

学校编码: 10384

分类号_____密级_____

学号: 15420071151202

UDC _____

廈門大學

硕士学位论文

竞争风险模型在女性乳腺癌预后
预测研究中的应用

The Application of Competing Risk Model on the Prediction
Study of Prognosis for Female Breast Cancer

胡海兰

指导教师姓名: 方亚教授

专业名称: 统计学

论文提交日期: 2010年 月

论文答辩日期: 2010年 月

学位授予日期: 2010年 月

答辩委员会主席: _____

评阅人: _____

201 年 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为(国家自然科学基金项目——乳腺癌预后的竞争风险性多终点生存概率预测模型研究)课题(组)的研究成果,获得(方亚)课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

- () 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。
- () 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

目的：通过研究女性乳腺癌患者的术后生存结果，分析评价临床预后因素对女性乳腺癌的影响，探索竞争风险模型在乳腺癌患者预后预测中的应用，并与传统生存分析结果进行比较。

方法：通过调查的方法，获取厦门市及武汉市 4 家医院 1998 年至 2008 年入院接受手术治疗的乳腺癌患者的临床资料。生存分析中的单变量分析以 K-M 方法进行，用 Log-rank 法检验生存差异，采用 Cox 比例风险模型进行多因素分析。在考虑竞争风险的情况下，采用非参数方法计算累积风险率。采用原因别风险模型进行多因素分析，并评价复发的影响因素。采用累积风险模型分析各个独立的预后因素分别对复发和转移的影响，并计算累积风险率。全部分析采用 SPSS16.0, EXCEL2007, R2.10.1 等统计软件实现。

结果：共调查收集 246 例乳腺癌患者资料，其 3 年、5 年的累积生存率分别为 84.1%、75.4%。单因素分析表明：肿瘤大小、手术方式、治疗方式、个人乳腺癌疾病史是影响乳腺癌死亡风险率的重要因素。Cox 比例风险模型多因素分析表明：肿瘤大小和手术方式是乳腺癌死亡风险率的独立预后因素。肿瘤大小“ $>5\text{cm}$ ”的患者的死亡危险率是大小为“ $\leq 2\text{cm}$ ”的患者的 5.461 倍。乳腺癌根治术的死亡危险率为乳腺癌改良根治术的 3.004 倍。在竞争风险下，非参估计法的计算结果为：复发的 3 年风险率为 20.73%，5 年风险率为 28.2%；转移的 3 年风险率为 17.2%，5 年风险率为 24.95%，K-M 方法的风险率高于竞争风险情况下的风险率。采用原因别风险模型对复发进行多因素分析，得到肿瘤大小和个人乳腺癌疾病史是复发的独立预后因素。最后运用累积风险回归模型分别对淋巴结转移、肿瘤大小、个人乳腺癌疾病史进行了回归，与原因别模型的结果大体一致。

结论：能有效提高乳腺癌治疗效果的决定性因素除手术方式、治疗方式、遗传以及女性生理因素、心理因素外，定期检查、早发现、早诊断、早治疗更为重要。生存分析的 K-M 法相较于竞争风险模型会高估风险率；累积风险模型的结果更直观、更容易解释；竞争风险模型在乳腺癌患者的预后预测中更合理、更客观。

关键词：乳腺癌，生存分析，竞争风险模型，原因别风险模型，累积风险模型

Abstract

Objectives: The aim of this study was to evaluate the survival results after operation and explore the prognostic factors which have an impact on the patients with breast cancer. The application of competing risk model on the prediction of prognosis study for female breast cancer was discussed and compared with the traditional survival analysis method.

Methods: Data of this study was collected from four hospitals in Xiamen and Wuhan, which was the clinical data of patients with breast cancer from 1998 to 2008. K-M survival curves, log-rank test and Cox proportional hazard model were used to evaluate the survival analysis and access the prognostic factors. In the presence of competing risks, non-parametric method, cause-specific hazard model, cumulative incidence functions were used to calculate the cumulative risk of relapse and metastasis rate and access the prognostic factors. SPSS16.0, EXCEL2007, R2.10.1 were used in the study.

Results: 246 cases of breast cancer were collected. 3-year and 5-year cumulative survival rates were 84.1% and 75.4%, respectively. Tumor size, type of surgery, type of treatment, personal history of breast diseases were the significant factors which impact the mortality rates of breast cancer, which was analyzed by K-M method in single factor analysis. Cox proportional hazard model confirmed that tumor size and type of surgery were the independent prognostic factors in multiple factor analysis. The patients whose tumor size is more than 5 cm had 5.461 times higher risk of mortality than patients with equal and less than 2 cm. The risk of mortality according to treatment was 3.004 for the modified radical mastectomy, in comparison with the radical mastectomy. In the presence of competing risks, the results from the Non-parameter estimation method showed: 3-year, 5-year risk rates of relapse and metastasis were 20.73%, 28.2%, 17.2%, 24.95%, respectively. The result from K-M method was higher than Non-parameter method in presence of competing risks. Cause-specific hazard

model confirmed that tumor size and personal history of breast diseases were the independent prognostic factors for relapse. Finally, lymph nodes, tumor sizes, personal history of breast cancer were respectively analyzed by cumulative incidence regression model, which is almost consistent with the results from cause-specific hazard model. Furthermore, the cumulative risk rates between different groups could be calculated by cumulative incidence model at any time points.

Conclusion: In addition to surgical methods, treatments, genetic factor and female physiological, psychological factors, which were the key factors to impact the survival quality, early detection, early diagnosis and early treatment were more important. K-M method compared to competing risks model will overestimate the risk rates. The results from cumulative risk models could be more intuitive and more easily explained, in comparison with cause-specific hazard model. The results showed that competing risk model was more rational than the traditional survival analysis method.

Key Words: Breast cancer, Survival analysis, Competing risks model, Cause-specific hazard model, Cumulative incidence model

目 录

第一章 绪论	1
1.1 研究背景.....	1
1.2 研究意义.....	2
1.3 研究目标.....	3
1.4 研究框架.....	4
第二章 文献综述	5
第三章 理论的基本介绍	10
3.1 生存分析的基本概念.....	10
3.2 生存分析的研究发展.....	13
3.3 生存分析常用的方法.....	14
3.4 竞争风险的基本概念.....	18
3.5 竞争风险的基本方法.....	19
第四章 实证分析	25
4.1 数据来源和调查方法.....	25
4.2 Kaplan-Meier 方法下的总体生存率.....	26
4.3 Kaplan-Meier 方法下的单因素分析.....	29
4.4 Cox 比例风险模型的多因素分析.....	43
4.5 竞争风险下的非参数估计方法.....	45
4.6 原因别风险模型.....	49
4.7 累积风险模型.....	52
第五章 传统生存分析与竞争风险模型结果的比较分析	57
第六章 讨论和建议	60
6.1 讨论.....	60
6.2 建议.....	63

参考文献..... 65

致 谢..... 69

厦门大学博硕士学位论文摘要库

Table of Contents

Chapter 1 Introduction.....	1
1.1 Research background	1
1.2 The significance of research	2
1.3 The goal of research	3
1.4 The framework of research	4
Chapter 2 Literature Review	5
Chapter 3 Introduction of the theory	10
3.1 The concept of survival analysis	10
3.2 The development of survival analysis	13
3.3 The methods of survival analysis.....	14
3.4 The concept of competing risk	18
3.5 The basic methods of competing risk.....	19
Chapter 4 Empirical analysis.....	25
4.1 Data sources and methods of the investigation	25
4.2 The general survival rate by Kaplan-Meier method	26
4.3 Single factor analysis by Kaplan-Meier method	29
4.4 Multiple factor analysis by Cox proportional hazard model.....	43
4.5 Non-parameter estimation method in presence of competing risk.....	45
4.6 Cause-specific hazard model.....	49
4.7 Cumulative incidence model	52
Chapter 5 The comparison of survival analysis and competingrisk	57
Chapter 6 Discussion and recommendation.....	60
6.1 Discussion	60
6.2 Recommendation	63

References..... 65

Acknowledgements 69

厦门大学博硕士学位论文摘要库

第一章 绪论

1.1 研究背景

乳腺癌是女性最常见的恶性肿瘤之一，是一种严重影响妇女身心健康甚至危及生命的最常见的恶性肿瘤之一。据资料统计，发病率占全身各种恶性肿瘤的7-10%。近年来，女性乳腺癌的发病率呈逐年上升趋势，在无症状女性疾病人群中，各种乳腺疾病患者竟高达52.4%，占女性慢性常见病发病首位。研究乳腺癌患者的预后及其影响因素显得尤为重要^[19]。

生存分析是在医学研究中一种比较常见的统计分析方法，如用于研究手术方式对乳腺癌症患者生存的影响，不仅关心死亡这个终点事件是否发生，而且还关心癌症患者死亡发生的时间。生存数据可以表示从观察起点到发生终点事件的时间跨度。它包括进入时间，事件时间及终点事件三个基本要素。例如，在研究乳腺癌手术后的发展变化中，生存数据就可以表示手术结束直到发生终点事件（例如死亡）的时间跨度。

生存数据分成完全数据和删失数据。完全数据是指在规定的观察期内观察到了从观察起点到终点事件的时间跨度。删失数据是指在规定的观察期内由于某种原因未能观察到生存数据的其中一个或多个要素。删失数据可以分成右删失，期内删失和左删失。右删失是指终点事件在观察结束时还没有出现；期内删失是指终点事件在观察期内发生而未能观察到其发生的时间或者在观察期内失访；左删失是指不知道其处在观察起点的时间。竞争风险是在生存分析的基础上发展而来的，它可以考虑多个不同的终点事件，在一般情况下，竞争风险只讨论右数据删失和完全数据^[20]。

一般来说，临床研究关注的终点事件不止一个^[21]。经典的生存分析一般只考虑一个终点事件，如果要考虑有多个终点事件，就要用到多结局的生存分析。竞争风险模型中不止一个终点事件，并且对任一个病例来讲，其中任何一个终点事件的发生就会影响其它所有的发生^[4]。上文的例子中原地复发和异地转移就是死亡的竞争风险。

竞争风险的研究最早可以追溯到 18 世纪, Bernoulli 研究天花的根除对死亡率的影响^[4],但竞争风险的快速发展是 Cox 在 1972 年提出比例风险模型之后^[5]。现在竞争风险的理论与方法丰富了许多,它的各个方面都有涉及,从样本量到方法研究,从一组数据分析到多组数据分析与比较都有许多研究成果。最近关于竞争风险研究发生的一个比较重要的变化就是:现在一般都选择直接模型化总累计风险函数而不是原因别风险函数,这样使得结果更加容易理解与解释^[4,5,8]。在医学研究中,竞争风险模型一出现就得到了科研者的青睐,并得到很大的发展,特别是在最近二三十年中发展尤其迅速,但是在我国这方面的研究比较少见。

1.2 研究意义

中国人口协会日前发布的《中国乳腺疾病调查报告》称,2003 年到 2009 年 7 年内,我国城市乳腺癌患者死亡率增长了 38.91%,已成为对妇女健康威胁最大的疾病之一。据资料统计,乳腺癌的发病通常与遗传有关,以及 40-60 岁之间、绝经期前后的妇女发病率较高^[22]。

本研究以在女性中恶性肿瘤排在首位的乳腺癌为例,乳腺癌是女性死亡率很高的一种疾病。多数患者死亡年龄在 45~79 岁之间,其中又以正处于生命力旺盛的 45~59 岁为高峰期。本文首先运用经典的生存分析 Kaplan-Meier 方法(简称 K-M 方法)对乳腺癌患者的总体生存率进行分析,并运用 K-M 方法对各个预后因素进行单因素分析,再通过 Cox 比例风险模型进行多因素分析,进一步认识影响乳腺癌的临床预后因素。如考察手术方式、治疗方式、年龄大小、肿瘤大小、病理类型、淋巴结情况对病人生存率的影响,从而对这些因素加以控制,提高病人的预后生存质量。

本文的生存分析实证研究中只考虑一个终点事件,即假设不存在竞争风险的情况。显然只考虑一个终点事件得出的结果是有偏差的。本文的第五章实证分析将考虑复发、转移两个终点事件,将复发、转移视作互为竞争事件,这样比只考虑一个终点事件更准确合理的分析生存率,并且能够分别考察在竞争风险下,各个预后因素对于不同结局的影响,从而更加合理的控制影响因素,帮助制定更加

合理的治疗方案，改善乳腺癌患者的预后，达到更好的治疗效果，提高患者的生存率和生存质量，为卫生工作者提供一定的参考依据。

1.3 研究目标

本文采用厦门市及武汉市 4 家医院 1998 年至 2008 年入院接受治疗的女性乳腺癌患者的数据。首先在不考虑竞争风险的情况下，运用 K-M 估计方法估算乳腺癌患者的总体生存率，并对各个潜在因素进行单因素分析，用 Log-rank 法检验比较不同组别的患者的生存曲线，检验生存差异，从而进行单因素分析。然后运用 Cox 比例风险模型对潜在的影响因素进行多因素分析，得出影响乳腺癌患者的预后生存指标。在考虑竞争风险的情况下，即视复发、转移互为竞争风险事件，运用原因别比例风险模型估计协变量的系数，然后在原因别比例风险模型的基础上计算累积风险率，与 K-M 法估算出的总体生存率进行对比，分析两种方法的优劣势。最后，在竞争风险模型的基础上分析各个预后因素对乳腺癌患者的影响，得出独立的预后因素，帮助制定合理的治疗方案，改善乳腺癌的预后，提高患者的生存率和生存质量，为卫生工作者提供一定的参考依据。

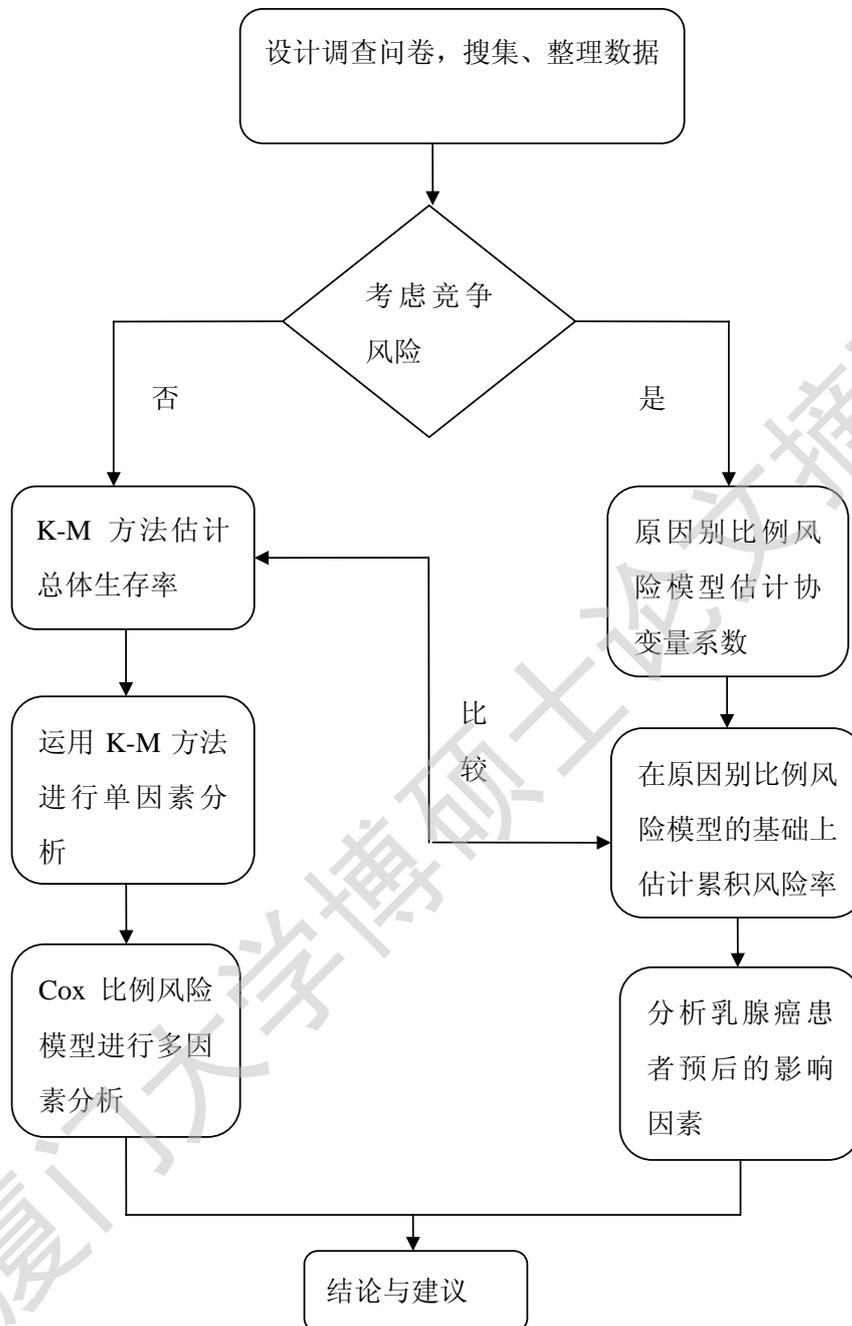
在不考虑竞争风险的情况下：

- (1) 用 K-M 方法估计乳腺癌患者的总体生存率；
- (2) 用 K-M 法对预后因素进行单因素分析，用 Log-rank 法检验统计差异性；
- (3) (3) 用 Cox 比例风险模型对潜在的影响因素进行多因素分析，分析影响乳腺癌患者生存率的预后因素。

在考虑竞争风险的情况下：

- (4) 用原因别比例风险模型估计协变量的系数；
- (5) 在原因别比例风险模型的基础上计算累积风险率；
- (6) 在竞争风险模型的基础上得出的生存率与 K-M 法估计的生存率进行比较；
- (7) 在竞争风险模型的基础上分析乳腺癌患者预后的影响因素。

1.4 研究框架



第二章 文献综述

在疾病的预后研究中，生存分析是比较常见的统计分析方法，如通过Kaplan-Meier方法（简称K-M法）估计生存概率，用Logrank方法比较两条或多条生存曲线^[1]，采用Cox 比例风险模型^[2]分析多个潜在因素对生存时间的影响等。上述经典的生存分析研究仅关心某一个终点的情况，然而，在医学研究中，观察的终点往往不是单一的^[3]，而是存在多个终点和竞争风险事件，若在可能会发生多个终点事件的情况下仍应用这些单终点分析方法，将会由于竞争风险的存在而导致对这些终点事件概率的估计偏差，同时，由于同一观察对象的不同终点之间存在相关，不宜用多次单一终点的生存分析方法进行分析，那么，在可能发生具有竞争风险性的多个终点事件的研究中，如何找到一种合理的分析方法以更科学地预测各终点发生的概率呢？

竞争风险模型就是一种用于处理具有竞争风险事件的分析技术，它可分析面临多种潜在结局的生存数据，这些数据包括失效的时间跨度和导致失效的终点事件，这种终点事件可能有多个，这些潜在的终点事件互被称为“竞争风险”事件。竞争风险仅仅关心每个观察的第一个发生的终点事件，而后发生的任何终点事件则成为删失事件^[4]。例如，在研究乳腺癌患者术后的发展变化中，以“死亡”为研究终点事件时，复发就成为死亡的竞争风险事件，但若一个病人先发生了复发，而后发生的死亡就成为了删失事件。这类模型都普遍用来进行可靠性研究和生存分析。在医学研究中，竞争风险模型一出现就得到了科研工作者的青睐，并有了很大发展，近二三十年的发展尤为迅速，但我国这方面的研究较少。本文就竞争风险模型的发展及其应用以及在该模型发展历程中出现的几个主要模型分别加以讨论和阐述。

一、模型的发展及应用

竞争风险的研究最早可以追溯至 18 世纪，当时 Bernoulli 研究根除天花对死亡率的影响^[4]。K-M 法于 1958 年出现后，迅速地应用在生存分析研究上，一段时间内也将它应用于竞争风险事件分析，但它是一种经验性研究，并非真正的预

后分析^[1]。竞争风险研究的快速发展是 Cox 于 1972 年提出比例风险模型之后^[2]，更确切的说是原因别风险模型出现之后，因为它标志着竞争风险研究从以前的经验性研究发展到实证预后研究。现在，竞争风险的理论与方法丰富了许多，从样本量研究到模型方法研究，从一组数据分析到多组数据分析与比较等都取得了许多研究成果。近年来，关于竞争风险研究发生了重要的变化，即以前都是从原因别风险模型入手解决问题和解释结果^[5-8]，而现在更多的是直接采用累积风险模型和边际风险模型^[9-12]，因为原因别风险模型的结果有时比较难以解释，而直接模型化总累积风险模型或者边际风险模型会使得结果更加容易理解与解释，且在结果的准确性上也更胜一筹^[10]，所以累积风险模型和边际风险模型现在的应用越来越广泛。

许多研究者应用比例风险模型获得了研究成果。如 Brent R.logan 和 Fohn P.Klein 等人利用比例风险模型分析了 630 名患者进行相同 HLA 的亲属骨髓移植和 143 名患者进行外周血干细胞移植后发生 GVHD（移植物抗宿主病）的情况，模型分析结果显示，进行外周血干细胞移植后发生 GVHD 的风险是骨髓移植的 2.25 倍^[10]。

原因别风险模型解决了比例风险模型的最大缺陷，即不能同时较准确地考虑多个终点事件，但原因别风险模型仍有一些不足，如它要求观察量两两之间及协变量之间独立，对结果的解释不是很直观等等。尽管它有这些缺点，但这并不影响它对竞争风险研究起的作用和影响，它的应用非常普遍，使得竞争风险研究前进了一大步。它的基本表达式为：

$$\lambda_k(t) = \lambda_{k0}(t) \exp(\beta_k^T \mathbf{X})$$

其中， $\lambda_k(t)$ 表示协变量 X 在时间 t 上发生第 k 种终点事件的风险率， $\lambda_{k0}(t)$ 表示第 k 种终点事件在时间 t 上的基础风险率， β_k 是协变量 X 的系数。

具有一特定协变量的观察就有随时间 t 变化的各种终点事件的风险率。生存率估计为：

$$S(t) = \exp \left\{ - \sum_{1 \leq k \leq K} \Lambda_{k0}(t) \exp(\beta_k^T \mathbf{X}) \right\}$$

其中， $\Lambda_{k0}(t)$ 表示累积基础风险函数。

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库