Quantifying the association between progression-free survival and overall survival in oncology trials using Kendall's $\tau$

Enya M. Weber | Andrew C. Titman

Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

Correspondence
Enya M. Weber, Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK.
Email: e.weber@lancaster.ac.uk

Funding information
European Union’s Horizon 2020 research and innovation programme, Grant/Award Number: 633567

This paper considers methods for estimating the association between progression-free and overall survival in oncology trials. Copula-based, nonparametric, and illness-death model-based methods are reviewed. In addition, the approach based on an underlying illness-death model is generalized to allow general parametric models. The performance of these methods, in terms of bias and efficiency, is investigated through simulation and also illustrated using data from a clinical trial of treatments for colon cancer. The simulations suggest that the illness-death model-based method provides good estimates of Kendall's $\tau$ across several scenarios. In some situations, copula-based methods perform well but their performance is sensitive to the choice of copula. The Clayton copula is most appropriate in scenarios, which might realistically reflect an oncology trial, but the use of copula models in practice is questionable.

KEYWORDS
copula, inverse probability of censoring weights, Kendall's $\tau$, multistate model, overall survival, progression-free survival

1 | INTRODUCTION

In assessing the performance of therapies in cancer trials, different clinical endpoints can be considered. The most reliable and commonly used primary endpoint is overall survival (OS), which is defined as the time from study entry to death from any cause. However, the occurrence of progression, defined as the growth or spread of cancer, can also be considered as a possible endpoint. Progression-free survival (PFS) is defined as the time from study entry until progression or death, depending on what occurs first. In some cancer therapies, PFS has been accepted as a suitable surrogate endpoint for OS, especially in earlier phases of the drug development.1,2 Defining PFS as the primary endpoint for analyzing survival times in a trial can be efficient in terms of costs and time because potentially long follow-up periods after disease progression can be avoided.

Quantifying the association between PFS and OS in cancer trials can provide an indication of the extent to which PFS may be an effective surrogate for OS. Estimating the correlation between PFS and OS may be a potential step for validating PFS as a surrogate endpoint for OS, usually accompanied by metaregression to also establish correlation between treatment effects across multiple studies.3,4 Existing approaches to quantifying the correlation between PFS and OS include parametric and semiparametric copula models5,6 as well as nonparametric methods based on inverse probability of censoring weights (IPCWs).7 Recently, there has also been a focus on illness-death model-based methods for quan-
The correlation between progression-free survival (PFS) and overall survival (OS) is a crucial aspect in clinical trials, especially in cancer research. Many studies have attempted to quantify the association between these two endpoints. However, traditional measures like Pearson’s correlation coefficient have limitations, particularly in handling non-normal distributions and censoring. Copula models offer a versatile framework to address these issues.

### 2 | Different Approaches to Estimating the Correlation Between Progression-Free Survival and Overall Survival

#### 2.1 | General setting

In this section, we describe existing approaches to quantifying the association between PFS and OS. Throughout, it is assumed that the desired measurement of the association between PFS and OS is the Kendall’s τ rank correlation. Let \( (S_1, T_1) \) and \( (S_2, T_2) \) be random variables representing the PFS and OS times for two independent patients 1 and 2. Kendall’s τ is defined as

\[
\tau = P\{(S_1 - S_2)(T_1 - T_2) > 0\} - P\{(S_1 - S_2)(T_1 - T_2) < 0\},
\]

where the first term refers to the probability of concordance and the second to the probability of discordance.

#### 2.2 | Copula models for bivariate survival data

An existing approach for the measurement of the dependence structure between failure times PFS and OS is through the use of copula models. Burzykowski et al. proposed a parametric copula method for quantifying the correlation between PFS and OS. A semiparametric approach is also possible using the methods developed by Shih and Louis for semiparametric bivariate survival copulas. Generally, the idea of these models is to estimate the marginal distributions for PFS and OS and to impose a particular dependence structure between these two endpoints.

Copulas are continuous multivariate distributions where each of its variables follows a uniform marginal distribution. According to Sklar’s theorem, any arbitrary multivariate joint distribution can be expressed through its marginal distribution and a copula function separately. The copula function includes all the information of the dependence between the endpoints independently from their marginal distribution.

The joint survival function of the failure endpoints \( (S, T) \), where \( S : \text{PFS} \) and \( T : \text{OS} \), can be expressed by

\[
S(s, t) = P(S \geq s, T \geq t) = C(S_s(s), S_T(t)), \quad s, t \geq 0,
\]

where \( S_s \) and \( S_T \) are the marginal survivor functions of \( S \) and \( T \), respectively, and \( C : [0, 1]^2 \rightarrow [0, 1] \) is a bivariate copula function. There are a wide range of copula families allowing different patterns of dependency. In this paper, we consider the Clayton, Hougaard, and Frank copula functions, all of which belong to the class of Archimedean copulas. This class is often used as it allows the dependence between the two variables to be defined by a single parameter, \( \delta \). In general, a bivariate distribution in terms of the Archimedean copula family is given by

\[
C(u, v) = \phi_\delta \left[ \phi_\delta^{-1}(u) + \phi_\delta^{-1}(v) \right], \quad 0 \leq u, v \leq 1,
\]

where \( 0 \leq \phi \leq 1, \phi(0) = 1, \phi' < 0, \) and \( \phi'' > 0 \). If \( \phi_\delta \) is a Laplace transform of some distribution, then the corresponding Archimedean copula is equivalent to a proportional frailty model. Three special cases of the proportional frailty model are of interest here. First, Clayton’s model can be represented as

\[
C_\delta(u, v) = \left( u^{(1-\delta)} + v^{(1-\delta)} - 1 \right)^{1/(1-\delta)}, \quad \delta > 1,
\]
where $\phi_\delta(x) = (1 + x)^{1/(1 - \delta)}$ is the Laplace transform of a gamma distribution with rate parameter 1 and shape parameter $1/(\delta - 1)$. $S$ and $T$ are positively associated when $\delta > 1$ and become independent as $\delta \to 1$. The second example is Hougaard’s model,\textsuperscript{13} where the function is given by

$$C_\delta(u, v) = \exp\left(-[-\log(u)^{(1/\delta)} - \log(v)^{(1/\delta)}]\right)^\delta, \quad 0 < \delta < 1,$$

where $\phi_\delta(x) = \exp(-x^\delta)$ is the Laplace transform of the positive stable distribution with density

$$-\frac{1}{\pi x^\delta} \sum_{k=1}^\infty \frac{\Gamma(k\delta + 1)}{k!} (-x^{-\delta})^k \sin(\delta k\pi), x > 0 \text{[13]}.$$

$S$ and $T$ are positively associated when $\delta$ is small and become independent when $\delta \to 1$.

The copula function of Frank’s model\textsuperscript{14} can be expressed as

$$C_\delta(u, v) = \log_\delta \left[1 + \frac{(\delta^u - 1)(\delta^v - 1)}{\delta - 1}\right], \quad \delta > 0,$$

where $\phi_\delta(x) = \log_\delta(1 - (1 - \delta) \exp(-x))$ is the Laplace transform of a logarithmic series distribution for $0 < \delta < 1$ and $\log_\delta$ represents the logarithm to the base $\delta$. $S$ and $T$ are positively associated for the case $\delta < 1$, negatively associated for the case $\delta > 1$, and become independent when $\delta \to 1$.

Burzykowski et al\textsuperscript{5} presented a fully parametric copula model. The marginal distributions are assumed to have a particular parametric form, for instance, they may be assumed to each have separate Weibull distributions. Combining the marginal distributions for survival with a copula function, a corresponding bivariate joint survival copula function based on Weibull distributions can be constructed. As shown in the work of Burzykowski et al,\textsuperscript{5} a likelihood function can be derived by taking all potential censoring cases in the datasets into account. The scale and shape parameters in the two hazard functions as well as the copula dependence parameter $\delta$ can be jointly estimated by using maximum likelihood estimation.

Shih and Louis\textsuperscript{6} proposed a semiparametric model in which a parametric copula is assumed for the dependence, but the marginal distributions are left unspecified. A two-stage approach is taken for estimation. The idea is to estimate the marginal survivor functions $(u, v)$ by the nonparametric Kaplan-Meier estimator in the first stage. After deriving the likelihood function incorporating the different cases of censoring,\textsuperscript{6} maximum likelihood can be used to estimate the unknown copula association parameter $\delta$ conditional on the values of the survival functions of PFS and OS.

A particularly useful aspect of Archimedean copula methods is that Kendall’s $\tau$ can be expressed directly as a function of $\phi_\delta^{-1}(x)$ as follows (see the work of Genest and MacKay\textsuperscript{15}):

$$\tau = 1 + 4 \int_0^1 \frac{\phi_\delta^{-1}(v)}{\partial \phi_\delta^{-1}(v)/\partial v} dv. \quad (8)$$

As a direct relationship exists between the Kendall rank correlation $\tau$ and $\delta$, an estimate of $\tau$ can be derived from the maximum likelihood estimator of $\delta$. Application of (8) for the respective generator functions leads to Kendall’s $\tau$ for Clayton’s copula of $\tau = \frac{1}{\delta + 1}$, for Hougaard’s copula $\tau = 1 - \delta$, and for Frank’s copula of $\tau = 1 - 4 \frac{D_1(-\log(\delta)^{-1})}{\log(\delta)},$ where $D_1$ represents the Debye function of order 1.\textsuperscript{16}

While copula methods are very convenient for modeling bivariate survival data, their specific use for modeling PFS and OS is potentially questionable. By definition, the PFS time must be less than or equal to the OS time. However, the copula model does not restrict the ordering of PFS and OS. Moreover, if PFS and OS are assumed to have continuous marginal distributions, the copula model assumes the values of PFS and OS coincide with probability 0, whereas, in fact, they will coincide whenever a patient dies before progression. As a consequence, the copula model is guaranteed to be somewhat misspecified, even if the marginal distributions are correctly specified. There is therefore potential that estimates of dependency from the copula model will be biased. Dejardin et al\textsuperscript{17} performed a limited simulation study to investigate the possible bias of applying a bivariate shared Gamma frailty model (equivalent to the Clayton copula model) to data on PFS and OS in which a three-state unidirectional model was assumed for the generation process. They identified a small bias in the estimate of Kendall’s $\tau$, with the magnitude of bias being greater for the scenario with lower $\tau$.

### 2.3 Nonparametric methods based on IPCW

Nonparametric estimation of Kendall’s $\tau$ for censored data is possible through the use of IPCW.\textsuperscript{7} The methods are applicable to general bivariate survival data, including PFS and OS times as a special case where some of the calculations are...
simplified if it is assumed that a common censoring time will apply to both the PFS and OS times for a given patient. Let \((S_i, T_i)\) for every \(i = (1, \ldots, n)\) be independent replications of the failure endpoint times \((S, T)\). A pair of two replications can be seen as the survival experience from two individuals. The concordance or discordance status of the pairs is required for the empirical calculation of Kendall’s \(\tau\). The concordance status for subjects \(i\) and \(j\) is given by

\[
C_{ij} = \begin{cases} 
1, & \text{if } (S_i - S_j)(T_i - T_j) > 0 \\
-1, & \text{if } (S_i - S_j)(T_i - T_j) < 0.
\end{cases}
\]

In the case of no censoring, the concordance status can be determined for all \(\binom{n}{2}\) possible pairs. Hence, Kendall’s \(\tau\) can be estimated by its sample version

\[
\tau = \left(\binom{n}{2}\right)^{-1} \sum_{i<j} C_{ij},
\]

where summing over \(i < j\) avoids taking a pair into account twice.

In the presence of censored data, the concordance or discordance status can only be determined for orderable pairs. Let \(R_{ij}\) be an indicator of whether the pair \((i, j)\) is orderable. Defining the respective censoring times \(C_i\) and \(C_j\) for subjects \(i\) and \(j\), then \(R_{ij}\) is given by

\[
R_{ij} = I(\tilde{S}_{ij} < \tilde{C}_{ij}, \tilde{T}_{ij} < \tilde{C}_{ij}) = \begin{cases} 
1, & \text{if pair orderable} \\
0, & \text{otherwise}.
\end{cases}
\]

where \(\tilde{S}_{ij} = \min(S_i, S_j), \tilde{T}_{ij} = \min(T_i, T_j),\) and \(\tilde{C}_{ij} = \min(C_i, C_j).\)

Oakes\(^{18}\) extended the estimator in (9) to an estimator for \(\tau\) by taking the sum over the orderable pairs only. The IPCW technique can also be applied to the Oakes estimator in order to correct the bias caused by the presence of missing data. The contribution of each orderable pair to the Kendall’s \(\tau\) is weighted by the inverse probability of being orderable. Then, the estimator can be represented as follows\(^7\):

\[
\tau = \frac{\sum_{i<j} R_{ij} C_{ij} W_{ij}}{\sum_{i<j} R_{ij} W_{ij}} \in [-1, 1],
\]

where \(W_{ij} = \frac{1}{p_{ij}}\) are the weights defined by the inverse estimated selection probabilities for orderable pairs \(p_{ij}\) as follows:

\[
p_{ij} = P(R_{ij} = 1|\tilde{S}_{ij}, \tilde{T}_{ij}) = P(\tilde{S}_{ij} < \tilde{C}_{ij}, \tilde{T}_{ij} < \tilde{C}_{ij}|\tilde{S}_{ij}, \tilde{T}_{ij}) = G(\max(\tilde{S}_{ij}, \tilde{T}_{ij}))^2,
\]

where \(G(\cdot)\) is the survival function of censoring, which can be estimated, under an assumption of random censoring, via a Kaplan-Meier estimate obtained by reversing the censoring indicator. While in the above we assume a common censoring time for PFS and OS, in many oncology trials, the progression time is effectively censored at the last screening time. The methods in the work of Lakhal et al.\(^7\) allow for different but dependent censoring times for the two event times. However, note that censoring for time-to-progression is not the same as PFS since patients who die after their last screening time will not be treated as censored. Hence, the case of differential censoring times is not easily accommodated.

The IPCW approach to estimating the association between PFS and OS is potentially quite attractive as it requires no assumptions to be made about either the dependence structure or the marginal distributions of the times to PFS and OS. However, for consistency, the IPCW method requires that \(P(R_{ij} = 1|\tilde{S}_{ij}, \tilde{T}_{ij}) > 0\) for all potential \(\tilde{S}_{ij}, \tilde{T}_{ij}\). Effectively, this means that there must be some chance of a pair of observations being orderable regardless of the times until progression or death. For this to be the case, the support of the distributions of PFS and OS must be contained within the support of the censoring distribution. Such an assumption is very unlikely to be plausible in most oncology trials, where at the time of analysis, the maximum follow-up time will typically be shorter than the longest possible survival time. The bias in the estimate of \(\tau\) for data with limited follow-up will depend on how much of the upper tails of the PFS and OS distributions are not observable and how representative the dependence in the body of the distribution is to that of the tails. In terms of future work, the extent of bias in estimation of \(\tau\) when follow-up is limited and how the bias depends on the type of the censored data could be explored. This behavior could be investigated through simulation by generating data in a greater range of censoring scenarios to understand how shorter follow-up times affect the estimate of \(\tau\).
2.4 Model-based methods

The survival process in patients with cancer can be expressed in terms of a three-state illness-death model where the states correspond to pre-progression (assumed to apply at the time of randomization), progression, and death. PFS corresponds to the time of exit from the pre-progression state, whereas OS corresponds to the time of entry into the death state.

Fleischer et al.\(^8\) presented a parametric multistate model describing the Pearson correlation between survival outcomes PFS and OS under the assumption that the transition intensities between the states in the underlying multistate model are constant. More recently, Li and Zhang\(^9\) extended the method by using Weibull hazard functions to describe the transition intensities, therefore allowing them to either be monotonically increasing or decreasing with time. Their model corresponds to a *homogeneous semi-Markov* model where the hazard of death after progression depends on time since progression rather than time since randomization. It may be considered as a special case of the *general semi-Markov* model\(^19\) where the transition intensity may depend both on the duration in the progression state and the time since randomization. The model has four parameters, \(\lambda_1, \lambda_2, \lambda_3,\) and \(\alpha,\) with the model in the work of Fleischer et al.\(^8\) arising as a special case where \(\alpha = 1.\) Li and Zhang expressed the model in terms of the distributions of latent event times: time to progression, survival before progression, and time from progression to death. However, expressing the model in terms of the transition intensities is more desirable since it avoids making untestable assumptions about the independence between time to progression and survival before progression.\(^20\)

The three-state model is depicted in Figure 1. In the model of Li and Zhang, the transition intensities are given by

\[
\begin{align*}
\pi_{01}(t) &= \lambda_1 \alpha t^{\alpha-1}, \\
\pi_{02}(t) &= \lambda_2 \alpha t^{\alpha-1}, \\
\pi_{12}(s) &= \lambda_3 \alpha s^{\alpha-1},
\end{align*}
\]

where \(t\) and \(s\) refer to time since randomization and to time since progression, respectively.

Estimation of the correlation between PFS and OS involves first estimating the parameters of the multistate model via maximum likelihood. Li and Zhang derived a closed-form expression for the Pearson correlation between PFS and OS for given parameters, into which the maximum likelihood estimates can be substituted.

In order to ensure the existence of a closed-form expression for the Pearson correlation, Li and Zhang assumed the same shape parameter \(\alpha\) for the three Weibull functions. However, the necessity for analytical tractability leads to a somewhat restrictive model. For instance, in real data examples, the hazard of progression may increase with time whereas the hazard of death before progression may be close to constant. Furthermore, PFS and OS as time-to-event outcomes will typically be highly positively skewed, and as a consequence, a nonlinear dependence between PFS and OS would be expected, meaning the Pearson coefficient is unlikely to be an appropriate measure of association.

Due to these aspects, we extend the method to allow estimation of the Kendall rank correlation coefficient for general parametric illness-death models.

2.5 Generalized model-based methods

In this section, we generalize the illness-death model–based approach of Li and Zhang\(^9\) to achieve more flexibility and to allow estimation of Kendall’s \(\tau\) rather than the Pearson correlation coefficient.

The modified approach continues to use a multistate illness-death model but allows any parametric formulation for the transition intensities between states. In particular, we can allow post-progression survival to depend on both time to progression, denoted by \(t_0,\) and time since progression, denoted by \(s.\) We assume \(\pi_{01}(t), \pi_{02}(t),\) and \(\pi_{12}(s; t_0)\) are parameterized by a vector of parameters \(\theta,\) which can be consistently estimated from data with a finite follow-up period through maximum likelihood estimation.
From the definition of Kendall’s \( \tau \) in (1) and under an assumption that the bivariate lifetime random variables \((S_n, T_n)_{n \in \mathbb{N}}\) representing the PFS and OS times of the \( n \) patients are independent and identically distributed, for the general illness-death model, the Kendall’s \( \tau \) implied by the model is as follows:

\[
\tau_{\text{mod}} = 4 \int_0^\infty \pi_{02}(s) \exp(-2\Pi_0(s)) ds + 4 \int_0^\infty \int_0^{s_1} \int_0^\infty \pi_{01}(s_1) \pi_{01}(s_2) \exp(-\Pi_0(s_1) - \Pi_0(s_2)) \\
\times \pi_{12}(s_3; s_1) \exp(-\Pi_{12}(s_3; s_1))[1 - \exp(-\Pi_{12}(s_1 + s_3 - s_2; s_2))] ds_3 ds_2 ds_1 \\
+ 4 \int_0^\infty \int_0^{s_1} \pi_{02}(s_1) \pi_{01}(s_2) \exp(-\Pi_0(s_1) - \Pi_0(s_2))(1 - \exp(-\Pi_{12}(s_1 - s_2; s_2))) ds_2 ds_1 - 1, \tag{10}
\]

where \( \Pi_0(t) = \int_0^t \pi_{01}(u) + \pi_{02}(u) du \) and \( \Pi_{12}(s; v) = \int_0^s \pi_{12}(u; v) du \). The first term in (10) refers to the case where one patient dies before progression, before the other has died or progressed. The second term refers to the case where patients 1 and 2 progress at times \( s_1 \) and \( s_2 \), respectively, where \( s_1 > s_2 \), and subsequently, patient 1 survives an additional \( s_3 \), whereas patient 2 dies within \( s_1 - s_3 + s_2 \) of progression. The third term refers to the case where patient 1 progresses and dies before patient 2, despite patient 2 dying without progression. A full derivation of (10) is given in the Appendix.

For the model of Fleischer et al, by substituting \( \pi_{01}(t) = \lambda_1, \pi_{02}(t) = \lambda_2, \) and \( \pi_{12}(t; t_0) = \lambda_3 \) into (10) and directly integrating, after some algebraic manipulation, we obtain

\[
\tau_{\text{mod}} = \frac{\lambda_1^2 + 2\lambda_1\lambda_2 + 2\lambda_1\lambda_3 + 2\lambda_2^2 + 2\lambda_2\lambda_3}{(\lambda_1 + \lambda_2)(\lambda_1 + \lambda_2 + \lambda_3)} - 1.
\]

However, the integrals are analytically intractable for the model of Li and Zhang. Nevertheless, the lack of a closed-form expression is not a major hindrance since \( \tau_{\text{mod}} \) can be obtained quite easily and with arbitrary accuracy via numerical or Monte Carlo methods. Moreover, making the underlying model more complex, for instance, by allowing separate Weibull shape parameters for each transition intensity, has little or no bearing on the computational difficulty of calculating \( \tau_{\text{mod}} \).

Monte Carlo methods provide a particularly convenient way of evaluating the model-based Kendall’s \( \tau \). We can use the fact that, for a model where \( S \) and \( T \) are continuous,

\[
\tau = 2P(S_1 > S_2, T_1 > T_2) - 2P(S_1 < S_2, T_1 > T_2) \\
= 2P(S_1 > S_2, T_1 > T_2) - (1 - 2P(S_1 > S_2, T_1 > T_2)) \\
= 4P(S_1 > S_2, T_1 > T_2) - 1. \tag{11}
\]

It is therefore only necessary to evaluate \( P(S_1 > S_2, T_1 > T_2) \), which can be achieved by simulating \( 2M \) pairs of \((S_i, T_i)\) and then taking

\[
\hat{P}(S_1 > S_2, T_1 > T_2) = M^{-1} \sum_{i=1}^M I(S_i > S_{i+M}, T_1 > T_{i+M}). \tag{12}
\]

Simulation for general illness-death models can be achieved using the methods in Beyersmann et al.\textsuperscript{21} The Monte Carlo standard error associated with the approximation is at most \( 1/2\sqrt{M} \). Typically, \( M = 1 \times 10^8 \) or \( 1 \times 10^7 \) samples can be generated using very little computation time, meaning the Monte Carlo standard error is negligible. A point estimate for \( \tau_{\text{mod}} \) can be obtained by simulating \( 2M \) independent pairs of PFS and OS times from the illness-death model with parameter estimates \( \hat{\theta} := (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3) \). The parameters of the parametric illness-death model are estimated as in the work of Li and Zhang\textsuperscript{9} via maximum likelihood. The only difference is that we used distinct shape parameters \( \alpha_1, \alpha_2, \) and \( \alpha_3 \) corresponding to each transition intensity instead of a common shape parameter \( \alpha \) for all transitions.

A simulation-based approach may also be used to obtain confidence intervals for \( \tau_{\text{mod}} \) using a variant of the simulation delta method.\textsuperscript{22} This involves firstly generating \( B \) samples

\[
\theta_{b}^1, \ldots, \theta_{b}^B \sim N(\hat{\theta}, I(\hat{\theta})^{-1}),
\]

where \( I(\hat{\theta}) \) is the observed Fisher information of the log-likelihood. For each of the \( B \) samples, a pair of \((S, T)\) from the illness-death process with parameters \( \theta_{b}^k \) is simulated \( 2M \) times. The next step is to estimate \( \tau_{\text{mod}} \) denoted by \( \tau_{\text{mod}}^b \) for every \( b \in [1, B] \) using (12). Confidence intervals can then be constructed based either upon the sample standard deviation or sample quantiles of \( \tau_{\text{mod}}^1, \ldots, \tau_{\text{mod}}^B \). A nonparametric bootstrap variant of this algorithm is also possible where \( B \)
bootstrap samples are generated by repeatedly resampling from the original data and the maximum likelihood estimates are recomputed to generate each $\theta^*_b$.

3 | SIMULATIONS

3.1 | Simulation setup

In this section, the performance of the methods is studied through simulation. It is assumed that the true underlying model is an illness-death model as this plausibly reflects the underlying disease process. For the first simulation scenario A, we assume a homogeneous semi-Markov model with Weibull transition intensities and take the values of the shape and scale parameters for each intensity to be those that best fitted to an external dataset from a trial of treatments for colon cancer. As it would usually be expected, there is a lower hazard of death before progression than after progression. However, in the second simulation scenario B, we design this mechanism to be the other way around and therefore refer to this as the “unrealistic scenario.” In both scenarios, we assume a multistate model with a homogeneous semi-Markov assumption, where the imminent future is only dependent on the time spent in the present state and not on other previous history.

For the final simulation scenario C, we seek to investigate sensitivity of the illness-death model–based method to mis-specification of a homogeneous semi-Markov assumption, by generating data in which time to progression also affects the hazard of death given progression. Specifically, we assume a general semi-Markov process where the Weibull hazard function of death after progression depends on the time of progression $t_0$ as well as time since progression. The sojourn time in the post-progression state depends on whether progression or not occurred before a fixed time point, eg, 2 months. Therefore, we use different shape and scale parameters for the Weibull hazard function of death after progression. If progression occurs before 2 months, it is expected that the hazard of death given progression is higher compared to the case where progression is experienced after 2 months. Table 1 shows the setting values of the scale parameter $\lambda_1, \lambda_2, \lambda_3$ and shape parameter $\alpha_1, \alpha_2, \alpha_3$ for scenarios A, B, and C, respectively. Based on those parameter values, the Kendall’s $\tau$ can be obtained for each scenario via simulation as mentioned in Section 2.5.

We consider four censoring cases for each simulation scenarios A, B, and C. In censoring cases 1 and 2, the censoring follows an exponential distribution. However, case 1 has 20% of patients whose OS time is censored while case 2 has 45% of OS times censored. In censoring cases 3 and 4, the same levels of censoring are used, but the censoring times arise from a uniform rather than an exponential distribution. In more detail, the upper limit of the uniform distribution of censoring implies the maximum follow-up time. Regarding the normal scenario with high censoring, the maximum follow-up time is 10.5 years, where 39% and 45% of PFS and OS, respectively, were unobservable. The non-Markov case is similar, as the maximum follow-up time is 12 years, where 36% of PFS and 45% of OS are beyond the range of the follow-up. In comparison to these two scenarios, the unrealistic case, maximum follow up time is 59 years, where 7% of PFS and 45% of OS are censored.

| TABLE 1 Parameter values in the three simulation scenarios |
|---|---|---|
| Parameter | Scenario A | Scenario B | Scenario C |
| $\lambda_1$ | 0.216 | 0.216 | 0.216 |
| $\lambda_2$ | 0.011 | 0.049 | 0.011 |
| $\lambda_3$ | 0.602 | 0.015 | 0.603, if $t_0 < 2$ 0.398, if $t_0 \geq 2$ |
| $\alpha_1$ | 0.675 | 0.675 | 0.675 |
| $\alpha_2$ | 1.088 | 1.008 | 1.088 |
| $\alpha_3$ | 1.009 | 1.080 | 1.008, if $t_0 < 2$ 1.085, if $t_0 \geq 2$ |
| $\tau$ | 0.835 | 0.120 | 0.815 |
### 3.2 Simulation results

Box plots of the estimates of the Kendall’s $\tau$ from each method using 1000 simulated datasets under uniform censoring are shown in Figure 2. Further details such as the estimates of the Kendall’s $\tau$, the respective bias, standard deviation, and mean squared error for every model can be found in the Web-based Supporting Materials.

The upper row of plots in Figure 2 corresponds to the realistic scenario (scenario A) and indicates that the model-based method and the IPCW method perform very well. In particular, the IPCW continues to perform quite well even when there is a higher rate of censoring. The results from the copula models are quite varied. While the Clayton copula-based estimator is almost unbiased, the Hougaard copula model captures the dependence rather poorly. As expected, the efficiency of the semiparametric copula models is lower compared to the fully parametric models. Both the Hougaard and Frank copula estimates are sensitive to the rate of censoring, with a greater degree of bias for higher rates of censoring. Overall, the model-based method has the lowest bias and lowest mean squared error; however, as shown in the Web-based Supporting Materials, the Clayton copula model also has a low mean squared error being very close to the mean squared error of the generalized model-based method. Hence, the Clayton copula seems to be a very good model in this case. Further, it is not surprising that the generalized model-based method performs very well, as it corresponds to the mechanism of the data generation.

It is noticeable that the estimate of Kendall’s $\tau$ is sensitive to the choice of copula function. The Clayton model performs very well, in contrast to Hougaard’s model that seems to be misspecified leading to large negative biases.

---

**FIGURE 2** Box plots of estimates of Kendall’s $\tau$ from eight methods. Dashed red line indicates the true value. A: normal scenario, where the parameters are used from an external dataset from a trial of treatments for colon cancer. B: unrealistic scenario. C: general semi-Markov scenario. sClayton: two stage semiparametric Clayton model. sHougaard: two-stage semiparametric Hougaard model. sFrank: two-stage semiparametric Frank model. [Colour figure can be viewed at wileyonlinelibrary.com]
Contour plots of the density functions of each of the copula models and also the implied copula for the illness-death model are presented in Figure 3 and offer an explanation of these differences. The Clayton, Hougaard, and Frank copulas assume lower tail, upper tail, and symmetric dependence, respectively. An approximation to the density of the implied copula for the true generating model is obtained through bivariate kernel density estimation from simulated PFS and OS times that are transformed by their respective marginal distribution functions. The implied copula for the true generating model indicates lower tail dependence, which is qualitatively similar to that of the Clayton model.

The middle row of Figure 2 shows the box plots of the estimates of Kendall's $\tau$ for each method in the unrealistic scenario (Scenario B). The model-based method continues to perform well. However, in this case, the IPCW estimator is biased with a larger bias for the higher censoring rate. In the low censoring case, it is noticeable that the performance of the copula models is different in comparison to the first simulation scenario. The bias and the standard deviation of the Clayton models have increased, but the Hougaard model seems less misspecified than in the first simulation scenario. Figure 4 shows the bivariate density plots for the unrealistic scenario and indicates why the performance of the copula
models is sensitive to the type of data. The top-left graphic represents the density of the implied copula for the true model. The mode in the upper-left part of the plot corresponds to patients with a quick time to progression, but then a long follow-up to death. The band of higher density in the lower part of the plot corresponds to cases where PFS and OS are equal. The bivariate copula models have problems incorporating the combination of negative and positive correlation between PFS and OS and consequently are a poor fit to this type of data.

In addition, it is somewhat surprising that some of the copula models perform better in the presence of high censoring in comparison to the low censoring case. However, this can be explained by higher censoring leading to less contribution from the mode corresponding to short PFS and long OS, which the copula models are unable to accommodate.

As the copula models are not able to fit this type of data well, the model-based method is preferred in this special case. However, as with Scenario A, the superior performance of the model-based method is somewhat to be expected since it was the model by which the data were generated.

FIGURE 4  Unrealistic simulation case: contour plots for the bivariate density function based on the model-based method and the survivor joint Clayton’s Copula model, Hougaard’s Copula model, and Frank’s Copula model. Kendall’s $\tau$ is 0.119
The results of the final simulation scenario C, based on a general nonhomogeneous semi-Markov illness-death model, are shown in the lower row of Figure 2. This scenario was used to investigate how incorrectly assuming a homogeneous semi-Markov model affects model performance of the method-based method. Indeed, there is some negative bias in this case, while the performance of other methods stays broadly similar to scenario A. Due to the similar setting in scenario A and scenario C as seen in Figure 1, the bivariate density functions in C are expected to show the same structure as in A. The contour plots for the general semi-Markov scenario can be found in the Web-based Supporting materials.

One way of choosing the most appropriate copula model in practice is to choose the one with the lowest Akaike information criterion (AIC). Note that we can only compare AIC between the parametric copula models or between the semiparametric copula models, but not between a parametric and semiparametric model. Tables 3 and 4 in the Supplementary materials give the percentage of simulation replications for which each copula model had the lowest AIC. Broadly similar results were obtained for the parametric and semiparametric models, except that the proportion of times the Frank copula was chosen as opposed to Clayton was higher for the semiparametric models. In all cases, the Clayton model was preferred the majority of the time, with the Hougaard model never chosen. While for scenarios A and C, the Frank copula was chosen as opposed to Clayton was higher for the semiparametric models. In all cases, the Clayton model is the least biased, and for scenario B in the 20% uniformly distributed censoring case, the Clayton model is slightly more biased than Hougaard and Frank. The results therefore seem to indicate that the best model with respect to AIC will not necessarily correspond to the model that best estimates Kendall’s \( \tau \).

4 APPLICATION

In order to illustrate the performance of the different methods, they were applied to data from a clinical trial of treatments for colon cancer.\(^2\) This trial was conducted to investigate the effectiveness of two adjuvant therapies in improving surgical cure rates in stage III colon cancer. Patients were randomized to observation, to the treatment levamisole plus fluorouracil. In terms of the measurements in this trial, the time to progression and time to death were observed in order to evaluate the difference in the hazards of recurrence and death between the treatment groups. The dataset contains the survival experiences of 929 patients who were followed up for 5 years or more (median follow-up, 6.5 years). During the trial, 425 individuals died, 54 were censored after progression and 423 were censored before progression. The maximum follow-up time was 9.1 years, by which point based on the Kaplan-Meier estimates, 43% and 46% would be yet to experience the PFS time and OS time, respectively. Hence, in this case, considerable extrapolation beyond the follow-up period is required to fully characterize the distributions.

In the analysis, we initially pool together data from all treatment arms when estimating Kendall’s \( \tau \) and we first considered models using either Weibull transition intensities or Weibull marginal distributions for the illness-death model–based method and the parametric copula models. There was clear evidence against the constrained Weibull model used in the method of Li and Zhang based on a likelihood ratio comparing the models with common and different Weibull shape parameters for the three intensities (LR = 41.6 on 2 degrees of freedom, \( p < 0.001 \)). However, none of the Weibull-based models represented an adequate fit to the data. Comparisons of the estimated cumulative marginal hazard of PFS and OS with the Nelson-Aalen estimates are given in Figure 3 of the Supporting Materials and indicate substantial discrepancies in all cases. To improve the model fit, we considered Royston-Parmar (RP) flexible parametric models.\(^2\) Specifically, we consider models for which the cumulative log-hazard function \( \log H(t) \) is modeled as a natural cubic spline \( s(x, \gamma) \) with respect to log time \( x := \log t \). Note that this formulation includes Weibull hazards as a special case when there is a linear relationship between \( \log H(t) \) and \( x \). Moreover, the specification of the natural spline to be linear beyond the range of the boundary knots implies a Weibull tail. The RP transition intensity models can be fitted directly using the \texttt{flexsurv} package in R.\(^25\)

Following the guidelines in the work of Royston and Parmar,\(^2\) we considered spline models of increasing complexity by placing the boundary knots \( k_{\text{min}} \) and \( k_{\text{max}} \) at the lowest and highest uncensored event times, respectively, and placing internal knots \( k_1, \ldots, k_m \) with \( k_1 > k_{\text{min}} \) and \( k_m < k_{\text{max}} \) at quantiles of the distribution of uncensored event times. Given these assumptions, a natural cubic spline can be expressed as

\[
s(x, \gamma) = \gamma_0 + \gamma_1 x + \gamma_2 v_1(x) + \cdots + \gamma_{m+1} v_m(x),
\]

where \( \gamma_l, l = 0, \ldots, m \), are unknown parameters to be estimated, and \( v_j(x) \) is the \( j \)th spline basis function defined for \( j = 1, \ldots, m \) as

\[
v_j(x) = (x - k_j)_+^3 - \lambda_j (x - k_{\text{min}})_+^3 - (1 - \lambda_j) (x - k_{\text{max}})_+^3,
\]

where \( \lambda_j = (k_{\text{max}} - k_j)/(k_{\text{max}} - k_{\text{min}}) \) and \( (x - a)_+ = \max(0, x - a) \).
For instance, for the model with one internal knot point, the knot is placed at the median uncensored survival time to allow equal information to inform the two periods. For the copula models, the best AIC was achieved by allowing each marginal distribution to have an RP form with one internal knot. For the illness-death–based models, it was only necessary to include one internal knot point for the transition to progression, with the RP model having a worse AIC for the other two transitions. Figure 5 shows a comparison of the fitted marginal cumulative hazard functions for PFS and OS and indicates an adequate fit compared to the Nelson-Aalen estimates.

In Table 2, the required parameters for our flexible RP model-based method are shown. The spline coefficients \( \gamma_0, \gamma_1, \gamma_2 \) refer to the nonparametric spline function representing the time to progression. The Weibull hazard function of death before progression is defined by the shape parameter \( \alpha_2 \) and scale parameter \( \lambda_2 \), while the Weibull hazard function of death after progression is given by the parameter \( \alpha_3 \) and \( \lambda_3 \). The model parameters for the copula-based models are provided in Table 7 of the Web-based Supplementary Materials.

Based on AIC, the Clayton copula is preferred among the copula models since \( AIC_{\text{Clayton}} = 4242.255 < AIC_{\text{Frank}} = 4256.965 < AIC_{\text{Hougaard}} = 4355.2 \).

Similarly, for the semiparametric copula models, the Clayton model is preferred with \( AIC_{\text{semi-Clayton}} = -24.96 < AIC_{\text{semi-Frank}} = -3.20 < AIC_{\text{semi-Hougaard}} = 181.79 \).

### Table 2

| Parameter | RP Model-Based Method | SE   |
|-----------|-----------------------|------|
| \( \gamma_0 \) | 0.5894 | 0.1507 |
| \( \gamma_1 \) | 2.4705 | 0.1555 |
| \( \gamma_2 \) | 0.0997 | 0.0077 |
| \( \alpha_2 \) | 1.0600 | 0.1478 |
| \( \alpha_3 \) | 0.9619 | 0.0376 |
| \( \lambda_2 \) | 0.0096 | 0.0042 |
| \( \lambda_3 \) | 0.6323 | 0.0317 |

**Abbreviations:** RP, Royston-Parmar; SE, standard error.

**FIGURE 5** Nelson-Aalen and model-based estimates of the marginal cumulative hazard functions for progression-free survival and overall survival where the parametric models are fitted using Royston-Parmar distributions [Colour figure can be viewed at wileyonlinelibrary.com]
Table 3 gives the estimates of Kendall’s \( \tau \) derived from each of the methods. The example partially supports the findings of the simulation results. One similarity is the high sensitivity to choice of copula, as the values of the Kendall’s \( \tau \) differ in the different copula models as well. There is also reasonably close agreement between the model-based and the Clayton copula estimates.

Note that, for this example, if we instead considered quantifying association based on Pearson correlation (as proposed by Fleischer et al and Li and Zhang), then using the flexible RP model-based method and estimating by simulation returns an estimate exceeding 0.999. This is due to the estimate being dominated by the upper tail where there is strong agreement between PFS and OS since long survival times will correspond to those who did not progress and because the hazard of progression is estimated to decrease. The result highlights that Pearson correlation is often not a useful measure of dependence for survival outcomes.

### 4.1 Separate estimates by treatment arm

After previously having pooled the treatment arms in our analysis, we can instead take treatment into account in our calculations. As mentioned above, the colon data contain three treatment arms: the observation group, the group with levamisole (Lev) alone, and the group with a combination of levamisole and fluorouracil (Lev+5FU). It is of interest whether Kendall’s \( \tau \) between PFS and OS is different for each treatment arm. Burzykowski et al incorporated covariates into the calculation of Kendall’s \( \tau \) for copula models by allowing covariates to affect the marginal distributions of PFS and OS. In this way, a common Kendall’s \( \tau \) applies to all groups. Following this procedure, assuming proportional hazards for PFS and OS between each of the treatment groups, the Kendall’s \( \tau \) under a Clayton copula is 0.8464 (SE = 0.0093), which is almost unchanged from the estimate ignoring treatment effects.

Alternatively, we can fit completely separate models to each treatment arm. Table 4 shows the Kendall’s \( \tau \) for each treatment arm based on the RP illness-death model–based method and RP Clayton model, respectively. There is evidence that the treatment affects the degree of dependence between PFS and OS. In particular, Lev and Lev+5FU decreases the hazard of progression and hence increases the proportion of patients for whom PFS equals OS.

### 5 DISCUSSION

The relationship between PFS and OS can be investigated by copula-based approaches, a nonparametric IPCW approach, or illness-death model–based methods. As the copula-based approaches ignore the fact that PFS cannot be longer than
OS and the nonparametric IPCW method requires a strong assumption about censoring, another approach without these above drawbacks is of interest. The illness-death model–based method proposed by Li and Zhang only offers a partial solution to these issues. In this paper, we generalized the model-based method to allow estimation of Kendall’s $\tau$ for general parametric models, potentially allowing dependence both on time since progression and time since randomization.

The simulation results in the previous section give insight on the issue of the most appropriate method for quantifying the association between PFS and OS in a series of scenarios. One notable result is the generally good performance of the Clayton copula model in realistic scenarios, being close to unbiased and having efficiency similar to the illness-death model approach.

However, while the Clayton copula appears to perform well, it is clear that, in general, the estimate of Kendall’s $\tau$ is sensitive to the choice of copula. These varied outcomes are primarily due to the different tail dependencies of the copulas. The Clayton copula may be most appropriate in cancer survival, as it focuses on the dependence in the lower tail of the bivariate density function, which captures the common situation where a quick time to progression leads to quick time to death. However, as seen in the unrealistic scenario, there are illness-death model scenarios where the Clayton copula model will be biased. If a copula-based approach is to be used in practice, it is important to consider several different copula models and adopt the one with the best fit. However, results from the simulation study suggested AIC is not necessarily a reliable criterion for finding the model with the least bias for Kendall’s $\tau$. Furthermore, it is questionable whether it is sensible to apply copula models even if they often produce reasonable estimates, given that they do not offer an admissible model for PFS and OS.

The generalized illness-death model method also relies upon the underlying model being close to correctly specified, with the simulations showing some bias arising if a homogeneous semi-Markov model is assumed when the true model is nonhomogeneous. Assessment of the appropriateness of the model should therefore be considered. In addition, parametric assumptions about the transition intensities have to be taken into account. A fully nonparametric approach is not possible as the hazard functions cannot be identified beyond the maximum follow-up time. The use of hazard functions based upon flexible natural cubic splines reduces the danger of estimates being sensitive to the particular parametric specification. However, they still rely on the assumption of log-linearity of the hazards beyond the last knot point.

An alternative nonparametric approach would be to aim to estimate a restricted version of Kendall’s $\tau$. For instance, rather than compute the unconditional probability of concordance and discordance, one could replace the probability terms in (1) with ones conditional on $S_1 \wedge T_1 < t^\ast, S_2 \wedge T_2 < t^\ast$ for some choice of $t^\ast$. A similar approach was employed in the context of estimating concordance odds in an Aalen additive hazards model.26 Provided $t^\ast$ is less than or equal to the maximum follow-up time, a consistent model-based estimate of the restricted Kendall’s $\tau$ can be obtained from, for instance, the Nelson-Aalen estimates of the transition intensities under a Markov assumption. Note, however, that restricting the definition of Kendall’s $\tau$ would also allow consistent estimation using IPCW with the added advantage of not requiring a particular Markov or semi-Markov assumption.

There are various ways in which covariates could be accommodated into the models for Kendall’s $\tau$. In the copula-based models, such as in the work of Burzykowski et al.,3 covariates are allowed to affect the marginal distributions of PFS and OS, but the dependence between PFS and OS is common across all patients. For the illness-death models, the natural way to accommodate covariates effects is to allow separate effects for each of the transition intensities. This results in a different estimated Kendall’s $\tau$ between PFS and OS depending on the value of covariates. In the colon cancer example, there was clear evidence of the treatment affecting the degree of dependence between PFS and OS, with a higher estimated Kendall’s $\tau$ in the Lev+5FU group. Since we would usually expect an effective treatment to alter the proportion of patients who progress before death, it seems reasonable that the Kendall’s $\tau$ between PFS and OS would also differ. Hence, copula-based methods, if pursued, should also incorporate covariate effects for $\tau$ itself.

Throughout this article, it has been assumed that both the time of death and time to progression can be continuously observed up to right censoring. While such an assumption is commonly applied, in practice, assessments of progression are intermittent, resulting in different right-censoring times for progression and death and interval-censored progression times.27 For the model-based approach, the model can be fitted using a likelihood that properly accounts for the intermittent observation.26,29 Adaptation of the other approaches is less straightforward since PFS is a composite measure of progression, which may be interval censored, and of OS, which is right censored.

ACKNOWLEDGEMENTS
This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant 633567.
REFERENCES

1. Eskens FALM, Verweij J. Clinical studies in the development of new anticancer agents exhibiting growth inhibition in models: facing the challenge of a proper study design. Crit Rev Oncol Hematol. 2000;34(2):33-88.

2. Stone A, Wheeler C, Barje A. Improving the design of phase II trials of cytostatic anticancer agents. Contemp Clin Trials. 2007;28(2):138-145.

3. Paoletti X, Oba K, Bang Y-J, et al. Progression-free survival as a surrogate for overall survival in advanced/recurrent gastric cancer trials: a meta-analysis. J Natl Cancer Inst. 2013;105(21):1667-1670.

4. Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. J Clin Oncol. 2007;25(33):5218-5224.

5. Burzykowski T, Molenberghs G, Buyse M, Geys H, Renard D. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. J Royal Stat Soc Ser C Appl Stat. 2001;50(4):405-422.

6. Shih JH, Louis TA. Inferences on the association parameter in copula models for bivariate survival data. Biometrics. 1995;51:1384-1399.

7. Li Y, Zhang Q. A Weibull multi-state model for the dependence of progression-free survival and overall survival. Statist Med. 2015;34(17):2497-2513.

8. Sklar M. Fonctions de répartition àn dimensions et leurs marges. Publications de l’Institut Statistique de l’Université Paris. 1959;8:229-231.

9. Oakes D. Bivariate survival models induced by frailties. J Am Stat Assoc. 1988;84(406):487-493.

10. Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika. 1978;65(1):141-151.

11. Hougaard P. Survival models for heterogeneous populations derived from stable distributions. Publications de l’Institut Statistique de l’Université Paris. 1986;73(2):387-396.

12. Frank MJ. On the simultaneous associativity of \( F(x, y) \) and \( x+y−F(x, y) \). Aequ Math. 1979;19(1):194-226.

13. Genest C, MacKay RJ. Copules archimédiennes et families de lois bidimensionnelles dont les marges sont données. Can J Stat. 1986;14(2):145-159.

14. Luke YL. The Special Functions and Their Approximations. New York, NY: Academic Press; 1969.

15. Dejardin D, Lesaffre E, Verbeke G. Joint modeling of progression-free survival and death in advanced cancer clinical trials. Stat Med. 2010;29(16):1724-1734.

16. Mandel M. Simulation-based confidence intervals for functions with complicated derivatives. Am Stat. 2013;67(2):76-81.

17. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Statist Med. 2002;21(15):2175-2197.

18. Jackson CH. flexsurv: a platform for parametric survival modeling in R. J Stat Softw. 2016;70(8):1-33.

19. Martinussen T, Piper CB. Estimation of odds of concordance based on the Aalen additive model. Lifetime Data Anal. 2013;19(1):100-116.

20. Zeng L, Cook R, Wen L, Boruvka A. Bias in progression-free survival analysis due to intermittent assessment of progression. Statist Med. 2015;34(24):3181-3193.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Tables and Figures referenced in Sections 3 and 4 and R code for fitting the parametric copula models, generalized model-based method, and the IPCW to the colon cancer dataset presented in this paper can be found in the Web-based Supporting Materials.
APPENDIX

DERIVATION OF THE FORMULA FOR MODEL-BASED KENDALL’S $\tau$

Under an assumption that patients are independent and identically distributed and that $S$ and $T$ are continuous random variables, from (11),

$$
\tau_{\text{mod}} = 4P(S_1 > S_2, T_1 > T_2) - 1,
$$

where $(S_1, T_1)$ and $(S_2, T_2)$ are two independent pairs of realizations of $(S, T)$.

The probability $P(S_1 > S_2, T_1 > T_2)$ can be expanded into three cases defined by which of patient 1 or patient 2 dies before progression. Specifically, we may write

$$
P(S_1 > S_2, T_1 > T_2) = P(S_2 = T_2 < S_1) + P(S_1 > S_2, T_1 > T_2, T_2 > S_2, T_1 > S_1) + P(S_1 = T_1, S_2 < S_1, S_2 < T_2 < T_1). \tag{A1}
$$

Note that $S_j$ can also be interpreted as the sojourn time in state 0 and that

$$
T_j = \begin{cases} 
S_j, & \text{if } \Delta_j = 0 \\
S_j + V_j, & \text{if } \Delta_j = 1,
\end{cases}
$$

where $\Delta_j$ is an indicator of whether progression occurs before death and $V_j$ is the sojourn time in the progression state.

Let $f_{S|T}(s) = \pi_{01}(t) \exp(-\Pi_0(t))$ and $f_{S|V}(s) = \pi_{02}(t) \exp(-\Pi_0(t))$ be the cause-specific densities for progression and death before progression, respectively. Also, let $S_j(t) = \exp(-\Pi_0(t))$ be the survivor function of the sojourn distribution in state 0 and $S_{V|S}(u) = \exp(-\Pi_{12}(s; u))$ and $f_{V|S}(s; u) = \pi_{12}(s; u) \exp(-\Pi_{12}(s; u))$ be the survivor function and density of the conditional sojourn distribution in state 1 given a sojourn of time $u$ in state 0. Then,

$$
P(S_2 = T_2 < S_1) = \int_0^{\infty} P(S_2 = s, \Delta_2 = 0)P(S_1 > s)ds
$$

$$
= \int_0^{\infty} f_{S2}(s)S_2(s)ds
$$

$$
= \int_0^{\infty} \pi_{02}(s) \exp(-2\Pi_0(s))ds,
$$

$$
P(S_1 > S_2, T_1 > T_2, T_2 > S_2, T_1 > S_1)
$$

$$
= \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} P(S_1 = s_1, \Delta_1 = 1)P(S_2 = s_2, \Delta_2 = 1)P(T_1 > T_2|S_1 = s_1, S_2 = s_2, \Delta_1 = 2, \Delta_2 = 1)ds_2ds_1
$$

$$
= \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} P(S_2 = s_2, \Delta_2 = 1)P(S_1 = s_1, \Delta_1 = 1)P(V_1 = S_3|S_1 = s_1)P(V_2 < s_1 + s_3 - s_2|S_2 = s_2)ds_3ds_2ds_1
$$

$$
= \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} f_{S1}(s_1)f_{S1}(s_2)f_{V|S}(s_3; s_1)(1 - S_{V|S}(s_1 + s_3 - s_2; s_2))ds_3ds_2ds_1
$$

$$
= \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} \pi_{01}(s_1)\pi_{01}(s_2)\exp(-\Pi_0(s_1) - \Pi_0(s_2))
$$

$$
\times \pi_{12}(s_3; s_1) \exp(-\Pi_{12}(s_3; s_1))[1 - \exp(-\Pi_{12}(s_1 + s_3 - s_2; s_2))]ds_3ds_2ds_1
$$

and
\[ P(S_1 = T_1, S_2 < S_1, S_2 < T_2 < T_1) \]
\[ = \int_0^\infty P(S_1 = s_1, \Delta_1 = 0)P(S_2 < T_2 < s_1)ds_1 \]
\[ = \int_0^\infty \int_0^{s_1} P(S_1 = s_1, \Delta_1 = 0)P(S_2 = s_2, \Delta_2 = 1)P(V_2 < s_1 - s_2|S_2 = s_2)ds_1ds_2 \]
\[ = \int_0^\infty \int_0^{s_1} f_{S_2}(s_1)f_{S_1}(s_2)(1 - S_{V|S}(s_1 - s_2))ds_2ds_1 \]
\[ = \int_0^\infty \int_0^{s_1} \pi_02(s_1)\pi_01(s_2) \exp\{-\Pi_0(s_1) - \Pi_0(s_2)\}(1 - \exp\{-\Pi_{12}(s_1 - s_2; s_2)\})ds_2ds_1. \]

Substituting these terms into (A1) and then into (11) produces the expression given in (10).