

# 用生存分析方法研究了美沙酮对戒毒所需时间的影响

蒋冰

学号：201322130002

## Abstract

本文主要根据生存分析中的方法，分析了 238 例美沙酮治疗海洛因成瘾者的戒毒数据，研究了美沙酮的剂量、治疗方法、是否有入狱激励对戒毒所需时间的影响规律。本文通过乘积极限法（PM 法）建立海洛因成瘾者的戒毒时间的生存率模型，然后用 Cox 比较危险率模型（PH 模型）分析了治疗方式、入狱记录、美沙酮剂量对戒毒所需时间的影响，并通过 clinic 前的系数发现，治疗方式 2 比治疗方式 1 要好。然后文章利用 ph 假定确定了两种治疗方式的区分是否会随时间改变。因此可以用对数 - 检验对两种不同的治疗方式进行了比较。通过对数 - 秩检验计算得到两种治疗方式对戒毒所需时间有着显著差异。综合可以得到，治疗方法 2 显著优于治疗方法 1。此外，本文还探讨了治疗方法和入狱记录、美沙酮剂量是否有协同作用。通过两个 COX PH 回归模型，做方差分析后，可以认为治疗方法和入狱记录、美沙酮剂量之间没有协同作用。

## 1 模型和方法

在世界上的众多国家中，禁止毒品的销售和使用一直是政府工作的重担之一。而毒品犹如“白色幽灵”弥散在全球各个角落，势头愈演愈烈。海洛因 (Heroin) 便是其中一种上瘾的药物，是最被滥用的药物之一，一般由吗啡处理而成，通常显示为白色或棕色粉末。全球鸦片产量自 1996 年以来每年均在 4000 吨以上，非法使用毒品的人数有 4.47 亿，占总人口的近 8%。禁毒也已成为各国政府无法回避的任务和人类面临的共同使命。因此，探索各种有效地戒毒方法也成为专家学者们研究的重点。例如，美沙酮 (methadone)，一种合成阿片类药物，能阻止海洛因和消除戒断症状的影响，已被证明成功的为海洛因成瘾的治疗。本文尝试借助生存分析模型，来分析美沙酮对治疗青少年海洛因成瘾的效果。

生存分析 (survival analysis) 也称之为风险模型 (hazard model) 或持续模型 (duration model)，是一种根据实验或调查数据，对生物、人以及具有类似于生存规律的其它事物的生存时间进行分析和推断的统计方法。目前生存分析已广泛应用于生物统计、医学、金融和工业工程等多个领域。

广义的生存时间指生物体存活的时间，或所关心的某种现象的持续时间。道路拥堵持续时间是指从交通拥堵现象产生开始，一直到拥堵结束为止的持续时间，属于广义生存时间范畴，可运用生存分析方法来进行研究。令  $T$  代表

生存的时间（寿命），它是一个非负的随机变量，令  $f(t)$  表示  $T$  的概率密度函数， $P$  代表概率， $T$  的分布函数为  $F(t) = P(T \leq t) = \int_0^t f(x)dx$ 。

生存函数  $S(t)$ ，也叫生存率，表示生存时间大于  $t$  的概率，它的表达式为  $S(t) = P(T > t) = 1 - F(t) = \int_t^\infty f(x)dx$ 。

生存分析中， $T$  可以用危险率函数  $h(t)$  来描述。危险率函数也叫条件生存率，它指病人从患病起持续了  $t$  时间后没有结束，但在接下来的一段很小的时间  $\Delta t$  内结束的概率，可表示为  $h(t) = \lim_{\Delta t \rightarrow \infty} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} =$

$\frac{f(t)}{S(t)} = \frac{-d \ln S(t)}{dt}$ 。它表示死亡速率的大小。如以  $t$  为横坐标， $f(t)$  为纵坐标作出的曲线称为密度曲线，由曲线上可看出不同时间的死亡速率及死亡高峰时间。纵坐标越大，其死亡速率越高，如曲线呈现单调下降，则死亡速率越来越小，如呈现峰值，则为死亡高峰。

## 2 数据和样本提取

戒毒数据来自于澳大利亚学者 Caplehorn et al. 在 2007 年进行的研究。他收集了 238 例子海洛因成瘾者的戒毒数据。为评价美沙酮的治疗效果，这些成瘾者分别接受了 2 种不同的治疗方式，即为 Clinc1 和 Clinc2。每名成瘾者的生存时间定义为从其接受治疗到他成功戒掉毒瘾（类似于一般生存分析中的“死亡”，为区别，记作“戒瘾”）或是由于死亡，其他疾病，可能的药物副作用，或个人原因不愿意继续下去以天数记的时间（称为失访）。我们把这些在跟踪调研内未到达调查终点的成瘾者的数据称为失访或截尾观察（censored observation）。

数据内容见附表，其中变量说明如下：

ID：成瘾者的调研号；

SURVT：成瘾者的生存时间，即其从其接受治疗到他成功戒掉毒瘾或是由于死亡，其他疾病，可能的药物副作用，或个人原因不愿意继续下去的时间（以天数记）；

STATUS：成瘾者的生存时间的指标，，取值 1 表示成瘾者成功戒掉毒瘾，取值 0 表示成瘾者由于死亡，其他疾病，可能的药物副作用，或个人原因不愿意继续下去；

CLINIC：成瘾者接受的治疗方式的指标，取值 1 表示成瘾者接受的是治疗方式 1，取值 2 表示成瘾者接受的是治疗方式 2；

PRISON：成瘾者是否有过入狱记录的指标，取值 1 表示成瘾者有过入狱记录，取值 0 表示成瘾者从未入狱；

DOSE：成瘾者平均每天接受美沙酮治疗的剂量 (mg/day)。

## 3 实证分析

### 3.1 乘积极限法 (Product-Limit Method)

乘积极限法又称为积限法或 PL 法，它是由统计学家 Kaplan 和 Meier 于 1958 年首先提出的，因此又称为 Kaplan-Meier 法，它是一种利用条件概率及概率的乘法原理计算生存率及其标准误的方法。

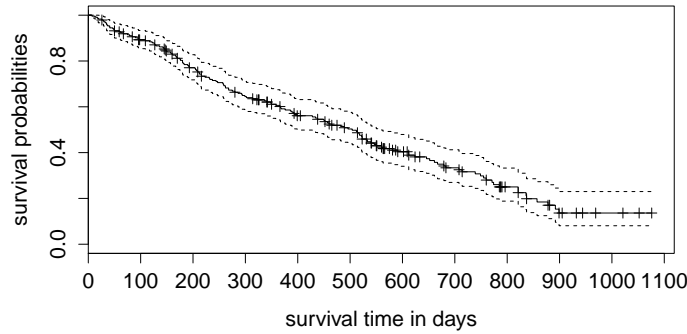
设  $S(t)$  表示  $t$  年的生存率,  $S(t_i|t_{i-1})$  表示活过  $t_{i-1}$  年后再活过  $t_i$  年的条件概率。根据条件概率的性质有  $S(t_2) = S(t_1)S(t_2|t_1)$ 。

设  $S_{i-1}$  是个体存活到  $t_{i-1}$  时没有失访的人数, 设在  $t_i$  时仍存活  $S_i$  人, 失败  $d_i$  人, 失访  $l_i$  人, 于是有  $S_{i-1} = S_i + d_i + l_i$ , 因此可以估计  $t_{i-1}$  时的存活着到时刻  $t_i$  时的生存概率为  $1 - \frac{d_i}{S_{i-1}} = 1 - \frac{d_i}{S_i + d_i + l_i}$ ,  $t_i$  时的失访者在  $t > t_i$  是不会对生存函数构成影响。所以在  $t_i$  上的生存概率  $S(t_i)$  的 Kaplan-Meier 估计量公式为:  $\hat{S}(t_i) = \left(1 - \frac{d_1}{S_0}\right) \left(1 - \frac{d_2}{S_1}\right) \cdots \left(1 - \frac{d_i}{S_{i-1}}\right)$ ,  $i = 1, 2, \dots, k$

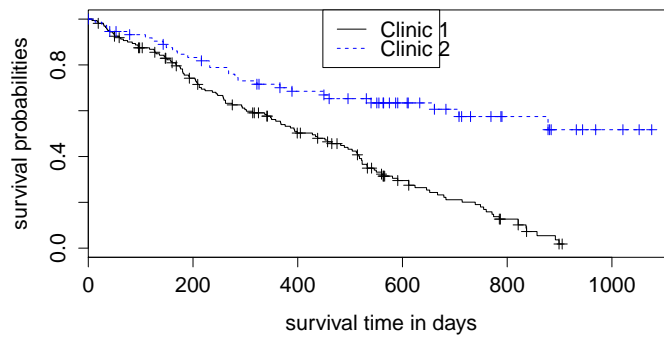
在引入  $Var\{\ln[\hat{S}(t)]\} = \sum_{j=1}^i \frac{d_j}{S_{j-1}(S_{j-1} - d_j)}$ , 于是可以解得在  $t_i$  上的

生存概率  $S(t_i)$  的双侧  $100\% \times (1 - \alpha)$  CI 即为  $(e^{c_L}, e^{c_R})$ , 则  $c_L = \ln[\hat{S}(t_i) - z_{1-\frac{\alpha}{2}} se\{\ln[\hat{S}(t_i)]\}]$ ,  $c_R = \ln[\hat{S}(t_i) + z_{1-\frac{\alpha}{2}} se\{\ln[\hat{S}(t_i)]\}]$ 。通过 KM 法计算可以得到各时间段的 KM 估计。

**KM survive curve**



把两种不同治疗方法的生存率分开计算, 可得下图。



### 3.2 比较危险率模型 (Proportional-Hazards Model)

对数秩检验是一个很有效的分析数据的方法，它把事件的发生跟时间联系在一起，这比简单地把时间部分时间只考虑发生与不发生要好。在这个检验中，对个体做追踪调查时，可允许数据中存在部分失访数据。

借助 logit 模型的构造思想，cox 比例风险模型为：

$$\ln \frac{h(t)}{h_0(t)} = (\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)$$

它另一个常见的形式为： $h(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)$

其中  $x_1, x_2, \cdots, x_k$  是一组独立变量，而  $h_0(t)$  是在基准状态下在  $t$  时刻上的基准危险率，它代表所有独立变量全取值 0 是的危险率吗，即不存在因素  $x_1, x_2, \cdots, x_k$  的影响的  $t$  时刻死亡率。

由于  $h_0(t)$  是未知的，所以又称为半参数模型。该模型通过建立偏似然函数，使对数偏似然函数最大并用最大似然法的 Newton-Raphson 迭代得到参数的估计值。该模型的系数意义为： $\beta_j$  表示存在因素  $x_j$  下，死亡的风险率是不存在因素  $x_j$  的  $e^{\beta_j}$  倍。 $\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$  为风险指数 (hazard index, 简称 HI)，又称预后指数 (prognostic index) 或预后得分 (prognostic score)。HI=0 代表风险处于平均水平；HI<0 代表风险处于低水平；HI>0 代表风险处于高水平。做出零假设为  $H_{j0}: \beta_j = 0$ ，备择假设为  $H_{j1}: \beta_j \neq 0$ ，计算检验统计量  $z = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$ ，当显著性水平为  $\alpha$  时，当  $z < z_{\frac{\alpha}{2}}$  或  $z > z_{1-\frac{\alpha}{2}}$  时，则拒绝  $H_{j0}$ ；若  $z_{\frac{\alpha}{2}} \leq z \leq z_{1-\frac{\alpha}{2}}$ ，则接受  $H_{j0}$ 。

经过计算的各项系数及其标准差如下表所示：

	coef	exp(coef)	se(coef)	z	Pr(> z )
prison	0.326555	1.386184	0.167225	1.953	0.0508 .
dose	-0.035369	0.965249	0.006379	-5.545	2.94E-08 ***
clinic	-1.009896	0.364257	0.214889	-4.7	2.61E-06 ***

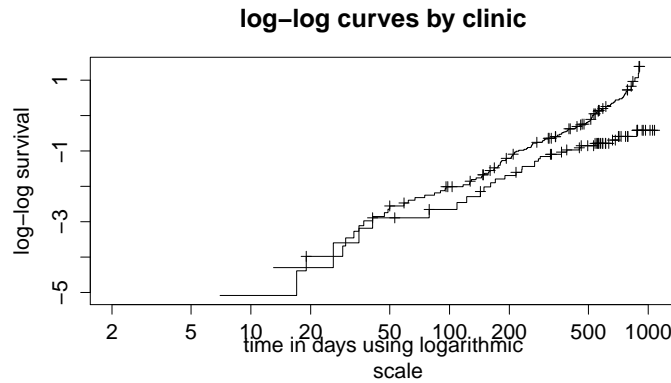
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

得到的 cox 风险比例模型为  $h(t) = h_0(t) \exp(0.326555prison - 0.035369dose - 1.009896clinic)$  可以看出，在  $\alpha=0.1$  时，所有系数均显著，且 prison 前面的系数前面符号为正，说明有入狱记录的成瘾者的风险更大，而 dose、clinic 前面的系数前面符号为负，分别说明美沙酮的平均每天剂量越大，成瘾者的戒瘾效率更好和治疗方式 2 的戒瘾效果要好于治疗方式 1。

### 3.3 PH 假定

PH 假定即比例风险假定 (proportional hazards assumption)，即一个比例风险模型具有不同个体有成比例的性质，即对于两个协变量  $X_1$  和  $X_2$ ， $h(t|X_1)/h(t|X_2)$  不随时间  $t$  改变，这是由比例风险模型的形式决定的。

同时，它意味着二次对数生存曲线 (log-log survival curves) 应该平行。即  $h(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)$  两边对数值加上负号，再取对数  $\ln(-\ln(h(t))) = \ln(-\ln(h_0(t))) - (\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k) = \ln(-\ln(h_0(t))) + bX$  这是二次对数生存率，记为 LML，可以发现，不同的 LML 相差只是个与  $t$  无关的常数： $LML_1 - LML_2 = b(X_1 - X_2)$ 。



从上图可以看出，两条曲线并没有平行，也就是不符合 PH 假设，即两种治疗方式之间  $h(t|X1)/h(t|X2)$  随时间  $t$  改变。

### 3.4 对数 -秩检验

我们虽然可以从 KM 生存曲线比较指定时间点上的两者疗法的生存率的大小。但要比较两条生存曲线，则需要对两曲线减的差异做比较。一般地，在  $t_i$  时危险率定义为  $\hat{h}(t_i) = \frac{d_i}{S_{i-1}}$ 。为区分两种治疗方式的危险率，定义  $h_1(t)$  = 治疗方式 1 中成瘾者在时间  $t$  的危险率， $h_2(t)$  = 治疗方式 2 中成瘾者在时间  $t$  的危险率。

因为我们假设危险率是符合 logistic 模型，因此可以假定这两个危险率之比是常数  $\exp(\beta)$ ，即  $\frac{h_1(t)}{h_2(t)} = \exp(\beta)$ 。要检验两种治疗方式是否有显著差异，做出生假设  $H_0: \beta = 0$ ，备择假设为  $H_1: \beta \neq 0$ 。若  $\beta = 0$ ，则表示两条生存曲线基本相同；若  $\beta < 0$ ，说明治疗方式 1 比治疗方式 2 的治疗效果更好；若  $\beta > 0$ ，说明治疗方式 2 比治疗方式 1 的治疗效果更好。

比较两种治疗方法的危险率的对数 -秩检验，设调研时间长度为  $T$ ，危险率随对眼时间而改变。把  $T$  划分成  $k$  个较小的时间区间，在每个小区间上，危险率被认为是齐性的。计算每个时间区间上的  $2 \times 2$  列联表，如下图所示

治疗方法	戒瘾	未戒瘾（失败或失访）	总数
治疗方法 1	$a_i$	$b_i$	$n_{i1}$
治疗方法 2	$c_i$	$d_i$	$n_{i2}$
总数	$a_i + c_i$	$b_i + d_i$	$n_i$

其中，各变量的定义如下：

- $a_i$  = 接受治疗方法 1 的成瘾者中，在第  $i$  个时间区间内成功戒瘾的个数；
- $b_i$  = 接受治疗方法 1 的成瘾者中，在第  $i$  个时间区间内未成功戒瘾的个数；
- $c_i$  = 接受治疗方法 2 的成瘾者中，在第  $i$  个时间区间内成功戒瘾的个数；
- $d_i$  = 接受治疗方法 2 的成瘾者中，在第  $i$  个时间区间内未成功戒瘾的个数；
- $n_{i1}$  = 在第  $i$  个时间区间开始时，接受治疗方法 1 的成瘾者的个数；
- $n_{i2}$  = 在第  $i$  个时间区间开始时，接受治疗方法 2 的成瘾者的个数；
- $n_i$  = 在第  $i$  个时间区间开始时，还在继续接受治疗的成瘾者的个数；

根据上面的列联表做 Mantel-Haenszel 检验, 即计算检验统计量  
 $x_{LR}^2 = \frac{(|O-E|-0.5)^2}{Var_{LR}}$  其中式中的各变量的计算公式如下:

$$O = \sum_{i=1}^k a_i$$

$$E = \sum_{i=1}^k E_i = \sum_{i=1}^k \frac{(a_i + b_i)(a_i + c_i)}{n_i}$$

$$Var_{LR} = \sum_{i=1}^k V_i = \sum_{i=1}^k \frac{(a_i + b_i)(c_i + d_i)(a_i + c_i)(b_i + d_i)}{n_i^2(n_i - 1)}$$

在  $H_0$  成立的条件下,  $X_{LR}^2$  服从自由度为 1 的卡方分布。

对于双侧显著性水平为  $\alpha$  的检验, 若  $X_{LR}^2 > \chi_{1-\alpha}^2(1)$ , 则拒绝  $H_0$ ; 若  $X_{LR}^2 \leq \chi_{1-\alpha}^2(1)$ , 则接受  $H_0$ 。一般要求  $Var_{LR} \geq 5$  时, 此检验才可以适用。通过计算可得,  $Var_{LR}=34.65794>5$ , 因此此检验适用。 $X_{LR}^2=27.9$ ,  $p$  值 =  $1.28e-07 < 0.05$ , 即说明两种治疗方式对戒毒的效果有着显著差异。

### 3.5 治疗方法和入狱记录、美沙酮剂量的协同作业分析

为检验治疗方法和入狱记录、美沙酮剂量是否有协同作用。做两个 COX PH 回归模型: 没有考虑协同作用的  $h_1(t) = h_0(t)\exp(\beta_1prison + \beta_2dose + \beta_3clinic)$  和考虑治疗方法和入狱记录、美沙酮剂量的协同作用的  $h_2(t) = h_0(t)\exp(\beta_1prison + \beta_2dose + \beta_3clinic + \beta_4clinic \times prison + \beta_5clinic \times dose)$ 。

回归结果为  $h_1(t) = h_0(t)\exp(0.326555prison - 0.035369dose - 1.009896clinic)$

$h_2(t) = h_0(t)\exp(1.19238prison + 0.9810dose + 1.1967clinic + 0.4779clinic \times prison + 0.9861clinic \times dose)$

做两者的方差分析,  $h_1(t)$  的极大似然值为 -673.26,  $h_2(t)$  的极大似然值为 -671.45, 卡方值为 3.6182,  $p$  值为 0.1638  $> 0.05$ , 即认为治疗方法和入狱记录、美沙酮剂量之间没有协同作用。两个方程之间差异不大, 就需要比较两个方程的 AIC 值,  $h_1(t)$  的 AIC=1351.86, 小于  $h_2(t)$  的 AIC1352.9, 所以认为方程 1 表示更好。

## 4 结论

本文主要根据生存分析中的方法, 分析了 238 例美沙酮治疗海洛因成瘾者的戒毒数据, 研究了不同的治疗方法对吸毒者戒毒的效果的影响。本文通过乘积极限法 (PM 法) 建立海洛因成瘾者的戒瘾时间的生存率模型, 然后用 Cox 比较危险率模型 (PH 模型) 分析了治疗方式、入狱记录、美沙酮剂量对戒毒所需时间的影响。模型结果为  $h(t) = h_0(t)\exp(0.326555prison - 0.035369dose - 1.009896clinic)$ 。通过 clinic 前的系数发现, 治疗方式 2 比治疗方式 1 要好。因此根据 ph 假定对两种治疗方式的区分是否随时间改变。实验结果表明两者两种治疗方式的区分随时间改变, 因此可以用对数-秩检验对两种不同的治疗方式进行了比较。通过对数-秩检验计算得到  $X_{LR}^2=27.9$ ,  $p$  值 =  $1.28e-07 < 0.05$ , 即说明两种治疗方式对戒毒对戒毒所需时间有着显著差异。由结合 clinic 的系数符号为正, 知道治疗方法 2 好于治疗方法 1。综合可以得到, 治疗方法 2 显著优于治疗方法 1。

此外, 本文还探讨了治疗方法和入狱记录、美沙酮剂量是否有协同作用。通过两个 COX PH 回归模型, 做方差分析知, 两个模型卡方值为 3.6182,  $p$  值为  $0.1638 > 0.05$ , 即认为治疗方法和入狱记录、美沙酮剂量之间没有协同作用。

本文只是用生存分析方法研究了美沙酮的剂量、治疗方法、是否有入狱激励对戒毒所需时间的影响规律, 如果能取到更多的数据, 还可以进一步深入分析影响戒毒所需时间的影响因素等问题。由此可以看出, 生存分析在医药分析, 疾病预防治疗控制等方面有着重要用途。