

竞争风险下Fine-Gray模型与Cox模型 的比较与评价

都业成, 侯文*

辽宁师范大学数学学院, 辽宁 大连

收稿日期: 2024年1月5日; 录用日期: 2024年1月31日; 发布日期: 2024年2月7日

摘要

在生存分析中经常会遇到存在竞争风险的情形, 在存在竞争风险情形下, 比较Cox模型与Fine-Gray模型估计的分析效果。本文介绍了在存在竞争风险情形下的Fine-Gray模型, 举例说明了Fine-Gray模型风险集的定义以及权重计算方法, 并利用Cox模型与Fine-Gray模型对实际存在竞争风险的数据集进行了实例分析, 结果表明, Fine-Gray模型与Cox模型分析结果不同, 尤其是Cox模型会高估事件的累计发生率。因此, 在生存分析中存在竞争风险时, 选择Fine-Gray模型进行建模是更合适的。

关键词

Fine-Gray模型, Cox模型, 竞争风险

Comparison and Evaluation of Fine-Gray and Cox Models under Competitive Risks

Yecheng Du, Wen Hou*

School of Mathematics, Liaoning Normal University, Dalian Liaoning

Received: Jan. 5th, 2024; accepted: Jan. 31st, 2024; published: Feb. 7th, 2024

Abstract

In survival analysis, scenarios with competing risks are frequently encountered. In the presence of competing risks, the analysis performance of the Cox model and the Fine-Gray model is compared. This paper introduces the Fine-Gray model in the context of competing risks, providing a detailed explanation of the definition of the Fine-Gray model's risk set and the calculation method for weights.

*通讯作者。

Through example analysis using actual data with competing risks, the paper demonstrates that the results of the Fine-Gray model and the Cox model differ, especially with the Cox model overestimating the cumulative incidence of events. Therefore, in survival analysis, when facing competing risks, choosing the Fine-Gray model for modeling is more suitable.

Keywords

Fine-Gray Model, Cox Model, Competing Risks

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在许多学科领域, 如医学、生物学、保险精算学、经济学以及人口统计学等多个领域都需要研究生存分析问题, 当所研究的问题只存在单一结局时, Kaplan 和 Meier 提出的 K-M 方法[1]和 Cox 提出的 Cox 模型[2]是最为常用的方法, Cox 模型是一种常用的生存分析模型, 用于探讨与生存时间相关的因素。它假设风险比在时间的推移过程中是恒定的, 不随时间变化。在被给出的协变量集合 $x_{1i}, x_{2i}, \dots, x_{mi}$ 下, 个体 i 的风险可以分解成两个部分, 一部分是包括协变量不包括时间, 一部分包括时间不包括协变量。Cox 模型基本形式如下:

$$h_i(t) = h_0(t) \exp\{\beta_1 x_{1j} + \dots + \beta_m x_{mj}\}$$

其中 $h_0(t)$ 是基线危险函数, $\beta_1, \beta_2, \dots, \beta_m$ 是需要估计的参数或系数。 $t_1 < t_2 < \dots < t_r$ 是惟一的有序感兴趣时间点, 回归系数 $\beta_1, \beta_2, \dots, \beta_m$ 由偏似然方程的最大值决定。累积危险函数为 $H(t) = \int_0^t h(u) du$, 累积分布函数 $F(t)$ 可由 $H(t)$ 导出, $F(t) = 1 - \exp(-H(t))$ 。

然而, 所研究的问题并不总是单一结局的, 因此在某些情况下, 上述方法可能并不适用, 其中一种情况就是存在竞争风险, Gelman [3]和 Gooley [4]对竞争风险理论进行了描述, Kalbfleisch 和 Prentice [5]提出了一种考虑竞争风险的方法, 在存在竞争风险条件下, 事件 i 发生在时间 t 之前的概率分布叫做子分布, 类型 i 事件的累计发生率函数被定义为 $F_i(t) = P(T \leq t, C = i)$, 记为 CIF, Fine 和 Gray [6]于 1999 年提出了 Fine-Gray 模型, 这是一种基于 CIF 建模的一种竞争风险模型, 然而, 所研究的问题并不总是单一结局的, 因此在某些情况下, 上述方法可能并不适用, 其中一种情况就是存在竞争风险, Gelman [3]和 Gooley [4]对竞争风险理论进行了描述, Kalbfleisch 和 Prentice [5]提出了一种考虑竞争风险的方法, 在存在竞争风险条件下, 事件 i 发生在时间 t 之前的概率分布叫做子分布, 类型 i 事件的累计发生率函数被定义为 $F_i(t) = P(T \leq t, C = i)$, 记为 CIF, Fine 和 Gray [6]于 1999 年提出了 Fine-Gray 模型, 这是一种基于 CIF 建模的竞争风险模型, Fine-Gray 模型在存在竞争风险的生存分析实际问题中有着广泛应用, 朱旭[7]和钱迪[8]等分别以胃癌和肝癌患者为样本构建 Fine-Gray 模型研究生存分析问题。

2. Fine-Gray 模型介绍

该模型基于

$$\gamma(t, x) = \gamma_0(t) e^{\beta x}$$

这里的 γ 是子分布危险函数, γ_0 是子分布的基线危险函数, x 是协变量向量, β 是回归系数。其中子分布危险函数为

$$\gamma_i(t) = \lim_{\delta t \rightarrow 0} \left\{ \frac{P(t < T < t + \delta t, C = i | T > t \text{ 或 } T \leq t \text{ 且 } C \neq i)}{\delta t} \right\}$$

偏似然方程为

$$L(\beta) = \prod_{j=1}^r \frac{\exp(\beta x_j)}{\sum_{i \in R_j} w_{ji} \exp(\beta x_i)}$$

在所有的感兴趣的事件被观察到的时间点取乘积, 其中 $t_1 < t_2 < \dots < t_r$ 。Fine-Gray 模型的偏似然形式和 Cox 比例风险模型相似, 与 Cox 比例风险模型相比, Fine-Gray 模型主要有两个差异, 风险集合 R_j 定义方式不同以及引入了删失生存权重 w_{ji} 。风险集合 R_j 由在时间 t 之前未经历事件的个体和在时间 t 之前经历了一个竞争风险事件的个体组成。

$$R_j(t) = \{i: T_i \geq t \text{ 或 } T_i \leq t \text{ 并且经历过一个竞争风险事件的个体}\}$$

因此, 那些经历过竞争风险事件的个体一直处于风险集中。在存在竞争风险的前提下, 一种类型 i 事件在时间 t 之前不发生的概率称为子生存函数, 定义为 $S_i(t) = P(T > t, C = i)$ 。

删失生存概率权重 w_{ji} 定义为

$$w_{ji} = \frac{\hat{S}(t_j)}{\hat{S}(\min(t_j, t_i))}$$

这里的 \hat{S} 是删失生存分布函数的一个 K-M 估计, 需要解释一下, 删失生存分布定义为 (T_i, C_i) , 这里的 T_i 定义为观察到第一个事件的时间, 如果未观察到事件则 $C_i = 1$, 而观察到事件时 $C_i = 0$ 。在每个观测到感兴趣事件发生的时间点(下标为 j), 风险集合由在时间 t 之前未经历事件的个体和在时间 t 之前经历了一个竞争风险事件的个体组成。对给出的部分似然的对数进行求导, 得到得分统计量:

$$U(\beta) = \sum_{j=1}^r \left\{ x_j - \frac{\sum_{i \in R_j} w_{ji} x_i \exp(x_i \beta)}{\sum_{i \in R_j} w_{ji} \exp(x_i \beta)} \right\}$$

$\hat{\beta}$ 表示回归系数 β 的估计, 是函数 $U(\beta) = 0$ 时 β 的取值。

为说明 Fine-Gray 模型的计算过程, 给出示例数据集见表 1。表 1 中包括 10 个观测个体, 其观测到的时间点, 事件类型, 每个个体协变量为分别在表 1 的第 1~4 列。

Table 1. Example dataset

表 1. 示例数据集

个体单元	t_i	事件类型	x	\hat{S}
1	t_1	III	12	0.90
2	t_2	III	10	0.80
3	t_3	I	8	0.80

续表

4	t_4	II	9	0.80
5	t_5	I	7	0.80
6	t_6	III	10	0.64
7	t_7	I	11	0.64
8	t_8	II	13	0.64
9	t_9	III	8	0.32
10	t_{10}	I	9	0.32

备注: 事件类型 I: 感兴趣事件; 事件类型 II: 竞争风险事件 ; 事件类型 III: 删失。

表 1 第 5 列给出了删失分布生存函数估计值 \hat{S} , 这里的 \hat{S} 由 K-M 估计计算可得, K-M 方法的过程如下: 假定 $t_1 < t_2 < \dots < t_r$ 是唯一的删失时间点, 设 d_j 为在 t_j 时间发生删失的个体数量, n_j 为在 t_j 之前有经历该事件风险的个体数量。则

$$\hat{S}(t) = \prod_{t_j \leq t} \frac{n_j - d_j}{n_j}$$

在 10 个观测个体中, 有 4 个个体发生删失。删失时间点分别在 $t = t_1, t_2, t_6, t_9$, 这 4 个删失时间点的生存分布 \hat{S} 值, 计算如下:

$$\hat{S}(t_1) = \frac{10-1}{10} = 0.90$$

$$\hat{S}(t_2) = \hat{S}(t_1) \times \frac{8}{9} = 0.80$$

$$\hat{S}(t_6) = \hat{S}(t_2) \times \frac{4}{5} = 0.64$$

$$\hat{S}(t_9) = \hat{S}(t_6) \times \frac{1}{2} = 0.32$$

在其他未发生删失 6 个时间点, 每个时间点的生存函数估计值 \hat{S} 与小于该时间点的前一个删失时间点的生存估计值相同。

在得到各个时间点的生存函数估计值后, 可以计算删失生存概率权重 w_{ji} , 计算过程见表 2。观测个体单元在第 1 行, 感兴趣(类型 I)事件发生的时间点在第 1 列, 表中数据为类型 I 事件在各阶段所在风险集中取得的权重 w_{ji} 。

Table 2. Calculation process of weights w_{ji} for deleted survival probabilities

表 2. 删失生存概率权重 w_{ji} 的计算过程

类型 I 时间点	1	2	3	4	5	6	7	8	9	10
t_3	-	-	$w_{33} = 1$	$w_{34} = \frac{\hat{S}(t_3)}{\hat{S}(t_3)} = 1$	$w_{35} = 1$	$w_{36} = 1$	$w_{37} = 1$	$w_{38} = 1$	$w_{39} = 1$	$w_{3,10} = 1$

续表

t_5	-	-	-	$w_{54} = \frac{\hat{S}(t_5)}{\hat{S}(t_4)} = 1$	$w_{55} = 1$	$w_{56} = 1$	$w_{57} = 1$	$w_{58} = 1$	$w_{59} = 1$	$w_{5,10} = 1$
t_7	-	-	-	$w_{74} = \frac{\hat{S}(t_7)}{\hat{S}(t_4)} = 0.8$	-	-	$w_{77} = 1$	$w_{78} = 1$	$w_{79} = 1$	$w_{7,10} = 1$
t_{10}	-	-	-	$w_{10,4} = \frac{\hat{S}(t_{10})}{\hat{S}(t_4)} = 0.4$	-	-	-	$w_{10,8} = \frac{\hat{S}(t_{10})}{\hat{S}(t_8)} = 0.5$	-	$w_{10,10} = 1$

备注：表中“-”表示该单元对应时间点 t_i 不在讨论的风险集 R_j 中。

对于感兴趣事件(类型 I)，竞争风险事件(类型 II)，删失(类型 III) 3 种不同类型事件，10 个观测个体他们在不同时刻的权重也不相同。

注意，在风险集 R_j 中，对于时间点 t_j ，当 $t_j \leq t_i$ 时，由 $\min(t_j, t_i) = t_j$ ，此时

$$w_{ji} = \frac{\hat{S}(t_j)}{\hat{S}(\min(t_j, t_i))} = 1$$

当 $t_j > t_i$ 时，权重小于 1。

以个体 4 为例，其对应的删失生存概率权重 w_{ji} 为

$$w_{4i} = \frac{\hat{S}(t_4)}{\hat{S}(\min(t_4, t_i))}$$

个体 1 和个体 2 在 $t = t_1$ 和 $t = t_2$ 时发生删失，在此之前没有观测到感兴趣事件，因此不在风险集中，个体 3 在 $t = t_3$ 时观测到感兴趣事件，它只参与了第一项的部分似然，权重为 1，个体 4 在 $t = t_4$ 时观测到竞争风险事件，在 t_4 之前，权重为 1，随着时间推移，权重逐渐降低。个体 5 在 $t = t_5$ 时观测到感兴趣事件，参与了前两项的部分似然，权重为 1。

对于本例，需要求解的得分统计量为：

$$\begin{aligned}
 U(\beta) &= x_3 - \frac{w_{33}x_3e^{x_3\beta} + w_{34}x_4e^{x_4\beta} + \dots + w_{3,10}x_{10}e^{x_{10}\beta}}{w_{33}e^{x_3\beta} + w_{34}e^{x_4\beta} + \dots + w_{3,10}e^{x_{10}\beta}} \\
 &+ x_5 - \frac{w_{54}x_4e^{x_4\beta} + w_{55}x_5e^{x_5\beta} + \dots + w_{5,10}x_{10}e^{x_{10}\beta}}{w_{54}e^{x_4\beta} + w_{55}e^{x_5\beta} + \dots + w_{5,10}e^{x_{10}\beta}} \\
 &+ x_7 - \frac{w_{74}x_4e^{x_4\beta} + w_{77}x_7e^{x_7\beta} + \dots + w_{7,10}x_{10}e^{x_{10}\beta}}{w_{74}e^{x_4\beta} + w_{77}e^{x_7\beta} + \dots + w_{7,10}e^{x_{10}\beta}} \\
 &+ x_{10} - \frac{w_{10,4}x_4e^{x_4\beta} + w_{10,8}x_8e^{x_8\beta} + w_{10,10}x_{10}e^{x_{10}\beta}}{w_{10,4}e^{x_4\beta} + w_{10,8}e^{x_8\beta} + w_{10,10}e^{x_{10}\beta}} \\
 &= 0
 \end{aligned}$$

将表 2 中的权重值和表 1 中的 x 代入可得

$$\begin{aligned}
 U(\beta) &= 8 - \frac{8e^{8\beta} + 9e^{9\beta} + 7e^{7\beta} + 10e^{10\beta} + 11e^{11\beta} + 13e^{13\beta} + 8e^{8\beta} + 9e^{9\beta}}{e^{8\beta} + e^{9\beta} + e^{7\beta} + e^{10\beta} + e^{11\beta} + e^{13\beta} + e^{8\beta} + e^{9\beta}} \\
 &+ 7 - \frac{9e^{9\beta} + 7e^{7\beta} + 10e^{10\beta} + 11e^{11\beta} + 13e^{13\beta} + 8e^{8\beta} + 9e^{9\beta}}{e^{9\beta} + e^{7\beta} + e^{10\beta} + e^{11\beta} + e^{13\beta} + e^{8\beta} + e^{9\beta}} \\
 &+ 11 - \frac{0.8 \times 9e^{9\beta} + 11e^{11\beta} + 13e^{13\beta} + 8e^{8\beta} + 9e^{9\beta}}{0.8e^{9\beta} + e^{11\beta} + e^{13\beta} + e^{8\beta} + e^{9\beta}} \\
 &+ 9 - \frac{0.4 \times 9e^{9\beta} + 0.5 \times 13e^{13\beta} + 9e^{9\beta}}{0.4 \times e^{9\beta} + 0.5 \times e^{13\beta} + e^{9\beta}} \\
 &= 0
 \end{aligned}$$

可以算出函数的解 $\hat{\beta} = -0.4363$ 。

3. 基于 Fine-Gray 模型的 CIF 估计

定义 $H(t)$ 为子分布的累积危险函数, 它的估计计算方法如下

$$\hat{H}(t; x_0, \hat{\beta}) = \sum_{t_j \leq t} \left\{ \frac{\exp(x_0 \hat{\beta})}{\sum_{i \in R_j} w_{ji} \exp(x_i \hat{\beta})} \right\}$$

其中 t 表示感兴趣的时间点, x_0 为代入的协变量, $\hat{\beta}$ 为回归系数的估计值。

通过累积危险函数 $H(t)$ 可计算 CIF 的估计值, 计算公式如下:

$$F(t) = 1 - \exp(-H(t))$$

利用表 1 中的示例数据集, 代入 4 个感兴趣时间点及对应协变量, 计算累积危险函数以及 CIF 的估计值。

$$\begin{aligned}
 \hat{H}(t = t_3; x_0 = 8, \hat{\beta} = -0.4363) &= \frac{e^{8\hat{\beta}}}{e^{8\hat{\beta}} + e^{9\hat{\beta}} + e^{7\hat{\beta}} + e^{10\hat{\beta}} + e^{11\hat{\beta}} + e^{13\hat{\beta}} + e^{8\hat{\beta}} + e^{9\hat{\beta}}} \\
 &= 0.1773
 \end{aligned}$$

$$\begin{aligned}
 \hat{H}(t = t_5; x_0 = 7, \hat{\beta} = -0.4363) &= \frac{e^{7\hat{\beta}}}{e^{8\hat{\beta}} + e^{9\hat{\beta}} + e^{7\hat{\beta}} + e^{10\hat{\beta}} + e^{11\hat{\beta}} + e^{13\hat{\beta}} + e^{8\hat{\beta}} + e^{9\hat{\beta}}} \\
 &+ \frac{e^{7\hat{\beta}}}{e^{9\hat{\beta}} + e^{7\hat{\beta}} + e^{10\hat{\beta}} + e^{11\hat{\beta}} + e^{13\hat{\beta}} + e^{8\hat{\beta}} + e^{9\hat{\beta}}} \\
 &= 0.2743
 \end{aligned}$$

$$\begin{aligned}
 \hat{H}(t = t_7; x_0 = 11, \hat{\beta} = -0.4363) &= \frac{e^{11\hat{\beta}}}{e^{8\hat{\beta}} + e^{9\hat{\beta}} + e^{7\hat{\beta}} + e^{10\hat{\beta}} + e^{11\hat{\beta}} + e^{13\hat{\beta}} + e^{8\hat{\beta}} + e^{9\hat{\beta}}} \\
 &+ \frac{e^{11\hat{\beta}}}{e^{9\hat{\beta}} + e^{7\hat{\beta}} + e^{10\hat{\beta}} + e^{11\hat{\beta}} + e^{13\hat{\beta}} + e^{8\hat{\beta}} + e^{9\hat{\beta}}} \\
 &+ \frac{e^{11\hat{\beta}}}{0.8e^{9\hat{\beta}} + e^{11\hat{\beta}} + e^{13\hat{\beta}} + e^{8\hat{\beta}} + e^{9\hat{\beta}}} \\
 &= 0.0479
 \end{aligned}$$

$$\begin{aligned}\hat{H}(t=t_{10}; x_0=9, \hat{\beta}=-0.4363) &= \frac{e^{9\hat{\beta}}}{e^{8\hat{\beta}} + e^{9\hat{\beta}} + e^{7\hat{\beta}} + e^{10\hat{\beta}} + e^{11\hat{\beta}} + e^{13\hat{\beta}} + e^{8\hat{\beta}} + e^{9\hat{\beta}}} \\ &+ \frac{e^{9\hat{\beta}}}{e^{9\hat{\beta}} + e^{7\hat{\beta}} + e^{10\hat{\beta}} + e^{11\hat{\beta}} + e^{13\hat{\beta}} + e^{8\hat{\beta}} + e^{9\hat{\beta}}} \\ &+ \frac{e^{9\hat{\beta}}}{0.8e^{9\hat{\beta}} + e^{11\hat{\beta}} + e^{13\hat{\beta}} + e^{8\hat{\beta}} + e^{9\hat{\beta}}} \\ &+ \frac{e^{9\hat{\beta}}}{0.4e^{9\hat{\beta}} + 0.5e^{13\hat{\beta}} + e^{9\hat{\beta}}} \\ &= 0.6724\end{aligned}$$

CIF 的估计值为

$$\begin{aligned}\hat{F}(t=t_3; x_0=8, \hat{\beta}=-0.4363) &= 1 - e^{-\hat{H}(t=t_3; x_0=8, \hat{\beta}=-0.4363)} = 0.1625 \\ \hat{F}(t=t_5; x_0=7, \hat{\beta}=-0.4363) &= 1 - e^{-\hat{H}(t=t_5; x_0=7, \hat{\beta}=-0.4363)} = 0.2399 \\ \hat{F}(t=t_7; x_0=11, \hat{\beta}=-0.4363) &= 1 - e^{-\hat{H}(t=t_7; x_0=11, \hat{\beta}=-0.4363)} = 0.0468 \\ \hat{F}(t=t_{10}; x_0=9, \hat{\beta}=-0.4363) &= 1 - e^{-\hat{H}(t=t_{10}; x_0=9, \hat{\beta}=-0.4363)} = 0.4895\end{aligned}$$

4. 实例分析

数据来自 R 中的数据集(survival 包中的 mgus2), 共计 1371 例单克隆丙种球蛋白病患者, 其中结局事件定义为“发生浆细胞恶性肿瘤”有 114 例, 而结局事件是在“发生浆细胞恶性肿瘤”之前因其他原因死亡患者有 855 例, 这些发生“其他死亡”的患者因无法观察到“发生浆细胞恶性肿瘤”的终点, 就被称为与“发生浆细胞恶性肿瘤”存在竞争风险的事件, 其他 402 例患者是在观测期间上述结局事件没有发生。研究指标变量定义与赋值见表 3。

Table 3. Definitions and assignments of variables

表 3. 各变量的定义与赋值

变量	赋值
事件类型	
类型 I: 发生浆细胞恶性肿瘤	$Z = 1$
类型 II: 发生浆细胞恶性肿瘤之前因为其他原因死亡	$Z = 2$
类型 III: 删失, 上述情况都不发生	$Z = 0$
血红蛋白水平	x_1
年龄	x_2
生存时间(单位: 月)	t

协变量 x 有两个, 分别为血红蛋白水平和年龄, 记为 x_1, x_2 , 创立协变量矩阵。生存时间 t 对类型 I 事件表示直至发展为浆细胞恶性肿瘤或最后一次访视时间, 对类型 II 事件表示直到死亡或最后一次接触的时间, 对类型 III 事件表示最后一次接触时间。

数据集 mgus2 中的各个变量的统计描述见表 4。

Table 4. Statistical description of variables
表 4. 变量的统计性描述

变量	最小值	中位数	最大值	均值	标准差
Z	0.00	2.00	2.00	1.33	0.90
x_1	5.70	13.50	18.90	13.30	2.02
x_2	24.00	72.00	96.00	70.48	12.17
t	1.00	81.00	424.00	93.34	71.90

利用 Fine-Gray 模型分析建模, 这里的协变量为 x_1, x_2 , 自变量为生存时间 t , 响应变量为回归系数 $\hat{\beta}$ 。

如果用 Cox 模型进行回归分析, 由于不考虑竞争风险, 因此在考虑类型 I 事件时, 将类型 I 事件为感兴趣事件, 类型 II 和类型 III 事件改为删失事件, 在考虑类型 II 事件时, 则将类型 II 事件作为感兴趣事件, 类型 I 和类型 III 事件改为删失事件。

利用 Fine-Gray 模型和 Cox 模型建模所得结果见表 5。

Table 5. Comparison between Fine-Gray model and Cox model
表 5. Fine-Gray 模型与 Cox 模型对比

事件分类	协变量	Fine-Gray 模型			Cox 模型		
		$\hat{\beta}$	s.e. ($\hat{\beta}$)	P	$\hat{\beta}$	s.e. ($\hat{\beta}$)	P
Z = 1	x_1	-0.0243	0.0435	0.5800	-0.1192	0.0502	0.0176
	x_2	-0.0173	0.0058	0.0029	0.0097	0.0082	0.2401
Z = 2	x_1	-0.0916	0.0218	0.0000	-0.1153	0.0180	0.0000
	x_2	0.0510	0.0038	0.0000	0.0569	0.0036	0.0000

根据表 5, 可得出以下结果:

在对类型 I 事件建模中, Fine-Gray 模型与 Cox 模型估计结果差异很大。Fine-Gray 模型协变量 x_1 的回归系数不显著, x_2 的回归系数显著。Cox 模型估计结果则与之相反。而在对于类型 II 事件建模中, Fine-Gray 模型与 Cox 模型估计结果非常接近。协变量 x_1 和 x_2 对应的回归系数都非常显著。说明利用 Fine-Gray 模型与 Cox 模型的建模区别主要体现在对类型 I 事件的估计结果上。主要由于 Cox 模型只针对样本中发生类型 I 事件 114 例的个体建模, 而 Fine-Gray 模型在考虑了发生类型 I 事件的 114 例个体基础上, 还考虑发生类型 II (竞争风险) 事件的 855 例个体, 在建模过程中, 通过表 2 中的权重计算方法, 算出竞争风险事件个体的权重, 并且随着时间推移, 竞争风险事件个体在风险集中的权重逐渐降低。

为了更直观的说明 Fine-Gray 模型与 Cox 模型的区别, 分别绘制类型 I 事件的累计发生率的曲线图。以 mgus2 数据集协变量的中位数 $x_1 = 13.5, x_2 = 72$, 编号为 135 的患者为例, 分别用 Fine-Gray 模型和 Cox 模型对发生浆细胞恶性肿瘤这一感兴趣事件的累计发生率进行预测, 结果见图 1。

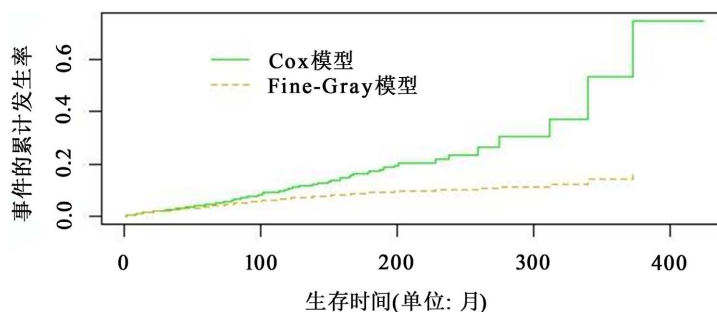


Figure 1. Predictive comparison between Fine-Gray model and Cox model

图 1. Cox 模型与 Fine-Gray 模型对比预测图

5. 结语

由第 3 部分实证对比分析可以看出, Fine-Gray 模型适用于具有多个终点的生存数据, 在存在竞争风险的情况下, 得到的结果更符合实际, 具有更好的拟合性。如临床上常见术后死亡患者无法获取关心终点, 故术后死亡与关心终点存在竞争风险。相比之下, 使用传统 Cox 模型会忽略竞争风险, 可能高估所研究事件的累计发生率。

基金项目

本研究由 2022 年度辽宁省研究生教育教学改革研究项目(2022-180-39510165)资助。

参考文献

- [1] Kaplan, E.L. and Meier, P. (1958) Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, **53**, 457-481. <https://doi.org/10.1080/01621459.1958.10501452>
- [2] Cox, D.R. (1972) Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, **34**, 187-202. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>
- [3] Gelman, R., Gelber, R., Henderson, I.C., et al. (1990) Improved Methodology for Analyzing Local and Distant Recurrence. *Journal of Clinical Oncology*, **8**, 548-555. <https://doi.org/10.1200/JCO.1990.8.3.548>
- [4] Gooley, T.A., Leisenring, W., Crowley, J., et al. (1999) Estimation of Failure Probabilities in the Presence of Competing Risks: New Representations of Old Estimators. *Statistics in Medicine*, **18**, 695-706. [https://doi.org/10.1002/\(SICI\)1097-0258\(19990330\)18:6<695::AID-SIM60>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-0258(19990330)18:6<695::AID-SIM60>3.0.CO;2-O)
- [5] Kalbfleisch, J.D. and Prentice, R.L. (2011) *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, Hoboken.
- [6] Fine, J.P. and Gray, R.J. (1999) A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*, **94**, 496-509. <https://doi.org/10.1080/01621459.1999.10474144>
- [7] 朱旭, 陈书, 魏歆然, 等. 基于 Fine-Gray 竞争风险模型的青年胃癌患者预后影响因素研究[J]. 第二军医大学学报, 2021, 42(10): 1140-1147. <https://doi.org/10.16781/j.0258-879x.2021.10.1140>
- [8] 钱迪, 涂博祥, 陈泉, 等. 基于 Fine-Gray 竞争风险模型的小肝癌患者预后模型构建与验证[J]. 中华肿瘤防治杂志, 2023, 30(23): 1435-1443. <https://doi.org/10.16073/j.cnki.cjcp.2023.23.07>