# 半竞争风险的双变量 copula 回归模型

Yinghui Wei[1]*、Ma-lgorzata Wojty´s[1],Lexy Sorrell[1] 和 Peter Rowe[2]

[1]普利茅斯大学工程、计算与数学学院数学科学中心，*PL4 8AA*[2]西南移植中心，普利茅斯大学医院 *NHS* 信托，*PL6 8DH*

arXiv:2304.02954v1 [stat.ME] 6 Apr 2023

**摘要**

当同一个体发生两个事件时，时间至事件的半竞争风险终点可能存在关联。这些事件及其间的关联也可能受到个体特征的影响。本文提出联结函数生存模型，用于估计协变量对非终末事件和终末事件风险比的影响，以及协变量对两个事件间关联的影响。我们采用正态、Clayton、Frank 和 Gumbel Copula，为非终末事件与终末事件之间提供多种关联结构。我们将所提方法用于建立肾移植患者移植物衰竭和死亡的半竞争风险模型。研究发现，在估计协变量对非终末事件风险比的影响时，联结函数生存模型优于 Cox比例风险模型。我们还发现，在 Copula模型的关联参数中纳入协变量，有助于提高风险比的估计精度。

关键词：Copula 模型、肾移植、半竞争风险、生存分析、风险比

## 1 引言

在医学研究中，失访的患者，或在研究期间未发生关注事件的患者，通常会留下一个删失观测。然而，对于半竞争风险终点，非终末事件也可能被终末事件删失（Fine 等，2001）。一

*通讯作者：YinghuiWei，普利茅斯大学工程、计算与数学学院数学科学中心，PL4 8AA，英国。电子邮箱：yinghui.wei@plymouth.ac.uk。

半竞争风险的一个例子（本文稍后将对此进行分析）是肾移植后的移植物衰竭和死亡，其中非终末事件是移植物衰竭，终末事件是死亡。

Cox比例风险模型在实践中被广泛用于分析事件发生时间结局。Cox模型最初是为分析全因死亡率而开发的（Cox, 1972），这类结局不存在竞争风险。在假设独立删失机制的情况下，使用Cox模型估计的风险比在分析非终末事件时可能存在偏倚（Berry et al., 2010）。一些作者使用Copula回归模型对非终末事件和终末事件进行联合建模。Peng and Fine (2007)、Hsieh et al. (2008) 和 Chen (2012) 提出了针对边际分布回归的半参数联结函数模型。终末事件的分析可以采用常见的生存分析方法。例如，Peng and Fine (2007) 在其Copula模型中对终末事件采用边际比例风险模型进行建模。Hsieh et al. (2008) 使用Copula模型，该模型允许协变量组之间存在不同的依赖结构。Fu等人（2013）开发了贝叶斯正态诱导联结函数估计模型。

Copula是多元分布函数，可以分别对边际分布以及它们之间的依赖结构进行建模。Sorrell等人（2022）引入了二元联结函数模型，以使用以下四种Copula函数——正态、Clayton、Frank和Gumbel Copula——估计半竞争风险终点之间的相关性。然而，非终末事件和终末事件的风险率，以及它们之间的关联参数，可能会受到协变量的影响。在生存终点相关性分析中纳入协变量，有助于理解个体特征可能如何影响这种关联。协变量也可以纳入边际分布分析，从而估计风险比，并进一步比较各组之间生存终点的风险。

在本文中，我们通过使用条件联结函数（Patton，2006）开发了联结函数生存回归模型，该模型允许生存终点之间的关联参数和风险率均取决于多个二元协变量。这也是对 Sorrell等人（2022）所引入的联结函数生存模型的扩展，在该模型中未纳入协变量。我们使用含二元协变量的Copula 生存回归模型估计风险比。我们还估计这些二元协变量对半竞争风险之间关联的影响。通过对非终末事件和终末事件进行联合建模，并纳入二者之间的相关性，我们希望改进关于非-

终末事件。

我们重点关注使用条件联结函数分析包含半竞争风险的生存数据时协变量的纳入。在半竞争风险情境下，Chen (2012) 开发了一种模型，允许 Copula 函数的关联参数随分类协变量变化。Hsieh etal.(2008) 提出了一种模型，允许每个协变量组采用单独的阿基米德Copula 函数。这样可以使用不同的依赖结构来描述每个协变量组中半竞争风险事件之间的关联。Ghosh (2006) 研究了离散协变量对非终末事件和终末事件的影响，以及半竞争风险事件之间关联参数的影响。Peng and Fine (2007)通过提出一种具有时间依赖 Copula 的模型，研究了协变量对非终末事件和关联参数的影响，并将这些方法应用于艾滋病诊断后的生存终点。类似地，Zhuetal. (2021)提出了一种Copula 回归方法，使用两阶段方法考察协变量对非终末事件和终末事件的影响。在第1阶段，研究协变量对边缘事件的影响；在第2阶段，估计 Copula 模型的关联参数。与Peng and Fine (2007)提出的一阶段方法相比，Zhu et al.(2021) 更倾向于采用两阶段方法，因为该方法能够分别识别边缘回归模型和 Copula 模型的错误设定。

本文其余部分的组织结构如下。第 2 节介绍本文的动机数据。随后，我们在第 3 节中介绍所提出的Copula回归生存分析方法，利用多种条件 Copula 函数估计协变量对半竞争风险事件和关联参数的影响。在第 4 节中，我们将所提出的方法应用于动机数据，随后在第 5 节中通过模拟研究比较 Copula模型。最后，我们将进行讨论。

# 2 动机数据

研究动机数据来自英国国家医疗服务体系（NHS）血液与移植管理的英国移植登记处（UKTR）。研究结局为肾移植后的移植物衰竭时间和死亡时间，二者属于半竞争风险。研究人群为在英国于1995年至2016年间接受首次且单次肾移植、并且移植时年龄、性别和供者类型信息已知的肾移植受者。我们通过使用所提出的 Copula 生存回归模型，展示了一种新颖的应用。

我们纳入了英国于1995年至2016年间进行的40,348例首次且单次肾移植。78.01%的患者移植物衰竭时间被删失，80.0%的患者死亡时间被删失，而64.3%的患者两项事件均被删失。

表1展示了我们所关注的基线特征。本文后文应用中将纳入的协变量包括供者类型（已故供者或活体供者）、受者性别（男性或女性）以及受者年龄（$> 50$ 岁或 $\leq 50$ 岁）。供者类型表示受者在接受移植时所获得的肾脏来自已故供者还是活体供者。根据以下文献，这些协变量之所以被纳入，是因为它们是影响移植结局的重要因素。正如Terasaki等人（1995）和Port等人（2004）所报告的，与已故供者受者相比，活体供者肾脏受者具有更好的生存。

此外，受者年龄已被证明是影响移植后生存的重要因素（Becker 等，2000；Fabrizii 等，2004；Keith 等，2004）。在 Fabrizii等（2004）的研究中，研究发现，5年生存受到受者年龄组的影响；然而，在控制混杂因素后，这些差异并不明显。

| **变量** | | $n$ (%) |
|---|---|---|
| **供者类型** | 已故 | 27，971（69.9%） |
| | 活体 | 12,377 (30.1%) |
| **受者性别** | Male | 24,945 (61.8%) |
| | 女性 | 15,403 (38.2%) |
| **受者年龄（年）** | $\leq 50$ | 23,208 (57.5%) |
| | $> 50$ | 17,140 (42.5%) |

表1：来自英国移植登记处数据集的单次及首次肾移植受者基线特征（$1995 - 2016$）。具有相应特征的受者人数及其百分比列于括号中。

# 3 方法

本节介绍如何在 Copula 函数中纳入协变量，以考察协变量对半竞争风险事件的影响。我们将 Copula 函数的关联参数以及边际生存分布的风险率设定为以协变量为条件。我们采用条件 Copula，这一方法由 Patton（2006）首次提出，他扩展了 Sklar定理（Sklar,

1959），以考虑随机变量以协变量为条件的情况。 .

令 $T_1$ 表示非终末事件发生时间， $T_2$ 表示终末事件发生时间，删失时间为 $C$。首次事件发生时间记为$X = \min(T_1, T_2, C)$，事件指示变量为 $d_1 = 1$（若 $T_1 \leq \min(T_2, C)$），否则为 $d_1 = 0$。第二次事件发生时间记为 $Y = \min(T_2, C)$，事件指示变量为 $d_2 = 1$ （若 $T_2 \leq C$），否则为 $d_2 = 0$。

我们令边际风险率和关联参数依赖于协变量，$\mathbf{W} = (W_1, \ldots, W_p)$。非终末事件和终末事件的边际生存函数分别记为$S_{T_1|\mathbf{W}}(t_1|\mathbf{w}) = P(T_1 > t_1|\mathbf{w})$ 和 $S_{T_2|\mathbf{W}}(t_2|\mathbf{w}) = P(T_2 > t_2|\mathbf{w})$。非终末事件和终末事件的边际概率密度函数（PDFs）分别记为$f_{T_1|\mathbf{W}}(t_1|\mathbf{w})$ 和 $f_{T_2|\mathbf{W}}(t_2|\mathbf{w})$。

我们使用 Copula 函数中的单变量生存函数，估计个体生存概率之间的关联。与 Patton（2006）中使用累积分布函数（CDF）相比，这改变了终点之间关联强度的解释（Renfro等，2015年）。我们使用 Copula 函数$C_\theta$表示联合生存函数，其关联参数为 $\theta$,

$$S_B(t_1, t_2|\mathbf{w}) = P(T_1 > t_1, T_2 > t_2|\mathbf{w}) = C_\theta(S_{T_1|\mathbf{W}}(t_1|\mathbf{w}), S_{T_2|\mathbf{W}}(t_2|\mathbf{w})|\mathbf{w}), \quad (1)$$

其中，下标 $B$ 用于表示该函数的二元性质。

然后，Copula密度函数如下,

$$c_\theta(S_{T_1|\mathbf{W}}(t_1|\mathbf{w}), S_{T_2|\mathbf{W}}(t_2|\mathbf{w})) = \frac{\partial^2 C_\theta(S_{T_1|\mathbf{W}}(t_1|\mathbf{w}), S_{T_2|\mathbf{W}}(t_2|\mathbf{w})|\mathbf{w})}{\partial S_{T_1|\mathbf{W}}(t_1|\mathbf{w}) \partial S_{T_2|\mathbf{W}}(t_2|\mathbf{w})}. \quad (2)$$

联合概率密度函数可以使用Copula密度函数和边缘概率密度函数表示,

$$f_B(t_1, t_2|\mathbf{w}) = c_\theta(S_{T_1|\mathbf{W}}(t_1|\mathbf{w}), S_{T_2|\mathbf{W}}(t_2|\mathbf{w})|\mathbf{w}) f_{T_1|\mathbf{W}}(t_1|\mathbf{w}) f_{T_2|\mathbf{W}}(t_2|\mathbf{w}). \quad (3)$$

我们考虑四种不同类型的 Copula：正态、Clayton、Frank 和 Gumbel，其定义见 Sorrell等人（2022）。这些 Copula 函数涵盖了多种依赖结构，并为公式（1）中的联合生存函数提供了不同的形状。对于上述每一种 Copula,

关联参数 $\theta$ 是单变量的。对于两种事件的边际分布，我们考虑指数分布、Weibull分布和Gompertz分布，这些都是常用的参数生存模型。我们用它们来说明这些方法及其应用；不过，这些方法也很容易扩展到其他参数生存分布。

## 3.1 似然

令 $\mathbf{\Theta}$ 为边际分布的参数和在协变量条件下的关联参数 $\theta$ 组成的向量。则，似然函数如下：

$$
\begin{aligned}
L(\mathbf{\Theta}) = \prod_{i=1}^{n} & \left(c_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i) f_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i) f_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)\right)^{d_{i,1}d_{i,2}} \\
& \left(\frac{\partial C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i)}{\partial S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i)} f_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i)\right)^{d_{i,1}(1-d_{i,2})} \\
& \left(\frac{\partial C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i)}{\partial S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)} f_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)\right)^{(1-d_{i,1})d_{i,2}} \\
& \left(C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i)\right)^{(1-d_{i,1})(1-d_{i,2})},
\end{aligned}
\tag{4}
$$

其中， $n$ 为个体总数。

下面，我们给出了使用四种 Copula（正态、Clayton、Frank 和 Gumbel）时的似然公式。表 2 总结了每种Copula 所使用的不同链接函数，这些函数用于确保关联参数处于相应藤函数内允许的取值范围之内。

对于正态Copula，关联参数 $\theta = \rho$，其中 $\rho$ 是皮尔逊

| Copula $C_\theta$ | $\theta$的链接函数 | $\theta$的取值范围 |
|---|---|---|
| 正态 | $\theta = \frac{\exp(2(b_0 + b_1W_1 + ... + b_pW_p)) - 1}{\exp(2(b_0 + b_1W_1 + ... + b_pW_p)) + 1}$ | $\theta \in [-1, 1]$ |
| Clayton | $\theta = \exp(b_0 + b_1W_1 + ... + b_pW_p)$ | $\theta \in (0, \infty)$ |
| Frank | $\theta = b_0 + b_1W_1 + ... + b_pW_p$ | $\theta \in (-\infty, \infty)$ |
| Gumbel | $\theta = \exp(b_0 + b_1W_1 + ... + b_pW_p) + 1$ | $\theta \in [1, \infty)$ |

表 2：关联参数 $\theta$ 的链接函数。

相关性系数，以及方程（4）的似然函数为

$$
\begin{aligned}
L_N(\mathbf{\Theta}) = \prod_{i=1}^{n} \Bigg( & \frac{1}{\sqrt{1-\rho^2}} \exp\Bigg( \frac{1}{2(1-\rho^2)} \\
& 2\rho\Phi^{-1}(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i))\Phi^{-1}(S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)) - \rho^2(\Phi^{-1}(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i))^2 \\
& +\Phi^{-1}(S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i))^2) \Bigg) f_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i) f_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i) \Bigg)^{d_{i,1}d_{i,2}} \\
& \left( \Phi\left( \frac{\Phi^{-1}(S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)) - \rho\Phi^{-1}(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i)))}{\sqrt{1-\rho^2}} \right) f_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i) \right)^{d_{i,1}(1-d_{i,2})} \\
& \left( \Phi\left( \frac{\Phi^{-1}(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i)) - \rho\Phi^{-1}(S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i))}{\sqrt{1-\rho^2}} \right) f_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i) \right)^{(1-d_{i,1})d_{i,2}} \\
& \left( C_\rho(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)) \right)^{(1-d_{i,1})(1-d_{i,2})},
\end{aligned}
\tag{5}
$$

其中 $\Phi$ 是标准正态分布的累积分布函数。

对于 Clayton Copula，似然函数为

$$
\begin{aligned}
L_C(\mathbf{\Theta}) = \prod_{i=1}^{n} \Bigg( & (1+\theta) \frac{(C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i))^{1+2\theta}}{(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i) S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i))^{1+\theta}} \\
& f_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i) f_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i) \Bigg)^{d_{i,1}d_{i,2}} \\
& \left( \left( \frac{C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i)}{S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i)} \right)^{1+\theta} f_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i) \right)^{d_{i,1}(1-d_{i,2})} \\
& \left( \left( \frac{C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i)}{S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)} \right)^{1+\theta} f_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i) \right)^{(1-d_{i,1})d_{i,2}}
\end{aligned}
$$

$$\left(C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i))\right)^{(1-d_{i,1})(1-d_{i,2})}. \tag{6}$$

对于 Frank Copula，似然函数为

$$\begin{aligned}
L_F(\mathbf{\Theta}) = \prod_{i=1}^{n} \Bigg( & \frac{\theta e^{\theta C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i)} (e^{\theta C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i)} - 1)}{(e^{\theta S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i)} - 1)(e^{\theta S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)} - 1)} \\
& f_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i) f_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)\Big)^{d_{i,1}d_{i,2}} \\
& \left(\frac{1 - e^{\theta C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i)}}{1 - e^{\theta S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i)}} f_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i)\right)^{d_{i,1}(1-d_{i,2})} \\
& \left(\frac{1 - e^{\theta C_\theta(S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i)}}{1 - e^{\theta S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)}} f_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)\right)^{(1-d_{i,1})d_{i,2}} \\
& \left(C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i)\right)^{(1-d_{i,1})(1-d_{i,2})}.
\end{aligned} \tag{7}$$

最后，对于 Gumbel Copula，似然函数为

$$\begin{aligned}
L_G(\mathbf{\Theta}) = \prod_{i=1}^{n} \Bigg( & \frac{1}{S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i) S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)(-\log(C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|w_i)))^{2\theta-1}} \\
& C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i)(-\log(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i)))^{\theta-1}(-\log(S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)))^{\theta-1} \\
& (\theta - 1 - \log(C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i))) f_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i) f_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)\Big)^{d_{i,1}d_{i,2}} \\
& \left(\frac{C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i)(-\log(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i)))^{\theta-1}}{S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i)(-\log(C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i))^{\theta-1}} f_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i)\right)^{d_{i,1}(1-d_{i,2})} \\
& \left(\frac{C_\theta(S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i)(-\log(S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)))^{\theta-1}}{S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)(-\log(C_\theta(S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)|\mathbf{w}_i))^{\theta-1}} f_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i)\right)^{(1-d_{i,1})d_{i,2}} \\
& \left(C_\theta(S_{T_1|\mathbf{W}}(t_{i,1}|\mathbf{w}_i), S_{T_2|\mathbf{W}}(t_{i,2}|\mathbf{w}_i))\right)^{(1-d_{i,1})(1-d_{i,2})}.
\end{aligned} \tag{8}$$

对于正态、Clayton 和 Gumbel Copula，关联参数 $\theta$ 的方差可以使用 Delta 方法进行近似。我们在补充材料中分别针对一个协变量以及多个协变量进行了描述：正态部分见第 1.2 节，Clayton 部分见第 1.3 节，Gumbel Copula 部分见第 1.4 节。

## 3.2 事件时间模型

我们考虑将指数分布、Weibull分布和Gompertz分布用于事件时间的边际分布。对于每个模型，我们都指定边际参数的链接函数。指数分布模型假设风险率为常数，而Weibull分布和Gompertz分布模型则允许风险率递增、保持不变或递减。这些方法可以轻松扩展到其他参数生存分布，例如对数正态分布或对数逻辑斯蒂分布。

**指数事件时间**

假设两个事件的边际分布都服从指数分布，对于非终末事件和终末事件，其风险率分别为 $\lambda_1$ 和 $\lambda_2$。那么，生存函数分别为 $S_{T_1|}$w($t_1$) = exp($-\lambda_1 t_1$)和$S_{T_2|}$w($t_2$) = exp($-\lambda_2 t_2$)，分别对应非终末事件和终末事件。我们将协变量 $W_1, \ldots, W_p$ 纳入两个事件的风险率中：

$$\lambda_1 = \exp(a_0 + a_1 W_1 + ... + a_p W_p), \tag{9}$$
$$\lambda_2 = \exp(c_0 + c_1 W_1 + ... + c_p W_p), \tag{10}$$

其中，$a_0, ..., a_p$是非终末事件的回归系数，而 $c_0, ..., c_p$是终末事件的回归系数。因此，二元协变量 $W_k$的风险比（其中 $k \in \{1, \ldots, p\}$）为。

$$\mathrm{HR}_{NT} = \frac{\exp(a_0 + a_k)}{\exp(a_0)} = \exp(a_k), \tag{11}$$
$$\mathrm{HR}_{T} = \frac{\exp(c_0 + c_k)}{\exp(c_0)} = \exp(c_k), \tag{12}$$

分别针对非终末事件和终末事件。

**Weibull分布事件时间**

假设两个事件的边际分布均服从Weibull分布，其概率密度函数为

$$f_{T_j|\mathbf{w}}(t_j) = \beta_j \alpha_j t_j^{\alpha_j - 1} \exp(-\beta_j t_j^{\alpha}),$$

对于 $j = 1$，2，其中 $\alpha_j$ 是形状参数，$\beta_j$ 是尺度参数，$t_j$分别表示非终末事件和终末事件的事件时间。

生存函数和风险函数如下所示。

$$S_{T_j|\mathbf{w}}(t_j) = \exp(-\beta_j t_j^{\alpha}),$$
$$h_{T_j|\mathbf{w}}(t_j) = \beta_j \alpha_j t_j^{\alpha_j - 1},$$

适用于半竞争风险的双变量 Copula 回归模型，$j = 1,\ 2.$

我们考虑形状参数 $\alpha_1$ 和 $\alpha_2$ 为常数，尺度参数 $\beta_1$ 和 $\beta_2$ 依赖于协变量 $\mathbf{W} = (W_1, \ldots, W_p)$ 的情形，并使用以下校准函数:

$$\beta_1 = \exp(a_0 + a_1 W_1 + ... + a_p W_p),$$
$$\beta_2 = \exp(c_0 + c_1 W_1 + ... + c_p W_p),$$

其中 $a_0$、.....、$a_p$是非终末事件的回归系数，$c_0$、.....、$c_p$是终末事件的回归系数。

因此，二元协变量 $W_k$ 的风险比（其中 $k \in \{1, \ldots, p\}$）为。

$$\mathrm{HR}_{NT} = \frac{\exp(a_0 + a_k)\alpha_1 t^{\alpha_1 - 1}}{\exp(a_0)\alpha_1 t^{\alpha_1 - 1}} = \exp(a_k),$$
$$\mathrm{HR}_{T} = \frac{\exp(c_0 + c_k)\alpha_2 t^{\alpha_2 - 1}}{\exp(c_0)\alpha_2 t^{\alpha_2 - 1}} = \exp(c_k),$$

分别如此。这些与方程(11)和 (12)中所述的指数事件时间具有相同形式。

**Gompertz事件时间**

假设两个事件的边际分布均遵循Gompertz分布，其概率密度函数为

$$f_{T_j|\mathbf{w}}(t_j) = \lambda_j \exp\left(\gamma_j t_j - \frac{\lambda_j}{\gamma_j}(\exp(\gamma_j t_j) - 1)\right),$$

对于 $j = 1$,2，其中 $\gamma_j$ 是形状参数，$\lambda_j$ 是速率参数，$t_j$分别表示非终末事件和终末事件的事件时间。

生存函数和风险函数如下所示

$$S_{T_j|\mathbf{w}}(t_j) = \exp\left(-\frac{\lambda_j}{\gamma_j}(\exp(\gamma_j t_j) - 1)\right),$$
$$h_{T_j|\mathbf{w}}(t_j) = \lambda_j \exp(\gamma_j t_j),$$

对于 $j = 1,\ 2.$

我们考虑形状参数 $\gamma_1$ 和 $\gamma_2$ 为常数，而速率参数 $\lambda_1$ 和 $\lambda_2$ 依赖于协变量 $W_1$, ... , $W_p$ 的情形，并使用以下校准函数:

$$\lambda_1 = \exp(a_0 + a_1 W_1 + ... + a_p W_p),$$
$$\lambda_2 = \exp(c_0 + c_1 W_1 + ... + c_p W_p),$$

其中 $a_0$，..... $a_p$ 是非终末事件的回归系数， $c_0$，..... $c_p$是终末事件的回归系数。

因此，对于二元协变量 $W_k$（其中 $k \in \{1, \ldots, p\}$），风险比为

$$\mathrm{HR}_{NT} = \frac{\exp(a_0 + a_k)\exp(\gamma_1 t)}{\exp(a_0)\exp(\gamma_1 t)} = \exp(a_k),$$
$$\mathrm{HR}_{T} = \frac{\exp(c_0 + c_k)\exp(\gamma_2 t)}{\exp(c_0)\exp(\gamma_2 t)} = \exp(c_k),$$

分别如此。这些与指数分布和Weibull分布事件时间的形式相同。

对于所有三个边际模型，风险比的方差可以近似为

$$\mathrm{Var}(\mathrm{HR}_{NT}) = \exp(2\hat{a}_k)\mathrm{Var}(\hat{a}_k), \tag{13}$$
$$\mathrm{Var}(\mathrm{HR}_{T}) = \exp(2\hat{c}_k)\mathrm{Var}(\hat{c}_k), \tag{14}$$

使用补充材料第 1.1 节所述的 Delta 方法。

## 3.3 估计

边际分布的回归系数和关联参数通过最大化对数似然（即，对数

（4）。在本文介绍的实际应用中，通过使用 `R` 软件包中的函数 `optim` （R Core Team，2021）实现这一点。95% 置信区间（CI）使用 Fisher 信息矩阵构造。

# 4 应用

我们将第 3 节所述的方法应用于第 2 节介绍的英国肾脏移植登记（UKTR）数据集。该数据集包含肾移植受者发生移植物衰竭的时间和死亡时间。为说明这些方法的使用，我们在两个生存终点的分析中纳入以下二元协变量：受者年龄组（$> 50$ 岁和 $\leq 50$ 岁）、受者性别（女性和男性）以及供者类型（活体供者和已故供者）。我们允许风险率和生存终点之间的关联均随这些协变量而变化。我们估计非终末事件和终末事件的风险比，以及参考组和协变量组的关联参数。

我们使用指数分布、Weibull分布和Gompertz分布对生存时间进行建模。我们应用四种不同的藤函数来描述生存终点之间的关联，并使用赤池信息准则（Akaike，1974）选择拟合效果最佳的模型。对于指数分布、Gompertz分布和Weibull分布生存分布，各协变量的估计风险比以及协变量关于关联参数的回归系数分别列于表3、表4和表5中。Cox模型的结果列于表3。此外，表3–5还报告了在配备 $11^{th}$ 代英特尔（R）酷睿（TM）i7-1165G7 CPU @ 2.80GHz处理器的计算机上各模型的计算时间。

## 4.1 风险比

### 4.1.1 移植后的移植物衰竭

在所有考察的模型中，未发现性别与移植物衰竭风险相关。与已故供者移植相比，活体供者移植与较低的移植物衰竭风险相关。在所有考察的 Copula 模型中，较高年龄（$> 50$ 年）与风险增加相关，但有两种情况例外：采用威布尔生存模型的正态 Copula 和 Gumbel Copula 未发现关联。这与 Cox模型的结果相反：在 Cox模型中，较高年龄组与较低的移植物衰竭风险相关（HR：0.911，95% CI：0.872 至 0.952，表3）。这可能部分是由于移植物衰竭被

Cox模型中因死亡而删失所致；在发生移植物衰竭之前死亡的老年人，可能会被视为发生移植物衰竭的可能性较低。
在拟合的移植物衰竭Cox模型中，有证据表明三个协变量均存在非比例风险。采用Grambsch和Therneau的方法诊断非比例性后，我们分别得到性别、年龄和供者类型的P值为0.091、0.017和 $<0.001$。

### 4.1.2 移植后死亡

对于所有考虑的模型，包括Cox模型和所有Copula模型，三个协变量：性别、年龄和供者类型，均被发现与移植后死亡相关。具体而言，女性、活体供者移植和较年轻的年龄（$\leq$ 50 岁）与较低的死亡风险相关。相比之下，男性、已故供者移植和较年长的年龄（$>$ 50 岁）与较高的死亡风险相关。这些发现适用于所有模型，且估计的风险比相当相似。

在拟合的死亡Cox模型中，有证据表明供者类型存在非比例风险（P$<0.001$），而没有足够证据表明性别存在非比例风险（P$=0.936$）和年龄存在非比例风险（P$=0.882$）。

## 4.2 移植物衰竭与死亡之间的关联

所有 Copula生存模型的结果均表明，与年轻年龄组相比，年龄较大的个体中移植物衰竭与死亡之间的关联更强。

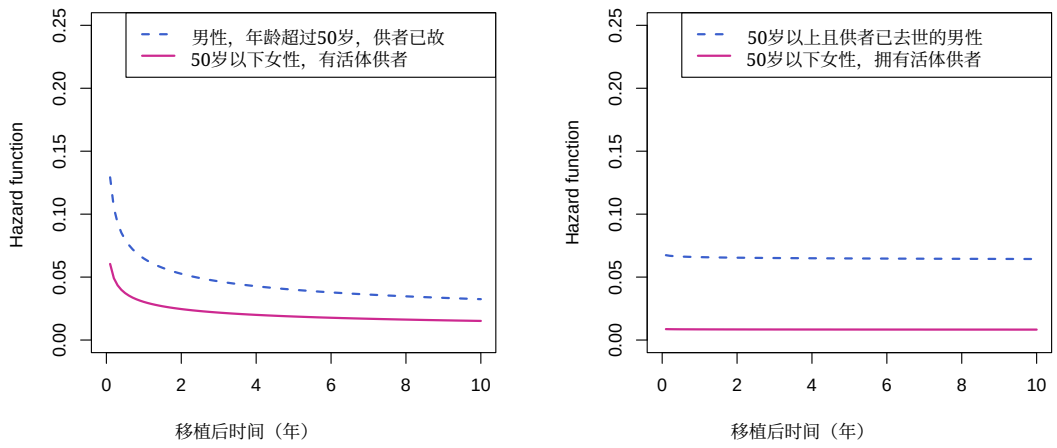
在大多数拟合的 Copula 模型中，我们观察到女性与男性受者在移植物衰竭与死亡之间的关联方面没有差异。例外的是 Clayton Copula 模型和 Frank Copula Gompertz分布模型，它们显示女性受者这两个终点之间的关联更强。同样，在大多数拟合的 Copula 模型中，我们观察到活体供者与已故供者受者在移植物衰竭与死亡之间的关联方面没有差异。例外的是 Clayton Copula 模型和 Frank Copula 指数分布模型，它们显示活体供者受者这两个终点之间的关联更强。

## 4.3 首选模型的结果

在实际数据分析中，我们报告了每个模型的赤池信息准则值。然而，我们没有使用赤池信息准则比较 Cox模型与 Copula 模型。这是因为 Cox模型分别对移植物衰竭和死亡进行拟合，且采用的是部分似然而非完整似然。相比之下，联结函数生存模型对移植物衰竭和死亡进行联合分析，并采用完整似然。因此，分别分析结局的 Cox模型与联合分析结局的基于 Copula 的参数生存模型，其赤池信息准则值不可比较。不过，我们使用赤池信息准则比较了不同的联结函数生存模型。对于每一种生存分布，根据赤池信息准则，Frank Copula 模型最优（表3、4、5）。此外，对于每一种 Copula 模型，Weibull分布的赤池信息准则值最低。

在 Frank Copula 威布尔生存模型中，年龄会影响移植物衰竭与死亡之间的关联；与较年轻年龄组相比，较年长年龄组的个体两者之间的关联更强。活体供者的移植物衰竭风险比为 0.560（95% CI：0.532 至 0.588），死亡风险比为 0.519（95% CI：0.490 至 0.549），表明活体供者受者组发生这两种事件的风险均较低。与男性相比，女性的死亡风险较低，风险比为 0.920（95% CI：0.882 至 0.957）。较年长年龄组的移植物衰竭风险比为 1.202（95% CI：1.152 至 1.251），死亡风险比为 3.734（95% CI：3.565 至 3.903），表明较年长年龄组发生移植物衰竭和死亡的风险较高。

Frank Copula 威布尔生存模型表明，总体而言，50岁以上且接受已故供者肾脏移植的男性患者，在肾移植后的死亡总体风险最高。与之相反，接受活体供者肾脏移植的年轻女性患者，死亡总体风险最低。图1分别展示了这两个亚组根据首选模型估算的风险函数。对于两个亚组，移植物衰竭风险在移植后立即达到最高，随后逐渐下降，并在4年内趋于稳定（图1a）。死亡风险随移植后时间保持稳定（图1b）。对于接受已故供者肾脏移植、年龄为$> 50$ 岁的男性受者，在肾移植后的前2年内，移植物衰竭风险高于死亡风险，此后两者的高低顺序则相反。




(a) 移植物衰竭　　(b) 死亡

图1：使用 Frank Copula 和 Weibull分布生存时间估计两组患者的风险函数：50岁以上、接受已故供者移植物的男性受者，以及50岁以下、接受活体供者移植物的女性受者。

# 5 模拟研究

我们开展了两项模拟研究。在模拟研究 1 中，我们旨在评估所提出的用于半竞争风险数据的双变量 copula 回归模型的性能。我们模拟数据以仿真肾移植数据的真实示例。我们比较了三个模型的性能：Cox比例风险模型；带有风险率预测变量的双变量 copula 回归模型（Copula 模型1）；以及同时带有风险率和关联参数预测变量的双变量 copula 回归模型（连接函数模型2）。

在模拟研究 2 中，我们旨在评估生存分布设定错误对模型参数估计的影响。我们使用赤池信息准则值为模拟数据选择拟合最佳的模型，并计算正确选中真实生存模型的比例。

## 5.1 设计

我们模拟具有已知非终末事件和终末事件边际风险率以及两个事件时间之间相关性的数据。我们评估

非终末事件和终末事件的风险比估计值的准确性，分别为 $HR_{NT}$ 和 $HR_T$。两项模拟研究中的重复次数均为 1000。模拟算法如下所示。

1. 我们生成二元协变量 $W_k$，其中 $k = 1, \ldots, p$，共 $p$ 个协变量，其比例与肾移植数据相符。2. 按照表 2 所述，分别为正态Copula生成 Pearson 相关系数，或为 Clayton、Frank 和 Gumbel Copula 生成关联参数，并选定 $b_k$。3. 采用选定的 $a_k$ 和 $c_k$，根据方程 (9) 和 (10) 分别生成非终末事件 $\lambda_1$ 和终末事件 $\lambda_2$ 的风险率。4. 我们使用条件分布法（Nelsen，2006）从指定的 Copula 模拟非终末事件和终末事件的发生时间。

(a) $(U, V)$ 具有所选 Copula 的联合分布函数，其中 $U$ 表示经过均匀变换的非终末事件发生时间，$V$ 表示经过均匀变换的终末事件发生时间。(b) 从 Uniform(0，1) 生成 $U$。(c) 根据 $C(v|u)$ 生成 $V$，其中 $C(v|u)$ 是给定 $U$ 时 $V$ 的条件Copula分布函数的逆函数。

5. 我们分别通过边际生存函数的逆函数获得 $T_1$ 和 $T_2$ ，

$$T_1 = S_{T_1}^{-1}(U),$$

$$T_2 = S_{T_2}^{-1}(V).$$

当两个事件的生存分布均为指数分布时，这里 $T_1 = -\log(U)\,\lambda_1$ 和 $T_2 = -\log(V)\,\lambda_2$。

6. 我们从均匀分布中独立模拟删失时间 $C$ .

7. 将非终末事件的时间设为 $X = \min(T_1, T_2, C)$，事件指示变量设为：若 $T_1 \leq \min(T_2, C)$，则为 $d_1 = 1$；否则为 $d_1 = 0$。

8. 将发生终末事件的时间设为 $Y = \min(T_2, C)$，事件指示变量$d_2 = 1$ 若 $T_2 \leq C$，否则为 $d_2 = 0$。

在模拟研究1中，使用Cox模型以及具有潜在Copula 函数的模型对模拟数据进行分析。例如，在分析使用Clayton Copula生成的模拟数据时，我们使用ClaytonCopula 模型。Sorrell等人（2022）开展了一项广泛的模拟研究，以评估藤函数错误设定的影响，本文将不再重复该模拟。

## 5.2 参数选择

我们生成包含 3000 名个体的数据集，其风险率和关联参数仿照基于 UKTR 数据集的真实数据分析。二元协变量根据伯努利分布生成，其参数仿照真实数据。具体而言，年龄组为 $> 50$ 岁的概率为 0.40，女性的概率为 0.38，活体供者受者的概率为 0.30。

对于模拟研究 1 和 2，回归系数分别设为补充材料表 S1 和表 S2 中的数值。删失时间 $C$ 独立地从均匀分布 Uniform(0，25) 中生成，代表最长 25 年的随访时间。

## 5.3 性能指标

我们使用 R 中的 `optim` 函数最大化第 3 节所述的对数似然，以估计边际参数和关联参数的回归系数。通过偏倚、均方误差（MSE）和覆盖概率评估最大似然估计的性能。

## 5.4 模拟研究 1 的结果：所提出的双变量 copula 回归模型的性能

对于非终末事件，如果数据分别来自正态、Clayton、Frank和 Gumbel Copula（表6）模拟，则从Cox模型估计的年龄组风险比具有以下覆盖概率：11.7%、0.0%、0.0%和18.9%。这些覆盖概率远低于名义水平。在将协变量纳入风险率但不纳入关联参数的Copula 回归模型中（Copula 模型1），相应的覆盖概率为0.0%、90.3%、56.7%、45.5%（表6）。使用连接函数模型2时，即将协变量纳入风险率和关联参数，

每种 Copula 的覆盖概率都接近名义水平，分别为 94.6%、95.1%、94.5%、95.2%（表6）。与 Cox模型相比，我们观察到连接函数模型2中年龄组的非终末风险比的偏倚和均方误差均有所降低。这解释了真实数据分析中 Cox模型与 Copula 模型的发现为何存在差异。

对于非终末事件中性别和供者类型的风险比，Cox模型和Copula模型2的覆盖概率均接近名义水平。不过，Copula模型的结果显示，偏倚和均方误差略有降低。例如，对于由正态Copula生成的数据，使用Cox模型时，协变量年龄 agre 对应的非终末事件风险比的偏倚和均方误差分别为 0.240 和 0.062。与Copula模型2的结果相比，偏倚为 $-0.003$，均方误差为 0.007。

终末事件风险比的结果在Cox模型、Copula模型1和2之间相似，覆盖概率接近95%。这是预期的，因为终末事件不会被非终末事件删失。然而，在数据由 Clayton Copula生成的情况下，对于年龄组的风险比，使用Copula 模型1时，覆盖概率为100.0%。在Copula 模型2中将协变量年龄组纳入关联参数后，我们发现覆盖概率提高到95.7%。

## 5.5 模拟研究 2 的结果：生存分布设定错误的影响

我们评估了采用其他生存分布的情况，以及生存分布设定错误的影响。为开展考察错误识别问题的模拟研究，我们使用四种不同的 Copula 分布模拟数据，以仿真真实数据集；真实值见补充材料表 S2。我们还评估了使用赤池信息准则选择生存分布的效果。我们模拟了以下生存分布与 Copula 函数的组合：指数分布、Weibull分布和Gompertz分布，以及正态、Clayton、Gumbel和 Frank Copula。

模拟研究 2 的结果见补充材料中的表 S3。下面总结本次模拟研究的发现。我们的模拟研究表明，AIC 可用于选择生存分布。对于威布尔分布，几乎 100% 的情况下都能正确选出生存分布；对于 Gompertz分布，正确率约为 91% 或更高；对于约 78% 或

上述指数分布。若未能正确识别潜在的指数分布，约有10%的情况下会将生存分布错误设定为Weibull分布或冈珀茨分布。然而，总体而言，对于将指数分布误识别为Weibull分布或冈珀茨分布，估计结果具有稳健性。按最低赤池信息准则所选模型的性能总结表明，覆盖概率接近名义水平。

# 6 讨论

我们提出了一组 Copula 回归模型，使风险率以及非终末事件和终末事件发生时间之间的关联均可随协变量变化。我们采用条件 Copula（Patton，2006），并为半竞争风险事件的风险率和关联参数设置预测变量。所提方法的优势在于，估计风险比时考虑了半竞争风险之间的相关性，同时能够灵活使用多种 Copula 描述生存终点之间不同的关联模式。

我们的研究表明，考虑半竞争风险之间相关性的重要性。在真实数据分析中，对于年龄组的影响得出了不同结论：Cox模型显示，年龄较大与较低的移植物衰竭风险相关（风险比：0.911，95%置信区间：0.872至0.952）；而我们提出的Copula模型均显示，年龄较大与较高的移植物衰竭风险相关。AIC最低的Frank Copula威布尔生存模型估计的风险比为1.202，95%置信区间为1.152至1.251（表5）。

我们开展模拟研究，以评估 Copula回归模型的表现，并考察使用赤池信息准则选择生存分布的效果。与 Cox模型相比，使用连接函数模型2能够改善非终末事件风险比的估计，其覆盖概率接近名义水平；而在某些情景下，Cox模型的覆盖概率为 0%（表6）。这些结果证实了 Leffondr´e 等人（2013）的发现。对于非终末事件风险比，与 Cox模型相比，使用 Copula回归模型可以降低偏倚和均方误差。我们的研究强调，在分析协变量对某一终点的影响时，若该协变量对某种竞争风险具有强烈影响，就必须考虑半竞争风险。

We 已考虑生存分布设定错误的影响。Ou r

模拟研究表明，在大多数情况下，可以使用赤池信息准则选择生存分布。在其余情况下，当基础分布为指数分布或Gompertz分布时，更有可能出现模型错设。然而，使用威布尔分布错设这两种分布，在参数估计值方面仍然具有良好的性质。进一步的研究还可以考察其他参数生存分布，或使用非参数方法对边际生存函数进行建模。

我们采用了完全极大似然法来估计模型参数。我们承认，对于Copula生存模型而言，采用两阶段估计程序估计关联参数可能更加高效（Shih和Louis，1995）。在两阶段方法中，关联参数和边际生存分布中的参数分别进行估计。第一阶段假设边际分布相互独立，并对其中的参数进行估计。第二阶段将两个边际分布固定为第一阶段的估计值，然后估计关联参数。这种方法在第一阶段忽略了依赖结构，但其优势在于实际应用中效率较高，尤其是在模型变得更加复杂时。在我们提出的模型中应用两阶段方法，并将其与单阶段完全极大似然法进行比较，是未来研究的一个方向。另一个可能的扩展是使用非参数方法（Li等人，2020）或Cox比例风险模型（Li等人，2021）来估计边际生存函数。此外，研究使用网格搜索方法寻找用于优化似然函数的初始值也很有意义。

在本文中，我们通过纳入协变量，扩展了近期关于双变量半竞争风险模型的研究。未来研究的另一个方向可以是扩展该模型，纳入脆弱性项，以解释未测量的协变量。由于我们拥有双变量半竞争风险数据，因此这一扩展将需要使用双变量分布对脆弱性项进行建模，以解释两个脆弱性项之间潜在的相关性。

我们使用指数分布、Weibull分布和Gompertz分布来说明所提出模型的方法与应用。然而，这些方法可轻松推广至其他参数生存分布。我们通过模拟研究评估了所提出的联结函数生存模型的性能，并使用赤池信息准则对它们进行比较。未来的研究可进一步探索如何评估联结函数生存模型的拟合优度。

由于本文重点介绍带有二元协变量的Copula回归模型，因此纳入代表年轻和年长年龄组的二分年龄组（$> 50$ 年或 $\leq 50$ 年）作为示例。我们承认，年龄最好使用更多类别进行分析，或作为连续变量进行分析。

在后续工作中，我们正在开发能够纳入更多年龄类别，或将年龄作为连续变量的Copula回归模型。

未来的研究可以探讨纳入连续协变量和分类协变量。我们考虑了使用线性函数将协变量纳入风险率和关联参数，但也可以考虑其他替代方案。Acar 等（2013）通过使用广义似然比检验，开发了比较潜在函数的方法，这些函数描述了Copula 函数关系中关联参数与协变量之间的关系。我们使用可用案例分析来说明我们方法的应用。未来的研究可以在使用我们提出的方法时采用多重插补来处理缺失数据。当使用正态Copula 模型时，这可能会耗费较长时间。

如表 3–5 所示，正态Copula 模型的似然函数优化需要 2–4 小时。未来的研究可以探讨如何加快单个数据集的分析优化过程，以及如何在 R 中使用并行计算开展模拟研究。

**致谢**

作者衷心感谢英国国家医疗服务体系血液与移植部门提供本项目所用的英国移植登记处数据访问权限。我们感谢英国所有为本项目提供数据的移植中心。LS 获得了普利茅斯大学工程、计算与数学学院博士生奖学金。YW 获得了英国研究与创新署医学研究委员会奖学金（MC/W021358/1）的资助，并获得了英国研究与创新署工程与物理科学研究委员会影响加速账户（EP/X525789/1）的资助。我们很高兴能够使用普利茅斯大学科学与工程学院的高性能计算集群开展模拟研究。

# 参考文献

Acar, E. F.,Craiu,R.V. 和Yao, F.（2013）。条件联结函数模型中协变量效应的统计检验。电子统计学期刊，7:2822–2850。Akaike,H.（1974）。重新审视统计模型识别。*IEEE* 自动控制汇刊，19(6):716–723。Becker, B.,Becker, Y.,Pintar, T., Collins, B., Pirsch, J., Friedman, A., Sollinger, H. 和 Brazy, P.（2000）。利用肾移植评估一种简单方法，以预测终末期肾病治疗对患者

生存：观察到的/预期寿命。《美国肾脏病杂志》，35(4):653–659。Berry, S. D.、Ngo, L.、Samelson, E. J. 和 Kiel, D. P.（2010）。死亡竞争风险：老年人研究中的一项重要考量。《美国老年医学会杂志》，58(4):783–787。Chen, Y. H.（2012）。利用半参数回归模型对半竞争风险数据进行最大似然分析。*Lifetime Data Analysis*，18(1):36–57。Cox, D. R.（1972）。回归模型与生命表。《英国皇家统计学会杂志：*B*辑（统计方法论）》，34(2):187–220。Fabrizii, V.、Winkelmayer, W.、Klauser, R.、Kletzmayr, J.、S¨aemann, M.、Steininger, R.、Kramar, R.、H¨orl, W. 和 Kovarik, J.（2004）。老年肾移植受者的患者和移植物存活：年龄是否重要？《美国肾脏病学会杂志》，15(4):1052–1060。Fine, J. P.、Jiang, H. 和 Chappell, R.（2001）。关于半竞争风险数据。《生物计量学》，88(4):907–919。Fu, H.、Wang, Y.、Liu, J.、Kulkarni, P. M. 和 Melemed, A. S.（2013）。利用贝叶斯正态诱导联结函数估计模型对无进展生存和总生存进行联合建模。《医学统计学》，32(2):240–254。Ghosh, D.（2006）。半竞争风险数据关联的半参数推断。《医学统计学》，25(12):2059–2070。Hsieh, J. J.、Wang, W. 和 Ding, A. A.（2008）。基于半竞争风险数据的回归分析。《英国皇家统计学会杂志：*B*辑（统计方法论）》，70(1):3–20。Keith, D. S.、Demattos, A.、Golconda, M.、Prather, J. 和 Norman, D.（2004）。供者与受者年龄匹配对首次接受已故供者肾移植后生存的影响。《美国肾脏病学会杂志》，15(4):1086– 1091。Leffondr´e, K.、Touraine, C.、Helmer, C. 和 Joly, P.（2013）。区间删失事件发生时间与死亡竞争风险：疾病-死亡模型是否比Cox模型更准确？《国际流行病学杂志》，42(4):1177–1186。Li, D.、Hu, X. 和 Wang, R.（2021）。评估受信息性删失观测影响的两个事件时间之间的关联。《美国统计学会杂志。

Li, D., Hu, X. J.,McBride, M.L.和 Spinelli, J. J.（2020）。信息性删失存在时的多个事件时间：通过 Copula 进行建模与分析。*Lifetime Data Analysis*, 26(3):573–602。Nelsen, R. B.（2006）。*An Introductionto Copulas*。Springer-Verlag柏林，海德堡，第2版。Patton,A. J.（2006）。非对称汇率依赖。*International Economic Review*，47(2):527–556。Peng,L.和Fine, J. P.（2007）。半竞争风险数据的回归建模。*Biometrics*，63(1):96–108。Port, F. K.、Dykstra, D.M.、Merion,R.M. 和Wolfe, R. A.（2004）。美国器官捐献与移植的趋势和结果。*AmericanJournal of Transplantation*，5(4Part 2):843–849。RCore Team（2021）。*R*：用于统计计算的语言和环境。R Foundationfor StatisticalComputing，维也纳，奥地利。Renfro,L. A.、Shang, H.和Sargent,D. J.（2015）。Copula方向设定对替代终点多试验评估的影响。*Journal ofBiopharmaceutical Statistics*，25(4):857–877。Shih,J. H. 和Louis, T.A.（1995）。双变量生存数据Copula模型中关联参数的推断。*Biometrics*，51(4):1384–1399。Sklar, A.（1959）。n维分布函数及其边缘分布。*PublicInstitute Statistical University*，8:229–231。Sorrell,L.、Wei,Y.、Wojtys, M. 和Rowe, P.（2022）。半竞争风险生存终点之间相关性的估计。*Biometrial Journal*，64(1):131– 145。Terasaki, P. I.、Cecka, J.、Gjertson,D.和Takemoto,S.（1995）。来自配偶和活体无关供者的肾移植具有较高的生存率。*NewEngland Journal of Medicine*，333(6):333–336。Zhu, H.、Lan, Y.、Ning, J. 和 Shen,Y.（2021）。半竞争风险数据的半参数Copula回归建模。*Communications in Statistics- TheoryandMethods*。

| Model/ Covariate | Hazard Ratio (95%CI) | | Regression coefficients on | AIC | Time |
|---|---|---|---|---|---|
| | Graft Failure (GF) | Death (D) | association parameter | | (minutes) |
| Cox model | | | | 174,873.6 (GF) | 0.004 |
| | | | | 153,814.9 (D) | 0.005 |
| Age group: $> 50$ | 0.911 (0.872, 0.952) | 3.926 (3.745, 4.116) | NA | | |
| Sex: Female | 1.029 (0.986, 1.073) | 0.881 (0.841, 0.922) | NA | | |
| Donor type: Living | 0.615 (0.584, 0.648) | 0.539 (0.506, 0.574) | NA | | |
| Normal copula Exponential survival model | | | | 147545.5 | 123.6 |
| Age group: $> 50$ | 1.121 (1.074, 1.169) | 3.748 (3.576, 3.920) | 0.285 (0.248, 0.321) | | |
| Sex: Female | 1.015 (0.973, 1.058) | 0.900 (0.860, 0.939) | 0.021 ($-0.015$, 0.058) | | |
| Donor type: Living | 0.600 (0.569, 0.630) | 0.522 (0.490, 0.553) | 0.027 ($-0.025$, 0.078) | | |
| Clayton copula Exponential survival model | | | | 146956.7 | 1.63 |
| Age group: $> 50$ | 1.383 (1.329, 1.437) | 3.863 (3.691, 4.036) | 1.093 (0.983, 1.203) | | |
| Sex: Female | 1.006 (0.967, 1.045) | 0.934 (0.898, 0.969) | 0.139 (0.039, 0.239) | | |
| Donor type: Living | 0.587 (0.559, 0.616) | 0.540 (0.512, 0.568) | 0.529 (0.384, 0.674) | | |
| Frank copula Exponential survival model | | | | 146660.5 | 2.19 |
| Age group: $> 50$ | 1.361 (1.309, 1.414) | 3.861 (3.689, 4.032) | 5.068 (4.600, 5.537) | | |
| Sex: Female | 1.00 (0.961, 1.039) | 0.930 (0.893, 0.966) | 0.352 ($-0.059$, 0.762) | | |
| Donor type: Living | 0.586 (0.557, 0.615) | 0.535 (0.506, 0.564) | 0.861 (0.211, 1.511) | | |
| Gumbel copula Exponential survival model | | | | 147827.8 | 4.29 |
| Age group: $> 50$ | 1.134 (1.086, 1.182) | 3.667 (3.498, 3.837) | 1.355 (1.204, 1.507) | | |
| Sex: Female | 1.015 (0.973, 1.057) | 0.899 (0.859, 0.939) | 0.056 ($-0.079$, 0.191) | | |
| Donor type: Living | 0.603 (0.573, 0.634) | 0.530 (0.498, 0.562) | $-0.039$ ($-0.231$, 0.153) | | |

Table 3: Cox model and Exponential survival copula models: Results from analysing the UK Transplant Registry. The binary covariates are defined as follows, sex: female vs male (reference), donor type: living vs deceased (reference), and age group: $> 50$ years vs $\leq 50$ (reference).

| Model/ Covariate | Hazard Ratio (95%CI) | | Regression coefficients on | AIC | Time |
|---|---|---|---|---|---|
| | Graft failure | Death | association prameter | | (minutes) |
| Normal copula Gompertz survival model | | | | 146872.6 | 250.8 |
| Age group: $> 50$ | 1.080 (1.033, 1.126) | 4.163 (3.969, 4.357) | 0.282 (0.246, 0.318) | | |
| Sex: Female | 1.018 (0.976, 1.061) | 0.893 (0.854, 0.932) | 0.024 ($-0.012$, 0.060) | | |
| Donor type: Living | 0.594 (0.563, 0.624) | 0.553 (0.520, 0.586) | 0.020 ($-0.03$, 0.071) | | |
| Clayton copula Gompertz survival model | | | | 146708.3 | 5.28 |
| Age group: $> 50$ | 1.296 (1.242, 1.349) | 3.935 (3.756, 4.114) | 0.947 (0.835, 1.059) | | |
| Sex: Female | 0.999 (0.959, 1.039) | 0.909 (0.873, 0.945) | 0.161 (0.059, 0.263) | | |
| Donor type: Living | 0.592 (0.562, 0.622) | 0.555 (0.524, 0.586) | 0.287 (0.118, 0.456) | | |
| Frank copula Gompertz survival model | | | | 146409.4 | 4.81 |
| Age group: $> 50$ | 1.232 (1.182, 1.283) | 3.949 (3.771, 4.127) | 2.070 (1.655, 2.486) | | |
| Sex: Female | 1.003 (0.963, 1.042) | 0.919 (0.881, 0.957) | 0.639 (0.233, 1.045) | | |
| Donor type: Living | 0.544 (0.516, 0.572) | 0.498 (0.469, 0.528) | 0.423 ($-0.240$, 1.087) | | |
| Gumbel copula Gompertz survival model | | | | 147036.8 | 6.67 |
| Age group: $> 50$ | 1.065 (1.020, 1.111) | 4.11 (3.918, 4.303) | 1.401 (1.251, 1.505) | | |
| Sex: Female | 1.012 (0.970, 1.054) | 0.891 (0.852, 0.931) | $-0.017$ ($-0.152$, 0.118) | | |
| Donor type: Living | 0.596 (0.566, 0.627) | 0.557 (0.524, 0.591) | $-0.096$ ($-0.29$, 0.097) | | |

Table 4: Gompertz survival copula models: Results from analysing the UK Transplant Registry. The binary covariates are defined as follows, sex: female vs male (reference), donor type: living vs deceased (reference), and age group: $> 50$ years vs $\leq 50$ (reference).

| Model/ Covariate | Hazard Ratio (95%CI) | | Regression coefficients | AIC | Time |
|---|---|---|---|---|---|
| | Graft failure | Death | on association parameter | | (minutes) |
| Normal copula Weibull survival model | | | | 145038.4 | 201.0 |
| Age group: $> 50$ | 1.005 (0.962, 1.048) | 3.692 (3.523, 3.860) | 0.283 (0.240, 0.326) | | |
| Sex: Female | 1.022 (0.979, 1.064) | 0.904 (0.865, 0.943) | 0.028 ($-0.014$, 0.071) | | |
| Donor type: Living | 0.573 (0.544, 0.602) | 0.520 (0.489, 0.551) | 0.034 ($-0.027$, 0.094) | | |
| Clayton copula Weibull survival model | | | | 145137.8 | 8.03 |
| Age group: $> 50$ | 1.212 (1.160, 1.264) | 3.699 (3.532, 3.867) | 0.907 (0.782, 1.032) | | |
| Sex: Female | 1.024 (0.983, 1.066) | 0.924 (0.887, 0.96) | 0.162 (0.054, 0.271) | | |
| Donor type: Living | 0.562 (0.533, 0.590) | 0.527 (0.498, 0.556) | 0.479 (0.314, 0.643) | | |
| Frank copula Weibull survival model | | | | 144704.7 | 6.16 |
| Age group: $> 50$ | 1.202 (1.152, 1.251) | 3.734 (3.565, 3.903) | 3.954 (3.488, 4.402) | | |
| Sex: Female | 1.005 (0.965, 1.045) | 0.920 (0.882, 0.957) | 0.171 ($-0.251$, 0.592) | | |
| Donor type: Living | 0.560 (0.532, 0.588) | 0.519 (0.490, 0.549) | 0.493 ($-0.174$, 1.160) | | |
| Gumbel copula Weibull survival model | | | | 145276.1 | 6.27 |
| Age group: $> 50$ | 0.992 (0.949, 1.034) | 3.600 (3.436, 3.765) | 1.073 (0.935, 1.211) | | |
| Sex: Female | 1.020 (0.978, 1.062) | 0.906 (0.867, 0.945) | 0.044 ($-0.08$, 0.169) | | |
| Donor type: Living | 0.581 (0.552, 0.611) | 0.520 (0.489, 0.551) | $-0.043$ ($-0.218$, 0.132) | | |

Table 5: Weibull survival copula models: results from analysing the UK Transplant Registry. The binary covariates are defined as follows, sex: female vs male (reference), donor type: living vs deceased (reference), and age group: $> 50$ years vs $\leq 50$ (reference).

| | Normal | | | Clayton | | | Frank | | | Gumbel | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Estimand | Bias | CP | MSE | Bias | CP | MSE | Bias | CP | MSE | Bias | CP | MSE |
| **Cox Model** | | | | | | | | | | | | |
| $\text{HR}_{NT}$, age | 0.240 | 11.7 | 0.062 | 0.552 | 0.0 | 0.309 | 0.505 | 0.0 | 0.260 | 0.223 | 18.9 | 0.055 |
| $\text{HR}_{NT}$, sex | −0.016 | 94.3 | 0.006 | −0.012 | 95.2 | 0.006 | −0.026 | 94.0 | 0.007 | −0.013 | 94.8 | 0.006 |
| $\text{HR}_{NT}$, donor | −0.019 | 92.8 | 0.003 | −0.019 | 94.0 | 0.003 | −0.024 | 93.0 | 0.003 | −0.006 | 94.4 | 0.003 |
| $\text{HR}_{T}$, age | −0.010 | 95.8 | 0.079 | 0.000 | 94.6 | 0.086 | −0.012 | 95.4 | 0.080 | −0.004 | 96.4 | 0.074 |
| $\text{HR}_{T}$, sex | −0.001 | 95.1 | 0.005 | −0.005 | 94.7 | 0.005 | −0.005 | 94.8 | 0.005 | 0.000 | 95.5 | 0.005 |
| $\text{HR}_{T}$, donor | 0.002 | 95.2 | 0.002 | 0.000 | 94.9 | 0.002 | −0.001 | 95.5 | 0.002 | 0.001 | 94.8 | 0.002 |
| **Copula model 1** | | | | | | | | | | | | |
| $\text{HR}_{NT}$, age | −0.155 | 0.000 | 0.024 | −0.042 | 90.3 | 0.012 | −0.147 | 56.7 | 0.030 | −0.146 | 45.5 | 0.027 |
| $\text{HR}_{NT}$, sex | 0.075 | 100.0 | 0.006 | 0.004 | 95.0 | 0.005 | 0.007 | 94.9 | 0.005 | 0.004 | 94.6 | 0.005 |
| $\text{HR}_{NT}$, donor | −0.018 | 100.0 | 0.000 | −0.007 | 92.5 | 0.003 | −0.002 | 94.0 | 0.002 | 0.000 | 94.2 | 0.003 |
| $\text{HR}_{T}$, age | −0.042 | 100.0 | 0.002 | 0.397 | 76.9 | 0.252 | 0.144 | 94.6 | 0.105 | −0.132 | 91.2 | 0.083 |
| $\text{HR}_{T}$, sex | 0.046 | 100.0 | 0.002 | 0.017 | 95.4 | 0.004 | −0.001 | 95.3 | 0.004 | −0.002 | 94.9 | 0.004 |
| $\text{HR}_{T}$, donor | −0.051 | 100.0 | 0.003 | 0.028 | 92.4 | 0.003 | 0.003 | 95.6 | 0.002 | −0.003 | 94.6 | 0.002 |
| **Copula model 2** | | | | | | | | | | | | |
| $\text{HR}_{NT}$, age | −0.003 | 94.6 | 0.007 | −0.002 | 95.1 | 0.009 | −0.002 | 94.5 | 0.010 | −0.003 | 95.2 | 0.007 |
| $\text{HR}_{NT}$, sex | −0.003 | 94.6 | 0.006 | −0.001 | 95.9 | 0.005 | −0.002 | 95.3 | 0.005 | −0.003 | 95.4 | 0.006 |
| $\text{HR}_{NT}$, donor | 0.002 | 94.1 | 0.003 | 0.002 | 93.6 | 0.002 | 0.002 | 94.0 | 0.003 | 0.002 | 94.4 | 0.003 |
| $\text{HR}_{T}$, age | −0.009 | 95.7 | 0.077 | −0.011 | 95.1 | 0.080 | −0.014 | 95.0 | 0.093 | −0.004 | 95.9 | 0.073 |
| $\text{HR}_{T}$, sex | −0.001 | 95.1 | 0.004 | −0.002 | 95.5 | 0.004 | −0.001 | 94.6 | 0.005 | 0.000 | 94.8 | 0.004 |
| $\text{HR}_{T}$, donor | 0.002 | 95.1 | 0.002 | 0.001 | 94.0 | 0.002 | 0.002 | 94.6 | 0.002 | 0.002 | 94.7 | 0.002 |
| $\beta_0$ | 0.005 | 94.3 | 0.005 | 0.017 | 95.5 | 0.031 | 0.018 | 94.2 | 0.253 | 0.066 | 96.6 | 0.123 |
| $\beta_1$ | −0.006 | 96.0 | 0.006 | −0.018 | 96.2 | 0.032 | −0.056 | 95.6 | 0.557 | −0.055 | 96.9 | 0.129 |
| $\beta_2$ | −0.004 | 95.4 | 0.007 | −0.004 | 95.1 | 0.025 | −0.033 | 94.9 | 0.445 | −0.007 | 96.4 | 0.077 |
| $\beta_3$ | 0.002 | 96.0 | 0.009 | −0.005 | 94.6 | 0.033 | −0.030 | 95.0 | 0.839 | 0.014 | 95.6 | 0.094 |

Table 6: Results of simulation study. Copula model 1 represents the copula model with the marginal hazard rates depending on the covariates. Copula model 2 represents the copula model with both the marginal hazard rates and association parameters depending on the covariates. Data are generated from the Normal, Clayton, Frank and Gumbel copulas and the true copula distribution is used for estimation. We generate 1000 data sets with 3000 individuals in each. The non-terminal hazard ratio for each covariate is given by $\text{HR}_{NT,\text{covariate}}$, and the terminal hazard ratio for each covariate is given by $\text{HR}_{T,\text{covariate}}$. MSE refers to the mean squared error and CP refers to the coverage probability, given as a percentage.