has a normal life at time t and $V(t) = 0$ if the subject is seriously suffering from the treatment (e.g., chemotherapy) at time t.

In many applications, one will observe additional information on the subject in terms of baseline covariates and time-dependent measurements (e.g., CD4 count in an AIDS application). Let $W(t) \in \mathbb{R}^k$, $t \in \mathbb{R}_{\geq 0}$ be a covariate process. We will denote the time-independent covariates with $W(0)$. Now, the full data process of interest is $X(t) = (R(t), V(t), W(t))$, where $R(t) = I(T \leq t)$ and the full data structure is $X = \bar{X}(T) = \{X(s) : s \leq t\}$. Note that observing X implies observing U. Therefore, if each subject is observed until death, then the natural nonparametric estimator of the quality-adjusted survival function $S_U(t) \equiv P(U > t) = 1 - F_U(t)$ is the empirical survival function based on $U_1, \ldots, U_n$. Due to limited follow-up time or other reasons, one will not always observe the process X up until time T; one observes the process X up until the minimum of a censoring time C and T, and one knows whether this minimum is either the censoring time or T. Thus, the observed data structure can be represented as

$$Y = (\tilde{T} \equiv C \wedge T, \Delta = I(\tilde{T} = T), \bar{X}(\tilde{T})), \tag{3.2}$$

which thus corresponds with the data structure studied in this chapter. We observe n independent and identically distributed observations $Y_1, \ldots, Y_n$ of Y. As in the previous example, possible parameters of interest are the marginal distribution of U and regression parameters in Cox proportional hazards models or linear regression models with outcome $\log(U)$. Gelber, Goldhirsch, and Cavalli (1991) discuss quality-of-life data concerning operable breast cancer. The study compares a single cycle of adjuvant chemotherapy compared with longer -duration chemotherapy for premenopausal women or chemoendrocrine therapy for postmenopausal women. To define the quality-adjusted lifetime, they considered three health states: time without systems and toxicity (TWiST), toxicity (TOX), and survival time after relapse (REL). They weighted the time spent in each category according to subjective judgments as to the quality of life in each state. Specifically, TWiST was weighted as 1, with the other states naturally having less weight. Their goal was to compare the efficacy of different treatment regimes on the weighted sum of the times, known as the quality-adjusted lifetime, and the parameter of interest to measure this was the mean quality-adjusted lifetime in each treatment group. In addition to Gelber, et al. (1991), many statistical analyses of lifetime data from clinical trials have been concerned with inference on the quality-adjusted lifetime distribution (e.g., Gelber, Gelman, and Holdhirsch, 1989; Gelber et al., 1992; Glasziou, Simes and Gelber, 1990; Goldhirsch et al., 1989; and Korn, 1993). In these studies, one observes for each subject a quality-of-life state process up to the minimum of chronological death, and censoring and weights are assigned to each of the states given a priori (an overview of this type of data is given by Cox et al., 1992). Zhao and Tsiatis (1997) proposed

an ad hoc estimator of the quality-adjusted lifetime distribution itself for the marginal data structure.

Locally efficient estimators of treatment-specific marginal quality-adjusted survival functions have been developed and implemented in van der Laan and Hubbard (1999), analogous to the methods in this chapter (and Chapter 6).

3.2.3 *Right-censored data on a survival time with reporting delay*

Consider a study in which the survival time T of a subject is of interest. In most central data registries of certain types of patients (e.g., AIDS, cancer), data on a subject are reported with delay (e.g., Bachetti, 1996). This means that if a subject has not been reported dead at time t, then that might either mean that the subject is still alive or that the subject died before time t but that the death has not been reported yet.

Let $R(t) = I(T \leq t)$ represent the vital status of a subject at time t. Let V_1 report at time t until when the process R has been observed; If, at time t, R has been observed up to time $s \leq t$, then $V_1(t) = s$. In particular, if, at time t, T is already observed, then we have $V_1(t) = t$. Thus V_1 is an increasing function with $V_1(t) \leq t$.

One can describe V_1 in terms of monitoring times $U_1 < U_2 < \ldots < U_{k-1} < T$ and reporting times $A_1 < A_2 < \ldots < A_{k-1}$ of the subject under study. Here A_j is the time at which the vital status of the subject at time U_j is reported, $j = 1, \ldots, k-1$, and let A_k be the time at which T is reported, so at time A_j we know that $T > U_j$, $j = 1, \ldots, k-1$ and at time A_k we know that $T = t$ for some $t \leq A_k$. In the context of reporting delays as described above, we have

$$V_1(t) = \begin{cases} U_j & \text{if } t \in [A_j, A_{j+1}) \\ t & \text{if } t \geq A_k. \end{cases}$$

In longitudinal studies, one will typically also collect time-dependent covariates over time. Let $W(t)$, $t \in \mathbb{R}_{\geq 0}$, be a covariate process that is assumed to be subject to the same reporting delays as the vital status of T. The function V_1 provides us with a natural definition of the data observed on a subject up to time t. Consider the process

$$X(t) = (R(V_1(t)), V_1(t), W(V_1(t))). \tag{3.3}$$

Thus, observing the process X up to time t corresponds with observing the vital status process R, the ascertainment process V_1, and the covariate process W up to time $V_1(t)$. Let $\bar{X}(t) = \{X(s) : s \leq t\}$ represent the sample path of X up to time t. The data analyst observes this process $X(t)$ in the sense that at time t the computer contains the process X up to time t, assuming censoring occurs after t.