

Technical Appendix

1 Data and Notations

- There are m subjects in the dataset.
- y_{i1j} is the j th BMI-z score for the i th subject obtained at time t_{i1j} , where $t_{i11} < \dots < t_{i1n_{i1}}$ and $1 \leq j \leq n_{i1}$. y_{i2j} is the j th HbA1c value for the i th subject obtained at time t_{i2j} , where $t_{i21} < \dots < t_{i2n_{i2}}$ and $1 \leq j \leq n_{i2}$.
- \mathbf{z}_i is the n_z -dimensional vector of covariates for subject i .
- There are n_s discrete states under consideration. The relationship between these states can be described by a directed graph (\mathbb{V}, \mathbb{E}) , where the vertices set \mathbb{V} consists of all the n_s states, and the edge set \mathbb{E} consists of all possible transitions between the states. More specifically, if we number the states by $1, 2, \dots, n_s$, then $\mathbb{V} = \{1, \dots, n_s\}$, and for $s_1, s_2 \in \mathbb{V}$, the ordered pair (s_1, s_2) is in \mathbb{E} if and only if subjects can transition from s_1 to s_2 . For simplicity we will restrict our discussions on cases where there are no cycles in the directed graph (\mathbb{V}, \mathbb{E}) although the proposed method can be extended to such cases.
- Let $\eta_i(t)$ be the right-continuous process denoting the state occupied by the i th subject at time t , and $\eta_i(t) = 0$ if the i th subject has not entered the study or is censored due to lost-to-follow-up.
- For $(s_1, s_2) \in \mathbb{E}$, we let $T_{i,s_1 \rightarrow s_2}$ denote the observed times when subject i transitions from state s_1 to s_2 , and $T_{i,s_1 \rightarrow s_2}$ is ∞ if the i th subject was not observed to make a transition from s_1 to s_2 . $T_{i,s_1 \rightarrow s_2}$ can be uniquely defined since there are no cycles in the directed graph. We let $\delta_{i,s_1 \rightarrow s_2} = 1$ if the i th subject is observed to transition from s_1 to s_2 and 0 otherwise. As usual survival models, if we let E_i and C_i denote the entry time and censoring times, and $T_{i,s_1 \rightarrow s_2}^*$ denote the true transition time from state s_1 to s_2 , then $T_{i,s_1 \rightarrow s_2}$ and $\delta_{i,s_1 \rightarrow s_2}$ can be

explicitly written as

$$T_{i,s_1 \rightarrow s_2} = \begin{cases} T_{i,s_1 \rightarrow s_2}^* & \text{if } E_i < T_{i,s_1 \rightarrow s_2}^* \leq C_i, \\ \infty & \text{otherwise} \end{cases}$$

$$\delta_{i,s_1 \rightarrow s_2} = \begin{cases} 1 & \text{if } E_i < T_{i,s_1 \rightarrow s_2}^* \leq C_i, \\ 0 & \text{otherwise} \end{cases}.$$

2 Model

2.1 Longitudinal Model

We let $m_{i1}(t)$ and $m_{i2}(t)$ be the underlying smoothed trajectories for BMI-z scores and HbA1c values. Then following longitudinal model is assumed

$$y_{ikj} = m_{ik}(t_{ikj}) + \epsilon_{ikj} \quad (1)$$

$$m_{ik}(t) = \mu_k(t) + v_k(t) \quad (2)$$

$$= \alpha_{k0} + \alpha_{k1}t + \alpha_{k2}t^2 + \gamma_{ik0} + \gamma_{ik1}t + \gamma_{ik2}t^2 \quad (3)$$

where $m_{ik}(t_{ikj})$ is a Gaussian process that can be decomposed into the sums of the mean function $\mu_k(t)$ and the variation from the mean $v(t)$, $\mu(t)$ and $v(t)$ can both be modeled by quadratic functions $\alpha_{k0} + \alpha_{k1}t + \alpha_{k2}t^2$ and $\gamma_{ik0} + \gamma_{ik1}t + \gamma_{ik2}t^2$, $\gamma_{ik} = (\gamma_{ki0}, \gamma_{ki1}, \gamma_{ki2})^\top$ are independent with multivariate $N(\mathbf{0}, \Sigma_\gamma)$, and ϵ_{ikj} are independent with univariate $N(0, \sigma_k^2)$ distribution.

2.2 Survival Model

We will let

$$h_{i,s_1 \rightarrow s_2}(t) = \lim_{dt \downarrow 0} P \{T_{i,s_1 \rightarrow s_2} \in (t, t + dt] | \eta_i(t) = s_1\} / dt$$

be the cause-specific hazard of transitioning from s_1 to s_2 for subject i . $h_{i,s_1 \rightarrow s_2}(t)$ is quantified by the following proportional hazards model

$$h_{i,s_1 \rightarrow s_2}(t) = h_{0,s_1}(t) \exp \left\{ \alpha_{s_1 \rightarrow s_2} + \beta^\top \mathbf{m}_i(t) + \gamma^\top \mathbf{z}_i \right\} \quad (4)$$

where $\alpha_{s_1 \rightarrow s_2}$ is a constant parameter that controls the rate of transitioning from s_1 to s_2 , β is a 2-dimensional regression coefficient for $\mathbf{m}_i(t)$ and γ is a n_z -dimensional regression coefficient for \mathbf{z}_i . By this formulation, the log hazard ratio of transitioning to s'_2 versus s_2 is

$$\log\{h_{i,s_1 \rightarrow s'_2}(t)/h_{i,s_1 \rightarrow s_2}(t)\} = \alpha_{s_1 \rightarrow s'_2} - \alpha_{s_1 \rightarrow s_2}.$$

3 Cohorts

Four different cohorts are studied in four separate analyses.

- The first analysis focused on the cohort of subjects who went through the state of MIAA positive (only). Those subjects are followed until entering one of the next states or censored: “MIAA & GADA”, “MIAA & IA2A”, “MIAA & ZNT8”, “3AB & more” or “T1D”. The corresponding directed graph is presented in Figure A1.
- The second analysis focused on the cohort of subjects who went through the state of GADA positive (only). Those subjects are followed until entering one of the next states or censored: “MIAA & GADA”, “GADA & IA2A”, “GADA & ZNT8”, “3AB & more” or “T1D”. The corresponding directed graph is presented in Figure A2.
- The third analysis focused on the cohort of subjects who went through the state of “MIAA & GADA”. Those subjects are followed until entering one of the next states or censored: “MIAA & GADA & IA2A”, “MIAA & GADA & ZNT8”, “MIAA & GADA & ZNT8 & IA2A” or “T1D”. The corresponding directed graph is presented in Figure A3.
- The last analysis focused on the cohort of subjects who went through the state of “MIAA & GADA & ZNT8”. Those subjects are followed until entering one of the next states or censored: “MIAA & GADA & IA2A”, “MIAA & GADA & ZNT8”, “MIAA & GADA & ZNT8 & IA2A” or “T1D”. The corresponding directed graph is presented in Figure A4.

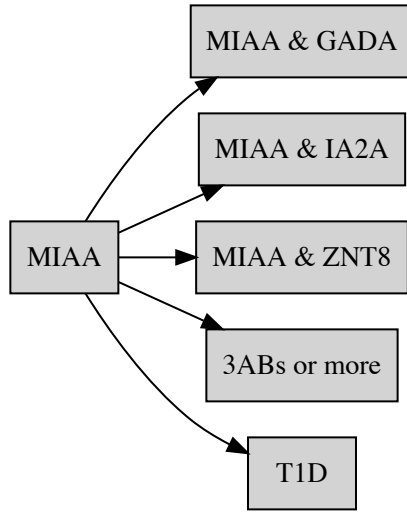


Figure A1: The directed graph considered for the MIAA cohort.

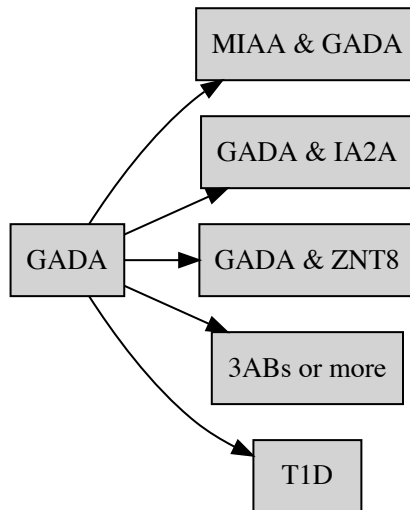


Figure A2: The directed graph considered for the GADA cohort.

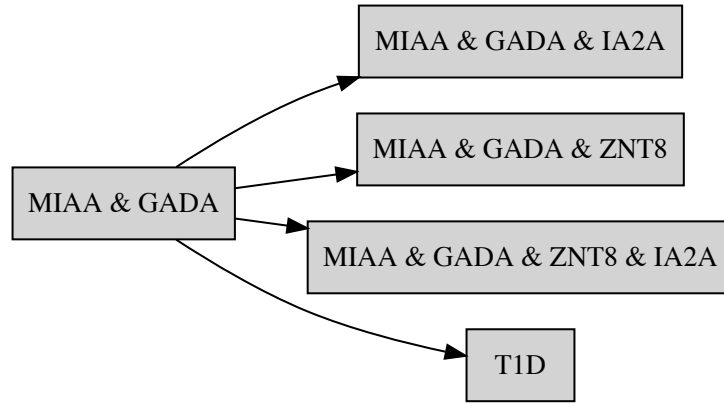


Figure A3: The directed graph considered for the GADA & MIAA cohort.

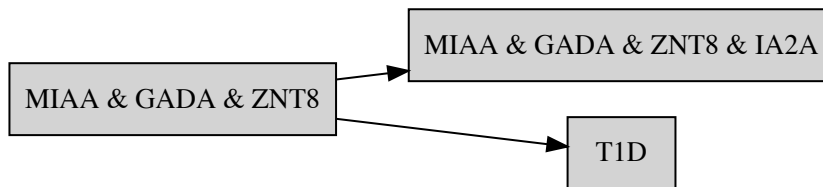


Figure A4: The directed graph considered for the GADA & MIAA & ZNT8 cohort.