that can only be detected at a monitoring time; for example, $T$ might be the time of onset of a tumor or the time at which the CD4 count of an AIDS patient drops below a particular value. At the monitoring time, one can find out whether $T$ happened, and if it happened one might be able to determine the precise value of $T$ (or use an extrapolated approximation) between the last and current monitoring times. In this case, we have no reporting delay for the $U_j$'s (i.e., $A_j = U_j$), $j = 1, \ldots, k - 1$, but $T$ is reported at the monitoring time $A_k$ following $T$. If the precise value of $T$ between two subsequent monitoring times can only be guessed, then in these kinds of applications it is also common practice to apply the Kaplan–Meier estimator to the guessed $T$'s and be satisfied with estimation of their distribution. In both situations, the Kaplan–Meier estimator can be expected to be biased due to the “reporting delay” of $T$, while the data structure above will acknowledge this reporting delay phenomenon.

In van der Laan and Hubbard (1998), locally efficient estimators of the survival function have been developed and implemented analogously to the methods presented in this chapter.

### 3.2.4 Univariately right-censored multivariate failure time data

Consider a longitudinal study in which various time variables $\bar{T} = (T_1, \ldots, T_k)$ on the subject are of interest. For example, consider a study in which HIV-infected subjects have been randomized to treatment groups. In such a study, one might be concerned with comparing the multivariate treatment-specific survival functions of time from the beginning of the study until AIDS diagnosis, death, and time until particular AIDS-related events (e.g., types of opportunistic infections). As a second example, one might be interested in the bivariate survival function of time until recurrence of cancer (measured from extraction of tumor) and time until death. In this setting, the researcher could also have interest in the estimation of functions of the joint distribution, such as the distribution of the gap time $T_2 - T_1$ from recurrence to death. We will not require that time variables $T_1, \ldots, T_k$ be ordered. Let $L(t)$ represent a time-dependent covariate process that one measures on the subject over time. This process includes the baseline covariates $L(0)$. The full data on a subject is defined as $X = (\bar{T}, L(T))$, where $T = \max(T_1, \ldots, T_k)$.

Let $C$ be the common right-censoring time, which could be the minimum of time until end of study or time until dropout of the subject. Each subject is observed until $\bar{T} = \min(T, C)$. Let $T_j = \min(T_j, C)$, $j = 1, \ldots, k$. Thus, the researcher observes, for each subject, the following data structure:

$$Y = (\bar{T}_1, \ldots, \bar{T}_k, \Delta_1 = (T_1 < C), \ldots, \Delta_k = (T_k \leq C), L(\bar{T})).$$ (3.5)