OXFORD GENETICS

GENETICS, 2023, 225(1), iyad119
https://doi.org/10.1093/genetics/iyad119
Advance Access Publication Date: 27 June 2023
Investigation

Characterization of direct and/or indirect genetic associations for multiple traits in longitudinal studies of disease progression

Myriam Brossard,[1,*] Andrew D. Paterson,[2,3] Osvaldo Espin-Garcia,[3,4,5,6] Radu V. Craiu,[5] Shelley B. Bull[1,3,*]

[1]Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto M5T 3L9, Ontario, Canada
[2]Program in Genetics and Genome Biology, Hospital for Sick Children Research Institute, Toronto M5G 1X8, Ontario, Canada
[3]Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto M5T 3M7, Ontario, Canada
[4]Department of Biostatistics, Princess Margaret Cancer Centre, Toronto M5G 2C1, Ontario, Canada
[5]Department of Statistical Sciences, University of Toronto, Toronto M5S 3G3, Ontario, Canada
[6]Department of Epidemiology and Biostatistics, Western University, London N6A 5C1, Ontario, Canada

*Corresponding authors: Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, 60 Murray Street, Box #18, Toronto M5T 3L9, Ontario, Canada.
Emails: bull@lunenfeld.ca, brossard@lunenfeld.ca.

Abstract

When quantitative longitudinal traits are risk factors for disease progression and subject to random biological variation, joint model analysis of time-to-event and longitudinal traits can effectively identify direct and/or indirect genetic association of single nucleotide polymorphisms (SNPs) with time-to-event. We present a joint model that integrates: (1) a multivariate linear mixed model describing trajectories of multiple longitudinal traits as a function of time, SNP effects, and subject-specific random effects and (2) a frailty Cox survival model that depends on SNPs, longitudinal trajectory effects, and subject-specific frailty accounting for dependence among multiple time-to-event traits. Motivated by complex genetic architecture of type 1 diabetes complications (T1DC) observed in the Diabetes Control and Complications Trial (DCCT), we implement a 2-stage approach to inference with bootstrap joint covariance estimation and develop a hypothesis testing procedure to classify direct and/or indirect SNP association with each time-to-event trait. By realistic simulation study, we show that joint modeling of 2 time-to-T1DC (retinopathy and nephropathy) and 2 longitudinal risk factors (HbA1c and systolic blood pressure) reduces estimation bias in genetic effects and improves classification accuracy of direct and/or indirect SNP associations, compared to methods that ignore within-subject risk factor variability and dependence among longitudinal and time-to-event traits. Through DCCT data analysis, we demonstrate feasibility for candidate SNP modeling and quantify effects of sample size and Winner's curse bias on classification for 2 SNPs identified as having indirect associations with time-to-T1DC traits. Joint analysis of multiple longitudinal and multiple time-to-event traits provides insight into complex traits architecture.

Keywords: joint models, longitudinal study, direct and/or indirect genetic association, pleiotropy, complex genetic architecture, multiple-trait analysis, random measurement error, quantitative trait trajectory, mixed model, frailty model

Introduction

Despite their known ability to improve inference in clinical and epidemiological studies, particularly in the presence of informative censoring/dropout or when longitudinal quantitative traits (QTs) are measured with biological random variation (Hogan and Laird 1998; Ibrahim *et al.* 2010; Chen *et al.* 2011), joint models for longitudinal QTs and time-to-event (TTE) traits have received limited attention in genetic association study design and analysis. Genome-wide association studies (GWAS) of QTs often require follow-up analyses to identify whether single nucleotide polymorphism (SNP) associations detected with a QT also affect clinical outcomes, such as disease complications, through direct and/or indirect effects induced by the QT (Fig. 1). Distinguishing between direct and/or indirect SNP effects can help to reveal genetic pathways in the etiology of disease progression with implications for the direction of ongoing investigations and development of new intervention strategies. However, within-patient variability in intermediate QTs (e.g. random biological variation) and unmeasured shared risk factors among longitudinal and TTE traits challenge accurate distinction between direct and/or indirect SNP associations.

Our objective is an integrated approach to investigate complex genetic architecture of disease progression and associated risk factors, motivated by genetic association analysis of individuals with type 1 diabetes (T1D). Risk of type 1 diabetes complications (T1DC), including diabetic retinopathy (DR) and diabetic nephropathy (DN), is hypothesized to depend on multiple genetic factors with direct and/or indirect effects induced via multiple shared and/or specific QT risk factors (Paterson and Bull 2012). The first GWAS in Diabetes Control and Complications Trial (DCCT) identified SNPs associated with hyperglycemia, as measured by hemoglobin A1c (HbA1c), at genome-wide significance and weaker

Received: March 30, 2023. **Accepted:** June 19, 2023

associations with time-to-DR and/or time-to-DN (Paterson *et al.* 2010). Other longitudinal QTs, also influenced by genetic factors, are postulated to have associations with T1DC. Integrated modeling therefore entails multiple longitudinal QT risk factors and multiple TTE traits, as well as multiple SNP trait associations (Figs. 2 and 3 for an illustration). In addition to genetic association with longitudinal QT risk factors, such QTs can be related to more than 1 complication, and genetic variants can affect risk of multiple complications directly and/or indirectly through those intermediate QTs. When a longitudinal QT associated with a TTE is ignored or measured with random errors, an indirect SNP–TTE association through the intermediate QT can be mistaken as a direct SNP–TTE association. Therefore, accounting for all the intermediate longitudinal QTs is essential to correctly distinguish between direct and/or indirect SNP–T1DC association. To this end, we formulate a general model extension for multiple longitudinal QTs and multiple TTE traits.

The multiple-trait extension stems from existing joint models that consist of a submodel for a single longitudinal QT linked to a submodel for a single right-censored TTE trait (Rizopoulos 2012; Wu *et al.* 2012; Asar *et al.* 2015). The longitudinal submodel describes the QT as an underlying smooth trajectory that depends on fixed effects of time and baseline covariates, as well as subject-specific random effects. The joint model association structure is induced via the functional dependence between the hazard of an event at time t and the longitudinal QT trajectory (Hickey *et al.* 2016; Papageorgiou *et al.* 2019). Specification of this relationship can be based on prior biological knowledge of the link between the longitudinal and TTE traits. This class of joint models provides interpretations of direct and/or indirect effects because the relationship between a baseline covariate, such as a SNP genotype, and each of the longitudinal and TTE traits, as well as the relationship between longitudinal and TTE traits, can be specified via model parameters corresponding to direct, indirect, and overall effects (Ibrahim *et al.* 2010).

Joint model extensions have been reviewed for multiple longitudinal QTs (Hickey *et al.* 2016; Papageorgiou *et al.* 2019) and for multiple TTE traits (Hickey *et al.* 2018a). Although a few extensions have been developed for both multiple longitudinal QTs and multiple TTE traits [(Zhu *et al.* 2012; Tang *et al.* 2014; Tang and Tang 2015), for example], these models are often formulated for a specific study question and can lack generalizability. In addition, multiple-trait extensions raise computational challenges for marginal likelihood maximization that integrates over the distribution of the multivariate random effects. Two-stage approaches for joint model fitting are computationally efficient and allow more flexible model formulations (Self and Pawitan 1992; Tsiatis *et al.* 1995; Bycott and Taylor 1998; Dafni and Tsiatis 1998). However, in some circumstances, inference can be mis-calibrated when parameter estimates and predictions from stage 1 are obtained from the longitudinal model without consideration of the TTE trait, or when the uncertainty in stage 1 estimates is ignored during stage 2 estimation (Wulfsohn and Tsiatis 1997), a problem known as propagation of errors.

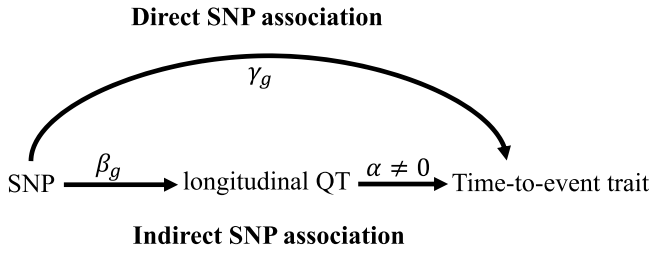

Fig. 1. Directed acyclic graph (DAG) illustrating the joint model parameters to characterize the direct SNP effect on the TTE trait and the indirect SNP effect via the intermediate longitudinal QT associated with the TTE trait. Figure adapted from (Ibrahim *et al.* 2010) which proposed a general joint model formulation for 1 longitudinal QT and 1 TTE trait to address questions specific to testing for treatment effects in randomized-controlled clinical trials.

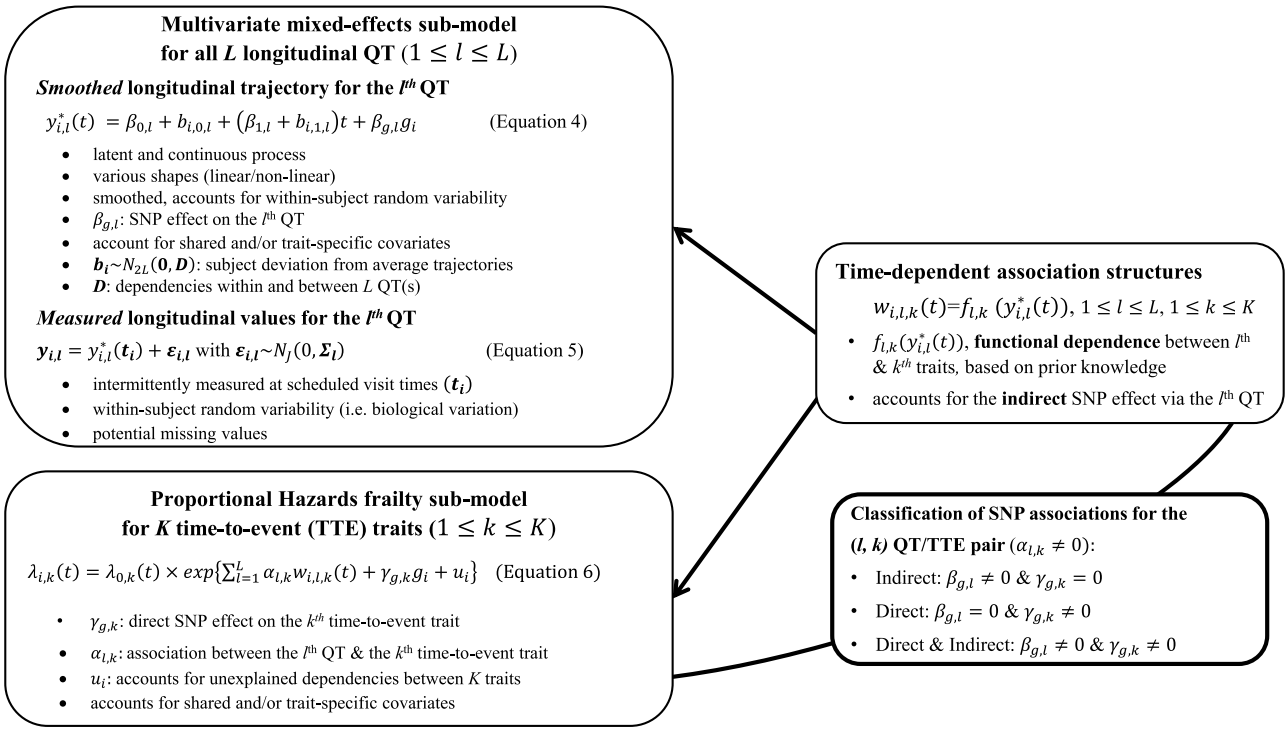

Fig. 2. Proposed joint modeling approach for characterization of complex genetic architecture of multiple disease progression.

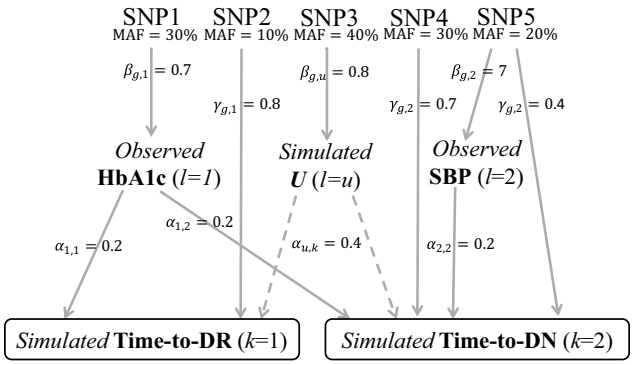

Fig. 3. Realistic DCCT data-based causal genetic scenario. We generated $R = 1{,}000$ replicates of $N = 667$ DCCT individuals with $M = 5$ causal variants and $K = 2$ TTE traits simulated under this *causal* genetic scenario and $R = 1{,}000$ replicates of $M = 5$ SNPs (with same MAFs as the causal ones) simulated under a *global null* genetic scenario where none of the SNPs is associated with any traits. The effects of sex on SBP and of T1D duration at baseline on both time-to-T1DC traits are not represented in this figure but are included in the data generating model; see Supplementary File 2 (sections 2 and 3) for details.

The *primary* contribution of the work we report is a general joint model specification for genetic association studies in which multiple longitudinal QTs are related to multiple TTE traits, correlated random effects, and a frailty address dependencies among QTs and among TTE traits. As a practical implementation, we develop inference for statistical genetic analysis using 2-stage estimation with bootstrap resampling for parameter estimation and hypothesis testing and then propose a novel procedure to classify SNP associations with each TTE trait as direct and/or indirect. A *second* contribution is the implementation of a data-informed simulation algorithm, under the postulated multiple-trait model for T1DC genetic architecture, to generate multiple causal SNPs with direct effects on simulated TTE traits and/or indirect effects via observed (measured) longitudinal QTs in DCCT and an unobserved (simulated) longitudinal QT. Evaluations using the implementation demonstrate that in comparison to separate joint models for each QT–TTE trait pair and approaches that ignore measurement error in longitudinal values, the proposed joint model extension reduces estimation bias and improves classification of direct and/or indirect SNP associations by (1) borrowing information shared among correlated traits, and (2) accounting for indirect genetic pathways via longitudinal QTs. The algorithm also provides a general approach to estimate power of the joint model analysis to detect association given study sample size and various direct/indirect genetic associations involving multiple observed or unobserved longitudinal QTs for time to disease complications. *Lastly*, we show computational feasibility and interpretation in an extended joint model application to DCCT genetic association analyses of selected SNPs. We classify 2 SNPs as having indirect association with 2 T1DC traits via the HbA1c longitudinal risk factor and obtain similar conclusions using alternative time-dependent association structures that account for cumulative and time-weighted effects of HbA1c on T1DC traits (Lind *et al.* 1995, 2010). Example R programs for data simulation and for application of the joint model are available on GitHub.

Materials and methods

Model formulation

We assume that a set of M SNPs has been genotyped, together with observation of K $(1 \le k \le K)$ unordered and noncompeting TTE traits, such as multiple disease complications, and L $(1 \le l \le L)$ longitudinal QTs (i.e. intermediate risk factors) measured in N unrelated individuals indexed by i $(1 \le i \le N)$. To characterize the genetic architecture of multiple longitudinal QTs and multiple TTE traits, we formulate a shared random-effects joint model that connects longitudinal and TTE submodels through specified time-dependent association structures. For ease of presentation, we simplify the model notation by assuming no adjusting covariates but note that trait-specific and/or shared covariates, such as confounding factors or ancestry-related principal components can be easily incorporated. In the application and the discussion, we comment on approaches to assess adequacy of the model specification. We first introduce the joint model for 1 longitudinal QT ($L = 1$) and 1 TTE trait ($K = 1$) and then present the extension for $L > 1$ QTs and $K > 1$ TTE traits.

Joint model for 1 longitudinal and 1 TTE trait

For each individual i, we define $\mathbf{y_i} = (y_{i,1}, \ldots, y_{i,j}, \ldots, y_{i,J})$, as the vector of QT measures collected over J visit times $\mathbf{t_i} = (t_{i,1}, \ldots, t_{i,j}, \ldots, t_{i,J})^T$ with $1 \le j \le J$ and $t_{i,1} \le \ldots \le t_{i,j} \le \ldots \le t_{i,J}$. We denote (T_i, δ_i) as the vector of right-censored event time T_i and event indicator δ_i for the TTE and assume $T_i = \min(T_i^*, C_i)$, where T_i^* is the latent (uncensored) event time and C_i is the censoring time (e.g. administrative censoring). We define $\delta_i = I(T_i^* \le C_i)$, with $\delta_i = 1$ if the event occurs during the observation period, and $\delta_i = 0$ otherwise.

Longitudinal submodel

Specification is by a linear mixed-effects model for the longitudinal QT (Laird and Ware 1982). The model assumes that for every individual in the sample, an underlying smooth trajectory of the longitudinal QT describes the subject-specific evolution dependent on time, SNP effect, and individual-level random-effects $\mathbf{b_i}$. To simplify the presentation, we assume a linear QT trajectory (Equation 1), but the longitudinal submodel can be specified with nonlinear trajectories using, for example, higher-order polynomials or splines to account for nonlinear time trends (Rizopoulos 2012, Bian 2020). The smooth linear trajectory is defined as:

$$y_i^*(t) = \beta_0 + b_{i,0} + (\beta_1 + b_{i,1})t + \beta_g g_i, \quad (1)$$

where:

- g_i, is the number of copies of the minor allele for each individual i for the SNP tested,
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_g)^T$ is the vector of fixed intercept, slope time, and SNP effects on the longitudinal QT,
- and $\mathbf{b_i} = (b_{i,0}, b_{i,1})^T$ is the subject-specific random intercept and slope time effects assuming $\mathbf{b_i} \sim N_2(0, \mathbf{D})$, where $\mathbf{D}$ is the variance–covariance matrix.

This trajectory cannot be observed directly; rather we observe longitudinal measurements $\mathbf{y_i}$ collected at discrete time points $\mathbf{t_i}$; measurements are subject to independent and identically distributed noise contamination variables $\boldsymbol{\varepsilon_i} \sim N_J(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\varepsilon_i} = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,j}, \ldots, \varepsilon_{i,J})^T$, $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I_J}$, and σ^2 is the residual variance of the QT. Then, the vector of observed QT values is:

$$\mathbf{y_i} = y_i^*(\mathbf{t_i}) + \boldsymbol{\varepsilon_i} \quad (2)$$

We assume that $\mathbf{b_i}$ and $\boldsymbol{\varepsilon_i}$ are independent (Laird and Ware 1982).

Equation 2 implies that $\mathbf{y_i}$ in $\Re^J$ follows a multivariate normal distribution with

$$E[\mathbf{y_i}] = \mathbf{X_i}\boldsymbol{\beta} \text{ and } \mathrm{Var}[\mathbf{y_i}] = \mathbf{Z_i}\mathbf{D}\mathbf{Z_i^T} + \boldsymbol{\Sigma}$$

where $\mathbf{X_i} = (\mathbf{1_J}, \mathbf{t_i}, g_i\mathbf{1_J})$ denotes the (J-by-3) design matrix for the fixed intercept, slope, and SNP effects and $\mathbf{Z_i} = (\mathbf{1_J}, \mathbf{t_i})$ is the (J-by-2) design matrix for the random intercept and slope effects, with $\mathbf{1_J} = (1, \ldots, 1, \ldots, 1)^T$. To improve robustness to misspecification of the variance–covariance matrix $\boldsymbol{D}$, we adopt an unstructured form for the random-effects variance, defined as $\mathbf{D} = \begin{pmatrix} \mathrm{Var}(b_{i,0}) & \mathrm{Cov}(b_{i,0}, b_{i,1}) \\ \mathrm{Cov}(b_{i,0}, b_{i,1}) & \mathrm{Var}(b_{i,1}) \end{pmatrix}$, which does not require additional constraints on serial dependence between the repeated measurements. This choice implies that the covariance between any pair of QT observations for individual i collected at 2 distinct visit times $t_{i,j} \neq t_{i,s}$ $(1 \leq j \leq J,\ 1 \leq s \leq J,\ j \neq s)$ is $\mathrm{Cov}(y_{i,j}, y_{i,s}) = t_{i,j}t_{i,s}\mathrm{Var}(b_{i,1}) + (t_{i,j} + t_{i,s})\mathrm{Cov}(b_{i,0}, b_{i,1}) + \mathrm{Var}(b_{i,0}) + \sigma^2$, with variance $\mathrm{Var}(y_{i,j}) = t_{i,j}^2\mathrm{Var}(b_{i,1}) + 2t_{i,j}\mathrm{Cov}(b_{i,0}, b_{i,1}) + \mathrm{Var}(b_{i,0}) + \sigma^2$, which is quadratic over time with positive curvature at $\mathrm{Var}(b_{i,1})$.

TTE submodel

We define a proportional hazard (PH) model in which the hazard function of the TTE is the instantaneous event rate in a small interval around T_i^* given that the event has not occurred before time t and depends on a function of the true unobserved longitudinal process up to time t, $W_i(t) = f(Y_i^*(t))$, with $Y_i^*(t) = \{ y_i^*(s),\ 0 \leq s \leq t\}$. The hazard function (Equation 3) specifies the SNP effect adjusted for association of the longitudinal QT risk factor with the TTE trait:

$$\lambda_i(t) = \lim_{dt \to 0} \Pr\{\ t \leq T_i^* < t + dt \,|T_i^* \geq t,\quad W_i(t) = f(Y_i^*(t)),\quad g_i\}/dt,$$

We assume:

$$\lambda_i(t) = \lambda_0(t) \times \exp\{\alpha w_i(t) + \gamma_g g_i\}, \tag{3}$$

where $\lambda_0(t)$ is a (parametric or nonparametric) baseline hazard function and $w_i(t) = f(y_i^*(t))$ specifies the function of the longitudinal QT trajectory values associated with risk of the event at time t. In the case of a *contemporaneous* parametrization, $\lambda_i(t)$ depends on the trajectory value at the same time t (i.e. $w_i(t) = y_i^*(t)$). Other functional forms weight earlier QT values according to prior knowledge of the relationship with the TTE trait (Hickey *et al.* 2016; Mauff *et al.* 2017; Papageorgiou *et al.* 2019). The parameters, α and γ_g correspond to the effect of the longitudinal QT on TTE (α), and to the SNP effect on the TTTE trait (γ_g) conditional on the longitudinal QT trajectory.

Interpretation

As depicted in Fig. 1, the joint model parameters characterize relationships among a SNP, an intermediate QT, and a TTE trait. Effects of a SNP on a TTE trait can be interpreted as indirect association induced via the SNP effect (β_g) on the longitudinal QT; direct association of the SNP (γ_g) independent of the QT (Ibrahim *et al.* 2010; Hickey *et al.* 2018b). Thus, when $\alpha \neq 0$, a SNP association with a TTE trait can be 1 of 3 types:

- Indirect association: the SNP has a nonnull effect on the longitudinal QT ($\beta_g \neq 0$) but no effect on TTE conditional on the QT trajectory ($\gamma_g = 0$); the overall SNP effect θ depends only on the indirect effect ($\theta = \mu_g$, with $\mu_g = \alpha\beta_g$).
- Direct association: the SNP has a nonnull effect on the TTE conditional on the QT trajectory ($\gamma_g \neq 0$), but no effect on the longitudinal QT ($\beta_g = 0$); the overall SNP effect depends only on the direct effect ($\theta = \gamma_g$).
- Both direct and indirect associations: the SNP has nonnull effects on the longitudinal QT ($\beta_g \neq 0$) and on the TTE conditional on the QT trajectory ($\gamma_g \neq 0$). In this case, the overall SNP effect θ aggregates the indirect and direct SNP effects ($\theta = \mu_g + \gamma_g$, with $\mu_g = \alpha\beta_g$).

When an associated longitudinal QT is omitted from the TTE model, the estimated SNP effect on the TTE is no longer conditional on the QT trajectory. This can occur in GWAS when the TTE analysis ignores an intermediate QT or when the TTE is associated with more than 1 such QTs. This observation also illustrates a limitation of the joint model for 1 longitudinal QT and 1 TTE trait, with the consequence that an indirect SNP association can be mistaken as a direct association when other longitudinal QTs are omitted.

Generalization of the joint model to multiple *longitudinal and* multiple *TTE traits*

To characterize the genetic architecture of a system of multiple longitudinal QT risk factors and multiple TTE traits, we extend the joint model to $L > 1$ longitudinal and $K > 1$ TTE traits (Fig. 2). We define $\mathbf{y_{i,l}} = (y_{i,l,1}, \ldots, y_{i,l,j}, \ldots, y_{i,l,J})$ as the observed longitudinal measures for each lth QT ($1 \leq l \leq L$) collected over J visit times $\mathbf{t_i}$ and $(T_{i,k}, \delta_{i,k})$ as the observed right-censored event time $T_{i,k}$ and event indicator $\delta_{i,k}$ for each kth TTE trait with $\delta_{i,k} = I(T_{i,k}^* \leq C_i)$. We assume the same censoring time C_i across all K TTE traits, but the model can be extended to situations where C_i varies for each TTE. Again, for ease of presentation, we assume no adjusting covariates, linear trajectories for all L longitudinal QTs, and contemporaneous effects of longitudinal QTs on the K TTE traits.

Multivariate longitudinal submodel

In the multivariate extension, we index subscripts in Equations 1 and 2 for lth longitudinal QT (Equations 4 and 5 in Fig. 2). The vector of observed repeated QT measures for the lth QT becomes:

$$\mathbf{y_{i,l}} = \mathbf{X_{i,l}}\boldsymbol{\beta_l} + \mathbf{Z_{i,l}}\mathbf{b_{i,l}} + \boldsymbol{\varepsilon_{i,l}},$$

where:

- $\mathbf{X_{i,l}} = (\mathbf{1_J}, \mathbf{t_i}, g_i\mathbf{1_J})$ and $\mathbf{Z_{i,l}} = (\mathbf{1_J}, \mathbf{t_i})$ are the design matrices for fixed and random effects,
- $\boldsymbol{\beta_l} = (\beta_{0,l},\ \beta_{1,l}, \beta_{g,l})^T$ and $\mathbf{b_{i,l}} = (b_{i,0,l}, b_{i,1,l})^T$ denote the QT-specific fixed and random-effects vectors,
- and $\boldsymbol{\varepsilon_{i,l}} = (\boldsymbol{\varepsilon_{i,1,l}}, \ldots, \boldsymbol{\varepsilon_{i,j,l}}, \ldots, \boldsymbol{\varepsilon_{i,J,l}})^T$ is the vector of residual error terms, with $\boldsymbol{\varepsilon_{i,l}} \sim \mathbf{N_J}(\mathbf{0},\ \boldsymbol{\Sigma_l})$, where $\boldsymbol{\Sigma_l} = \sigma_l^2\mathbf{I_J}$ and σ_l^2 is the residual variance for the lth QT; we assume independent $\boldsymbol{\varepsilon_{i,l}}$ for all L QTs.

To account for dependence among longitudinal QTs, we assume the overall random-effects vector $\mathbf{b_i} = (\mathbf{b_{i,1}}, \ldots, \mathbf{b_{i,l}}, \ldots, \mathbf{b_{i,L}})^T \sim N_{2L}(\mathbf{0}, \mathbf{D})$, where $\mathbf{D} = \begin{pmatrix} \mathbf{D_{1,1}} & \cdots & \mathbf{D_{1,L}} \\ \vdots & \mathbf{D_{l,l}} & \vdots \\ \mathbf{D_{L,1}} & \cdots & \mathbf{D_{L,L}} \end{pmatrix}$ is the variance–covariance matrix for L QTs, accounting for serial dependencies within each QT, i.e. $\mathbf{D_{l,l}} = \begin{pmatrix} \mathrm{Var}(b_{i,0,l}) & \mathrm{Cov}(b_{i,0,l}, b_{i,1,l}) \\ \mathrm{Cov}(b_{i,0,l}, b_{i,1,l}) & \mathrm{Var}(b_{i,1,l}) \end{pmatrix}$, and accounting for cross-dependencies between each pair of QTs ($l \neq m$),

i.e. $\boldsymbol{D}_{\mathbf{l},\boldsymbol{m}} = \begin{pmatrix} \mathrm{Cov}(b_{i,0,l}, b_{i,0,m}) & \mathrm{Cov}(b_{i,0,l}, b_{i,1,m}) \\ \mathrm{Cov}(b_{i,1,l}, b_{i,0,m}) & \mathrm{Cov}(b_{i,1,l}, b_{i,1,m}) \end{pmatrix}$.

This formulation implies that the vector of stacked repeated measures of L QTs for individual i, $\mathbf{y_i} = (\mathbf{y_{i,1}}, \ldots, \mathbf{y_{i,l}}, \ldots, \mathbf{y_{i,L}})^T$ in $\Re^{J\times L}$ follows a multivariate normal distribution with mean $E[\mathbf{y_i}] = \boldsymbol{X_i\beta}$ and variance $\mathrm{Var}[\mathbf{y_i}] = \mathbf{Z_i D Z_i^T} + \boldsymbol{\Sigma}$, where $\mathbf{X_i} = \mathrm{diag}(\mathbf{X_{i,1}}, \ldots, \mathbf{X_{i,l}}, \ldots, \mathbf{X_{i,L}})$ and $\mathbf{Z_i} = \mathrm{diag}(\mathbf{Z_{i,1}}, \ldots, \mathbf{Z_{i,l}}, \ldots, \mathbf{Z_{i,L}})$ are the overall (JL-by-$3L$) and (JL-by-$2L$) block diagonal design matrices for fixed and random effects, $\boldsymbol{\beta} = (\boldsymbol{\beta_1}, \ldots, \boldsymbol{\beta_l}, \ldots, \boldsymbol{\beta_L})^T$ is the ($3L$-by-1) stacked vector of fixed effects, $\boldsymbol{D}$ is the ($2L$-by-$2L$) covariance matrix for random effects $\mathbf{b_i} = (\mathbf{b_{i,1}}, \ldots, \mathbf{b_{i,l}}, \ldots, \mathbf{b_{i,L}})^T$, and $\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma_l}, \ldots, \boldsymbol{\Sigma_L})$ is the (JL-by-JL) block diagonal matrix of residual variances.

Assuming an unstructured variance–covariance matrix $\mathbf{D_{l,l}}$ for lth QT, the variance at each visit time $t_{i,j}$ is $\mathrm{Var}(y_{i,j,l})$, and the covariance function $\mathrm{Cov}(y_{i,j,l}, y_{i,s,l})$ between 2 visit times $t_{i,j} \neq t_{i,s}$ is analogous to those defined above for 1 longitudinal QT and 1 TTE trait. The multivariate mixed model accounts for dependencies of each QT pair (l, m; $l \neq m$) via random-effects covariance functions in $\mathbf{D_{l,m}}$ where the covariance between observations of 2 QTs measured at $t_{i,j}$ is $\mathrm{Cov}(y_{i,j,l}, y_{i,j,m}) = t_{i,j}^2\mathrm{Cov}(b_{i,1,l}, b_{i,1,m}) + t_{i,j}(\mathrm{Cov}(b_{i,0,l}, b_{i,1,m}) + \mathrm{Cov}(b_{i,0,m}, b_{i,1,l})) + \mathrm{Cov}(b_{i,0,l}, b_{i,0,m})$, which is quadratic over time and the covariance between 2 longitudinal QTs measured at $t_{i,j} \neq t_{i,s}$ is $\mathrm{Cov}(y_{i,j,l}, y_{i,s,m}) = t_{i,j}t_{i,s}\mathrm{Cov}(b_{i,1,l}, b_{i,1,m}) + t_{i,j}\mathrm{Cov}(b_{i,1,l}, b_{i,0,m}) + t_{i,s}\mathrm{Cov}(b_{i,0,l}, b_{i,1,m}) + \mathrm{Cov}(b_{i,0,l}, b_{i,0,m})$. Thus, joint analysis of correlated longitudinal QTs is expected to improve power over separate analysis of each QT by borrowing information through implied dependency structures among the random effects.

Multivariate TTE submodel

Finally, we extend Equation 3 to a multivariate PH frailty TTE submodel, with a subject-specific random effect (frailty term, u_i) to capture unexplained dependencies among the TTE traits (e.g. due to unmeasured baseline shared factors). The frailty term is assumed to follow a nonnegative distribution (e.g. a gamma distribution which corresponds to a well-understood and widely used frailty model, with available closed form solution; Balan and Putter 2020). In Fig. 2 (Equation 6), $\lambda_{0,k}(t)$ and $\gamma_{g,k}$ correspond to the baseline hazard function and SNP effect on the kth TTE trait ($1 \leq k \leq K$), accounting for association of lth QT with the TTE trait k ($\alpha_{l,k}$, $1 \leq l \leq L$). Equation 6 (Fig. 2) can be expressed as

$$\lambda_{i,k}(t) = \lambda_{0,k}(t) \times \exp\{\boldsymbol{\alpha_k w_{i,k}}(t) + \gamma_{g,k}g_i + u_i\},$$

where $\boldsymbol{\alpha_k} = (\alpha_{1,k}, \ldots, \alpha_{l,k}, \ldots., \alpha_{L,k})$ is the vector of L QT effects on the kth TTE trait and $\boldsymbol{w_{i,k}}(t) = (w_{i,1,k}(t), \ldots., w_{i,l,k}(t), \ldots, w_{i,L,k}(t))^T$ specifies the corresponding QT association profile. We note $w_{i,l,k}(t) = f_{l,k}(y^*_{i,l}(t))$, where $y^*_{i,l}(t)$ denotes the lth QT trajectory value ($1 \leq l \leq L$) at time t, which depends on the fixed and random effects $\boldsymbol{\beta_l}$ and $\mathbf{b_{i,l}}$.

Comparisons with joint model of 1 longitudinal and 1 TTE trait

In the proposed extension for multiple longitudinal and multiple TTE traits, the direct, indirect, and overall SNP effects defined above for 1 longitudinal QT and 1 TTE trait are interpreted similarly. However, there are important practical differences between analysis of a QT–TTE trait pair and the multiple-trait extension. *First*, because the extension can account for multiple longitudinal QTs associated with 1 (or more) TTE(s), it improves SNP association inference and SNP classification accuracy, particularly when a TTE depends on multiple QTs. This is illustrated in our numerical experiments that follow below. *Second*, in the multivariate longitudinal submodel, the variance–covariance matrix $\boldsymbol{D}$ for the random effects specifies nonzero covariance terms in $\mathbf{D_{l,m}}$ for each pair of QTs ($1 \leq l \leq L$ and $1 \leq m \leq L$, $l \neq m$). In contrast, under the assumption of null covariance terms in $\mathbf{D_{l,m}}$ for all QT pairs, the multivariate submodel reduces to independent submodels for each longitudinal QT and thus, separate analyses of each QT. When longitudinal QTs are correlated, assuming null covariances can fail to make use of information borrowed through the $\mathbf{b_i}$ and reduces efficiency of the parameter estimates in the longitudinal trajectories (Shah *et al.* 1997; Jensen and Ritz 2018). This, in turn, can adversely affect estimation in the TTE model. *Third*, without a frailty term u_i, the multivariate TTE submodel (Equation 6, Fig. 2) reduces to separate submodels for each TTE trait. *Thus*, through use of a shared u_i, the extended joint model accounts for residual dependency among the K TTE traits, not explained by the covariates shared by the TTE submodels. *Overall*, the joint model for multiple longitudinal QTs and multiple TTE traits can improve inference by accounting for intermediate longitudinal QT(s) and their dependencies, as well as dependencies among the TTE traits, and thereby improve classification accuracy of direct and/or indirect SNP associations.

Implementation

Effect estimation and test statistic construction

To address computational obstacles in maximization of the joint likelihood and allow more flexible inference, we estimate the parameters using a 2-stage approach (see Appendix). We work within the original framework of Tsiatis and colleagues (Tsiatis *et al.* 1995; Wulfsohn and Tsiatis 1997; Dafni and Tsiatis 1998; Tsiatis and Davidian 2001, 2004); the latter specify conditions that guarantee consistent and asymptotically normal 2-stage estimators. In the spirit of subsequent authors (Ye *et al.* 2008; Yuen *et al.* 2018; Arisido *et al.* 2019), we fit a multivariate mixed model (Equation 5, Fig. 2) in stage 1; we use the *mvlme*() function from the R package JoineRML (Hickey *et al.* 2018c; version 0.4.2) to estimate the parameters of the longitudinal trajectories of the QTs and obtain fitted values of the smoothed trajectories. In stage 2, we fit a Cox PH frailty TTE model (Equation 6, Fig. 2) adjusting for functions of the smoothed trajectories as time-dependent covariates using the *coxph*() function from the R survival package (Therneau and Grambsch 2000; Therneau 2020; versions 3.2.7 and 3.2.13). We specify nonparametric baseline hazard functions for each TTE using the strata argument in *coxph*(). We assume a gamma frailty distribution, that is, $u_i \sim \Gamma(a, b)$ with a, $b > 0$. This assumption implies that dependence among the K TTE traits for each individual i is stronger for late events. As shown in the DCCT application, sensitivity to alternative specification of the frailty such as a Gaussian distribution can be addressed in *coxph*().To account for propagation of errors, due to uncertainty in stage 1 estimates not accounted for in stage 2 (Wulfsohn and Tsiatis 1997) and to empirically estimate the covariance matrix of SNP–QT trajectory ($\beta_{g,l}$) and SNP–TTE effects ($\gamma_{g,k}$), we apply a nonparametric bootstrap. The bootstrap also provides reliable standard error estimates in the joint TTE submodel needed when using an unspecified baseline hazard (Hsieh *et al.* 2006; Gould *et al.* 2015; Furgal *et al.* 2019). For each bootstrap sample b ($1 \leq b \leq B$, B is the total number of bootstrap repetitions), we generate a new data set by randomly sampling N individuals with replacement and then refit the joint model. We compute the empirical covariance matrix for the vector of all $\widehat{\beta_{g,l}}$, $\widehat{\gamma_{g,k}}$, and $\widehat{\alpha_{l,k}}$ using B bootstrap estimates. Wald statistics for each $\beta_{g,l}$ are

computed as $\widehat{Z_{\beta_g}} = \widehat{\beta_{g,l}}/se_{\beta_{g,l}}$ using the empirical bootstrap standard errors $se_{\beta_{g,l}}$ (to test H0: $\beta_{g,l} = 0$ vs. H1: $\beta_{g,l} \neq 0$), and similarly for each $\gamma_{g,k}$ as $\widehat{Z_{\gamma_g}} = \widehat{\gamma_{g,k}}/se_{\gamma_{g,k}}$ (to test H0: $\gamma_{g,k} = 0$ vs. H1: $\gamma_{g,k} \neq 0$).

In contrast to the 2-stage joint model, a conventional 1-stage analysis to assess whether a SNP–TTE trait association is independent from the QT–TTE association relies on regression adjustment using observed QT values as time-dependent covariates in a Cox-PH model (Paterson *et al.* 2010; Deng and Pan 2017). This approach, based only on the TTE model, does not provide information about SNP–QT effects ($\beta_{g,l}$) and interprets the SNP as having a direct association with the TTE trait when the test of SNP–TTE effect ($\gamma_{g,k}$) is declared significant, given the observed QT. Inference for $\alpha_{l,k}$ under this approach can be biased or inefficient when the QT is measured with random error or high within-subject variability (Faucett and Thomas 1996; Wulfsohn and Tsiatis 1997; Xu and Zeger 2001; Song *et al.* 2002; Brown and Ibrahim 2003), and inference for $\gamma_{g,k}$ may also be affected. Although estimates of $\beta_{g,l}$ obtained from mixed model QT analysis, fitted separately, may be used to distinguish between direct alone versus both direct and indirect SNP association, unlike the joint model, this conditional approach ignores measurement error in the observed QT values.

Procedure to classify direct and/or indirect SNP associations

In Table 1, we present a practical procedure to classify a SNP association with TTE trait k as direct and/or indirect, accounting for the SNP association with the lth longitudinal QT. This procedure requires 2 significance thresholds, $P^*_{\beta_g}$ and $P^*_{\gamma_g}$, for hypothesis tests of $\beta_{g,l}$ and $\gamma_{g,k}$ respectively, to be specified prior to the analysis and adjusted for the number of SNPs tested. Depending on the research question, we can choose thresholds for P^*_{β} and P^*_{γ} to be different or the same ($P^* = P^*_{\beta_g} = P^*_{\gamma_g}$). The latter is applicable, for instance, to systematically classify direct and/or indirect association for a set of M SNPs and the former to assess which SNPs, among those reported to be associated with the longitudinal risk factor, have a direct effect on a TTE trait. To our knowledge, no comparable procedure to classify direct and/or indirect SNP association based on SNP effect estimates from joint models has been proposed for studies with longitudinal QTs and TTE traits. A key feature of the proposed joint model extension to multiple longitudinal and multiple TTE traits is inference for SNP effects on each of the traits in an integrated statistical model, while accounting for within-subject QT variability and trait dependencies.

Table 1 . Procedure to classify a SNP as having an association with a TTE trait k, indirectly through an associated longitudinal QT risk factor l and/or directly with TTE trait k, based on hypothesis tests of SNP effects $\beta_{g,l}$ and $\gamma_{g,k}$.

		SNP association with the longitudinal QT risk factor l	
		$P_{\beta_{g,l}} \leq P^*_{\beta_g}$	$P_{\beta_{g,l}} > P^*_{\beta_g}$
SNP association with the TTE trait k	$P_{\gamma_{g,k}} \leq P^*_{\gamma_g}$	**Direct and indirect** ($\beta_{g,l} \neq 0$ and $\gamma_{g,k} \neq 0$)	**Direct** ($\beta_{g,l} = 0$ and $\gamma_{g,k} \neq 0$)
	$P_{\gamma_{g,k}} > P^*_{\gamma_g}$	**Indirect** ($\beta_{g,l} \neq 0$ and $\gamma_{g,k} = 0$)	**Not direct and not indirect** ($\beta_{g,l} = 0$ and $\gamma_{g,k} = 0$)

$P_{\beta_{g,l}}$ and $P_{\gamma_{g,k}}$ are P-values from Wald tests ($1df$) for each of SNP effects $\beta_{g,l}$ and $\gamma_{g,k}$. For the former, we test H0: $\beta_{g,l} = 0$ vs. H1: $\beta_{g,l} \neq 0$. For the latter, H0: $\gamma_{g,k} = 0$ vs. H1: $\gamma_{g,k} \neq 0$. $P^*_{\beta_g}$ and $P^*_{\gamma_g}$ are the corresponding classification thresholds. For example, if H0: $\beta_{g,l} = 0$ is rejected and H0: $\gamma_{g,k} = 0$ is rejected by the corresponding test statistics, then the SNP is classified as being both indirectly and directly associated with the TTE trait k. The classification procedure is based on the 2 separate test statistics and does not require a joint test statistic of the overall SNP effect; it thus partitions the 2D parameter space into 4 mutually exclusive quadrants. In the simulations and application, we use $P^* = P^*_{\beta_g} = P^*_{\gamma_g}$, although different thresholds can be specified for $P^*_{\beta_g}$ and $P^*_{\gamma_g}$.

The focus of the statistical simulation study and DCCT data application that follows is to evaluate the SNP classification procedure applied in extended joint model analysis. As demonstrated by the DCCT (The Diabetes Control and Complications Trial Research Group 1993), intensive insulin treatment to control the HbA1c level to a normal range prevents and delays progression of long-term T1DC. DCCT GWAS identified 2 SNPs associated with within-patient mean HbA1c at genome-wide significance in the conventional treatment arm; rs10810632 in *BNC2*/9p22.2 and rs1358030 near *SORCS1*/10q25.1 (Paterson *et al.* 2010). Additional SNP associations with DR and DN were reported with potential pleiotropic effects (Hosseini *et al.* 2015). Other longitudinal QTs are postulated to have genetic associations with T1DC, for example, association of systolic blood pressure (SBP) with DN. Because the goal of intensive therapy in DCCT was to reduce HbA1c into the nondiabetic range, which produced treatment differences in HbA1c values, we base our joint model evaluation and application on $N = 667$ unrelated individuals of European ancestry from the conventional treatment group (see Supplementary Files 1 and 2). We incorporate observed longitudinal measurements for HbA1c and SBP which were recorded irrespective of the occurrence of any complication event(s) at up to 39 quarterly visits.

Simulation study

Design of the DCCT-data-based simulation study

We generate $R = 1{,}000$ replicated data sets simulated under a *complex genetic architecture* informed by the DCCT Genetics Study data (Fig. 3), which involves $N = 667$ subjects from the conventional treatment group, $M = 5$ simulated causal SNPs with direct effects on $K = 2$ simulated time-to-T1DC (with ~54% DR events and ~25% DN events on average), and/or indirect effects via $L = 3$ longitudinal QTs. Two longitudinal QTs as measured in DCCT (HbA1c and SBP) and another simulated unmeasured QT (U) are designed to induce shared dependency among the T1DC traits. We assume effects of sex on SBP and effects of T1D duration (at baseline) on both T1DC traits, as estimated in the original DCCT data, and specify contemporaneous association structures for the association of HbA1c and SBP on T1DC traits. We specify effect sizes and minor allele frequencies (MAFs) of the causal SNPs, as well as other parameter values according to the DCCT Genetics Study and the T1DC literature (Fig. 3). For SBP and DN, we inflate the typical SNP effects reported in the literature to achieve power sufficient to detect SNP associations given the available DCCT sample size. Under the *global null* genetic scenario in which none of the SNPs are associated with any traits, we also simulate M SNPs with the same MAFs as the causal SNPs, independently of the traits.

Algorithm for realistic data generation under a complex genetic architecture

To generate a data structure that combines observed and simulated traits, we formulate a genotype–phenotype multiple-trait model including (1) $L = 3$ linear mixed models linking each SNP with an indirect effect to a longitudinal risk factor and (2) $K = 2$ nonindependent parametric TTE models depending on fitted

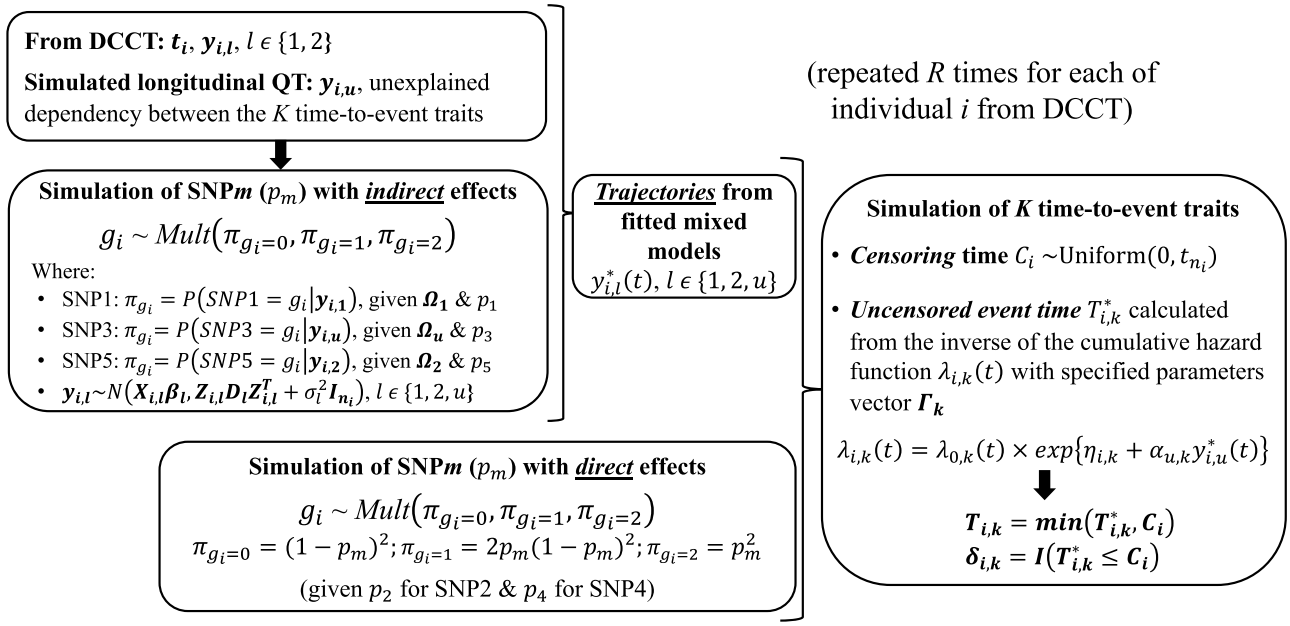

Fig. 4. Illustration of the procedure developed for DCCT-based simulation study under the scenario from Fig. 3. For each individual i, with $\{\mathbf{t}_i, \mathbf{y}_{i,1}, \mathbf{y}_{i,2}\}$ observed in DCCT, the algorithm simulates latent longitudinal QT values for U ($\mathbf{y}_{i,u}$), genetic data at M causal SNPs (with $\boldsymbol{p}$, the vector of p_m MAFs), and time-to-T1DC data ($T_{i,k}$, $\delta_{i,k}$) for $K = 2$ time-to-T1DC traits. The SNPs are simulated under Hardy–Weinberg and linkage equilibrium assumptions. SNPs with indirect effects are simulated from a multinomial distribution with calculated conditional genotype probabilities for individual i ($\pi_{g_i=0}$, $\pi_{g_i=1}$, $\pi_{g_i=2}$) based on $\mathbf{y}_{i,l}$. Each $\mathbf{y}_{i,l}$ is assumed to follow a multivariate normal distribution with $\mathbf{X}_{i,l}$ and $\mathbf{Z}_{i,l}$ the specified fixed and random-effect design matrices in longitudinal QT models and $\Omega_l = \{\beta_l, \mathbf{D}_l, \sigma_l^2\}$ the vector of specified parameter values for each lth QT. SNPs with direct effects are simulated from the population probabilities that depend only on their MAF. The specified hazard function for each kth TTE trait depends on the effects of the longitudinal QT trajectories and on the SNPs with direct effects in $\eta_{i,k}$, with $\eta_{i,1} = \gamma_{g,1}$ SNP2$_i$ + $\alpha_{1,1} y^*_{i,1}(t)$ for DR and $\eta_{i,2} = \gamma_{g,2}$SNP4$_i$ + $\gamma'_{g,2}$ SNP5$_i$ + $\alpha_{1,2}$ $y^*_{i,1}(t) + \alpha_{2,2}$ $y^*_{i,2}(t)$ for DN, as well as the effect of the shared latent QT trajectory $y^*_{i,u}(t)$ used to induce dependencies between the TTE traits. We define Γ_k as the vectors of specified parameter values for each kth TTE trait. The uncensored event time $T^*_{i,k}$ is simulated by calculating the inverse of the cumulative specified hazard function using the Brent univariate root-finding method (Brent 1973; Crowther and Lambert 2013). To simplify the exposition of the simulation procedure, we do not show the effects of sex on SBP or T1D duration on T1DC traits, but they are included in the data generating model; see Supplementary File 2 for details. Parameters for the *causal* genetic scenario are shown in Fig. 3 and Supplementary File 2 (section 4).

longitudinal QT trajectories and SNPs with direct effects. For each DCCT individual i with observed longitudinal measures for HbA1c and SBP and observed baseline covariates (sex and T1D duration), we simulate genotypes at M causal SNPs with MAF vector $\boldsymbol{p}$, longitudinal trait values $\mathbf{U}_i$, and $K = 2$ TTE traits ($(T_{i,k}, \delta_{i,k})$, $k = 1$ and 2 for DR and DN), using the algorithm illustrated in Fig. 4 and detailed in Supplementary File 2 (sections 1–5). All SNP genotypes are generated under Hardy–Weinberg and linkage equilibrium assumptions. SNPs with indirect effects through longitudinal QTs associated with DR and/or DN are generated from the observed (SNP1 and SNP5) or simulated (SNP3) QTs, while SNPs with direct effects (SNP2 and SNP4) are generated independently of the longitudinal QTs and are included in the hazard function used to generate each TTE trait (Fig. 4).

Scenario for DCCT-based complex genetic architecture

Overall, the simulated complex genetic architecture represents multiple types of SNP trait associations (Fig. 3): direct association with each T1DC trait (SNP2 and SNP4), indirect association with both T1DC traits via measured (SNP1) and unmeasured (SNP3) longitudinal QTs, and direct and indirect association via a measured longitudinal QT (SNP5); all longitudinal QTs exhibit within-subject random variability. Except for SNP3, all other SNP scenarios represent SNP association with a longitudinal QT (SNP1 and SNP5) or a TTE trait (SNP2, SNP4, and SNP5) testable in a single-trait GWAS. Particularly, SNP1, SNP3, and SNP5 have indirect effects on T1DC traits induced via intermediate QT(s), such that their associations with each T1DC trait are detectable in discovery analysis using Cox PH TTE models fitted separately for each TTE trait and ignoring the longitudinal QT(s) (Supplementary File 2, section 8). SNP1 corresponds to rs10810632 and rs1358030 associations reported in the motivating DCCT GWAS of HbA1c (Paterson *et al.* 2010), while SNP5 represents a strong signal that would be detected in separate GWAS analysis of each longitudinal QT and TTE trait.

Analysis of the simulated data

To evaluate statistical performance and classification accuracy of direct and/or indirect SNP associations under the complex genetic architecture outlined above, we assess the merits of extended joint model analysis over simpler methods that do not fully exploit the data structure. These alternative approaches include joint model analyses limited to 2 longitudinal QTs and 1 TTE trait, joint model analyses of 1 longitudinal QT and 1 TTE trait, and a conditional analysis of 2 TTE traits adjusted for observed values of 2 longitudinal QTs. In each replicated data set, each SNP is separately analyzed using the following methods:

- JM-cmp: a *completely* specified joint model analysis that includes observed (HbA1c and SBP) and unobserved (U) longitudinal QTs as well as baseline covariates (sex and T1D duration) used in the data simulation. Due to the latent nature of unobserved U, JM-cmp cannot be fitted in practice, but we include it as a benchmark for comparison against analysis ignoring the unobserved longitudinal U.

- JM-mis: includes the same variables as in JM-cmp but excludes U.
- JM-sep: analysis of 2 longitudinal QTs and 1 TTE trait that do not account for dependency between the TTE traits [where JM-sep($l = 1$ and 2; $k = 1$) denotes the joint model for DR and JM-sep($l = 1$ and 2; $k = 2$) the joint model for DN] and joint models of 1 longitudinal and 1 TTE trait that do not account for dependence between the longitudinal traits, nor between the TTE traits [referred to as JM-sep($l = 1$; $k = 1$), JM-sep($l = 1$; $k = 2$), and JM-sep($l = 2$; $k = 2$)]. Altogether, comparisons with JM-sep analyses assess the merits of the extended joint model methods JM-mis and JM-cmp.
- CM-obs: Cox PH frailty analysis that includes both TTE traits (DR and DN) and the same variables as in JM-mis but adjusts for the *observed* longitudinal QT values as time-dependent covariates; this corresponds to the conditional analysis approach mentioned above. Here, to classify SNPs as indirectly associated, we fit a linear mixed model to test the SNP effects on the QT(s) adjusted for the same covariates as used for the joint models. Comparisons of estimation, hypothesis testing, and classification results based on CM-obs to those based on JM-mis and JM-cmp assess the impact of within-subject QT variation/measurement errors on hypothesis testing and classification results for each SNP.

In each of these analyses, we compute empirical covariance matrices for effect estimates using 500 bootstrap iterations and construct large sample test statistics for SNP effects. Under 2-stage JM inference, we fit a bivariate ($l = 1$ and 2) or univariate ($l = 1$ or $l = 2$) longitudinal QT model and test $\beta_{g,l}$ for indirect SNP association in stage 1. Then in stage 2, we fit the specified TTE submodels with the QT trajectories from stage 1 and test $\gamma_{g,k}$ for direct SNP association. Given hypothesis test results for a pair of QT–TTE traits for each SNP, we apply the procedure defined in Table 1 to classify the SNP–TTE association. Although there are differences between single and multiple QT test results in stage 1, differences in classification arise largely through differences in test results in stage 2 TTE analysis. Under 2-stage and conditional independence assumptions (Appendix), $\widehat{\beta_{g,l}}$ and $\widehat{\gamma_{g,k}}$ and corresponding Z test statistics (Z_{β_g} and Z_{γ_g}) are expected to be uncorrelated under the *global null* scenario, but this may not necessarily hold under the *genetic alternative* scenario when the analysis model is misspecified or when the TTE estimation uses observed longitudinal QT values. We compute empirical correlations in each replicate under both scenarios.

Evaluation criteria

We compare type 1 error (T1E) and power of hypothesis tests of SNP–QT ($\beta_{g,l}$) and SNP–TTE ($\gamma_{g,k}$) association among the alternative analyses (JM-cmp, JM-mis, JM-sep, and CM-obs) at $P^* = 5$ and 1% critical values under the *global null* and *causal* genetic scenarios, analyzing each of the SNPs separately. We assess estimation accuracy using mean bias for $\widehat{\beta_{g,l}}$ and $\widehat{\gamma_{g,k}}$ and confidence interval coverage across replicates and similarly examine the distribution and mean of bootstrap standard errors and correlations for all the compared models.

For each SNP, we evaluate accuracy of SNP classification for each TTE trait as direct and/or indirect association (Table 1). Specifically, under the *global null* and *causal* genetic scenarios, we compare empirical and expected classification frequencies under the assumption of indirect and/or direct association built into the generating model. The empirical frequencies are tabulated from the distribution of simulation replicates in the 4 classification categories from Table 1 (direct, indirect, direct and indirect, and not direct and not indirect), specifying classification thresholds such that $P^*_\beta = P^*_\gamma = P^*$, with $P^* = 0.05$. To allow for potential dependence between the SNP effect tests, we calculate expected frequencies under the assumption that large sample statistics (Z_{β_g}, Z_{γ_g}), constructed from $\widehat{\beta_{g,l}}$ and $\widehat{\gamma_{g,k}}$, and their bootstrap variances follow a bivariate normal distribution with bootstrap covariances (see Supplementary File 2, section 9). We judge the classification procedure for a SNP association in each QT–TTE trait pair to have high accuracy when the empirical frequencies are consistent with those expected, and we compare accuracy among the different analysis models. We also assess empirical classification frequencies for causal SNPs with nonnull effects by increasing stringency of P^* up to 10^{-5}.

Results

Simulation study

SNP association test validity and power

Under the *global null* scenario of no genetic association with any of the traits, the T1E of SNP tests ($\beta_{g,l}$ and $\gamma_{g,k}$) is reasonably well controlled (Table 2), and P-values P_{β_g} and P_{γ_g} from the joint models show no departure from the expected large sample distributions (χ^2 with 1 degree of freedom, Supplementary File 2, section 7). Under the alternative multi-SNP *causal* genetic scenario (Table 2), T1Es for null $\beta_{g,l}$ are close to the nominal level of 5% for most analysis models (exceptions are SNP2 and SNP5), while tests of SNPs with effects on intermediate *measured* longitudinal QTs [SNP1 ($\beta_{g,1}$) and SNP5 ($\beta_{g,2}$)] reach 100% power for all the analysis models (Table 2).

Under the genetic *alternative* simulation scenario, T1E control and power in $\gamma_{g,k}$ tests are improved in JM-cmp and JM-mis compared to JM-sep($l = 1$ or 2; $k = 1$ or 2, i.e. for 1 QT–TTE pair) and CM-obs. Improvement in the multiple-trait models [including JM-sep($l = 1$ and 2; $k = 1$ or 2)] is largely explained by bias reduction and more efficient estimation (Supplementary File 2, section 6). The nonnull QT effect estimates in the TTE analysis ($\widehat{\alpha_{l,k}}$) are similarly improved. Both JM-sep (for a single QT–TTE pair) and CM-obs analysis exhibit elevated T1E and/or loss of power to detect SNP–TTE association and biased effect estimates when important QTs are not taken into account properly (Table 2 and Supplementary File 2, section 6). This includes SNP1–DR when HbA1c is ignored, SNP3–DR and DN when unmeasured U is ignored, as well as SNP4–DN and SNP5–DN when SBP is ignored. Overall, the relative ranking of empirical power among the analysis models persists for $\gamma_{g,k}$ tests across P^* varying down to 10^{-5} (Supplementary File 2, section 8) with steeper power reduction for SNP4-DN in JM-sep($l = 1$; $k = 2$).

Finally, we find little evidence for correlation between Z scores of $\widehat{\beta_{g,l}}$ and $\widehat{\gamma_{g,k}}$. The average bootstrap correlation $\widehat{\rho_{l,k}}$ across replicates is low for each QT–TTE trait pair (l; k) under the genetic *alternative* (Tables 3–7) and *global null* simulation scenarios (Supplementary File 2, section 9) for all joint model analyses ($|\widehat{\rho_{l,k}}| < 0.05$). However, we see larger $|\widehat{\rho_{l,k}}|$ values in CM-obs, particularly for SBP/DN, perhaps due to larger random variation in SBP. Advantages of the multiple-trait analyses (JM-cmp and JM-mis) for parameter estimation and hypothesis testing translate into improved performance of the classification procedure.

Classification of direct and/or indirect SNP associations with TTE traits

Under the *global null* scenario, empirical classification frequencies for direct and/or indirect SNP association with each T1DC trait at significance level $P^* = 0.05$ agree with those expected for all SNP association categories and all models (Supplementary File 2, section

Table 2. Empirical type-1 error and power (%) for SNP hypothesis tests of each of $\beta_{g,l}$ and $\gamma_{g,k}$ based on the complete joint model and compared models, assessed using $R = 1{,}000$ replicates of $N = 667$ DCCT subjects, with SNPs simulated under *global genetic null* and *genetic alternative* scenarios.

		Global null at $P^* = 5\%$				Genetic alternative at $P^* = 5\%$			
SNPs[a]	**Analysis models**	**HbA1c**	**SBP**	**DR**	**DN**	**HbA1c**	**SBP**	**DR**	**DN**
SNP1 (MAF = 30%)		$\beta_{g,1}=0$	$\beta_{g,2}=0$	$\gamma_{g,1}=0$	$\gamma_{g,2}=0$	$\beta_{g,1}=0.7$	$\beta_{g,2}=0$	$\gamma_{g,1}=0$	$\gamma_{g,2}=0$
	JM-cmp	5.2	5.8	4.4	4.8	**100**	5.1	4.1	5.4
	JM-mis	5.2	5.8	3.6	4.9	**100**	5.1	4.2	4.6
	JM-sep($l=1$ and 2; $k=1$)	5.2	5.8	3.8	.	**100**	5.1	4.7	.
	JM-sep($l=1$ and 2; $k=2$)	5.2	5.8	.	5.2	**100**	5.1	.	4.6
	JM-sep($l=1$; $k=1$)	5.6	.	4.0	.	**100**	.	4.6	.
	JM-sep($l=1$; $k=2$)	5.6	.	.	4.6	**100**	.	.	5.7
	JM-sep($l=2$; $k=2$)	.	5.9	.	4.6	.	5.2	.	11.2
SNP2 (MAF = 10%)		$\beta_{g,1}=0$	$\beta_{g,2}=0$	$\gamma_{g,1}=0$	$\gamma_{g,2}=0$	$\beta_{g,1}=0$	$\beta_{g,2}=0$	$\gamma_{g,1}=0.8$	$\gamma_{g,2}=0$
	JM-cmp	5.5	6.0	5.8	3.9	6.6	5.4	**100**	3.9
	JM-mis	5.5	6.0	4.5	3.8	6.6	5.4	**100**	4.6
	JM-sep($l=1$ and 2; $k=1$)	5.5	6.0	5.0	.	6.6	5.4	**100**	.
	JM-sep($l=1$ and 2; $k=2$)	5.5	6.0	.	4.4	6.6	5.4	.	5.2
	JM-sep($l=1$; $k=1$)	5.0	.	5.0	.	6.1	.	**100**	.
	JM-sep($l=1$; $k=2$)	5.0	.	.	5.1	6.1	.	.	4.6
	JM-sep($l=2$; $k=2$)	.	6.7	.	4.9	.	5.2	.	4.7
SNP3[a] (MAF = 40%)		$\beta_{g,1}=0$	$\beta_{g,2}=0$	$\gamma_{g,1}=0$	$\gamma_{g,2}=0$	$\beta_{g,1}=0$	$\beta_{g,2}=0$	$\gamma_{g,1}=0$	$\gamma_{g,2}=0$
	JM-cmp	4.0	6.7	4.3	4.5	4.5	4.6	4.7	3.8
	JM-mis	4.0	6.7	4.6	3.9	4.5	4.6	89.9	58.4
	JM-sep($l=1$ and 2; $k=1$)	4.0	6.7	4.7	.	4.5	4.6	89.7	.
	JM-sep($l=1$ and 2; $k=2$)	4.0	6.7	.	4.4	4.5	4.6	.	57.7
	JM-sep($l=1$; $k=1$)	4.1	.	4.6	.	4.8	.	89.7	.
	JM-sep($l=1$; $k=2$)	4.1	.	.	4.6	4.8	.	.	33.0
	JM-sep($l=2$; $k=2$)	.	6.5	.	4.8	.	4.4	.	56.4
SNP4 (MAF = 30%)		$\beta_{g,1}=0$	$\beta_{g,2}=0$	$\gamma_{g,1}=0$	$\gamma_{g,2}=0$	$\beta_{g,1}=0$	$\beta_{g,2}=0$	$\gamma_{g,1}=0$	$\gamma_{g,2}=0.7$
	JM-cmp	4.8	6.3	3.9	4.9	5.4	4.8	6.0	**100**
	JM-mis	4.8	6.3	4.9	4.7	5.4	4.8	5.4	**100**
	JM-sep($l=1$ and 2; $k=1$)	4.8	6.3	4.9	.	5.4	4.8	5.1	.
	JM-sep($l=1$ and 2; $k=2$)	4.8	6.3	.	5.1	5.4	4.8	.	**99.9**
	JM-sep($l=1$; $k=1$)	5.0	.	4.6	.	6.1	.	5.1	.
	JM-sep($l=1$; $k=2$)	5.0	.	.	5.0	6.1	.	.	**93.9**
	JM-sep($l=2$; $k=2$)	.	5.6	.	4.8	.	4.8	.	**99.9**
SNP5 (MAF = 20%)		$\beta_{g,1}=0$	$\beta_{g,2}=0$	$\gamma_{g,1}=0$	$\gamma_{g,2}=0$	$\beta_{g,1}=0$	$\beta_{g,2}=7.0$	$\gamma_{g,1}=0$	$\gamma_{g,2}=0.4$
	JM-cmp	5.0	5.5	4.7	5.6	2.8	**100**	5.5	**66.4**
	JM-mis	5.0	5.5	4.6	4.5	2.8	**100**	4.6	**64.5**
	JM-sep($l=1$ and 2; $k=1$)	5.0	5.5	5.2	.	2.8	**100**	4.8	.
	JM-sep($l=1$ and 2; $k=2$)	5.0	5.5	.	5.0	2.8	**100**	.	**63.2**
	JM-sep($l=1$; $k=1$)	5.1	.	5.3	.	2.8	.	4.6	.
	JM-sep($l=1$; $k=2$)	5.1	.	.	4.9	2.8	.	.	**100**
	JM-sep($l=2$; $k=2$)	.	5.6	.	5.4	.	**100**	.	**64.9**

Values are computed for each SNP and each genetic association parameter as the proportion of replicates that reject the null hypothesis at significance threshold $P^* = 0.05$. Results at other significance levels P^*, and also for CM-obs, are shown in Supplementary File 2, under the *global null* scenario at $P^* \leq 0.01$ (section 7), and under the genetic *alternative* scenario at $P^* \leq 10^{-5}$ (section 8). Power for each test of nonnull SNP effects on the trait simulated under the genetic alternative scenario (Fig. 3) is shown in bold.

[a] Except for JM-cmp, none of the analyses account for indirect SNP3 genetic pathways via the longitudinal risk factor U under the genetic *alternative* scenario; this translates into elevated T1E in the direct genetic effects tests of $\gamma_{g,k}$ for both time-to-T1DC traits.

9 and Table 9.2-2); this observation confirms accuracy of classification (empirical classification frequencies closer to those expected for all 4 categories of association). Empirical frequencies for the category of direct and indirect associations are less than those for direct or indirect categories because rejection of both single-parameter hypotheses in SNP tests of $\beta_{g,l}$ and $\gamma_{g,k}$ is required (Table 1).

Under the genetic *alternative* simulation scenario, when SNPs have direct and/or indirect effects via measured longitudinal QTs, the proposed multivariate joint models JM-cmp and JM-mis [and JM-sep($l = 1$ and 2; k with $k = 1$ or 2)] correctly classify the SNP associations for each QT–TTE pair in more than 88% of replicates for SNP1, SNP2, and SNP4 (Tables 3–5) and in more than 61% of replicates for SNP5 (Table 6) at specified significance level $P^* = 0.05$. Classification accuracy is improved relative to single-pair QT–TTE analysis: both JM-sep($l = 1$ or 2; $k = 1$ or 2) and CM-obs models exhibit larger differences between empirical and expected classification frequencies and lower ability to correctly distinguish between direct and/or indirect association and potential for classification error. This is clearly evident for SNPs with bias in $\gamma_{g,k}$ and low T1E control or power for $\gamma_{g,k}$ test in JM-sep($l = 1$ or 2; $k = 1$ or 2) and/or CM-obs analyses, notably for:

- SNP1 has indirect association with both T1DC traits via HbA1c (Table 3); empirical classification frequencies are lower than expected for CM-obs in the correct classification category with HbA1c/DR (indirect association) and for JM-sep ($l = 2$; $k = 2$) in the correct category with SBP/DN (not direct and not indirect association).
- SNP4 has a direct association with DN (Table 5): JM-sep($l = 1$; $k = 2$) has lower than expected empirical classification frequency for the correct direct association category with HbA1c/DN.
- SNP5 has direct and indirect effects on DN via SBP (Table 6): CM-obs has larger than expected empirical classification

Table 3. Classification frequencies for SNP1 association with each of the QT–TTE trait pairs based on the complete joint model and compared models at significance threshold $P^* = 0.05$, using $R = 1{,}000$ replicates of $N = 667$ DCCT subjects, with SNPs simulated under the genetic *alternative* scenario from Fig. 3.

SNP	Trait pairs	Analysis model	Mean bias[a]		Mean bootstrap SE's and correlation[b]			Classification frequencies (%) *Expected*[c] (above) vs. *empirical*[d] values			
SNP1 (MAF = 30%)	**HbA1c($l = 1$)/DR($k = 1$)**		$\beta_{g,1} = 0.7$	$\gamma_{g,1} = 0$	$\widehat{\sigma_{\beta_{g,1}}}$	$\widehat{\sigma_{\gamma_{g,1}}}$	$\widehat{\rho_{1,1}}$	**Indirect** **95.0**[c]	**Direct** **0**	**Direct and indirect** **5.0**	**Not Direct and not indirect** **0**
		JM-cmp	−0.051	0.002	0.006	0.009	−0.003	**95.9**[d]	0	4.1	0
		JM-mis	−0.051	<0.001	0.006	0.010	−0.002	**95.8**	0	4.2	0
		JM-sep($l = 1$ and 2; $k = 1$)	−0.051	<0.001	0.006	0.008	−0.002	**95.3**	0	4.7	0
		JM-sep($l = 1$; $k = 1$)	−0.047	<0.001	0.006	0.008	−0.002	**95.4**	0	4.6	0
		CM-obs	−0.051	0.032	0.006	0.009	0.041	**93.6**	0	6.4	0
	HbA1c($l = 1$)/DN($k = 2$)		$\beta_{g,1} = 0.7$	$\gamma_{g,2} = 0$	$\widehat{\sigma_{\beta_{g,1}}}$	$\widehat{\sigma_{\gamma_{g,2}}}$	$\widehat{\rho_{1,2}}$	**Indirect** **95.0**[c]	**Direct** **0**	**Direct and indirect** **5.0**	**Not Direct and not indirect** **0**
		JM-cmp	−0.051	−0.011	0.006	0.021	−0.012	**94.6**[d]	0	5.4	0
		JM-mis	−0.051	−0.014	0.006	0.024	−0.010	**95.4**	0	4.6	0
		JM-sep($l = 1$ and 2; $k = 2$)	−0.051	−0.011	0.006	0.021	−0.009	**95.4**	0	4.6	0
		JM-sep($l = 1$; $k = 2$)	−0.047	−0.004	0.006	0.019	−0.005	**94.3**	0	5.7	0
		CM-obs	−0.051	0.018	0.006	0.016	0.018	**94.3**	0	5.7	0
	SBP($l = 2$)/DN($k = 2$)		$\beta_{g,2} = 0$	$\gamma_{g,2} = 0$	$\widehat{\sigma_{\beta_{g,2}}}$	$\widehat{\sigma_{\gamma_{g,2}}}$	$\widehat{\rho_{2,2}}$	**Indirect** **4.75**	**Direct** **4.75**	**Direct and indirect** **0.25**	**Not Direct and not indirect** **90.25**[c]
		JM-cmp	−0.053	−0.011	0.233	0.021	−0.021	5.0	5.3	0.1	**89.6**[d]
		JM-mis	−0.053	−0.014	0.233	0.024	−0.012	5.1	4.6	0	**90.3**
		JM-sep($l = 1$ and 2; $k = 2$)	−0.053	−0.011	0.233	0.021	−0.009	5.1	4.6	0	**90.3**
		JM-sep($l = 2$; $k = 2$)	−0.053	0.095	0.232	0.019	−0.007	4.3	10.3	0.9	**84.5**
		CM-obs	−0.053	0.018	0.233	0.016	0.313	4.5	5.1	0.6	**89.8**

[a] Mean bias in the SNP1 effects, $\widehat{\beta_{g,l}}$ and $\widehat{\gamma_{g,k}}$, estimated by each of the analysis models.

[b] Mean bootstrap SE's, $\widehat{\sigma_{\beta_{g,l}}}$ and $\widehat{\sigma_{\gamma_{g,k}}}$, for respectively $\beta_{g,l}$ and $\gamma_{g,k}$, and mean correlation $\widehat{\rho_{l,k}}$ over the $R = 1{,}000$ replicates.

[c] *Expected* classification frequencies (%) for each category of association calculated as the probability to classify SNP1 association as Indirect, Direct, Direct and Indirect, or Not Direct and not Indirect, under the assumption that $\widehat{Z_g} = (\widehat{Z_{\beta_{g,l}}}, \widehat{Z_{\gamma_{g,k}}})^T$, constructed from the SNP effect estimates ($\widehat{\beta_{g,l}}$ and $\widehat{\gamma_{g,k}}$) and their bootstrap standard errors ($\widehat{\sigma_{\beta_{g,l}}}$ and $\widehat{\sigma_{\gamma_{g,k}}}$) as $\widehat{Z_{\beta_{g,l}}} = \widehat{\beta_{g,l}} / \widehat{\sigma_{\beta_{g,l}}}$ and $\widehat{Z_{\gamma_{g,k}}} = \widehat{\gamma_{g,k}} / \widehat{\sigma_{\gamma_{g,k}}}$ from JM-cmp, asymptotically follows a bivariate normal distribution, that is $\widehat{Z_g} \sim N_2(E[Z_g], H)$, where $E[Z_g]$ is the vector of expectations, and $H = \begin{pmatrix} 1 & \rho_{l,k} \\ \rho_{l,k} & 1 \end{pmatrix}$ with $\rho_{l,k} = \mathrm{Cor}(Z_{\beta_{g,l}}, Z_{\gamma_{g,k}}) = \sigma_{\beta_{g,l}\gamma_{g,k}} / (\sigma_{\beta_{g,l}} \times \sigma_{\gamma_{g,k}})$ and $\sigma_{\beta_{g,l}\gamma_{g,k}}$, the covariance between $\beta_{g,l}$ and $\gamma_{g,k}$ (see Supplementary File 2, section 9 for details). Values of expected classification frequencies are shown in bold italic in the row underneath the classification category labels.

[d] *Empirical* classification frequencies (shown as a percentage, %) for each SNP association under the genetic *alternative* simulation for $\beta_{g,l}$ and $\gamma_{g,k}$ hypothesis tests using the procedure described in Table 1 at $P^* = P^*_\beta = P^*_\gamma = 0.05$. The results for the correct classification category under the genetic *alternative* simulation scenario from Fig. 3 are shown in bold. Empirical classification frequencies at significance levels $P^* \leq 10^{-5}$, and their 95% confidence intervals are presented in Supplementary File 2 (section 11).

Table 4. Classification frequencies for SNP2 association with each QT–TTE trait pairs based on the complete joint model and compared models at significance threshold $P^* = 0.05$, using $R = 1{,}000$ replicates of $N = 667$ DCCT subjects, with SNPs simulated under the genetic *alternative* scenario from Fig. 3.

SNP	Trait pairs	Analysis model	Mean bias[a]		Mean bootstrap SE's and correlation[b]			Classification frequencies[c] (%) *Expected* (above) vs. *empirical* values			
SNP2 (MAF = 10%)	**HbA1c(l = 1)/DR(k = 1)**		$\beta_{g,1} = 0$	$\gamma_{g,1} = 0.8$	$\widehat{\sigma_{\beta_{g,1}}}$	$\widehat{\sigma_{\gamma_{g,1}}}$	$\widehat{\rho_{1,1}}$	**Indirect** **0**	**Direct** **95.0**	**Direct and indirect** **5.0**	**Not Direct and not indirect** **0**
		JM-cmp	−0.003	0.003	0.014	0.015	−0.005	0	**93.4**	6.6	0
		JM-mis	−0.003	−0.023	0.014	0.019	−0.006	0	**93.4**	6.6	0
		JM-sep(l = 1 and 2; k = 1)	−0.003	−0.099	0.014	0.015	−0.007	0	**93.4**	6.6	0
		JM-sep(l = 1; k = 1)	−0.003	−0.099	0.014	0.015	−0.006	0	**93.9**	6.1	0
		CM-obs	−0.003	−0.077	0.014	0.017	0.038	0	**93.4**	6.6	0
	HbA1c(l = 1)/DN(k = 2)		$\beta_{g,1} = 0$	$\gamma_{g,2} = 0$	$\widehat{\sigma_{\beta_{g,1}}}$	$\widehat{\sigma_{\gamma_{g,2}}}$	$\widehat{\rho_{1,2}}$	**Indirect** **4.75**	**Direct** **4.75**	**Direct and indirect** **0.25**	**Not Direct and not indirect** **90.25**
		JM-cmp	−0.003	<0.001	0.014	0.043	−0.010	6.4	3.7	0.2	**89.7**
		JM-mis	−0.003	−0.003	0.014	0.048	−0.011	6.3	4.3	0.3	**89.1**
		JM-sep(l = 1 and 2; k = 2)	−0.003	0.002	0.014	0.042	−0.011	6.2	4.8	0.4	**88.6**
		JM-sep(l = 1; k = 2)	−0.003	−0.006	0.011	0.038	0.011	6.1	4.6	0	**89.3**
		CM-obs	−0.003	−0.004	0.014	0.034	0.023	6.6	4.5	0	**88.9**
	SBP(l = 2)/DN(k = 2)		$\beta_{g,2} = 0$	$\gamma_{g,2} = 0$	$\widehat{\sigma_{\beta_{g,2}}}$	$\widehat{\sigma_{\gamma_{g,2}}}$	$\widehat{\rho_{2,2}}$	**Indirect** **4.75**	**Direct** **4.75**	**Direct and indirect** **0.25**	**Not Direct and not indirect** **90.25**
		JM-cmp	−0.004	<0.001	0.513	0.043	−0.018	5.4	3.9	0	**90.7**
		JM-mis	−0.004	−0.003	0.513	0.048	−0.012	5.4	4.6	0	**90.0**
		JM-sep(l = 1 and 2; k = 2)	−0.004	0.002	0.513	0.042	−0.008	5.3	5.1	0.1	**89.5**
		JM-sep(l = 2; k = 2)	−0.004	0.001	0.512	0.042	−0.005	5.0	4.5	0.2	**90.3**
		CM-obs	−0.004	−0.004	0.513	0.034	0.330	5.0	4.1	0.4	**90.5**

[a] Mean bias in the SNP2 effects, $\widehat{\beta_{g,l}}$ and $\widehat{\gamma_{g,k}}$, estimated by each of the analysis models.

[b] Mean bootstrap SE's, $\widehat{\sigma_{\beta_{g,l}}}$ and $\widehat{\sigma_{\gamma_{g,k}}}$, for respectively $\beta_{g,l}$ and $\gamma_{g,k}$, and their correlation $\widehat{\rho_{l,k}}$ over the R = 1,000 replicates.

[c] *Expected* and *empirical* frequencies (shown as percentage, %) to classify SNP2 association as Indirect, Direct, Direct and Indirect, or Not Direct and not Indirect, calculated as described in the footnote of Table 3. *Empirical* classification frequencies at significance levels $P^* \leq 10^{-5}$, and their 95% confidence intervals are presented in Supplementary File 2 (section 11). Values of the *expected* classification frequencies are shown in bold italic in the row underneath the classification category labels; values of the *empirical* classification frequencies that correspond to the correct classification category under the simulated *genetic alternative* scenario are shown in bold.

Table 5. Classification frequencies for SNP4 association with each pair of QT–TTE traits based on the complete joint model and compared models at significance threshold $P^* = 0.05$, using $R = 1{,}000$ replicates of $N = 667$ DCCT subjects, with SNPs simulated under the genetic *alternative* scenario from Fig. 3.

SNP	Trait pairs	Analysis model	Mean bias[a]		Mean bootstrap SE's and correlation[b]			Classification frequencies[c] (%) *Expected* (above) vs. *empirical* values			
SNP4 (MAF = 30%)	**HbA1c(l = 1)/DR(k = 1)**		$\beta_{g,1} = 0$	$\gamma_{g,1} = 0$	$\widehat{\sigma_{\beta_{g,1}}}$	$\widehat{\sigma_{\gamma_{g,1}}}$	$\widehat{\rho_{1,1}}$	**Indirect** **4.75**	**Direct** **4.75**	**Direct and indirect** **0.25**	**Not Direct and not indirect** **90.25**
		JM-cmp	0.001	−0.001	0.006	0.007	<0.001	4.9	5.5	0.5	**89.1**
		JM-mis	0.001	0.001	0.006	0.009	−0.003	5.0	5.0	0.4	**89.6**
		JM-sep(l = 1 and 2; k = 1)	0.001	0.001	0.006	0.007	−0.003	5.1	4.8	0.3	**89.8**
		JM-sep(l = 1; k = 1)	0.001	0.001	0.006	0.007	−0.002	5.7	4.7	0.4	**89.2**
		CM-obs	0.001	0.001	0.006	0.008	0.038	5.0	4.2	0.4	**90.4**
	HbA1c(l = 1)/DN(k=2)		$\beta_{g,1} = 0$	$\gamma_{g,2} = 0.7$	$\widehat{\sigma_{\beta_{g,1}}}$	$\widehat{\sigma_{\gamma_{g,2}}}$	$\widehat{\rho_{1,2}}$	**Indirect** **0**	**Direct** **94.98**	**Direct and indirect** **5.0**	**Not Direct and not indirect** **0.02**
		JM-cmp	0.001	−0.014	0.006	0.017	−0.010	0	**94.6**	5.4	0
		JM-mis	0.001	−0.028	0.006	0.019	−0.010	0	**94.6**	5.4	0
		JM-sep(l = 1 and 2; k = 2)	0.001	−0.076	0.006	0.017	−0.009	0	**94.5**	5.4	0.1
		JM-sep(l = 1; k = 2)	0.001	−0.281	0.006	0.015	0.007	0.3	**88.1**	5.8	5.8
		CM-obs	0.001	−0.216	0.006	0.013	0.025	0	**93.5**	5.4	1.1
	SBP(l = 2)/DN(k=2)		$\beta_{g,2} = 0$	$\gamma_{g,2} = 0.7$	$\widehat{\sigma_{\beta_{g,2}}}$	$\widehat{\sigma_{\gamma_{g,2}}}$	$\widehat{\rho_{2,2}}$	**Indirect** **0**	**Direct** **94.98**	**Direct and indirect** **5.0**	**Not Direct and not indirect** **0.02**
		JM-cmp	0.012	−0.014	0.215	0.017	−0.026	0	**95.2**	4.8	0
		JM-mis	0.012	−0.028	0.215	0.019	−0.021	0	**95.2**	4.8	0
		JM-sep(l = 1 and 2; k = 2)	0.012	−0.076	0.215	0.017	−0.020	0	**95.1**	4.8	0.1
		JM-sep(l = 2; k = 2)	0.012	−0.088	0.215	0.016	−0.019	0	**95.1**	4.8	0.1
		CM-obs	0.012	−0.216	0.215	0.013	0.332	0.1	**94.2**	4.7	1.0

[a] Mean bias in the SNP4 effects, $\widehat{\beta_{g,l}}$ and $\widehat{\gamma_{g,k}}$, estimated by each of the analysis models.

[b] Mean bootstrap SE's, $\widehat{\sigma_{\beta_{g,l}}}$ and $\widehat{\sigma_{\gamma_{g,k}}}$, for respectively $\beta_{g,l}$ and $\gamma_{g,k}$, and their correlation $\widehat{\rho_{l,k}}$ over the $R = 1{,}000$ replicates.

[c] *Expected* and *empirical* frequencies (shown as percentage, %) to classify SNP4 association as Indirect, Direct, Direct and Indirect, or Not Direct and not Indirect, calculated as described in the footnote of Table 3. Empirical classification frequencies at significance levels $P^* \leq 10^{-5}$, and their 95% confidence intervals are presented in Supplementary File 2 (section 11). Values of the *expected* classification frequencies are shown in bold italic on the row underneath the classification category labels; values of the *empirical* classification frequencies that correspond to the correct classification category under the simulated genetic alternative scenario are shown in bold.

Table 6. Classification frequencies for SNP5 association with each pair of QT–TTE traits based on the complete joint model and compared models at significance threshold $P^* = 0.05$, using $R = 1{,}000$ replicates of $N = 667$ DCCT subjects, with SNPs simulated under the genetic *alternative* scenario from Fig. 3.

SNP	Trait pairs	Analysis model	Mean bias[a]		Mean bootstrap SE's and correlation[b]			Classification frequencies[c] (%) *Expected* (above) vs. *empirical* values			
SNP5 (MAF = 20%)	**HbA1c(l = 1)/DR(k = 1)**		$\beta_{g,1} = 0$	$\gamma_{g,1} = 0$	$\widehat{\sigma_{\beta_{g,1}}}$	$\widehat{\sigma_{\gamma_{g,1}}}$	$\widehat{\rho_{1,1}}$	**Indirect** 4.75	**Direct** 4.75	**Direct and indirect** 0.25	**Not Direct and not indirect** 90.25
		JM-cmp	−0.021	<0.001	0.010	0.012	0.001	2.4	5.1	0.4	**92.1**
		JM-mis	−0.021	0.017	0.010	0.015	<0.001	2.4	4.2	0.4	**93.0**
		JM-sep(l = 1 and 2; k = 1)	−0.021	0.013	0.010	0.012	<0.001	2.4	4.4	0.4	**92.8**
		JM-sep(l = 1; k = 1)	−0.019	0.013	0.010	0.012	−0.003	2.6	4.4	0.2	**92.8**
		CM-obs	−0.021	0.019	0.010	0.013	0.043	2.4	4.4	0.4	**92.8**
	HbA1c(l = 1)/DN(k = 2)		$\beta_{g,1} = 0$	$\gamma_{g,2} = 0.4$	$\widehat{\sigma_{\beta_{g,1}}}$	$\widehat{\sigma_{\gamma_{g,2}}}$	$\widehat{\rho_{1,2}}$	**Indirect** 1.44	**Direct** 67.67	**Direct and indirect** 3.56	**Not Direct and not indirect** 27.33
		JM-cmp	−0.021	−0.020	0.010	0.025	−0.016	1.1	**64.7**	1.7	32.5
		JM-mis	−0.021	−0.009	0.010	0.029	−0.013	0.8	**62.5**	2.0	34.7
		JM-sep(l = 1 and 2; k = 2)	−0.021	−0.036	0.010	0.026	−0.014	1.2	**61.6**	1.6	35.6
		JM-sep(l = 1; k = 2)	−0.019	**0.765**	0.010	0.021	−0.008	0	**97.2**	2.8	0
		CM-obs	−0.021	**0.505**	0.010	0.022	0.017	0	**97.2**	2.8	0
	SBP(l = 2)/DN(k = 2)		$\beta_{g,2} = 7.0$	$\gamma_{g,2} = 0.4$	$\widehat{\sigma_{\beta_{g,2}}}$	$\widehat{\sigma_{\gamma_{g,2}}}$	$\widehat{\rho_{2,2}}$	**Indirect** 28.77	**Direct** 0	**Direct and indirect** 71.23	**Not Direct and not indirect** 0
		JM-cmp	0.046	−0.020	0.385	0.025	−0.050	33.6	0	**66.4**	0
		JM-mis	0.046	−0.009	0.385	0.029	−0.047	35.5	0	**64.5**	0
		JM-sep(l = 1 and 2; k = 2)	0.046	−0.036	0.385	0.026	−0.047	36.8	0	**63.2**	0
		JM-sep(l = 2; k = 2)	0.039	−0.029	0.384	0.025	−0.041	35.1	0	**64.9**	0
		CM-obs	0.046	**0.505**	0.385	0.022	0.280	0	0	**100**	0

[a] Mean bias in the SNP5 effects, $\widehat{\beta_{g,l}}$ and $\widehat{\gamma_{g,k}}$, estimated by each of the analysis models.

[b] Mean bootstrap SE's, $\widehat{\sigma_{\beta_{g,l}}}$ and $\widehat{\sigma_{\gamma_{g,k}}}$, for respectively $\beta_{g,l}$ and $\gamma_{g,k}$, and their correlation $\widehat{\rho_{l,k}}$ over the $R = 1{,}000$ replicates.

[c] *Expected* and *empirical* frequencies (shown as percentage, %) to classify SNP5 association as Indirect, Direct, Direct and Indirect, or Not Direct and not Indirect, calculated as described in the footnote of Table 3. Empirical classification frequencies at significance levels $P^* \leq 10^{-5}$, and their 95% confidence intervals are presented in Supplementary File 2 (section 11). Values of expected classification frequencies are shown in bold italic in the row underneath the classification category labels; values of empirical classification frequencies that correspond to the correct classification category under the simulated genetic *alternative* scenario are shown in bold.

Table 7. Classification frequencies for SNP3 association with each pair of QT–TTE traits based on the complete joint model and compared models at significance threshold $P^* = 0.05$, using $R = 1{,}000$ replicates of $N = 667$ DCCT subjects, with SNPs simulated under the genetic *alternative* scenario from Fig. 3.

SNP	Trait pairs	Analysis model	Mean bias[a]		Mean bootstrap SE's and correlation[b]			Classification frequencies[c] (%) *Expected* (above) vs. *empirical* values			
SNP3[d] (MAF = 40%)	**HbA1c($l = 1$)/DR($k = 1$)**		$\beta_{g,1} = 0$	$\gamma_{g,1} = 0$	$\widehat{\sigma_{\beta_{g,1}}}$	$\widehat{\sigma_{\gamma_{g,1}}}$	$\widehat{\rho_{l,1}}$	**Indirect** **4.75**	**Direct** **4.75**	**Direct and indirect** **0.25**	**Not Direct and not indirect** **90.25**
		JM-cmp	−0.012	0.001	0.006	0.009	−0.001	4.3	4.5	0.2	**91.0**
		JM-mis	−0.012	0.299	0.006	0.009	−0.004	0.7	86.1	3.8	**9.4**
		JM-sep($l = 1$ and 2; $k = 1$)	−0.012	0.274	0.006	0.007	−0.004	0.6	85.8	3.9	**9.7**
		JM-sep($l = 1$; $k = 1$)	−0.014	0.274	0.006	0.007	−0.003	0.5	85.4	4.3	**9.8**
		CM-obs	−0.012	0.280	0.006	0.008	0.035	0.7	85.5	3.8	**10.0**
	HbA1c($l = 1$)/DN(k=2)		$\beta_{g,1} = 0$	$\gamma_{g,2} = 0$	$\widehat{\sigma_{\beta_{g,1}}}$	$\widehat{\sigma_{\gamma_{g,2}}}$	$\widehat{\rho_{1,2}}$	**Indirect** **4.75**	**Direct** **4.75**	**Direct and indirect** **0.25**	**Not Direct and not indirect** **90.25**
		JM-cmp	−0.012	0.004	0.006	0.021	−0.012	4.5	3.8	0	**91.7**
		JM-mis	−0.012	0.305	0.006	0.021	−0.016	1.7	55.6	2.8	**39.9**
		JM-sep($l = 1$ and 2; $k = 2$)	−0.012	0.291	0.006	0.019	−0.016	1.6	54.8	2.9	**40.7**
		JM-sep($l = 1$; $k = 2$)	−0.014	0.197	0.006	0.017	0.005	3.2	31.4	1.6	**63.8**
		CM-obs	−0.012	0.228	0.006	0.015	0.019	2.3	46.0	2.2	**49.5**
	SBP($l = 2$)/DN($k = 2$)		$\beta_{g,2} = 0$	$\gamma_{g,2} = 0$	$\widehat{\sigma_{\beta_{g,2}}}$	$\widehat{\sigma_{\gamma_{g,2}}}$	$\widehat{\rho_{2,2}}$	**Indirect** **4.75**	**Direct** **4.75**	**Direct and indirect** **0.25**	**Not Direct and not indirect** **90.25**
		JM-cmp	0.157	0.004	0.211	0.021	−0.017	4.4	3.6	0.2	**91.8**
		JM-mis	0.157	0.305	0.211	0.021	−0.017	1.8	55.6	2.8	**39.8**
		JM-sep($l = 1$ and 2; $k = 2$)	0.157	0.291	0.211	0.019	−0.015	1.8	54.9	2.8	**40.5**
		JM-sep($l = 2$; $k = 2$)	0.156	0.282	0.212	0.018	−0.009	1.8	53.8	2.6	**41.8**
		CM-obs	0.157	0.228	0.211	0.015	0.316	1.4	45.0	3.2	**50.4**

[a] Mean bias in the SNP3 effects, $\widehat{\beta_{g,l}}$ and $\widehat{\gamma_{g,k}}$, estimated by each of the analysis models.

[b] Mean bootstrap SE's, $\widehat{\sigma_{\beta_{g,l}}}$ and $\widehat{\sigma_{\gamma_{g,k}}}$, for respectively $\beta_{g,l}$ and $\gamma_{g,k}$ and their correlation $\widehat{\rho_{l,k}}$ over the $R = 1{,}000$ replicates.

[c] *Expected* and *empirical* frequencies (shown as percentage, %) to classify SNP3 association as Indirect, Direct, Direct and Indirect, or Not Direct and not Indirect calculated as described in the footnote of Table 3. Empirical classification frequencies at significance levels $P^* \leq 10^{-5}$, and their 95% confidence intervals are presented in Supplementary File 2 (section 11). Values of the *expected* classification frequencies are shown in bold italic in the row underneath the classification category labels; values of the *empirical* classification frequencies that correspond to the correct classification category under the simulated genetic alternative scenario are shown in bold.

[d] All the models, except JM-cmp, do not account for the indirect genetic pathways via the intermediate longitudinal risk factor U. This translates into a large bias in the direct genetic effects $\gamma_{g,k}$ on both time-to-T1DC traits and increased misclassification rates for SNP3 as a direct association.

frequencies for the correct classification category with HbA1c/DN (direct association) and with SBP/DN (direct and indirect association).

On the other hand, the presence of an unmeasured QT is misleading for all practical analysis methods: SNP3 has indirect effects on both T1DC traits via the unmeasured longitudinal QT (U) but no effects through HbA1c or SBP. Except for JM-cmp which exhibits accurate (and high) classification frequencies for HbA1c and SBP as neither direct nor indirect, all the other compared models (which ignore U) exhibit poor classification accuracy (Table 7) and tend to mistakenly classify SNP3 as a direct association with each T1DC trait in ∼30–86% of the replicates. As stringency of P^* increases up to 10^{-5} for the same effect sizes, empirical classification frequencies decrease in the correct classification category, while misclassification increases in the other categories due to reduced power; for example, JM-cmp tends to mistakenly classify SNP5 as an indirect association with SBP/DN (Supplementary File 2, section 11).

In summary, for SNPs directly associated with a TTE trait or indirectly associated via a measured longitudinal QT, our simulations show that by using smoothed trajectories and accounting for all QT associations, the extended joint model improves parameter inference and classification accuracy compared to CM-obs or JM-sep of only 1 QT–TTE pair. Reduced classification accuracy translates to increased risk of under- or over-classifying a SNP as direct and/or indirect association [e.g. under-classification of SNP4 and SNP5 associations with HbA1c/DN in JM-sep($l = 1$; $k = 2$) as direct associations and over-classification of SNP5 associations with SBP/DN as direct and indirect associations in CM-obs]. In addition, when a SNP has both direct and indirect effects on a TTE trait, such as for SNP5, classification frequencies can be low when $\gamma_{g,k}$ test power is low (where power depends on effect size, MAF, and trait distribution). As a result, a SNP with both direct and indirect associations can be misclassified as an indirect association. Finally, when a SNP has an indirect effect on both T1DC traits via an unmeasured QT, as for SNP3, the testing procedure based on JM-mis that captures some of the unmodeled dependency between TTE traits through the frailty term does not prevent misclassification as a direct association. This observation also demonstrates the importance of the joint model extension which allows analysis of all the intermediate QT(s), as opposed to JM-sep, and reduces misclassification of direct and/or indirect SNP associations.

Application in the DCCT Genetics Study data

We demonstrate feasibility of the 2-stage extended joint model method by analysis in the DCCT individuals from the conventional treatment group genotyped on HumanCoreExome Bead Array (Illumina, San Diego, CA, USA) with ungenotyped autosomal SNPs imputed using 1,000 genomes data phase 3 (The 1000 Genomes Project Consortium *et al.* 2015), as detailed in Supplementary File 1. We use time to mild retinopathy and time to persistent microalbuminuria, for DR and DN outcomes respectively, as previously defined in the motivating GWAS of HbA1c (Paterson *et al.* 2010); see Supplementary File 3, section 1. We analyze $N = 516$ conventionally treated subjects, without mild to moderate nonproliferative retinopathy or DN at DCCT baseline. By the time of the DCCT close-out visit, 297 (57.6%) had a DR event, and 61 (11.8%) had a DN event, including 47 subjects (9.1%) that experienced both events. After SNP filtering and pruning on linkage disequilibrium (Supplementary File 3, section 2), we analyze 307 SNPs reported as associated with HbA1c, SBP, or multiple definitions of DR and/or DN (Paterson *et al.* 2010; Grassi *et al.* 2011; Sandholm *et al.* 2012; Hosseini *et al.* 2015; Wheeler *et al.* 2017; Evangelou *et al.* 2018; Pollack *et al.* 2019); see Supplementary File 4 for the full list of SNPs.

Joint model fitting

We fit the joint model for each SNP at a time, including baseline covariates, longitudinal HbA1c and SBP, and T1DCs DR and DN (Supplementary File 3, section 3). In the Cox PH frailty TTE submodel, PH assumptions are well satisfied when the baseline hazard is stratified on the cohort variable; overall conclusions are equivalent when cohort is included as a covariate. Given prior evidence for long-term HbA1c effects on T1DC, we present results under the time-weighted cumulative specification for HbA1c association with time-to-T1DC traits, which exhibits stronger prior association with T1DC in the DCCT individuals analyzed here, but we obtained similar classification results under alternative specifications for HbA1c association (i.e. contemporaneous and updated mean, Supplementary File 3, section 4). Application of diagnostic tools in the joint model analysis, including residual analysis in the longitudinal and TTE submodels, finds little evidence for model misspecification (Supplementary File 3, section 5). As shown, martingale residuals are consistent with assumption of linear relationships between QTs and each time-to-T1DC trait in the Cox PH frailty model. Conclusions are unchanged when we assume a Gaussian rather than a gamma frailty distribution.

In Fig. 5a, we show that rs1358030 and rs10810632, discovered in a previous GWAS of HbA1c in DCCT (Paterson *et al.* 2010), are classified as indirect associations with both T1DC traits via their shared association with HbA1c ($P_{\beta_g} \leq P^*$ and $P_{\gamma_g} > P^*$, $P^* = 1.7 \times 10^{-4}$ after Bonferroni correction for the 289.02 effective SNP tests; Li and Ji 2005). Although the other candidate SNPs were selected from larger meta-analysis in T1D and/or independent studies in general populations (Paterson *et al.* 2010; Grassi *et al.* 2011; Sandholm *et al.* 2012; Hosseini *et al.* 2015; Wheeler *et al.* 2017; Evangelou *et al.* 2018; Pollack *et al.* 2019), the SNP trait association tests do not reach Bonferroni-corrected significance thresholds in our analysis (Supplementary File 3, section 4), potentially due to effect heterogeneity and low power to detect these variants or by inherent heterogeneity in the study designs and/or phenotypic definitions. Thus, classification for these candidate SNPs is uninformative.

Sample size and power considerations

Compared to the simulation, the number of DN events is lower in the DCCT data application. We thus expect lower power to detect direct SNP association with DN; this implies reduced accuracy to distinguish between indirect, direct, and both direct and indirect association. To assess how genetic association and classification for rs10810632 and rs1358030 may be affected by increasing sample size, we applied parametric resampling (see Supplementary File 3, section 6) to draw data sets with sample size up to 5 times the DCCT sample of $N = 516$, and we then extrapolated the classification beyond $N = 2580$. This analysis demonstrates decreasing variances of the SNP effects for both SNPs as sample size increases and narrowing of their 95 and 99% confidence intervals (Supplementary File 3, section 6); Fig. 6a illustrates the corresponding shift in classification of rs10810632 association with DR as indirect via HbA1c toward classification as both indirect and direct. On the other hand, given that both SNPs were previously discovered in a GWAS of HbA1c in DCCT, SNP effect estimates for HbA1c association ($\widehat{\beta_{g,1}}$) may be overestimated due to Winner's curse bias (Kraft 2008; Sun *et al.* 2011), which would

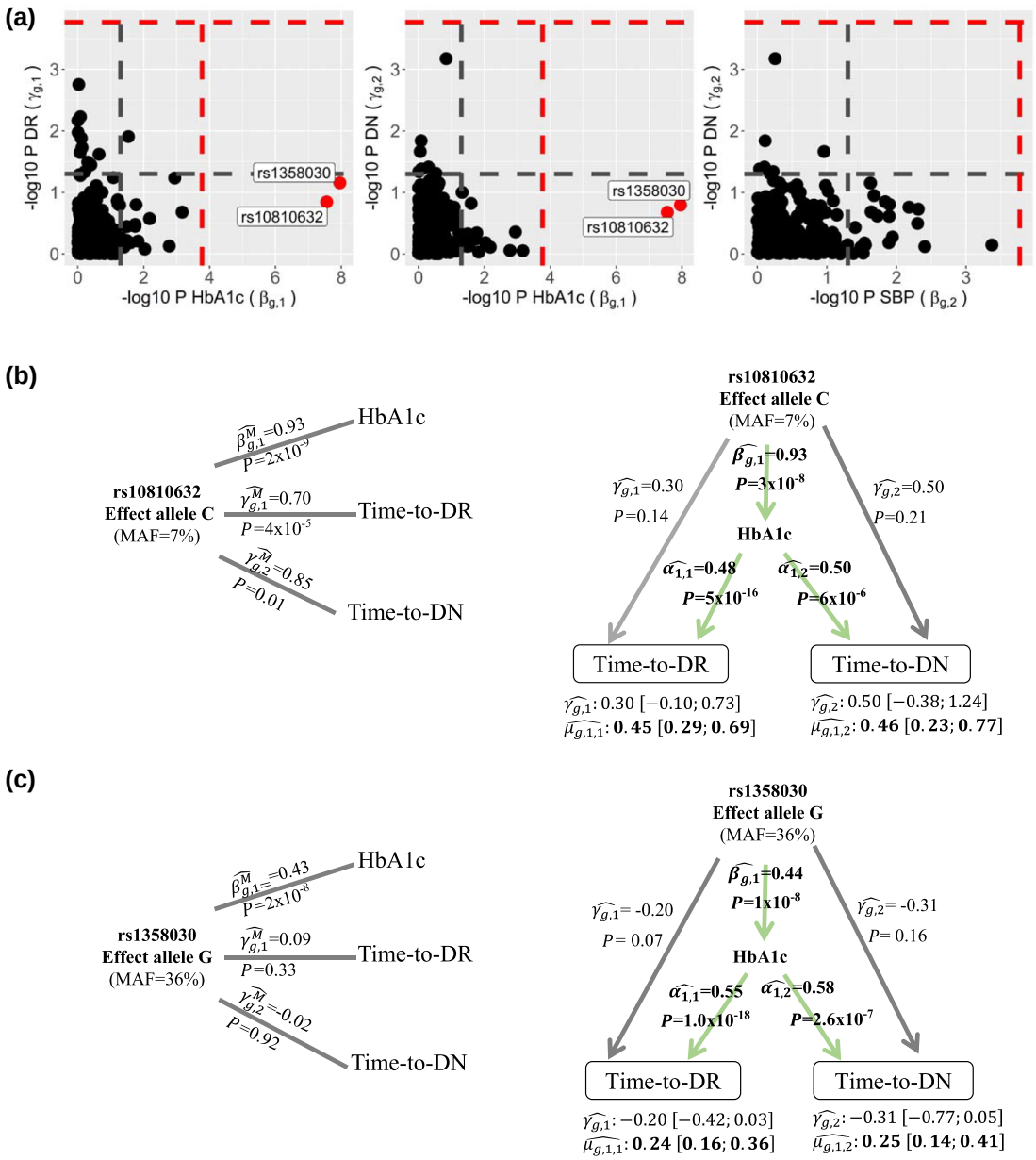

Fig. 5. Classification of direct and/or indirect SNP associations in the DCCT Genetics Study data. a) Scatter plots of the P-values (-log10) for tests of $\beta_{g,1}$ (H0: $\beta_{g,1} = 0$ vs. H1: $\beta_{g,1} \neq 0$) on the X axis and $\gamma_{g,k}$ (H0: $\gamma_{g,k} = 0$ vs. H1: $\gamma_{g,k} \neq 0$) on the Y axis for HbA1c/DR, HbA1c/DN, and SBP/DN trait pairs. Significance levels $P^* = 1.7 \times 10^{-4}$ and $P^* = 0.05$ are indicated by horizontal and vertical dashed lines. b) and c) represent association results for rs10810632 and rs1358030 detected as indirect associations at $P^* = 1.7 \times 10^{-4}$. Left panels present results from separate analysis of each trait (i.e. longitudinal model for each QT and Cox PH TTE model without adjusting for the longitudinal traits) as used in naïve discovery GWAS; and right panels show results from the joint model with bootstrap 95% confidence intervals for the direct and indirect SNP effects. Results are presented using time-weighted cumulative HbA1c effects on T1DC traits.

impact the classification. We repeated the sample size analyses specifying an adjusted SNP effect size for the HbA1c association equal to 50% of its effect estimate in our DCCT analysis. In this scenario, the test of SNP–HbA1c effect falls just below the P^* threshold in the resample size $N = 516$, although power improves in larger sample sizes (Fig. 6). Here again, classification tends to shift from indirect toward both direct and indirect SNP associations with both T1DC traits (Supplementary File 3, section 6, for complete results for both SNPs); however, a much larger sample size is needed to cross the P^* threshold for a powerful test of direct association.

In the DCCT application, we take advantage of the largest available clinical study of T1D complications with long-term follow-up and high-density longitudinal QT measurements. This highlights the dearth of longitudinal studies with both a large number of individuals and long-term clinical follow-up, as well as the related challenges in joint model analysis. In prospective study designs with both longitudinal and TTE traits, an inherent imbalance can exist among traits in detection of SNP associations, in that power for QT(s) depends on the number of measurements while power for TTE traits depends on the number of events and duration of follow-up. Although there is currently a trade-off between epidemiological studies with large sample sizes but low density of longitudinal measurements and clinical studies with more modest sample sizes but high measurement densities, we anticipate informative application of joint model analysis in large biobanks,

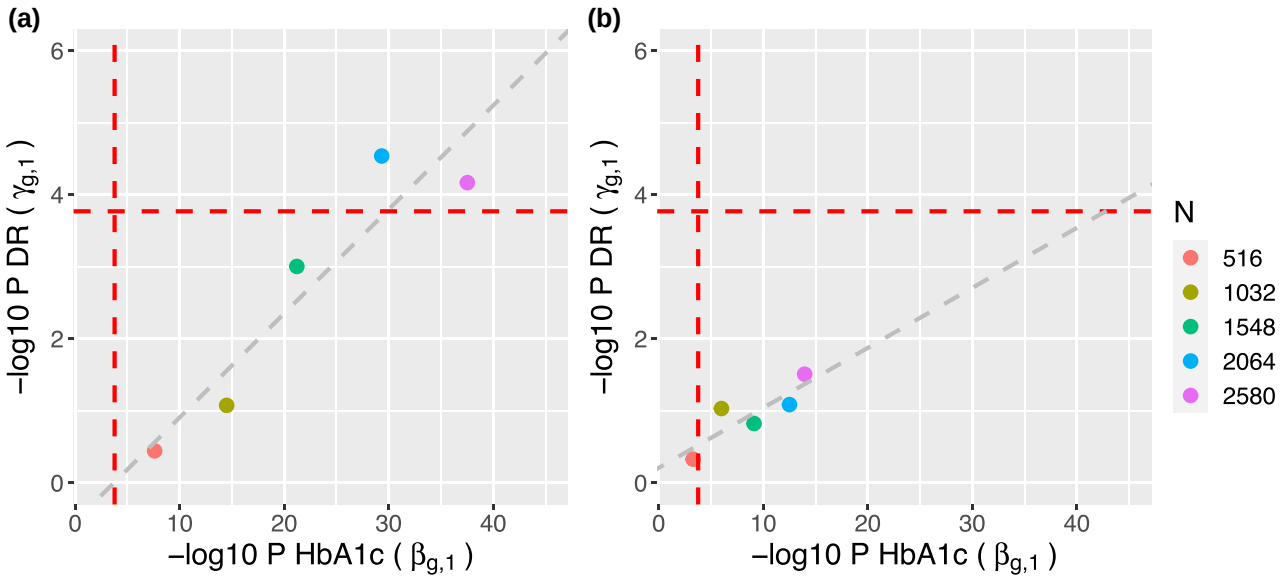

Fig. 6. Change in classification results of rs10810632 for HbA1c/DR with (a) increasing sample size and (b) winner's curse bias for the SNP effect on HbA1c, investigated using parametric resampling (Supplementary File 3, section 6). We resampled data sets with sample size up to 5 times the DCCT sample size of $N = 516$ and then extrapolated the classification results beyond $N = 2580$. The X axes of a) and b) show the P-values (-log10) for the test of rs10810632 effect on HbA1c (H0: $\beta_{g,1} = 0$ vs. H1: $\beta_{g,1} \neq 0$), while the Y axes show the P-values (-log10) for the test of rs10810632 effect on DR (H0: $\gamma_{g,1} = 0$ vs. H1: $\gamma_{g,1} \neq 0$) for the sample sizes investigated (shown by different colors). We fitted a regression line on each plot to project the trend of the classification beyond $N =$ 2,580 individuals. These plots illustrate the corresponding shift in classification of rs10810632 association with DR as indirect via HbA1c toward classification as both indirect and direct association. Complete results with HbA1c/DN are shown in Supplementary File 3 (section 6).

for example (Scholtens *et al.* 2015; Bycroft *et al.* 2018; Dummer *et al.* 2018), now assembling longitudinal measures jointly with binary outcomes and genetic data.

Discussion

We present new, more informative methods for statistical genetic analysis under a joint model specification of multiple longitudinal risk factors and multiple TTE traits, designed to characterize the complex genetic architecture of related traits in longitudinal studies of disease progression. The proposed extended model is formulated to deal with dependencies within and between traits and can account for trait-specific and shared covariates, within-subject random variability in the longitudinal traits, as well as effects of unobserved baseline confounding factors between the TTE traits through a subject-specific frailty term. We introduce a realistic data-based simulation algorithm to assess joint model performance that can also be used to estimate achievable power in clinical studies such as DCCT characterized by extensive longitudinal follow-up but limited sample size.

Evaluation by realistic simulation study of complex T1DC genetic architecture shows that accounting for trait dependencies and measurement errors in longitudinal QT risk factors using the joint model extension improves classification accuracy of SNP as direct and/or indirect association compared to (1) separate joint model analysis of each TTE trait with 1 or multiple longitudinal QTs and (2) Cox-PH frailty model analysis adjusted for multiple observed longitudinal QT values. This improvement in classification accuracy under the joint models of multiple longitudinal and multiple TTE traits results from improved T1E control for single-parameter tests of SNP effects ($\beta_{g,l}$ and $\gamma_{g,k}$) and improved power to detect SNP association, largely explained by reduced parameter estimation bias. However, we also observe that estimation bias and misclassification can be severe in the presence of SNP association with a longitudinal risk factor that is unmeasured or absent from the joint model, and misclassification may be nonnegligible when the study has limited power to detect only 1 of the SNP effects ($\beta_{g,l}$ or $\gamma_{g,k}$), as in the DCCT study.

In the application, we use parametric resampling to evaluate how study sample size or Winner's curse bias affects classification accuracy and anticipate that this approach may also help inform replication study design with sufficient power. Nevertheless, we conclude that joint model analysis in longitudinal studies of disease progression, such as in the DCCT Genetics Study, improves classification of direct and/or indirect SNP association. By integrative analysis of multiple SNP–QT–TTE pathways, classification of direct versus indirect association helps prioritize relevant genomic regions, phenotypes, and tissues for functional dissection. As outlined in a recent review (Rao et al. 2021), experimental approaches to obtain biological insight require identification of variants and target genes, with functional studies of biological mechanisms to link variants and genes to phenotypes. The relationships among lipid variants and genes, lipid levels, and coronary artery disease is a well-recognized instance of how genetic associations identified in GWAS led to subsequent experimental validation of indirect effects (Musunuru *et al.* 2010). Although such experimental studies are beyond the scope of this manuscript, analytic methods that more deeply mine existing GWAS data can inform functional studies to accelerate translational impact of GWAS findings.

Although the primary aim in this report is to develop statistical methods to distinguish among direct and/or indirect SNP associations with each TTE trait, the multiple-trait extension of the joint model lends itself to development of multiple-trait SNP association testing for SNP discovery. In Supplementary File 5, we present a joint-parameter test for global SNP association with all the longitudinal and TTE traits, based on a generalized Wald statistic. In application to the simulated DCCT-based complex genetic

architecture, we observe good T1E control under the global genetic null scenario, improved power for SNP discovery when a SNP has multiple-trait effects, and power maintenance in other SNP association scenarios.

The extended joint model can be applied to studies where longitudinal QT measurements are missing at some of the visits, but choice of the joint model estimation method depends on the missing data mechanism. In the presence of informative missing data mechanisms, we recommend sensitivity analysis using existing joint likelihood implementations that assume the time-to-dropout mechanism depends on missing longitudinal QT values through the posterior distribution of the random effects; this corresponds to an informative missing data mechanism (Rizopoulos 2012). To our knowledge, these implementations exist only for simple joint model formulations either with 1 longitudinal and 1 TTE trait (Rizopoulos 2010) or with multiple longitudinal QTs and 1 TTE trait (Rizopoulos 2016; Hickey *et al.* 2018c). Alternatively, multiple imputation methods have been implemented for 2-stage estimation (Rubin 1987; Moreno-Betancur *et al.* 2018); these impute missing values for multiple longitudinal QTs using the conditional distribution of each QT trait given a TTE trait and other QTs. More generally for missing data, further development of methods to maximize the extended joint likelihood, for example, using Bayesian methods, would alleviate numerical challenges with increasing model complexity in multivariate extensions of joint models (Gould *et al.* 2015); but this would require the design of an efficient sampling algorithm to study the posterior distribution.

We acknowledge several features of the joint model approach that warrant examination in further work. *Firstly*, to reduce computational complexity and improve model flexibility, we use 2-stage parameter estimation. In some circumstances, this approach can produce biased effect estimates in the longitudinal and/or TTE submodels as well as underestimation of their standard errors (Tsiatis *et al.* 1995; Wulfsohn and Tsiatis 1997; Dafni and Tsiatis 1998; Ye *et al.* 2008; Albert and Shih 2010; Sweeting and Thompson 2011; Ye and Wu 2017; Huong *et al.* 2018; Mauff *et al.* 2020). For instance, estimates are biased when data are informatively missing in the presence of nonrandom censoring of the longitudinal trait values, e.g. due to the occurrence of an event or informative dropout (Ye *et al.* 2008; Albert and Shih 2010; Mauff *et al.* 2020). In the DCCT application, characterized by administrative censoring and a high completion rate, these biases are of less concern because longitudinal QT values continued to be recorded on a prespecified visit schedule regardless of the occurrence of any T1DC events; we estimated the trajectories using all the available measurements. Furthermore, we obtain robust nonparametric bootstrap estimates of the covariance matrix for the SNP effects, and simulation results under the null do not show deviation from expected distributions.

Secondly, because the joint model integrates longitudinal and TTE submodels, model misspecification can occur in multiple ways. Misspecification of the QT submodel (such as by ignoring nonlinear time trends) can induce bias in the TTE submodel (Arisido *et al.* 2019) and mislead inference. In the DCCT analysis, we account for nonlinearities in HbA1c trajectories by including an additional short-term trend in the first 6 months after DCCT entry. Data-adaptive splines in longitudinal QT trajectories to improve robustness were found to be feasible for a joint model with 1 longitudinal QT (HbA1c) and 1 TTE trait (time-to-DR) in a DCCT analysis (Bian 2020), but evaluation for multiple QTs is warranted. Model comparisons (based on Akaike or Bayesian information criteria) and diagnostic analyses may also help to detect departures from joint model assumptions. In Supplementary File 3 (section 5), we provide an illustration of diagnostic analyses in the DCCT application which can serve as guidance in other applications. Overall, we recommend careful consideration of 2-stage model specification and data-based simulation studies to evaluate classification robustness of direct and/or indirect associations.

Thirdly, patient visits were scheduled with high frequency in DCCT, so we ignored the modest degree of interval censoring in the current implementation of the joint model; when there are longer gaps between visits, alternative methods are needed to account for interval censoring in the TTE submodel with additional simulation studies to assess impact on joint model estimation, testing, and classification.

Computationally, joint model fitting can be very demanding, particularly for genetic association studies that test millions of variants. In the DCCT application, it took ~1 min to fit the joint model for each SNP and ~18 more min to estimate the covariance matrix with 500 bootstraps run in parallel on 4 nodes (each node with 40 CPU and 202 GB RAM). While analysis at the genome-wide level, involving, for example, ~9 million imputed autosomal SNPs in DCCT study (Roshandel *et al.* 2018), is computationally unrealistic at present, a screening approach without bootstrap to select informative SNPs, followed by bootstrap refinement, would reduce the computational burden to feasible levels. Recently, computationally efficient algorithms have been developed to improve feasibility of linear mixed models (Sikorska *et al.* 2018) and Cox PH models (Rizvi *et al.* 2019; Bi *et al.* 2020) for genome-wide genetic association studies, but to date, they remain underdeveloped for multivariate outcomes.

Lastly, Liu *et al.* (2018) discuss various formulations and interpretations of joint models in the context of mediation analysis, with shared random effects to account for potential unmeasured baseline confounding factors between 1 longitudinal and 1 TTE traits. In clinical trial applications, they illustrate interpretation of sensitivity analysis to unmeasured baseline confounders. Adaptation of the joint model we propose for multiple longitudinal QTs and multiple TTE traits for mediation analysis requires extension of the mediation assumptions (Sobel 1982; Mackinnon *et al.* 1995) to the case of multiple mediators and multiple TTE traits. Specific evaluations of the joint model extension under these assumptions are also warranted.

In conclusion, we anticipate application of joint model methods in large biobank data sets to be informative in characterization of the genetic architecture of complex traits. Some extensions of joint models have been proposed to account for additional challenges raised by large biobanks, for example, informative visiting processes (Gasparini *et al.* 2020). By providing more efficient SNP effect estimates and increased precision in polygenic risk score development, results of such analysis also have potential to contribute to the translation of human genetic findings into personalized medicine (Young *et al.* 2019), including dynamic subject-specific risk assessment (Papageorgiou *et al.* 2019; Bull *et al.* 2020) as well as optimization of intervention strategies (Sweeting and Thompson 2011; Yuen *et al.* 2018).

Data availability

DCCT data are available to authorized users at https://repository.niddk.nih.gov/studies/edic/ and https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000086.v3.p1 (IRB #07-0208-E). Example R codes for DCCT data-based simulation and analysis of the simulated data are on GitHub (https://github.com/brossardMyriam/Joint-model-for-multiple-trait-genetics). Supplementary files are available on Figshare. Supplementary File 1

describes the DCCT data set and lists DCCT/EDIC Research Group members; Supplementary File 2 is supplemental information for the DCCT-based simulations; Supplementary File 3 is supplemental information for the DCCT Genetics Study data analysis; Supplementary File 4 lists SNPs analyzed in DCCT; Supplementary File 5 details multiple traits SNP association testing under the joint model framework. Compute Canada provided high performance computing (see https://docs.computecanada.ca/wiki/Niagara#Niagara_hardware_specifications and https://docs.scinet.utoronto.ca/index.php/Niagara_Quickstart for the hardware specifications and characteristics). Figshare DOI: https://doi.org/10.25386/genetics.23315162

Acknowledgments

This study uses data provided by the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Research Group which is sponsored through research contracts from the National Institute of Diabetes, Endocrinology and Metabolic Diseases of the National Institute of Diabetes and Digestive Kidney Diseases (NIDDK), and the National Institutes of Health (NIH). The authors are grateful to the subjects in the DCCT/EDIC cohort for their long-term participation. A complete list of the individuals and institutions participating in the DCCT/EDIC Research Group can be found in Supplementary File 1. Computations were performed on the Niagara supercomputer at the SciNet HPC Consortium.

Supplemental material available at GENETICS online.

Funding

This project was supported by: Canadian Institutes of Health Research Grants (#MOP-84287, #PJT-159509, #PJT-159463), Canadian Statistical Sciences Institute, Collaborative Research Team in "Statistical methods for the analysis of genetic data with survival outcomes", Canadian Statistical Sciences Institute postdoctoral fellowship (MB), and Canadian Institutes of Health Research, Strategic Training Grant fellowships (MB and OEG, #GET-101831). SciNet is funded by the Canada Foundation for Innovation, the Government of Ontario, Ontario Research Fund: Research Excellence, and the University of Toronto.

Conflicts of interest

The author(s) declare no conflict of interest.

Literature cited

Albert PS, Shih JH. On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. Biometrics. 2010;66(3):983–987. doi:10.1111/j.1541-0420.2009.01324_1.x.

Arisido MW, Antolini L, Bernasconi DP, Valsecchi MG, Rebora P. Joint model robustness compared with the time-varying covariate Cox model to evaluate the association between a longitudinal marker and a time-to-event endpoint. BMC Med Res Methodol. 2019;19(1):222. doi:10.1186/s12874-019-0873-y.

Asar Ö, Ritchie J, Kalra PA, Diggle PJ. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. Int J Epidemiol. 2015;44(1):334–344. doi:10.1093/ije/dyu262.

Balan TA, Putter H. A tutorial on frailty models. Stat Methods Med Res. 2020;29(11):3424–3454. doi:10.1177/0962280220921889.

Bi W, Fritsche LG, Mukherjee B, Kim S, Lee S. A fast and accurate method for genome-wide time-to-event data analysis and its application to UK Biobank. Am J Hum Genet. 2020;107(2):222–233. doi:10.1016/j.ajhg.2020.06.003.

Bian Y. 2020. Hypothesis testing in joint models for longitudinal and time-to-event outcomes. University of Toronto. https://tspace.library.utoronto.ca/handle/1807/103512.

Brent RP. Algorithms for Minimization Without Derivatives, Chapter 4. Englewood Cliffs (NJ): Prentice- Hall; 1973. .

Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. Biometrics. 2003;59(2):221–228. doi:10.1111/1541-0420.00028.

Bull L, Lunt M, Martin G, Hyrich K, Sergeant J. Harnessing repeated measurements of predictor variables for clinical risk prediction: a review of existing methods. Diagnostic Progn Res. 2020;4(1):9. doi:10.1186/s41512-020-00078-z.

Bycott P, Taylor J. A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. Stat Med. 1998;17(18):2061–2077. doi:10.1002/(sici)1097-0258(19980930)17:18<2061::aid-sim896>3.0.co;2-o.

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–209. doi:10.1038/s41586-018-0579-z.

Chen LM, Ibrahim JG, Chu H. Sample size and power determination in joint modeling of longitudinal and survival data. Stat Med. 2011;30(18):2295–2309. doi:10.1002/sim.4263.

Crowther MJ, Lambert PC. Simulating biologically plausible complex survival data. Stat Med. 2013;32(23):4118–4134. doi:10.1002/sim.5823.

Dafni UG, Tsiatis AA. Evaluating surrogate markers of clinical outcome when measured with error. Biometrics. 1998;54(4):1445–1462. doi:10.2307/2533670.

Deng Y, Pan W. Conditional analysis of multiple quantitative traits based on marginal GWAS summary statistics. Genet Epidemiol. 2017;41(5):427–436. doi:10.1002/gepi.22046.

Dummer TJB, Awadalla P, Boileau C, Craig C, Fortier I, Goel V, Hicks JMT, Jacquemont S, Knoppers BM, Le N, et al. The Canadian partnership for tomorrow project: a pan-Canadian platform for research on chronic disease prevention. CMAJ. 2018;190(23):E710–E717. doi:10.1503/cmaj.170292.

Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, Ntritsos G, Dimou N, Cabrera CP, Karaman I, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet. 2018;50(10):1412–1425. doi:10.1038/s41588-018-0205-x.

Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. Stat Med. 1996;15(15):1663–1685. doi:10.1002/(SICI)1097-0258(19960815)15:15<1663::AID-SIM294>3.0.CO;2-1.

Furgal AKC, Sen A, Taylor JMG. Review and comparison of computational approaches for joint longitudinal and time-to-event models. Int Stat Rev. 2019;87(2):insr.12322. doi:10.1111/insr.12322.

Gasparini A, Abrams KR, Barrett JK, Major RW, Sweeting MJ, Brunskill NJ, Crowther MJ. Mixed-effects models for health care longitudinal data with an informative visiting process: a Monte Carlo simulation study. Stat Neerl. 2020;74(1):5–23. doi:10.1111/stan.12188.

Gould AL, Boye ME, Crowther MJ, Ibrahim JG, Quartey G, Micallef S, Bois FY. Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. Stat Med. 2015;34(14):2181–2195. doi:10.1002/sim.6141.

Grassi MA, Tikhomirov A, Ramalingam S, Below JE, Cox NJ, Nicolae DL. Genome-wide meta-analysis for severe diabetic retinopathy. Hum Mol Genet. 2011;20(12):2472–2481. doi:10.1093/hmg/ddr121.

Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. BMC Med Res Methodol. 2016;16(1):117. doi:10.1186/s12874-016-0212-5.

Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. Joint models of longitudinal and time-to-event data with more than one event time outcome: a review. Int J Biostat. 2018a;14(1):1-19. doi:10.1515/ijb-2017-0047

Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial. J R Stat Soc Ser A Stat Soc. 2018b;181(4):1105–1123. doi:10.1111/rssa.12348.

Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. joineRML: a joint model and software package for time-to-event and multivariate longitudinal outcomes. BMC Med Res Methodol. 2018c;18(1):50. doi:10.1186/s12874-018-0502-1.

Hogan JW, Laird NM. Increasing efficiency from censored survival data by using random effects to model longitudinal covariates. Stat Methods Med Res. 1998;7(1):28–48. doi:10.1177/096228029800700104.

Hosseini SM, Boright AP, Sun L, Canty AJ, Bull SB, Klein BEK, Klein R, Paterson AD. The association of previously reported polymorphisms for microvascular complications in a meta-analysis of diabetic retinopathy. Hum Genet. 2015;134(2):247–257. doi:10.1007/s00439-014-1517-2.

Hougaard P. Frailty models for survival data. Lifetime Data Anal. 1995;1(3):255–273. doi:10.1007/BF00985760.

Hsieh F, Tseng YK, Wang JL. Joint modeling of survival and longitudinal data: likelihood approach revisited. Biometrics. 2006;62(4):1037–1043. doi:10.1111/j.1541-0420.2006.00570.x.

Huong PTT, Nur D, Pham H, Branford A. A modified two-stage approach for joint modelling of longitudinal and time-to-event data. J Stat Comput Simul. 2018;88(17):3379–3398. doi:10.1080/00949655.2018.1518449.

Ibrahim JG, Chen M-H, Sinha D. Joint models for longitudinal and survival data. Bayesian Survival Analysis. 2001:262–289. https://link.springer.com/chapter/10.1007/978-1-4757-3447-8_7

Ibrahim JG, Chu H, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. J Clin Oncol. 2010;28(16):2796–2801. doi:10.1200/JCO.2009.25.0654.

Jensen SM, Ritz C. A comparison of approaches for simultaneous inference of fixed effects for multiple outcomes using linear mixed models. Stat Med. 2018;37(16):2474–2486. doi:10.1002/sim.7666.

Kraft P. Curses–winner's and otherwise–in genetic epidemiology. Epidemiology. 2008;19(5):649–651. discussion 657-8. doi:10.1097/EDE.0b013e318181b865.

Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982;38(4):963–974. doi:10.2307/2529876.

Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb). 2005;95(3):221–227. doi:10.1038/sj.hdy.6800717.

Lind M, Odén A, Fahlén M, Eliasson B. The relationship of glycemic exposure (HbA1c) to the risk of development and progression of retinopathy in the Diabetes Control and Complications Trial. Diabetes. 1995;44(8):968–983. doi:10.2337/diab.44.8.968.

Lind M, Odén A, Fahlén M, Eliasson B. The shape of the metabolic memory of HbA1c: re-analysing the DCCT with respect to time-dependent effects. Diabetologia. 2010;53(6):1093–1098. doi:10.1007/s00125-010-1706-z.

Liu L, Zheng C, Kang J. Exploring causality mechanism in the joint analysis of longitudinal and survival data. Stat Med. 2018;37(26):3733–3744. doi:10.1002/sim.7838.

Mackinnon DP, Warsi G, Dwyer JH. A simulation study of mediated effect measures. Multivariate Behav Res. 1995;30(1):41. doi:10.1207/s15327906mbr3001_3. .

Mauff K, Steyerberg E, Kardys I, Boersma E, Rizopoulos D. Joint models with multiple longitudinal outcomes and a time-to-event outcome: a corrected two-stage approach. Stat Comput. 2020;30(4):999–1014. doi:10.1007/s11222-020-09927-9.

Mauff K, Steyerberg EW, Nijpels G, van der Heijden AAWA, Rizopoulos D. Extension of the association structure in joint models to include weighted cumulative effects. Stat Med. 2017;36(23):3746–3759. doi:10.1002/sim.7385.

Moreno-Betancur M, Carlin JB, Brilleman SL, Tanamas SK, Peeters A, Wolfe R. Survival analysis with time-dependent covariates subject to missing data or measurement error: multiple imputation for joint modeling (MIJM). Biostatistics. 2018;19(4):479–496. doi:10.1093/biostatistics/kxx046.

Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, *et al.* From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature. 2010;466(7307):714–719. doi:10.1038/NATURE09266.

Papageorgiou G, Mauff K, Tomer A, Rizopoulos D. An overview of joint modeling of time-to-event and longitudinal outcomes. Annu Rev Stat Its Appl. 2019;6(1):223–240. doi:10.1146/annurev-statistics-030718-105048.

Paterson AD, Bull SB. Does familial clustering of risk factors for long-term diabetic complications leave any place for genes that act independently? J Cardiovasc Transl Res. 2012;5(4):388–398. doi:10.1007/s12265-012-9385-4.

Paterson AD, Waggott D, Boright AP, Hosseini SM, Shen E, Sylvestre MP, Wong I, Bharaj B, Cleary PA, Lachin JM, *et al.* A genome-wide association study identifies a novel major locus for glycemic control in type 1 diabetes, as measured by both A1C and glucose. Diabetes. 2010;59(2):539–549. doi:10.2337/db09-0653.

Pollack S, Igo RP, Jensen RA, Christiansen M, Li X, Cheng C-Y, Ng MCY, Smith A V, Rossin EJ, Segrè A V, *et al.* Multiethnic genome-wide association study of diabetic retinopathy using liability threshold modeling of duration of diabetes and glycemic control. Diabetes. 2019;68(2):441–456. doi:10.2337/db18-0567.

Rao S, Yao Y, Bauer DE. Editing GWAS: experimental approaches to dissect and exploit disease-associated genetic variation. Genome Med. 2021;13(41):1-20. doi:10.1186/S13073-021-00857-3

Rizopoulos D. JM : an R package for the joint modelling of longitudinal and time-to-event data. J Stat Softw. 2010;35(9):9. doi:10.18637/jss.v035.i09.

Rizopoulos D. 2012. Joint models for longitudinal and time-to-event data. Chapman and Hall/CRC. https://www.taylorfrancis.com/books/9781439872871.

Rizopoulos D. The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. J Stat Softw. 2016;72(7). doi:10.18637/jss.v072.i07.

Rizvi AA, Karaesmen E, Morgan M, Preus L, Wang J, Sovic M, Hahn T, Sucheston-Campbell LE. . Gwasurvivr: an R package for genome-wide survival analysis. Bioinformatics. 2019;35(11):1968–1970. doi:10.1093/bioinformatics/bty920.

Roshandel D, Gubitosi-Klug R, Bull SB, Canty AJ, Pezzolesi MG, King GL, Keenan HA, Snell-Bergeon JK, Maahs DM, Klein R, *et al.* Meta-genome-wide association studies identify a locus on chromosome 1 and multiple variants in the MHC region for

serum C-peptide in type 1 diabetes. Diabetologia. 2018;61(5):1098–1111. doi:10.1007/s00125-018-4555-9.
Rubin DB. Multiple Imputation for Nonresponse in Surveys. Hoboken (NJ): John Wiley & Sons, Inc. (Wiley Series in Probability and Statistics); 1987.
Sandholm N, Salem RM, McKnight AJ, Brennan EP, Forsblom C, Isakova T, McKay GJ, Williams WW, Sadlier DM, Mäkinen V-P, *et al.* New susceptibility loci associated with kidney disease in type 1 diabetes.PLoS Genet. 2012;8(9):e1002921. doi:10.1371/journal.pgen.1002921.
Scholtens S, Smidt N, Swertz MA, Bakker SJL, Dotinga A, Vonk JM, Van Dijk F, Van Zon SKR, Wijmenga C, Wolffenbuttel BHR, *et al.* Cohort profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol. 2015;44(4):1172–1180. doi:10.1093/ije/dyu229.
Self S, Pawitan Y. Modeling a marker of disease progression and onset of disease. In: AIDS Epidemiology. Boston (MA): Birkhäuser Boston; 1992. p. 231–255.
Shah A, Laird N, Schoenfeld D. A random-effects model for multiple characteristics with possibly missing data. J Am Stat Assoc. 1997;92(438):775. doi:10.2307/2965726.
Sikorska K, Lesaffre E, Groenen PJF, Rivadeneira F, Eilers PHC. Genome-wide analysis of large-scale longitudinal outcomes using penalization—gALLOP algorithm. Sci Rep. 2018;8(1):6815. doi:10.1038/s41598-018-24578-7.
Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. Chapter 7 in Sociol Methodol. 1982;13:290–312. doi:10.2307/270723
Song X, Davidian M, Tsiatis AA. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. Biometrics. 2002;58(4):742–753. doi:10.1111/j.0006-341x.2002.00742.x.
Sun L, Dimitromanolakis A, Faye LL, Paterson AD, Waggott D, Bull SB; DCCT/EDIC Research Group. BR-squared: a practical solution to the winner's Curse in genome-wide scans. Hum Genet. 2011;129(5):545–552. doi:10.1007/s00439-011-0948-2.
Sweeting MJ, Thompson SG. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. Biometrical J. 2011;53(5):750–763. doi:10.1002/bimj.201100052.
Tang AM, Tang NS. Semiparametric Bayesian inference on skew-normal joint modeling of multivariate longitudinal and survival data. Stat Med. 2015;34(5):824–843. doi:10.1002/sim.6373.
Tang NS, Tang AM, Pan DD. Semiparametric Bayesian joint models of multivariate longitudinal and survival data. Comput Stat Data Anal. 2014;77(2014):113–129. doi:10.1016/j.csda.2014.02.015
The 1000 Genomes Project Consortium; Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. Nature. 2015;526(7571):68–74. doi:10.1038/nature15393. http://www.ncbi.nlm.nih.gov/pubmed/26432245.
The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med. 1993;329(14):977–986. doi:10.1056/NEJM199309303291401.
Therneau TM. 2020. A Package for survival analysis in R. R package version 3.2-7. http://CRAN.R-project.org/package=survival. https://cran.r-project.org/package=survival.
Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. New York (NY): Springer New York (Statistics for Biology and Health); 2000.
Tsiatis AA, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. Biometrika. 2001;88(2):447–458. doi:10.1093/biomet/88.2.447.
Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. Stat Sin. 2004;14(3):809–834.
Tsiatis AA, Degruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. J Am Stat Assoc. 1995;90(429):27–37. doi:10.1080/01621459.1995.10476485.
Wheeler E, Leong A, Liu C-TT, Hivert M-FF, Strawbridge RJ, Podmore C, Li M, Yao J, Sim X, Hong J, *et al.* Impact of common genetic determinants of hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: a transethnic genome-wide meta-analysis. PLoS Med. 2017;14(9):e1002383. doi:10.1371/journal.pmed.1002383.
Wu L, Liu W, Yi GY, Huang Y. Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. J Probab Stat. 2012;2012:e640153. doi:10.1155/2012/640153.
Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. Biometrics. 1997;53(1):330–339. doi:10.2307/2533118.
Xu J, Zeger SL. Joint analysis of longitudinal data comprising repeated measures and times to events. J R Stat Soc Ser C (Applied Stat). 2001;50(3):375–387. doi:10.1111/1467-9876.00241.
Ye Q, Wu L. Two-step and likelihood methods for joint models of longitudinal and survival data. Commun Stat—Simul Comput. 2017;46(8):6019–6033. doi:10.1080/03610918.2016.1188208.
Ye W, Lin X, Taylor JMG. Semiparametric modeling of longitudinal measurements and time-to-event data—a two-stage regression calibration approach. Biometrics. 2008;64(4):1238–1246. doi:10.1111/j.1541-0420.2007.00983.x.
Young AI, Benonisdottir S, Przeworski M, Kong A. Deconstructing the sources of genotype-phenotype associations in humans. Science. 2019;365(6460):1396–1400. doi:10.1126/science.aax3710.
Yuen HP, Mackinnon A, Hartmann J, Amminger GP, Markulev C, Lavoie S, Schäfer MR, Polari A, Mossaheb N, Schlögelhofer M, *et al.* Dynamic prediction of transition to psychosis using joint modelling. Schizophr Res. 2018;202:333–340. doi:10.1016/j.schres.2018.07.002.
Zhu H, Ibrahim JG, Chi YY, Tang N. Bayesian Influence measures for joint models for longitudinal and survival data. Biometrics. 2012;68(3):954–964. doi:10.1111/j.1541-0420.2012.01745.x

Editor: H. Huang

Appendix

Joint likelihood function of the joint model of 1 longitudinal QT and 1 TTE trait ($L = K = 1$)

Under the following assumptions (A1–A3),

(A1) $\boldsymbol{b_i} \sim N_2(0, \boldsymbol{D})$, where $\boldsymbol{b_i} = (b_{i,0}, b_{i,1})^T$ are subject-specific random effects

(A2) $\varepsilon_{i,j} \sim N(0, \sigma^2)$ and $\varepsilon_{i,j} \perp \varepsilon_{i,s}$ between visit times $t_{i,j} \neq t_{i,s}$, with $j \neq s$, for all $1 \leq j \leq J$ and $1 \leq s \leq J$

(A3) $\boldsymbol{b_i} \perp \boldsymbol{\varepsilon_i}$, where $\boldsymbol{\varepsilon_i} = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,j}, \ldots, \varepsilon_{i,J})^T$, $\boldsymbol{\varepsilon_i} \sim N_J(0, \sigma^2 \boldsymbol{I_J})$.

Then, conditional on the random effects $\boldsymbol{b_i}$ and fixed effects $\boldsymbol{\Omega}$, it is appropriate to assume that the repeated measurements in the longitudinal process are independent (Laird and Ware 1982), in other words that the serial correlation is taken into account; and

the longitudinal and TTE submodels are independent (Ibrahim *et al.* 2001). Under these conditional independence assumptions, the joint likelihood function of the joint model parameters $\boldsymbol{\Omega}$ given the observed data is

$$L(\boldsymbol{\Omega}|\mathbf{y_i}, T_i, \delta_i) = \prod_{i=1}^{N} \int f_1(\mathbf{y_i}|\ \mathbf{b_i}, \boldsymbol{\Omega})\ \times f_2(T_i, \delta_i\ |\ \mathbf{b_i},\ \boldsymbol{\Omega})\ \times f_3(\mathbf{b_i}|\boldsymbol{\Omega}) d\mathbf{b_i},$$

where

- $f_1(\mathbf{y_i}|\ \mathbf{b_i}, \boldsymbol{\Omega}) = (2\pi\sigma^2)^{-J/2} \prod_{j=1}^{J} \exp\left(\frac{(y_{i,j}-y_i^*(t_{i,j}))^2}{2\sigma^2}\right)$,
- $f_2(T_i, \delta_i\ |\ \mathbf{b_i},\ \boldsymbol{\Omega}) = [\lambda_i(T_i|\ \mathbf{b_i},\ \boldsymbol{\Omega})\ ]^{\delta_i}\ S_i(T_i|\ \mathbf{b_i},\ \boldsymbol{\Omega})$,
- $S_i(T_i|\mathbf{b_i}, \boldsymbol{\Omega}) = \exp\left[-\int_0^{T_i} \lambda_0(s)\exp\{\alpha w_i(s)+\gamma_g g_i\}ds\right]$ and $w_i(t) = f(y_i^*(t))$, where $y_i^*(t)$, denotes the lth QT trajectory at time t for $1 \le l \le L$ which depends on the fixed and random effects $\boldsymbol{\beta}$ and $\mathbf{b_i}$. The survival function depends on the whole QT history of the true unobserved longitudinal process up to time t, denoted as $Y_i^*(t) = \{\ y_i^*(s),\ 0 \le s \le t\}$;
- $f_3(\mathbf{b_i}|\boldsymbol{\Omega}) = (2\pi)^{-q/2}|\mathbf{D}|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{b_i}^T\mathbf{D}^{-1}\mathbf{b_i}\right\}$, where q is the dimension of the $\mathbf{D}$ matrix.

The event indicator δ_i is used to distinguish the contribution of the individuals who experience the event during the observation period from the individuals who are still at risk up to that time point but do not experience the event. Individuals who experience the event ($\delta_i = 1$) contribute to the cumulative hazard function and to the hazard function both evaluated at the T_i; the individuals who do not experience the event ($\delta_i = 0$) contribute to the hazard function only.

Joint model parameter estimation can be performed by maximization of the full joint likelihood function directly or by Bayesian computation. Direct maximization of the joint likelihood function can be performed using the expectation–maximization (EM) algorithm, treating the random effects as missing data (Wulfsohn and Tsiatis 1997). However, integrals with respect of time in the definition of the survival function, as well as the integral with respect to the $\mathbf{b_i}$ do not have an analytical solution; therefore, numerical approaches, such as adaptive Gauss–Hermite quadrature, are needed. Implementations of this model using maximization of the above joint likelihood function are available in R packages (see Furgal *et al.* 2019, for a review). However, as the dimension of the $\mathbf{b_i}$ increases, the integral over the $\mathbf{b_i}$ becomes computationally burdensome, and Bayesian approaches can be employed instead, where the $\mathbf{b_i}$ are also considered model parameters obtained as a posterior sample, and thus the integral over the $\mathbf{b_i}$ is no longer required.

Joint model of multiple longitudinal QTs and multiple TTE traits

Joint likelihood function

Extending the previous joint model assumptions (A1–A3) to $L > 1$ and $K > 1$, we have

(A1) $\mathbf{b_i} \sim N_2(0, \mathbf{D})$, where $\mathbf{b_i} = (\mathbf{b_{i,1}}, \ldots, \mathbf{b_{i,l}}, \ldots, \mathbf{b_{i,L}})^T$ are subject-specific random effects for all L QTs

(A2) $\boldsymbol{\varepsilon_{i,l}} \sim N_J(0,\ \sigma_l^2\mathbf{I_J})$ for lth QT with $1 \le l \le L$

(A3) $\mathbf{b_{i,l}} \perp \boldsymbol{\varepsilon_{i,l}}$ for lth QT with $1 \le l \le L$, and

(A4) $u_i \perp \mathbf{b_{i,l}}$, where u_i is a shared subject-specific frailty for K TTE traits.

Then, conditional on the random effects $\mathbf{b_i}$, the frailty u_i, and fixed effects $\boldsymbol{\Omega}$, we further assume the following: $\mathbf{b_i}$ accounts for association among the L longitudinal QTs (Shah *et al.* 1997) and association between the longitudinal and TTE traits (Ibrahim *et al.* 2001); and u_i accounts for residual dependence among the TTE traits (Hougaard 1995). Under these conditional independence assumptions, the joint likelihood function of the joint model parameters $\boldsymbol{\Omega}$ given the observed data is

$$L(\boldsymbol{\Omega}|\mathbf{y_i}, \mathbf{T_i}, \boldsymbol{\delta_i}) = \prod_{i=1}^{N} \int f_1(\mathbf{y_i}|\ \mathbf{b_i},\ \boldsymbol{\Omega})\ \times f_2(\mathbf{T_i}, \boldsymbol{\delta_i}\ |\ \mathbf{b_i},\ u_i, \boldsymbol{\Omega})\ \times f_3(\mathbf{b_i}|\boldsymbol{\Omega}) \times f_4(u_i|\boldsymbol{\Omega})\ du_i d\mathbf{b_i},$$

where:

- $(\mathbf{T_i}, \boldsymbol{\delta_i}) = ((T_{i,1}, \delta_{i,1}), \ldots, (T_{i,k}, \delta_{i,k}), \ldots, (T_{i,K}, \delta_{i,K}))^T$ is defined as the vector of K stacked TTE traits for individual i,
- $f_1(\mathbf{y_i}|\ \mathbf{b_i}, \boldsymbol{\Omega}) = \prod_{l=1}^{L} f(\mathbf{y_{i,l}}|\ \mathbf{b_{i,l}}, \boldsymbol{\Omega_l})$ with $f(\mathbf{y_{i,l}}|\ \mathbf{b_{i,l}}, \boldsymbol{\Omega_l}) = (2\pi\sigma_l^2)^{-\frac{J}{2}}\exp\left[-\frac{1}{2\sigma_l^2}\prod_{j=1}^{J}(y_{i,j,l} - y_{i,l}^*(t_{i,j}))^2\right]$,
- $f_2(\mathbf{T_i}, \boldsymbol{\delta_i}\ |\ \mathbf{b_i},\ u_i, \boldsymbol{\Omega}) = \prod_{k=1}^{K}[\lambda_{i,k}(T_{i,k}|\ \mathbf{b_i},\ u_i, \boldsymbol{\Omega})]^{\delta_{i,k}}\ S_{i,k}(T_{i,k}|\ \mathbf{b_i}, u_i, \boldsymbol{\Omega})$, with $S_{i,k}(T_{i,k}|\ \mathbf{b_i},\ u_i, \boldsymbol{\Omega}) = \exp\left[-\int_0^{T_{i,k}} \lambda_{0,k}(s)exp\left\{\sum_{l=1}^{L}\alpha_{l,k}w_{i,l,k}(s)\right) + \gamma_{g,k}g_i + u_i\}\ ds\right]$ and $w_{i,l,k}(s) = f_{l,k}(y_{i,l}^*(t))$, where $y_{i,l}^*(t)$ denotes the lth QT trajectory at time t for $1 \le l \le L$ which depends on the fixed and random effects $\boldsymbol{\beta_l}$ and $\mathbf{b_{i,l}}$.
- $f_3(\mathbf{b_i}|\boldsymbol{\Omega}) = (2\pi)^{-q/2}|\mathbf{D}|^{-\frac{1}{2}}\ exp\left[-\frac{1}{2}\mathbf{b_i}^T\mathbf{D}^{-1}\mathbf{b_i}\right]$, where q is the dimension of the $\mathbf{D}$ matrix
- $f_4(u_i|\boldsymbol{\Omega}) = \frac{u_i^{a-1}\exp(-u_i/b)}{\Gamma(a)b^a}$, i.e. we assume $u_i \sim \text{Gamma}(a, b)$ with $u_i > 0$, a corresponds to the shape parameter and b to the scale parameter, and $a, b > 0$. $\Gamma(a)$ is the gamma function evaluated at a.

We are not aware of any existing implementations of full likelihood maximization of the extended model in the literature. Calculation of the full likelihood requires multivariate integration with respect to the random effects distribution, which can lead to demanding computation. When the random effect vector $\mathbf{b_i}$ has a small dimension, say less than 3, the integral can be evaluated via Gaussian quadrature which approximates the integral by a weighted sum of the target function evaluated at prespecified sample points. However, when the dimension is larger, it is demanding to calculate the integrals with satisfactory approximation accuracy. Although a full likelihood specification enables rigorous study of asymptotic properties, its large sample approximation may not be accurate when sample size is small. In comparison, the Bayesian paradigm does not require asymptotic approximations, but the design of an efficient sampling algorithm to study the posterior distribution is challenging. Because of these limitations, we implement a 2-stage approach to estimation of fixed effect parameters in the extended multiple-trait model that is reasonable for the GWAS application of interest; in particular, the longitudinal measurements in DCCT are taken according to a prespecified schedule and are not terminated by the observation of diabetes complications, loss to follow-up and mortality are minimal, censoring is administrative, and each individual has a dense and nearly complete set of measurements.

Likelihood functions under the 2-stage approximation

Let $\boldsymbol{\Omega}_{\mathbf{Long}}$ and $\boldsymbol{\Omega}_{\mathbf{Surv}}$ be the vectors containing all fixed parameters from the longitudinal and TTE submodels respectively.

Stage 1: Multivariate mixed model

$$L(\boldsymbol{\Omega}_{\mathbf{Long}}|\mathbf{y}_i) = \prod_{i=1}^{N} \int_{\mathbf{b}_i} f_1(\mathbf{y}_i|\ \mathbf{b}_i,\ \boldsymbol{\Omega}_{\mathbf{Long}}) \times f_3(\mathbf{b}_i|\ \boldsymbol{\Omega}_{\mathbf{Long}}) d\mathbf{b}_i,$$

where

- $f_1(\mathbf{y}_i|\ \mathbf{b}_i,\ \boldsymbol{\Omega}_{\mathbf{Long}}) = \prod_{l=1}^{L} f(\mathbf{y}_{i,l}|\ \mathbf{b}_i, \boldsymbol{\Omega}_{\mathbf{Long}})$ with $f(\mathbf{y}_{i,l}|\ \mathbf{b}_i, \boldsymbol{\Omega}_{\mathbf{Long}}) = (2\pi\sigma_l^2)^{-\frac{J}{2}} \exp\left[-\frac{1}{2\sigma_l^2}\prod_{j=1}^{J}(y_{i,j,l} - y^*_{i,l}(t_{i,j}))^2\right]$,
- $f_3(\mathbf{b}_i|\boldsymbol{\Omega}_{\mathbf{Long}}) = (2\pi)^{-q/2}|\mathbf{D}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\mathbf{b}_i^T\mathbf{D}^{-1}\mathbf{b}_i\right]$, where q is the dimension of the $\mathbf{D}$ matrix.

The fixed-effect and random-effect parameters are estimated jointly for all longitudinal QTs using all available repeated measurements, without using the TTE information. Then fitted trajectory values are obtained by plugging the parameter estimates into

$$\mathbf{y}^*_{i,l}(t) = \mathbf{X}_{i,l}(t)\boldsymbol{\beta}_l + \mathbf{Z}_{i,l}(t)\mathbf{b}_{i,l},$$

where $\mathbf{X}_{i,l}(t) = (1,\ t,\ g_i)$ and $\mathbf{Z}_{i,l}(t) = (1,\ t)$

Stage 2: Multivariate Cox PH model adjusted for fitted trajectory values for the vector of L longitudinal QTs:

$$L(\boldsymbol{\Omega}_{\mathbf{Surv}}|\mathbf{T}_i,\ \boldsymbol{\delta}_i,\ \widehat{w_i(T_i)}) = \prod_{i=1}^{N} \int_{u_i} f_2(\mathbf{T}_i,\ \boldsymbol{\delta}_i\ |u_i,\ \widehat{\mathbf{w}_i(\mathbf{T}_i)},\ \boldsymbol{\Omega}_{\mathbf{Surv}}) \times f_4(u_i|\boldsymbol{\Omega}_{\mathbf{Surv}}) \times du_i,$$

with

$f_2(\mathbf{T}_i,\ \boldsymbol{\delta}_i\ |u_i,\ \widehat{\mathbf{w}_i(\mathbf{T}_i)},\ \boldsymbol{\Omega}_{\mathbf{Surv}}) = \prod_{k=1}^{K}[\lambda_{i,k}(T_{i,k}|\ u_i,\ \widehat{\mathbf{w}_{i,k}(\mathbf{T}_{i,k})},\ \boldsymbol{\Omega}_{\mathbf{Surv}})]^{\delta_{i,k}} \times S_{i,k}(T_{i,k}|u_i,\ \widehat{\mathbf{w}_{i,k}(\mathbf{T}_{i,k})},\ \boldsymbol{\Omega}_{\mathbf{Surv}}),\ S_{i,k}(T_{i,k}|\ u_i,\ \widehat{\mathbf{w}_{i,k}(\mathbf{T}_{i,k})},\ \boldsymbol{\Omega}_{\mathbf{Surv}}) = exp\left[-\int_0^{T_{i,k}} \lambda_{0,k}(s)\exp\left\{\sum_{l=1}^{L}\alpha_{l,k}\ \widehat{w_{i,l,k}(s)} + \gamma_{g,k}\ g_i + u_i\right\}ds\right]$, where $\widehat{w_{i,l,k}(s)}$ is obtained by plugging fitted trajectory values into $w_{i,l,k}(t) = f_{l,k}(y^*_{i,l}(t))$.

Unlike the *joint likelihood function*, where the shared random effects $\mathbf{b}_i$ account for the dependencies between the longitudinal QTs and the TTE traits, the 2-stage approach accounts for the dependencies between the longitudinal and TTE traits via the fitted values of the longitudinal trajectories. This approximation can produce biased estimates and/or underestimated standard errors for longitudinal and survival model parameters, when there is nonrandom censoring of the longitudinal QT values due to the occurrence of an event or from informative dropout (Ye *et al.* 2008; Albert and Shih 2010) and because of propagation errors of stage 1 parameter estimates into stage 2 (Wulfsohn and Tsiatis 1997). Under longitudinal model misspecification and estimation bias, the conditional independence assumptions fail, undermining the accurate of trajectory estimates. Because the TTE processes are related to length of follow-up, informative missingness/dropouts can lead to differential follow-up between subjects with and without an event, and thus the random effects $\mathbf{b}_i$ can depend on the event times (e.g. patients who have an event early are more likely to have positive random slopes). However, as we show in the simulation studies, in absence of model misspecification and informative dropouts/missingness, this approach has low bias and is computationally feasible for genetic association studies.