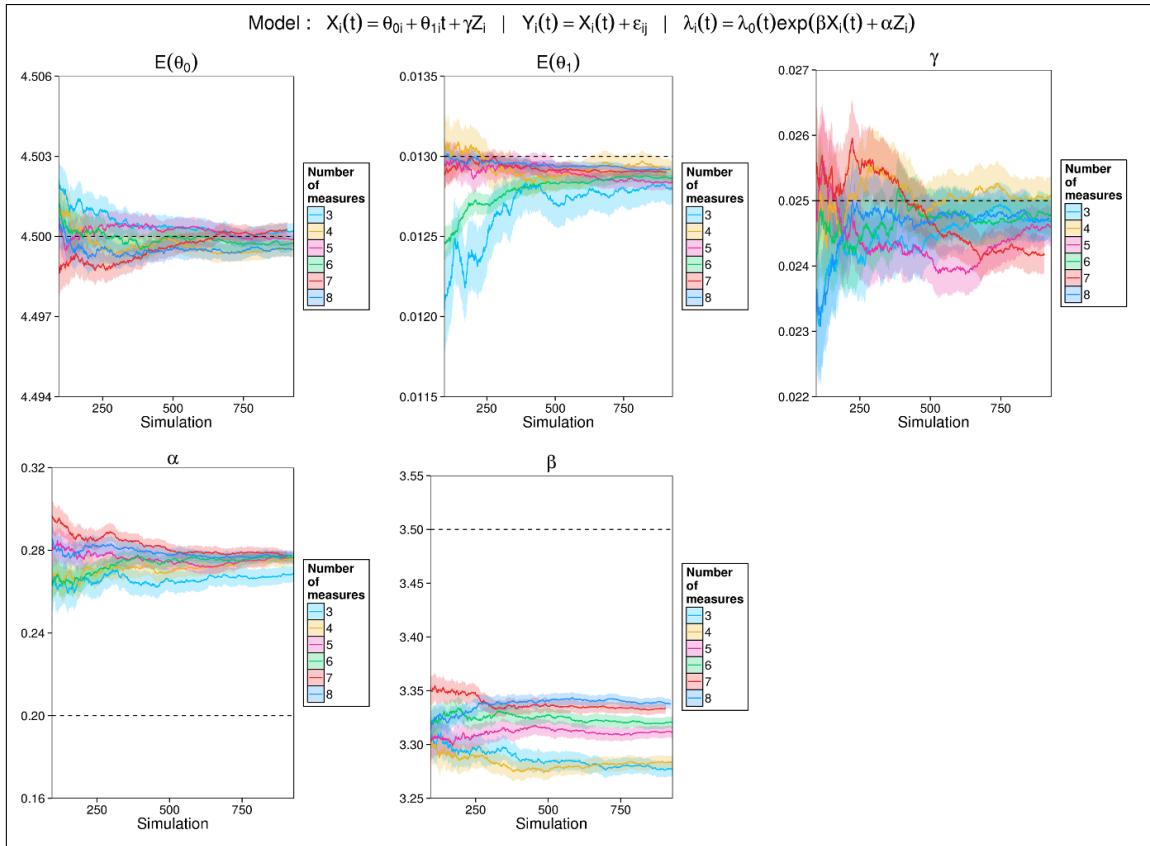


Figure 6: Estimation par `joineR` (Scénario 2).

part déjà été implémentés et seront étudiés prochainement.

Les implémentations actuelles sont peu appropriées en terme de temps de calcul et ce, malgré l'utilisation d'un serveur de calcul. En effet, 1000 simulations pour les scénarios présentés en [Section 5.2.1 et 5.2.2](#) ont demandé environ 15 heures en pleine charge (80 Coeurs). Au cours de la thèse, un effort plus important sera fourni dans l'amélioration du temps de calcul.

## 6 Perspectives

### 6.1 En cours / À court terme

1. Ecriture d'un article de revue de littérature sur les méthodes d'analyses de données longitudinales dans le contexte des études pangénomiques.
2. Développement et raffinement des méthodes de simulation dans un contexte génétique :
  - (a) fonction de risque par intervalles pour la composante de survie ;
  - (b) extension aux modèles de Weibull, Gompertz, etc ;
  - (c) caractérisation du biais de l'estimation des paramètres dans **JM** et **joineR**.

### 6.2 À long terme

Développement et implémentation de solutions :

- pour réduire le temps de calcul ;
- pour réduire le biais des estimations.

## 7 Formations

### 7.1 Passées

- Formation Collège doctoral : *Gérer efficacement sa documentation avec ZOTERO* (20 février 2015).
- Participation (poster) et obtention d'une bourse SFD-Lilly au *Congrès annuel de la SFD* (Société Francophone du Diabète) du 24 au 27 mars 2015, à Bordeaux.
- Formation Collège doctoral : *Veille et stratégie de recherche documentaire* (08 avril 2015).
- Participation et animation d'un atelier Julia lors des *JDEV 2015 (Journées nationales du DEVeloppement logiciel de l'Enseignement Supérieur et Recherche)*, du 20 juin au 3 juillet, à Bordeaux.
- Ecole doctorale : participation au *comité d'organisation de la Journée André Verbert 2015* (14 septembre 2015).

Nombre de crédits obtenus à ce jour : 9/60 crédits.

### 7.2 Futures

- Formation Collège doctoral : *LaTex Niveau avancé*.
- Formation Collège doctoral : *Déposer, signaler et diffuser sa thèse : ce qu'il faut savoir*.
- Formation Collège doctoral : *Améliorer ses chances d'être publié : Pourquoi ? Comment ?*
- *Congrès de la SFD 2016* (Société Francophone du Diabète)
- *Congrès IGES 2016* (International Genetic Epidemiology Society)
- *Rencontres R 2016*
- *useR 2016*.
- Formation Collège doctoral : *Accroître son aisance à l'oral grâce aux approches théâtrales*.
- ...

## Références

Beyene, J. and Hamid, J. S. (2014). Longitudinal Data Analysis in Genome-Wide Association Studies. *Genetic Epidemiology*, 38(S1) :S68–S73.

DIAGRAM (2015). DIAGRAM (diabetes genetics replication and meta-analysis) Consortium.

Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A. U., Wheeler, E., Glazer, N. L., Bouatia-Naji, N., Gloyn, A. L., Lindgren, C. M., Mägi, R., Morris, A. P., Randall, J., Johnson, T., Elliott, P., Rybin, D., Thorleifsson, G., Steinhorsdottir, V., Henneman, P., Grallert, H., Dehghan, A., Hottenga, J. J., Franklin, C. S., Navarro, P., Song, K., Goel, A., Perry, J. R. B., Egan, J. M., Lajunen, T., Grarup, N., Sparsø, T., Doney, A., Voight, B. F., Stringham, H. M., Li, M., Kanoni, S., Shrader, P., Cavalcanti-Proença, C., Kumari, M., Qi, L., Timpson, N. J., Gieger, C., Zabena, C., Rocheleau, G., Ingelsson, E., An, P., O'Connell, J., Luan, J., Elliott, A., McCarroll, S. A., Payne, F., Roccasecca, R. M., Pattou, F., Sethupathy, P., Ardlie, K., Ariyurek, Y., Balkau, B., Barter, P., Beilby, J. P., Ben-Shlomo, Y., Benediktsson, R., Bennett, A. J., Bergmann, S., Bochud, M., Boerwinkle, E., Bonnefond, A., Bonnycastle, L. L., Borch-Johnsen, K., Böttcher, Y., Brunner, E., Bumpstead, S. J., Charpentier, G., Chen, Y.-D. I., Chines, P., Clarke, R., Coin, L. J. M., Cooper, M. N., Cornelis, M., Crawford, G., Crisponi, L., Day, I. N. M., de Geus, E., Delplanque, J., Dina, C., Erdos, M. R., Fedson, A. C., Fischer-Rosinsky, A., Forouhi, N. G., Fox, C. S., Frants, R., Franzosi, M. G., Galan, P., Goodarzi, M. O., Graessler, J., Groves, C. J., Grundy, S., Gwilliam, R., Gyllensten, U., Hadjadj, S., Hallmans, G., Hammond, N., Han, X., Hartikainen, A.-L., Hassanali, N., Hayward, C., Heath, S. C., Hercberg, S., Herder, C., Hicks, A. A., Hillman, D. R., Hingorani, A. D., Hofman, A., Hui, J., Hung, J., Isomaa, B., Johnson, P. R. V., Jørgensen, T., Jula, A., Kaakinen, M., Kaprio, J., Kesaniemi, Y. A., Kivimaki, M., Knight, B., Koskinen, S., Kovacs, P., Kyvik, K. O., Lathrop, G. M., Lawlor, D. A., Le Bacquer, O., Lecoeur, C., Li, Y., Lyssenko, V., Mahley, R., Mangino, M., Manning, A. K., Martínez-Larrad, M. T., McAteer, J. B., McCulloch, L. J., McPherson, R., Meisinger, C., Melzer, D., Meyre, D., Mitchell, B. D., Morken, M. A., Mukherjee, S., Naitza, S., Narisu, N., Neville, M. J., Oostra, B. A., Orrù, M., Pakyz, R., Palmer, C. N. A., Paolisso, G., Pattaro, C., Pearson, D., Peden, J. F., Pedersen, N. L., Perola, M., Pfeiffer, A. F. H., Pichler, I., Polasek, O., Posthuma, D., Potter, S. C., Pouta, A., Province, M. A., Psaty, B. M., Rathmann, W., Rayner, N. W., Rice, K., Ripatti, S., Rivadeneira, F., Roden, M., Rolandsson, O., Sandbaek, A., Sandhu, M., Sanna, S., Sayer, A. A., Scheet, P., Scott, L. J., Seedorf, U., Sharp, S. J., Shields, B., Sigurðsson, G., Sijbrands, E. J. G., Silveira, A., Simpson, L., Singleton, A., Smith, N. L., Sovio, U., Swift, A., Syddall, H., Syvänen, A.-C., Tanaka,

- T., Thorand, B., Tichet, J., Tönjes, A., Tuomi, T., Uitterlinden, A. G., van Dijk, K. W., van Hoek, M., Varma, D., Visvikis-Siest, S., Vitart, V., Vogelzangs, N., Waeber, G., Wagner, P. J., Walley, A., Walters, G. B., Ward, K. L., Watkins, H., Weedon, M. N., Wild, S. H., Willemsen, G., Witteman, J. C. M., Yarnell, J. W. G., Zeggini, E., Zelenika, D., Zethelius, B., Zhai, G., Zhao, J. H., Zillikens, M. C., Borecki, I. B., Loos, R. J. F., Meneton, P., Magnusson, P. K. E., Nathan, D. M., Williams, G. H., Hattersley, A. T., Silander, K., Salomaa, V., Smith, G. D., Bornstein, S. R., Schwarz, P., Spranger, J., Karpe, F., Shuldiner, A. R., Cooper, C., Dedoussis, G. V., Serrano-Ríos, M., Morris, A. D., Lind, L., Palmer, L. J., Hu, F. B., Franks, P. W., Ebrahim, S., Marmot, M., Kao, W. H. L., Pankow, J. S., Sampson, M. J., Kuusisto, J., Laakso, M., Hansen, T., Pedersen, O., Pramstaller, P. P., Wichmann, H. E., Illig, T., Rudan, I., Wright, A. F., Stumvoll, M., Campbell, H., Wilson, J. F., Hamsten, A., Bergman, R. N., Buchanan, T. A., Collins, F. S., Mohlke, K. L., Tuomilehto, J., Valle, T. T., Altshuler, D., Rotter, J. I., Siscovick, D. S., Penninx, B. W. J. H., Boomsma, D., Deloukas, P., Spector, T. D., Frayling, T. M., Ferrucci, L., Kong, A., Thorsteinsdottir, U., Stefansson, K., van Duijn, C. M., Aulchenko, Y. S., Cao, A., Scuteri, A., Schlessinger, D., Uda, M., Ruokonen, A., Jarvelin, M.-R., Waterworth, D. M., Vollenweider, P., Peltonen, L., Mooser, V., Abecasis, G. R., Wareham, N. J., Sladek, R., Froguel, P., Watanabe, R. M., Meigs, J. B., Groop, L., Boehnke, M., McCarthy, M. I., Florez, J. C., and Barroso, I. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature genetics*, 42(2) :105–116.
- Fisher, L. D. and Lin, D. Y. (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Review of Public Health*, 20 :145–157.
- Ibrahim, J. G., Chu, H., and Chen, L. M. (2010). Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data. *Journal of Clinical Oncology*, 28(16) :2796–2801.
- MAGIC (2015). MAGIC (the meta-analyses of glucose and insulin-related traits consortium) Consortium.
- McCrink, L. M., Marshall, A. H., and Cairns, K. J. (2013). Advances in Joint Modelling : A Review of Recent Developments with Application to the Survival of End Stage Renal Disease Patients : Advances in Joint Modelling. *International Statistical Review*, 81(2) :249–269.
- Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segrè, A. V., Steinhorsdottir, V., Strawbridge, R. J., Khan, H., Grallert, H., Mahajan, A., Prokopenko, I., Kang, H. M., Dina, C., Esko, T., Fraser, R. M., Kanoni, S., Kumar, A., Lagou, V., Langenberg, C., Luan, J., Lindgren, C. M., Müller-Nurasyid, M., Pechlivanis, S., Rayner, N. W., Scott, L. J., Wiltshire, S., Yengo, L., Kinnunen, L., Rossin, E. J., Raychaudhuri, S., Johnson, A. D., Dimas, A. S., Loos,

- R. J. F., Vedantam, S., Chen, H., Florez, J. C., Fox, C., Liu, C.-T., Rybin, D., Couper, D. J., Kao, W. H. L., Li, M., Cornelis, M. C., Kraft, P., Sun, Q., van Dam, R. M., Stringham, H. M., Chines, P. S., Fischer, K., Fontanillas, P., Holmen, O. L., Hunt, S. E., Jackson, A. U., Kong, A., Lawrence, R., Meyer, J., Perry, J. R., Platou, C. G., Potter, S., Rehnberg, E., Robertson, N., Sivapalaratnam, S., Stančáková, A., Stirrups, K., Thorleifsson, G., Tikkannen, E., Wood, A. R., Almgren, P., Atalay, M., Benediktsson, R., Bonnycastle, L. L., Burtt, N., Carey, J., Charpentier, G., Crenshaw, A. T., Doney, A. S. F., Dorkhan, M., Edkins, S., Emilsson, V., Eury, E., Forsen, T., Gertow, K., Gigante, B., Grant, G. B., Groves, C. J., Guiducci, C., Herder, C., Hreidarsson, A. B., Hui, J., James, A., Jonsson, A., Rathmann, W., Klopp, N., Kravic, J., Krjutškov, K., Langford, C., Leander, K., Lindholm, E., Lobbens, S., Männistö, S., Mirza, G., Mühleisen, T. W., Musk, B., Parkin, M., Rallidis, L., Saramies, J., Sennblad, B., Shah, S., Sigurðsson, G., Silveira, A., Steinbach, G., Thorand, B., Trakalo, J., Veglia, F., Wennauer, R., Winckler, W., Zabaneh, D., Campbell, H., van Duijn, C., Uitterlinden89, A. G., Hofman, A., Sijbrands, E., Abecasis, G. R., Owen, K. R., Zeggini, E., Trip, M. D., Forouhi, N. G., Syvänen, A.-C., Eriksson, J. G., Peltonen, L., Nöthen, M. M., Balkau, B., Palmer, C. N. A., Lyssenko, V., Tuomi, T., Isomaa, B., Hunter, D. J., Qi, L., Shuldiner, A. R., Roden, M., Barroso, I., Wilsgaard, T., Beilby, J., Hovingh, K., Price, J. F., Wilson, J. F., Rauramaa, R., Lakka, T. A., Lind, L., Dedoussis, G., Njølstad, I., Pedersen, N. L., Khaw, K.-T., Wareham, N. J., Keinanen-Kiukaanniemi, S. M., Saaristo, T. E., Korpi-Hyövälti, E., Saltevo, J., Laakso, M., Kuusisto, J., Metspalu, A., Collins, F. S., Mohlke, K. L., Bergman, R. N., Tuomilehto, J., Boehm, B. O., Gieger, C., Hveem, K., Cauchi, S., Froguel, P., Baldassarre, D., Tremoli, E., Humphries, S. E., Saleheen, D., Danesh, J., Ingelsson, E., Ripatti, S., Salomaa, V., Erbel, R., Jöckel, K.-H., Moebus, S., Peters, A., Illig, T., de Faire, U., Hamsten, A., Morris, A. D., Donnelly, P. J., Frayling, T. M., Hattersley, A. T., Boerwinkle, E., Melander, O., Kathiresan, S., Nilsson, P. M., Deloukas, P., Thorsteinsdottir, U., Groop, L. C., Stefansson, K., Hu, F., Pankow, J. S., Dupuis, J., Meigs, J. B., Altshuler, D., Boehnke, M., and McCarthy, M. I. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics*, 44(9) :981–990.
- Pal, A. and McCarthy, M. I. (2013). The genetics of type 2 diabetes and its clinical relevance. *Clinical Genetics*, 83(4) :297–306.
- Philipson, P., Diggle, P., Sousa, I., Kolamunnage-Dona, R., Williamson, P., and Henderson, R. (2012). joineR : Joint modelling of repeated measurements and time-to-event data.
- R Core Team (2014). *R : A Language and Environment for Statistical Computing*. R Foundation

for Statistical Computing, Vienna, Austria.

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.

Scott, R. A., Lagou, V., Welch, R. P., Wheeler, E., Montasser, M. E., Luan, J., Mägi, R., Strawbridge, R. J., Rehnberg, E., Gustafsson, S., Kanoni, S., Rasmussen-Torvik, L. J., Yengo, L., Lecoeur, C., Shungin, D., Sanna, S., Sidore, C., Johnson, P. C. D., Jukema, J. W., Johnson, T., Mahajan, A., Verweij, N., Thorleifsson, G., Hottenga, J.-J., Shah, S., Smith, A. V., Sennblad, B., Gieger, C., Salo, P., Perola, M., Timpson, N. J., Evans, D. M., Pourcain, B. S., Wu, Y., Andrews, J. S., Hui, J., Bielak, L. F., Zhao, W., Horikoshi, M., Navarro, P., Isaacs, A., O'Connell, J. R., Stirrups, K., Vitart, V., Hayward, C., Esko, T., Mihailov, E., Fraser, R. M., Fall, T., Voight, B. F., Raychaudhuri, S., Chen, H., Lindgren, C. M., Morris, A. P., Rayner, N. W., Robertson, N., Rybin, D., Liu, C.-T., Beckmann, J. S., Willems, S. M., Chines, P. S., Jackson, A. U., Kang, H. M., Stringham, H. M., Song, K., Tanaka, T., Peden, J. F., Goel, A., Hicks, A. A., An, P., Müller-Nurasyid, M., Franco-Cereceda, A., Folkersen, L., Marullo, L., Jansen, H., Oldehinkel, A. J., Bruinenberg, M., Pankow, J. S., North, K. E., Forouhi, N. G., Loos, R. J. F., Edkins, S., Varga, T. V., Hallmans, G., Oksa, H., Antonella, M., Nagaraja, R., Trompet, S., Ford, I., Bakker, S. J. L., Kong, A., Kumari, M., Gigante, B., Herder, C., Munroe, P. B., Caulfield, M., Antti, J., Mangino, M., Small, K., Miljkovic, I., Liu, Y., Atalay, M., Kiess, W., James, A. L., Rivadeneira, F., Uitterlinden, A. G., Palmer, C. N. A., Doney, A. S. F., Willemsen, G., Smit, J. H., Campbell, S., Polasek, O., Bonnycastle, L. L., Hercberg, S., Dimitriou, M., Bolton, J. L., Fowkes, G. R., Kovacs, P., Lindström, J., Zemunik, T., Bandinelli, S., Wild, S. H., Basart, H. V., Rathmann, W., Grallert, H., Maerz, W., Kleber, M. E., Boehm, B. O., Peters, A., Pramstaller, P. P., Province, M. A., Borecki, I. B., Hastie, N. D., Rudan, I., Campbell, H., Watkins, H., Farrall, M., Stumvoll, M., Ferrucci, L., Waterworth, D. M., Bergman, R. N., Collins, F. S., Tuomilehto, J., Watanabe, R. M., de Geus, E. J. C., Penninx, B. W., Hofman, A., Oostra, B. A., Psaty, B. M., Vollenweider, P., Wilson, J. F., Wright, A. F., Hovingh, G. K., Metspalu, A., Uusitupa, M., Magnusson, P. K. E., Kyvik, K. O., Kaprio, J., Price, J. F., Dedoussis, G. V., Deloukas, P., Meneton, P., Lind, L., Boehnke, M., Shuldiner, A. R., van Duijn, C. M., Morris, A. D., Toenjes, A., Peyser, P. A., Beilby, J. P., Körner, A., Kuusisto, J., Laakso, M., Bornstein, S. R., Schwarz, P. E. H., Lakka, T. A., Rauramaa, R., Adair, L. S., Smith, G. D., Spector, T. D., Illig, T., de Faire, U., Hamsten, A., Gudnason, V., Kivimaki, M., Hingorani, A., Keinanen-Kiukaanniemi, S. M., Saaristo, T. E., Boomsma, D. I., Stefansson, K., van der Harst, P., Dupuis, J., Pedersen, N. L., Sattar, N., Harris, T. B., Cucca, F., Ripatti, S., Salomaa, V., Mohlke, K. L., Balkau, B., Froguel, P., Pouta, A., Jarvelin, M.-R., Wareham, N. J., Bouatia-Naji, N., McCarthy, M. I., Franks, P. W., Meigs, J. B.,

- Teslovich, T. M., Florez, J. C., Langenberg, C., Ingelsson, E., Prokopenko, I., and Barroso, I. (2012). Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature genetics*, 44(9) :991–1005.
- Self, S. and Pawitan, Y. (1992). Modeling a Marker of Disease Progression and Onset of Disease. In Jewell, N. P., Dietz, K., and Farewell, V. T., editors, *AIDS Epidemiology*, pages 231–255. Birkhäuser Boston.
- Sikorska, K. (2014). *Computationally Fast Approaches to Genome-Wide Association Studies*. PhD thesis.
- Sikorska, K., Montazeri, N. M., Uitterlinden, A., Rivadeneira, F., Eilers, P. H., and Lesaffre, E. (2015). GWAS with longitudinal phenotypes : performance of approximate procedures. *European Journal of Human Genetics*.
- Sikorska, K., Rivadeneira, F., Groenen, P. J., Hofman, A., Uitterlinden, A. G., Eilers, P. H., and Lesaffre, E. (2013). Fast linear mixed model computations for genome-wide association studies with longitudinal data. *Statistics in Medicine*, 32(1) :165–180.
- Sitlani, C. M., Rice, K. M., Lumley, T., McKnight, B., Cupples, L. A., Avery, C. L., Noordam, R., Stricker, B. H. C., Whitsel, E. A., and Psaty, B. M. (2015). Generalized estimating equations for genome-wide association studies using longitudinal phenotype data. *Statistics in Medicine*, 34(1) :118–130.
- Voight, B. F., Kang, H. M., Ding, J., Palmer, C. D., Sidore, C., Chines, P. S., Burtt, N. P., Fuchsberger, C., Li, Y., Erdmann, J., Frayling, T. M., Heid, I. M., Jackson, A. U., Johnson, T., Kilpeläinen, T. O., Lindgren, C. M., Morris, A. P., Prokopenko, I., Randall, J. C., Saxena, R., Soranzo, N., Speliotes, E. K., Teslovich, T. M., Wheeler, E., Maguire, J., Parkin, M., Potter, S., Rayner, N. W., Robertson, N., Stirrups, K., Winckler, W., Sanna, S., Mulas, A., Nagaraja, R., Cucca, F., Barroso, I., Deloukas, P., Loos, R. J. F., Kathiresan, S., Munroe, P. B., Newton-Cheh, C., Pfeufer, A., Samani, N. J., Schunkert, H., Hirschhorn, J. N., Altshuler, D., McCarthy, M. I., Abecasis, G. R., and Boehnke, M. (2012). The Metabochip, a Custom Genotyping Array for Genetic Studies of Metabolic, Cardiovascular, and Anthropometric Traits. *PLoS Genet*, 8(8) :e1002793.
- Voight, B. F., Scott, L. J., Steinhorsdottir, V., Morris, A. P., Dina, C., Welch, R. P., Zeggini, E., Huth, C., Aulchenko, Y. S., Thorleifsson, G., McCulloch, L. J., Ferreira, T., Grallert, H., Amin, N., Wu, G., Willer, C. J., Raychaudhuri, S., McCarroll, S. A., Langenberg, C., Hofmann, O. M., Dupuis, J., Qi, L., Segrè, A. V., Hoek, M. v., Navarro, P., Ardlie, K., Balkau, B.,

Benediktsson, R., Bennett, A. J., Blagieva, R., Boerwinkle, E., Bonnycastle, L. L., Boström, K. B., Bravenboer, B., Bumpstead, S., Burtt, N. P., Charpentier, G., Chines, P. S., Cornelis, M., Couper, D. J., Crawford, G., Doney, A. S. F., Elliott, K. S., Elliott, A. L., Erdos, M. R., Fox, C. S., Franklin, C. S., Ganser, M., Gieger, C., Grarup, N., Green, T., Griffin, S., Groves, C. J., Guiducci, C., Hadjadj, S., Hassanali, N., Herder, C., Isomaa, B., Jackson, A. U., Johnson, P. R. V., Jørgensen, T., Kao, W. H. L., Klopp, N., Kong, A., Kraft, P., Kuusisto, J., Lauritzen, T., Li, M., Lieverse, A., Lindgren, C. M., Lyssenko, V., Marre, M., Meitinger, T., Midthjell, K., Morken, M. A., Narisu, N., Nilsson, P., Owen, K. R., Payne, F., Perry, J. R. B., Petersen, A.-K., Platou, C., Proença, C., Prokopenko, I., Rathmann, W., Rayner, N. W., Robertson, N. R., Rocheleau, G., Roden, M., Sampson, M. J., Saxena, R., Shields, B. M., Shrader, P., Sigurdsson, G., Sparsø, T., Strassburger, K., Stringham, H. M., Sun, Q., Swift, A. J., Thorand, B., Tichet, J., Tuomi, T., Dam, R. M. v., Haeften, T. W. v., Herpt, T. v., Vliet-Ostaptchouk, J. V. v., Walters, G. B., Weedon, M. N., Wijmenga, C., Witteman, J., Investigators, T. M., Consortium, T. G., Bergman, R. N., Cauchi, S., Collins, F. S., Gloyn, A. L., Gyllensten, U., Hansen, T., Hide, W. A., Hitman, G. A., Hofman, A., Hunter, D. J., Hveem, K., Laakso, M., Mohlke, K. L., Morris, A. D., Palmer, C. N. A., Pramstaller, P. P., Rudan, I., Sijbrands, E., Stein, L. D., Tuomilehto, J., Uitterlinden, A., Walker, M., Wareham, N. J., Watanabe, R. M., Abecasis, G. R., Boehm, B. O., Campbell, H., Daly, M. J., Hattersley, A. T., Hu, F. B., Meigs, J. B., Pankow, J. S., Pedersen, O., Wichmann, H.-E., Barroso, I., Florez, J. C., Frayling, T. M., Groop, L., Sladek, R., Thorsteinsdottir, U., Wilson, J. F., Illig, T., Froguel, P., Duijn, C. M. v., Stefansson, K., Altshuler, D., Boehnke, M., and McCarthy, M. I. (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genetics*, 42(7) :579–589.

World Health Organization (2015). Diabetes Fact sheet No 312.

Wu, Z., Hu, Y., and Melton, P. E. (2014). Longitudinal Data Analysis for Genetic Studies in the Whole-Genome Sequencing Era. *Genetic Epidemiology*, 38(S1) :S74–S80.

Yaghootkar, H. and Frayling, T. M. (2013). Recent progress in the use of genetics to understand links between type 2 diabetes and related metabolic traits. *Genome Biology*, 14(3) :203.