## A. Supplementary Information of MIMIC-III Real Experiments

Below in Table 9, we list the complete first-order logic rules defined for sepsis patients in MIMIC-III. These logic rules are introduced following the guidance of experts (i.e., doctors), and serve as prior knowledge to our model.

### A.1. First-order Logic Rules defined for Sepsis Patients in MIMIC-III

*Table 9.* First-order Logic Rules defined for Sepsis Patients in MIMIC-III

| |
|---|
| $\texttt{GoodSurvivalCondition}(t) \leftarrow \texttt{Use-Drug}(t') \wedge \texttt{Before}(t', t)$, where $\texttt{Drug} = \{$Levofloxacin, Ceftriaxone, Meropenem, Ceftazidime, Tobramycin, Metronidazole, Vancomycin, Azithromycin, Ciprofloxacin, Piperacillin, Metoprolol Diltiazem, Norepinephrine, Furosemide$\}$ |
| $\texttt{GoodSurvivalCondition}(t) \leftarrow \texttt{Use-Drug1}(t') \wedge \texttt{Use-Drug2}(t'') \wedge \texttt{Before}(t', t) \wedge \texttt{Before}(t'', t) \wedge \texttt{Before}(t', t'')$, where $\texttt{Drug1} = \{$Levofloxacin, Ceftriaxone, Ceftazidime, Metronidazole, Azithromycin, Ciprofloxacin, Piperacillin$\}$ where $\texttt{Drug2} = \{$Meropenem, Tobramycin, Vancomycin$\}$ |
| $\texttt{NormalHeartRate}(t) \leftarrow \texttt{Use-Drug3}(t') \wedge \texttt{Before}(t', t)$ where $\texttt{Drug3} = \{$Metoprolol, Diltiazem, Norepinephrine$\}$ |
| $\texttt{NormalUrineOutput}(t) \leftarrow \texttt{Use-Furosemide}(t') \wedge \texttt{Before}(t', t)$ |
| $\texttt{NormalTemperature}(t) \leftarrow \texttt{Use-Acetaminophen}(t') \wedge \texttt{Before}(t', t)$ |
| $\texttt{GoodSurvivalCondition}(t) \leftarrow \bigwedge (\texttt{NormalSymptom}(t') \wedge \texttt{Equal}(t', t))$ where $\texttt{Symptom} = \{$BloodPressure, HeartRate, Temperature, RespiratoryRate, UrineOutput$\}$ |
| $\texttt{NormalSymptom}(t) \leftarrow \texttt{Use-Norepinephrine}(t') \wedge \texttt{Before}(t', t)$ where $\texttt{Symptom} = \{$HeartRate, BloodPressure$\}$ |
| $\neg\texttt{GoodSurvivalCondition}(t) \leftarrow \neg\texttt{NormalBloodPressure}(t') \wedge \texttt{Equal}(t', t)$ |

## B. Supplementary Information of USCD-FICO Fraud Detection Experiments

### B.1. First-order Logic Rules for Fraud Credit Card Transactions in UCSD-FICO

The list of first-order logic rules for fraud credit card transactions is shown in Table 10. The rules are summarized from domain knowledge (Delamaire et al., 2009) and preliminary data analyses of our data.

*Table 10.* First-order Logic Rules for Fraud Credit Card Transactions in UCSD-FICO

| |
|---|
| $\texttt{Is-Fraud}(t) \leftarrow \texttt{Has-FraudHistory}(t') \wedge \texttt{Before}(t', t)$ |
| $\texttt{Is-Fraud}(t) \leftarrow \texttt{Is-ZeroAmount}(t') \wedge \texttt{Equal}(t', t)$ |
| $\texttt{Is-Fraud}(t) \leftarrow \texttt{Is-SmallAmount}(t') \wedge \texttt{Equal}(t', t)$ |
| $\texttt{Is-Fraud}(t) \leftarrow \texttt{Has-ZeroAmountHistory}(t') \wedge \texttt{Before}(t', t)$ |
| $\texttt{Is-Fraud}(t) \leftarrow \texttt{Has-MultiZip}(t') \wedge \texttt{Before}(t', t)$ |
| $\texttt{Is-Fraud}(t) \leftarrow \texttt{Has-MultiCreditCard}(t') \wedge \texttt{Before}(t', t)$ |
| $\texttt{Is-Fraud}(t) \leftarrow \texttt{Is-LargeAmount}(t') \wedge \texttt{Equal}(t', t)$ |
| $\texttt{Is-Fraud}(t) \leftarrow \texttt{Has-LargeTransactionTimeGap}(t') \wedge \texttt{Before}(t', t)$ |

### B.2. More Exploratory Data Analyses of USCD-FICO data

By exploratory data analyses of USCD-FICO dataset, we also observe: (i) the "one-day transaction frequency " is distributed similarly between fraud data and normal data (shown in Figure 9); we therefore didn't include it as a factor, (ii) the "time gap between two transactions" exhibits subtle differences between fraud data and normal data (displayed in Figure 10); we incorporated this factor to the model but the learned logic weight didn't reflect its significance.

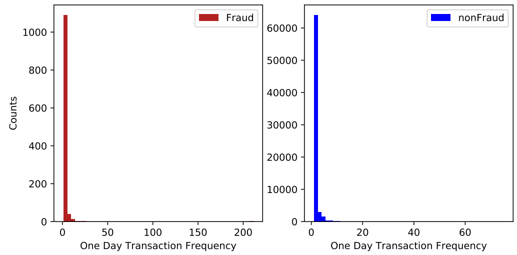


*Figure 9.* One-Day Transaction Frequency: Fraud v.s. Normal

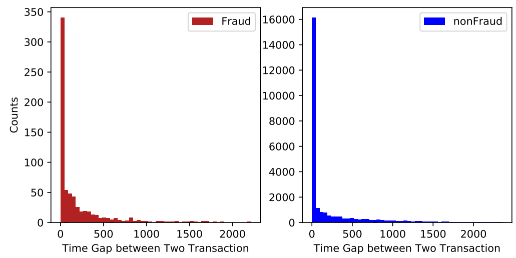


*Figure 10.* Time Gap Between Two Transactions: Fraud v.s. Normal.

## C. Proofs of the likelihood

For predicate $C$, given a realization of the process up to time $t$, as in Fig. 3(b), i.e., starts in state 0 and ends in state 1, the likelihood $\mathcal{L}(\{x_C(t)\}_{t\geq 0})$ can be derived as follows.

Let $p(t_{n+1}|\mathcal{H}_{t_n}, x_C(t_n) = 0)$ and $p(t_{n+1}|\mathcal{H}_{t_n}, x_C(t_n) = 1)$ be the conditional density function of the time of the next event $t_{n+1}$ given the history of previous events $(t_0, t_1, \cdots, t_n)$ while $x_C(t_n) = 0$, and $x_C(t_n) = 1$ respectively. Let $F(t|\mathcal{H}_{t_n}, x_C(t_n) = 0)$, and $F(t|\mathcal{H}_{t_n}, x_C(t_n) = 1)$ be the corresponding cumulative distribution function for any $t > t_n$. Based on the definition of the conditional transition intensity, we have

$$\lambda^*_C(t) = \frac{p(t|\mathcal{H}_{t_n}, x_C(t_n) = 0)}{1 - F(t|\mathcal{H}_t, x_C(t_n) = 0)},$$
$$\text{and} \quad \mu^*_C(t) = \frac{p(t|\mathcal{H}_{t_n}, x_C(t_n) = 1)}{1 - F(t|\mathcal{H}_{t_n}, x_C(t_n) = 1).} \tag{18}$$

From (18), we have

$$\lambda^*_C(t) = -\frac{d}{dt}\log(1 - F(t|\mathcal{H}_t, x_C(t_n) = 0)),$$
$$\mu^*_C(t) = -\frac{d}{dt}\log(1 - F(t|\mathcal{H}_t, x_C(t_n) = 1)).$$

Integrating both sides, we can get the conditional density and cumulative distribution function,

$$p(t|\mathcal{H}_{t_n}, x_C(t_n) = 0) = \lambda^*_C(t)\exp\left(-\int_{t_n}^{t}\lambda^*_C(s)ds\right),$$
$$F(t|\mathcal{H}_{t_n}, x_C(t_n) = 0) = 1 - \exp\left(-\int_{t_n}^{t}\lambda^*_C(s)ds\right),$$
$$p(t|\mathcal{H}_{t_n}, x_C(t_n) = 1) = \mu^*_C(t)\exp\left(-\int_{t_n}^{t}\mu^*_C(s)ds\right),$$
$$F(t|\mathcal{H}_{t_n}, x_C(t_n) = 1) = 1 - \exp\left(-\int_{t_n}^{t}\mu^*_C(s)ds\right).$$

Let $t_0 = 0$. Given the initial state $x_C(t_0) = 0$, and the history of the trajectory $(t_1, t_2, \ldots, t_n)$, where $x_C(t_n) = 1$, the likelihood function can be factorized into all the conditional densities of each points given all points before it, i.e., $\mathcal{L}$ is

$$p(t_1|\mathcal{H}_{t_0}, x_C(t_0) = 0)p(t_2|\mathcal{H}_{t_1}, x_C(t_1) = 1)\cdots$$
$$p(t_n|\mathcal{H}_{t_{n-1}}, x_C(t_{n-1}) = 0)(1 - F(t|\mathcal{H}_{t_n}, x_C(t_n) = 1))$$

Combined with the above conditional density and cumulative distribution function, the likelihood $\mathcal{L}(\{x_C(t)\}_{t\geq 0})$ is expressed as,

$$\mathcal{L}(\{x_C(t)\}_{t\geq 0}) = p(t_1|\mathcal{H}_{t_0}, x_C(t_0)=0)p(t_2|\mathcal{H}_{t_1}, x_C(t_1)=1)\cdots p(t_n|\mathcal{H}_{t_{n-1}}, x_C(t_{n-1})=0)(1-F(t|\mathcal{H}_{t_n}, x_C(t_n)=1))$$

$$= \lambda_C^*(t_1)\exp\left(-\int_0^{t_1}\lambda_C^*(s)ds\right)\cdot\mu_C^*(t_2)\exp\left(-\int_{t_1}^{t_2}\mu_C^*(s)ds\right)\cdots\lambda_C^*(t_n)\exp\left(-\int_{t_{n-1}}^{t_n}\lambda_C^*(s)ds\right)\cdot\exp\left(-\int_{t_n}^{t}\mu_C^*(s)ds\right),$$

which completes the proof.

Given other realizations of the predicate process $\{x_C(t)\}_{t\geq 0}$, such as (1) starts in state 0 and ends in state 0, or (2) starts in state 1 and ends in state 0, or (3) starts in state 1 and ends in state 1, the likelihood function can be derived accordingly.

The above derivations focus on the the process of only one predicate $x_C$. Furthermore, given joint processes of predicates, such as $\{x_A(t)\}_{t\geq 0}$, $\{x_B(t)\}_{t\geq 0}$, and $\{x_C(t)\}_{t\geq 0}$, the joint likelihood would be a product of $\mathcal{L}(\{x_A(t)\}_{t\geq 0})$, $\mathcal{L}(\{x_B(t)\}_{t\geq 0})$, and $\mathcal{L}(\{x_C(t)\}_{t\geq 0})$.

## D. Data preprocessing in MIMIC-III for state definition

For drugs, 1 means the treatment is applied, and 0 otherwise. For symptoms, the thresholds to distinguish normal and abnormal states are: abnormal Heart Rate (> 100/min), abnormal Urine Volume (< 0.8L/24h), abnormal Temperature (> 37.2°C or < 36.1°C ), abnormal Respiratory Rate (> 20/min), and Low Blood Pressure (< 70mmHg (MAP)). We removed the patients whose drug-usage was less are 10 times, and the dataset was reduced to 5,002 patients (with mean age $66.6$, $43.6\%$ female). We randomly split the data as training and test data. For the survival condition, 1 indicates survival and 0 otherwise.

## E. Algorithm

Given the observed predicate processes, the likelihood function can be evaluated using the three modules:

(*i*) *Compute Template*, an offline computational component. This computational component outputs the valid state combinations of evidence, which are informed by the first-order logic rules without the temporal predicates. The templates are independent of data and thus can be computed offline without data. The obtained templates will be used to extract evidence from data to further compute formula effect $\delta_{f_i}(t)$.

(*ii*) *Compute Intensity*, an online computational component, which computes the conditional transition intensity $\lambda^*(t)$ or $\mu^*(t)$.

(*iii*) *Compute Likelihood*, an online computational component, which computes the objective function of our method.

The pseudocode of the three computational components is summarized in the following Algorithm boxes respectively. For *Compute Intensity* and *Compute Likelihood*, we will use MIMIC-III as an example to illustrate the algorithm. To reduce the computational burden, we consider a time-window version to compute the intensity and the likelihood, which means we only search for evidence within a prespecified time window.

**Algorithm 1** Compute Template

**Input:** The set of logic formulas $\mathcal{F} = \{f_1, f_2, ..., f_n\}$ without considering the temporal constraints. For each logic $f_i$, the head predicate, denoted as $X_h^{f_i}$; and the corresponding body predicate(s), denoted as $\{X_b^{f_i}\}$.
**Output:** `Template`, a dictionary with key formula $f_i$.

**for** $f_i$ in $\mathcal{F}$ **do**
  $\mathcal{C} \leftarrow \{0,1\}^{|\{X_b^{f_i}\}|}$ ▷ Obtain all possible state combinations of body predicates in $f_i$
  $x_h^{f_i} = \{0\}$ ▷ Set head predicate's current state either 0 or 1
  **for** each body predicate combination $\{x_b^{f_i}\} \in \mathcal{C}$ **do**
    $\Delta_{f_i} \leftarrow f_i(1 - x_h^{f_i}, \{x_b^{f_i}\}) - f(x_h^{f_i}, \{x_b^{f_i}\})$ ▷ Compute the formula effect, without considering temporal predicates
    **if** $\Delta_{f_i} \neq 0$ (either 1 or -1) **then**
      `Template`[$f_i$][Body Predicate] = $X_b^{f_i}$
      `Template`[$f_i$][Body Predicate's Valid State Combination] = $x_b^{f_i}$
    **end if**
  **end for**
**end for**
Return `Template`

**Algorithm 2** Compute Likelihood (MIMIC-III Example)

**Input** A list of event data recorded as $\mathcal{D} = \{(u, X_d, t_n, x_d(t_n))\}$ where $u$ is patient ID, $X_d$ is predicate ID, $t_n$ is the transition time, and $x_d(t_n)$ is current state at time $t_n$. Logic weight parameters $\{w_f\}$ and temporal predicate parameters $\beta$, $\gamma$. Head predicate set $\mathcal{X}_h$. The set of logic formulas $\mathcal{F}$ (with temporal constraints). `Template` dictionary obtained from Algorithm 1. Integral grid $y$. Initialize log-likelihood $\ell = 0$. Time window $t_w$.

**for** each patient $u$ **do**
  **for** each head predicate $X_h \in \mathcal{X}_h$ **do**
    ▷ Compute the intensity at transition times:
    **for** $t$ in transition times $\{t_1, t_2, \ldots, t_n, \ldots\}$ **do**
      $\lambda_{X_h}(t_n) \leftarrow$ ComputeIntensity $(u, X_h, t_n, x_h(t_n))$
      ▷ Update log-likelihood:
      $\ell \leftarrow \ell + \log(\lambda_{X_h}(t_n))$
    **end for**
    ▷ Compute the intensity at grid points to approximate integration:
    **for** $t$ in grid points $\{0, y, 2y, \ldots\}$ within horizon **do**
      $\lambda_{X_h}(t) \leftarrow$ ComputeIntensity $(u, X_h, t, x_h(t))$
      ▷ Update log-likelihood:
      $\ell \leftarrow \ell$ - $\lambda_{X_h}(t) y$
    **end for**
  **end for**
**end for**
Return $\ell$

**ComputeIntensity** $(u, X_h, t, x_h(t))$:
**for** $f_i$ in $\mathcal{F}_{X_h}$ **do**
  $\{X_b\} \leftarrow$ `Template`[$f_i$][Body Predicate] ▷ Obtain body predicates
  $\{x_b\} \leftarrow$ `Template`[$f_i$][Body Predicate's Valid State Combination] ▷ Obtain body predicate's valid states
  $\phi_{f_i}(t) = 0$ ▷ Initialize feature
  **for** $X \in \{X_b\}$ **do**
    $t_{valid}^X \leftarrow (t - t_w < t_n^X < t) \wedge x(t_n^X) = x_b$ ▷ $t_n^X$ is the transition time of predicate $X$ for patient $u$
  **end for**
  $\mathcal{T} \leftarrow \bigotimes\{t_{valid}^X\}$ ▷ Obtain all combinations of valid transition times
  **for** $\{t_b\} \in \mathcal{T}$ **do**
    $\delta_{f_i}(t \mid \{t_b\}, \{x_b\}) = f_i(1 - x_h(t), t, \{x_b\}, \{t_b\}) - f_i(x_h(t), t, \{x_b\}, \{t_b\})$
    $\phi_{f_i}(t) \leftarrow \phi_{f_i}(t) + \delta_{f_i}(t \mid \{t_b\}, \{x_b\})$
  **end for**
**end for**
$\lambda_{X_h}(t) \leftarrow \exp(\sum_{f_i \in \mathcal{F}_{X_h}} w_i \phi_{f_i}(t))$
Return $\lambda_{X_h}(t)$