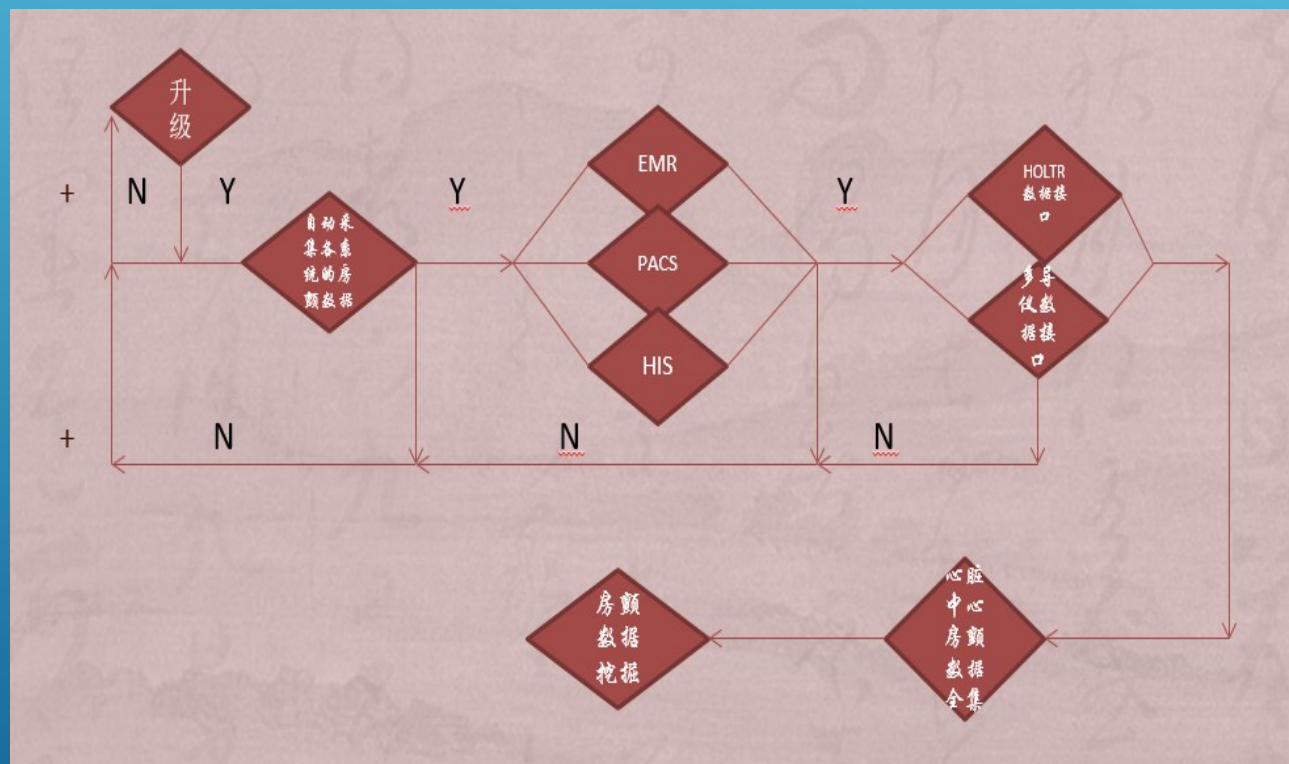


医学数据挖掘的成果与实践

邵学杰 李强 王艳铭

北京杰杰科技有限公司

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S |
|---|------|----|----|-------------|----|----|----|------|-------|------|-------|-----|--------|---|----|------|-------|------|-------|
| 1 | 一般情况 | | | 试验情况 (化验数据) | | | | | | | | | | | | | | | |
| 2 | 年龄 | 性别 | 籍贯 | 身高 | 体重 | 病史 | 心率 | 心率变异 | ST段分析 | 室性节律 | 室上性节律 | 监护仪 | HOLTER | | 心率 | 心率变异 | ST段分析 | 室性节律 | 室上性节律 |
| 3 | | | | | | | | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | | | | | | | | |
| 7 | | | | | | | | | | | | | | | | | | | |
| 8 | | | | | | | | | | | | | | | | | | | |



| 数据库 | |
|--------|--------------------|
| 采集项目名称 | |
| 患者信息 | 血液检查 |
| 患者体检信息 | HOLTER 机器诊断 |
| 临床诊断 | HOLTER 临床诊断 |
| 超声心动图 | 特殊检查 |
| 常规心电图 | 血清酶指标 |
| 用药情况 | 共计: 113 项 |

案例1: 房颤与肾功能的关联性



MIT-BIH Database Distribution

Harvard-MIT Division of Health Sciences and Technology



Welcome!

We invite you to visit [PhysioNet](#), the on-line component of the Research Resource for Complex Physiologic Signals, where you will find the data, software, and reference materials previously posted here or included on our CD-ROMs, and much more.

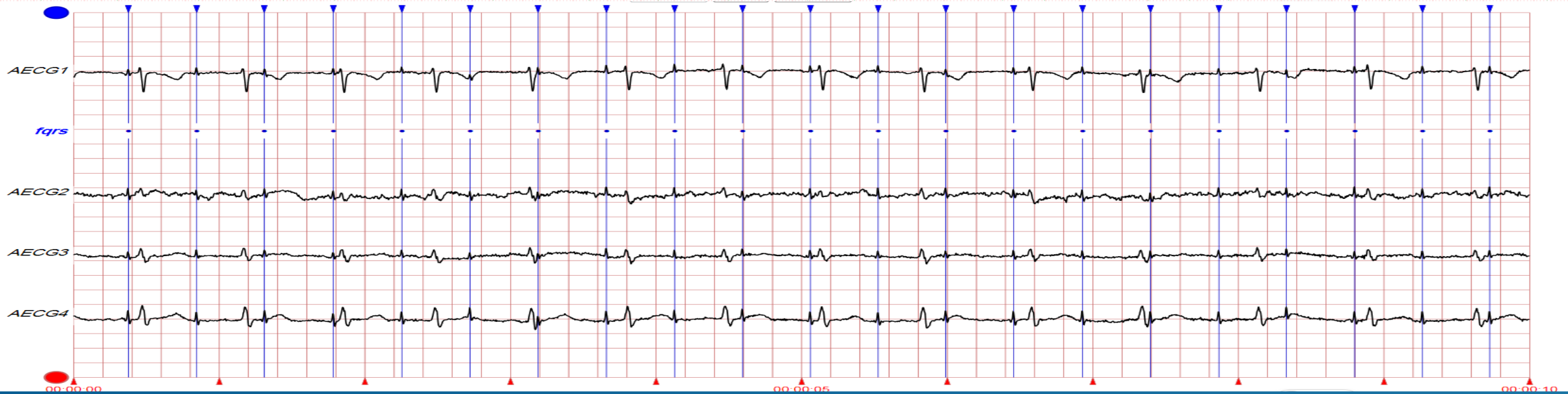
Since 1999, with the support of the National Center for Research Resources of the National Institutes of Health, we have been able to make a large amount of our data and software available freely via PhysioNet. Our CD-ROMs contained only a small fraction of the [data](#) and [software](#) freely available on PhysioNet. [Everything previously available on our CD-ROMs](#) is now downloadable without cost from PhysioNet, and we encourage you to use this service.

Links:

- ⌘ Our laboratory, the [Laboratory for Computational Physiology](#), is part of the [Harvard-MIT Division of Health Sciences and Technology](#).
- ⌘ We collaborate closely with colleagues at
 - the [Margret & H.A. Rey Laboratory for Nonlinear Dynamics in Medicine](#) at Boston's [Beth Israel Deaconess Medical Center](#)
 - the [Center for Polymer Studies](#) at [Boston University](#)
 - Boston's [Hebrew Rehabilitation Center for Aged](#)
 - the [Laboratory of Biomedical Computer Systems and Imaging](#) at the [University of Ljubljana](#) (Slovenia)
 - the [National Research Council \(CNR\) Institute of Clinical Physiology](#) in [Pisa](#) (Italy)
 - the [Center for Nonlinear Dynamics in Physiology and Medicine](#) at [McGill University](#)
- ⌘ Here is a brief exploration of [heart rate time series](#), with examples.
- ⌘ Meet us at [Computers in Cardiology!](#) We hosted the 1986 and 2000 annual conferences and attend CinC regularly.

Your comments and suggestions are welcome; please send them to:

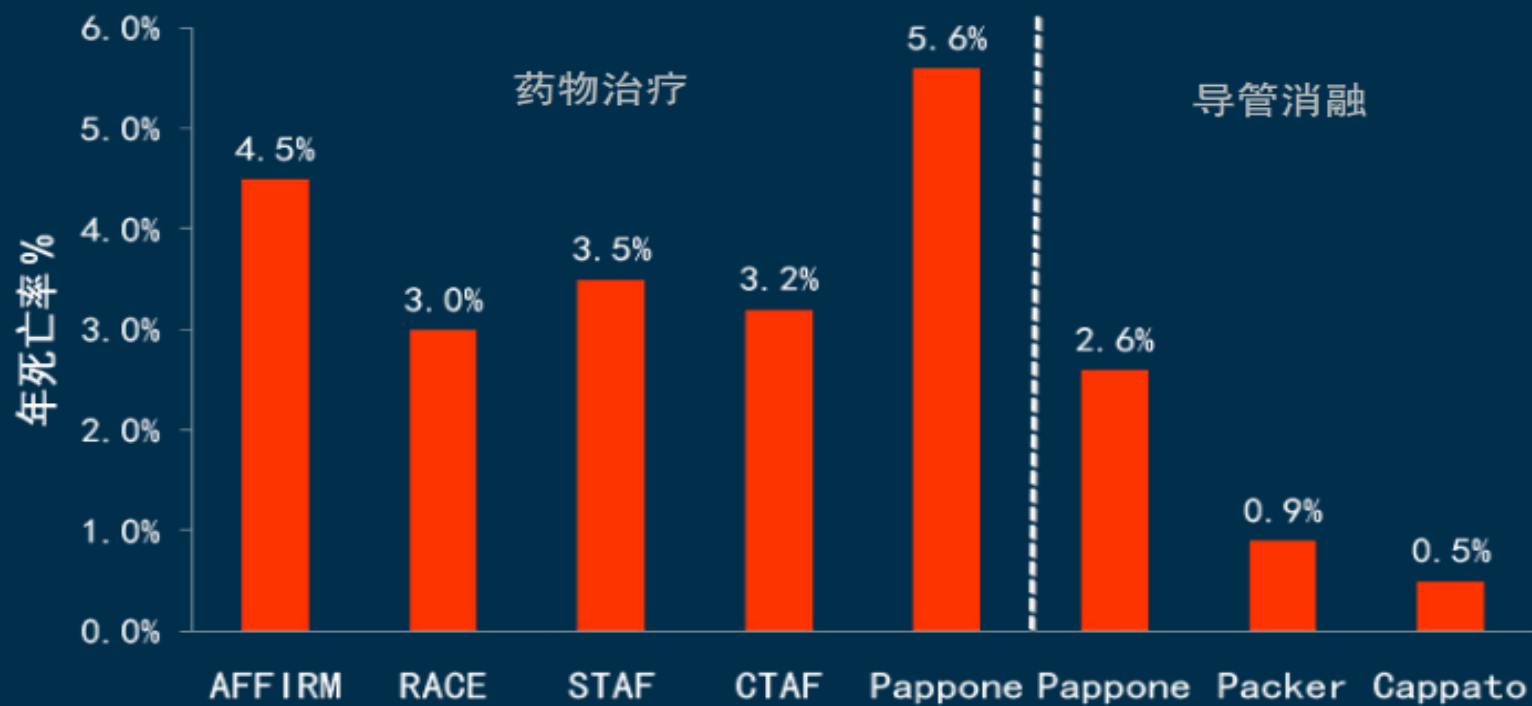
[George B. Moody \(george@mit.edu\)](mailto:george@mit.edu)
Harvard-MIT Division of Health Sciences and Technology
MIT Room E25-505A
Cambridge, MA 02139 USA
Updated 22 July 2005.



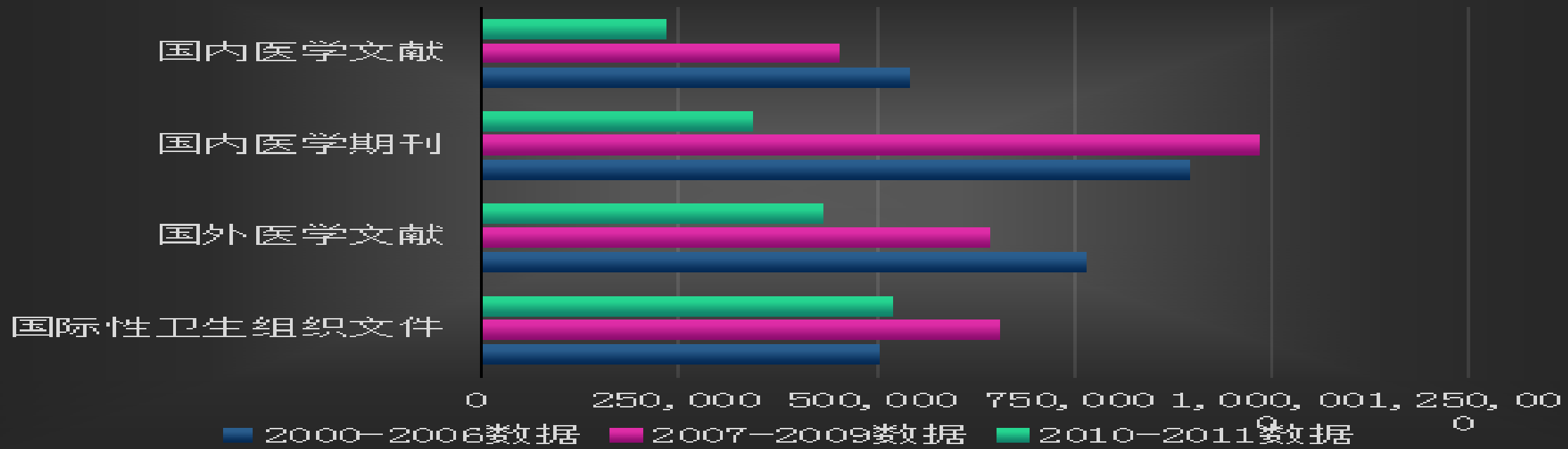
- 解释性数据分析
- 描述性建模
- 预测性建模
- 知识性的发现
- 序列模式发现
- 依赖关系的发现
- 异常与趋势发现

医学数据挖掘的7种模式

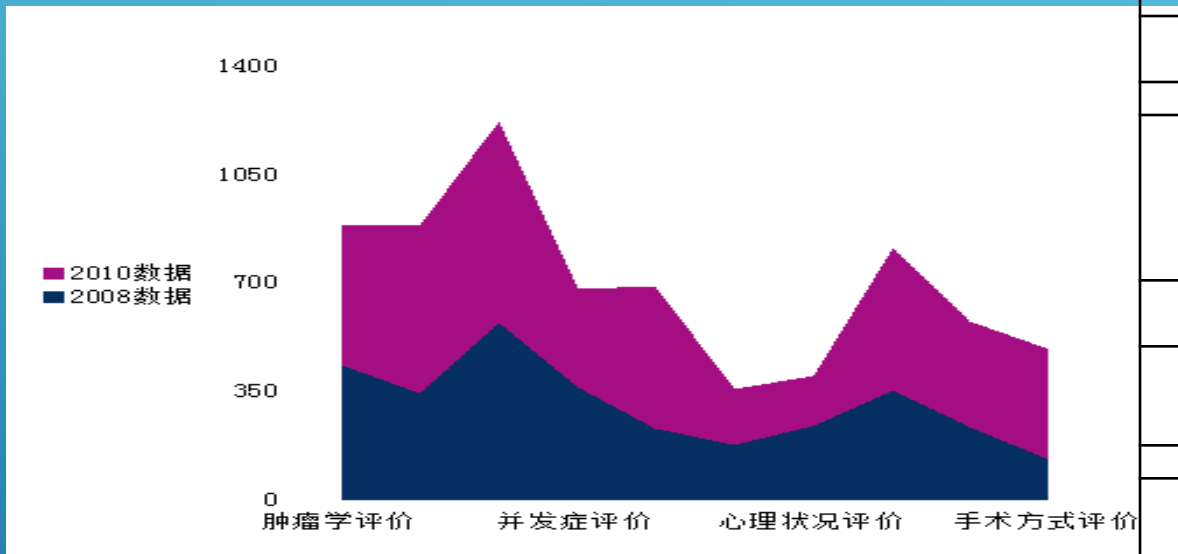
死亡率：药物 vs 导管消融



老年肺癌文献



案例2：老年肺癌研究（双盲实验）



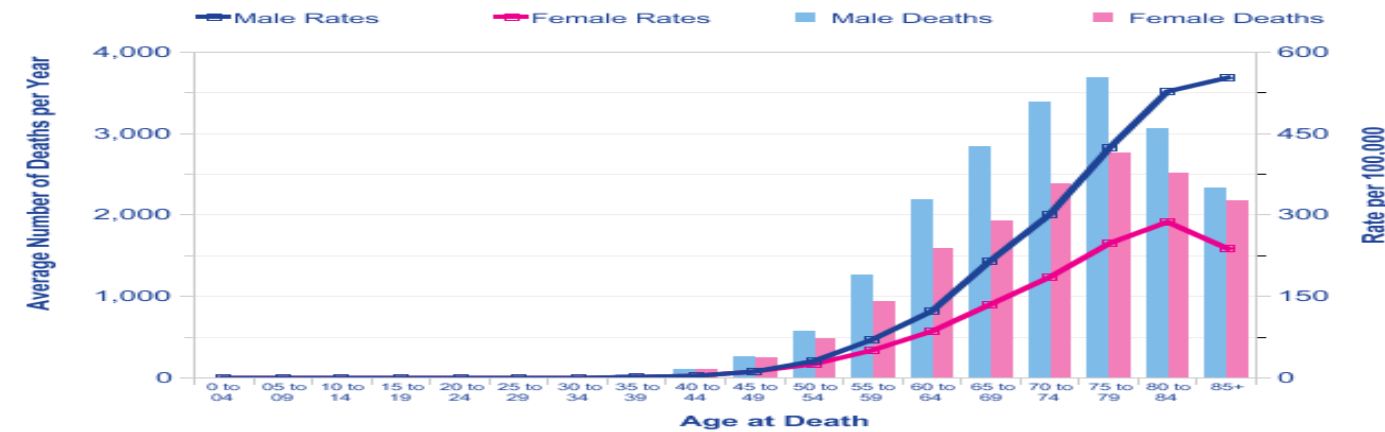
| | |
|-------------|---|
| 肿瘤学评价 | 一次标准的肿瘤学检查包括但不限于：彻底的病史及体检，支气管镜，CT/PET C T，血液检查，组织活检 |
| 老龄因素评价 | 身体功能，并发症探测，认知，心理，围术期护理，药物 |
| 身体功能 并发症 | 主要是与手术相关的心血管，肺功能测试 老年人过往病史研判，体检结果研判。能否预测哪些肺癌患者术后效果好、生存时间较长些？可用公用的预后评估模型察尔森合并症严重度指标 (Charlson comorbidity index,CCI) 为基础。 |
| 认知能力 | 认知缺陷对术前检查很重要，对术后康复也有重大影响 |
| 营养状况 | 老龄人口手术风险主要是耐受性。老年人营养差，身体弱会对手术，化疗的耐受性产生重大影响。 |
| 心理状况 | 30%的老年肿瘤患者都有心理疾病 |
| 手术方式选择 | 按照美国SEER数据库显示：右全肺切除要尽量避免，这与术后生存率息息相关。数据挖掘后还发现：年轻人肺段切和楔形切的生存率大大低于肺叶切除术，老年人则无此差别，就局部复发而言，肺段切高于肺叶切。 |
| 社会支持因素 | 老人家庭的支持，围术期护理至关重要 |

手术适用模型

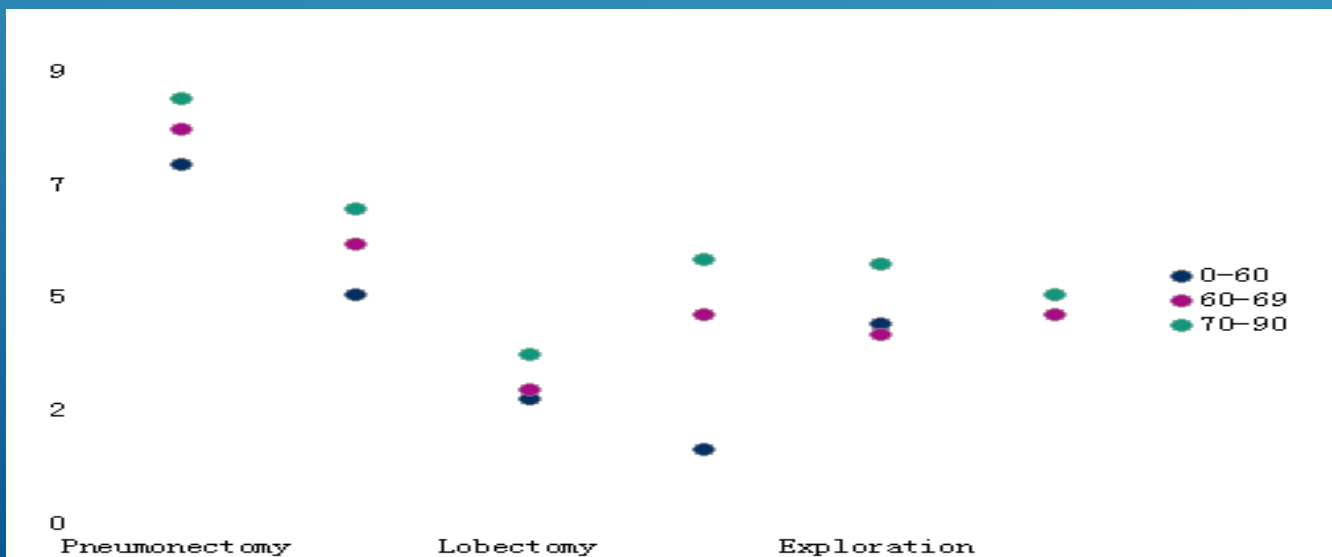
VATS Lobectomy Open Segmentectomy Significance

| | | | | |
|----------------------------|---------|-------------|--------------|--------|
| Mean % Predicted FEV1 | (Range) | 54(31-69) | 51(34-69) | p=0.76 |
| Mean Operating Time (mins) | (Range) | 204(80-270) | 195(114-266) | p=0.9 |
| ITU admission | | 3 (5.7%) | 3 (5.7%) | p=1 |
| 30 day mortality | | 3 (5.7%) | 3 (5.7%) | p=1 |
| Length of Stay (days) | (Range) | 8.8 (3-67) | 10.4 (3-32) | p=0.97 |
| Mean Survival | Years | 6 | 5.4 | p=0.98 |

资料表明电视胸腔镜（VATS）技术在早期肺癌切除中并没有优势，这也是很多胸外科医生仍然偏爱小切口开胸手术的原因。



从1971年-2009年，英国每十万人中肺癌死亡率综合成缓降趋势，然而男性与女性有较大的差别，男性死亡率呈下降趋势而女性呈缓慢增长的趋势。女性肺癌患者30年来比较平稳的死亡率表明英国社会肺癌的环境因素在长达30年的过程中没有太大的变化，人口老龄化，职业女性，外来移民女性吸烟习惯的增加是重要的因素



10万例肺癌手术中，Pneumonectomy死亡率最高7.2%，Segmental/wedge最低1.5%，Bilobectomy4.6%高于Lobectomy2.5%，这深刻的表明肺癌生长部位，手术切除部位与方法对患者存活率有重大影响，老人肺全切可能导致呼吸系统衰竭。

在许多假设机制中，心率升高可能直接影响心血管风险，多数与心肌需氧增加、能量缺乏、动脉粥样硬化进展或斑块破裂风险升高有关。

如果上述假设为真，心率数据隐含着什么真相？心率与哪些风险相关？

冠心病患者的风险临界点能否用单一心率指标作风险提示？
心率升高与结局之间的定量关系如何描述？

能否用数据挖掘的手段给出答案？

案例3：心率数据的模式识别

表1 患者基线特征

| | 心率<70次/min (n=2745) | 心率≥70次/min (n=2693) | P |
|-----------------------------|------------------------|------------------------|---------|
| 人口统计学特征 | | | |
| 年龄(岁) | 65.6(8.2) | 64.4(8.6) | <0.0001 |
| 性别(男) | 2298(84%) | 2209(82%) | 0.098 |
| 当前吸烟者 | 353(13%) | 481(18%) | <0.0001 |
| 体重指数(kg/m ²) | 28.3(4.1) | 28.7(4.7) | 0.0016 |
| 既往病史 | | | |
| 高血压病史 | 1911(70%) | 1927(72%) | 0.12 |
| 糖尿病史 | 864(31%) | 1155(43%) | <0.0001 |
| 血脂异常病史 | 2155(79%) | 2123(79%) | 0.77 |
| 既往心肌梗死 | 2468(90%) | 2349(87%) | 0.0019 |
| 经皮冠状动脉介入或冠状动脉搭桥术 | 1464(53%) | 1360(51%) | 0.037 |
| 卒中史 | 468(17%) | 503(19%) | 0.12 |
| 周围动脉疾病史 | 346(13%) | 402(15%) | 0.013 |
| 心脏参数 | | | |
| 心率(次/min) | 64.1(2.8) | 79.2(8.7) | .. |
| 收缩压(mmHg) | 127.2(15.2) | 128.5(15.7) | 0.0017 |
| 舒张压(mmHg) | 76.7(9.2) | 78.3(9.2) | <0.0001 |
| 左室射血分数(%) | 32.7(5.3) | 31.9(5.7) | <0.0001 |
| NYHA心力衰竭分级I级 | 467(17%) | 373(14%) | <0.0001 |
| NYHA心力衰竭分级II级 | 1744(64%) | 1615(60%) | |
| NYHA心力衰竭分级III级 | 534(19%) | 705(26%) | |
| 随机分组时的治疗药物 | | | |
| 阿司匹林或抗血小板药物 | 2596(95%) | 2507(93%) | 0.023 |
| 血管紧张素转换酶抑制剂和(或)血管紧张素II受体抑制剂 | 2452(89%) | 2421(90%) | 0.049 |
| β受体阻断剂 | 2465(90%) | 2273(84%) | <0.0001 |
| 他汀类药物 | 2087(76%) | 1945(72%) | 0.0014 |
| 利尿剂(除醛固酮拮抗剂) | 1490(54%) | 1704(63%) | <0.0001 |
| 硝酸酯类药物 | 1133(41%) | 1202(45%) | 0.0123 |
| 醛固酮拮抗剂 | 666(24%) | 800(30%) | <0.0001 |

除非特别说明,数据以n(%)或 $\bar{x}(s)$ 表示。NYHA=纽约心脏协会。1 mmHg ≈ 0.133 kPa

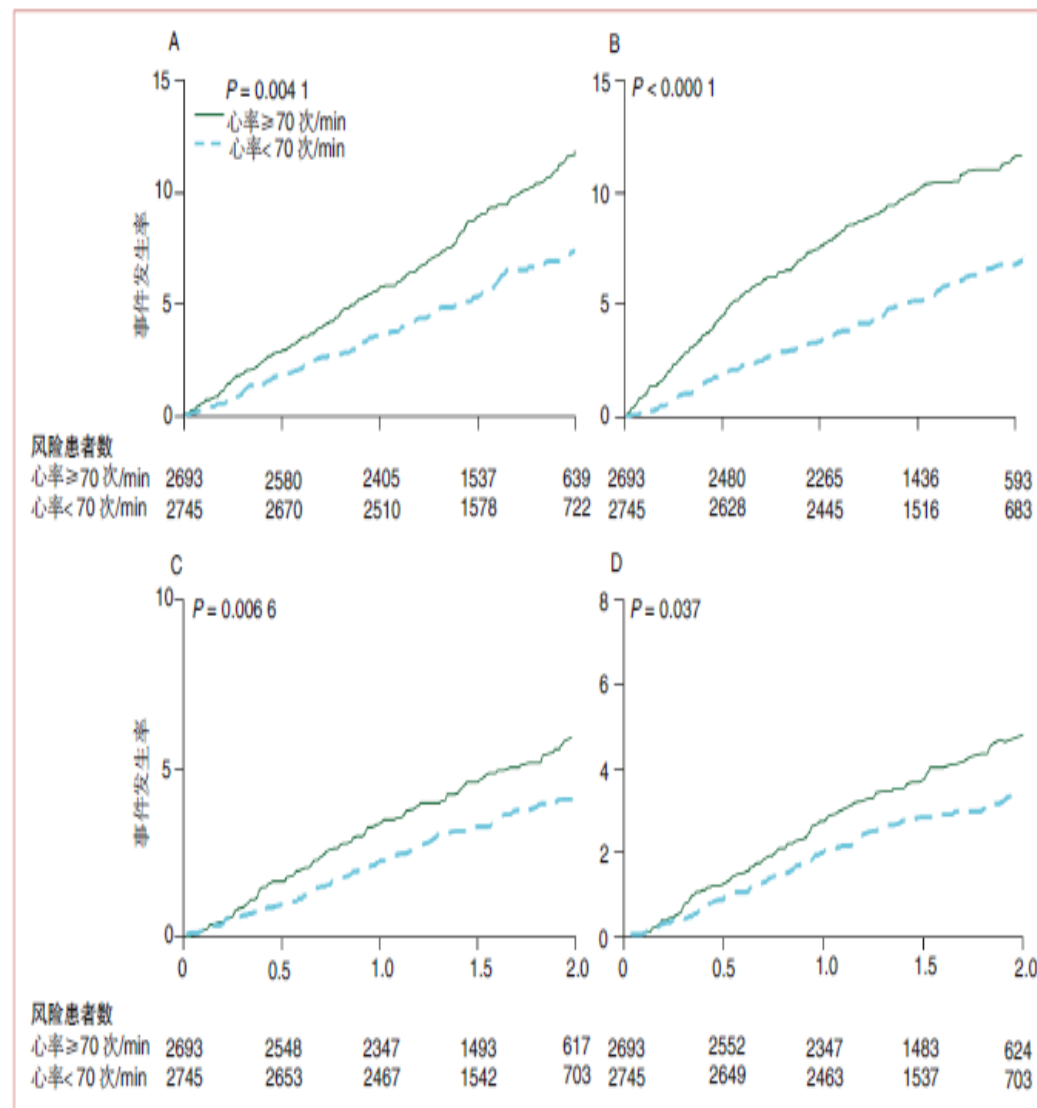


图1 Kaplan-Meier 时间-事件曲线,分别为心率与心血管死亡(A)、心力衰竭入院(B)、心肌梗死入院(C)及冠状动脉血运重建(D)

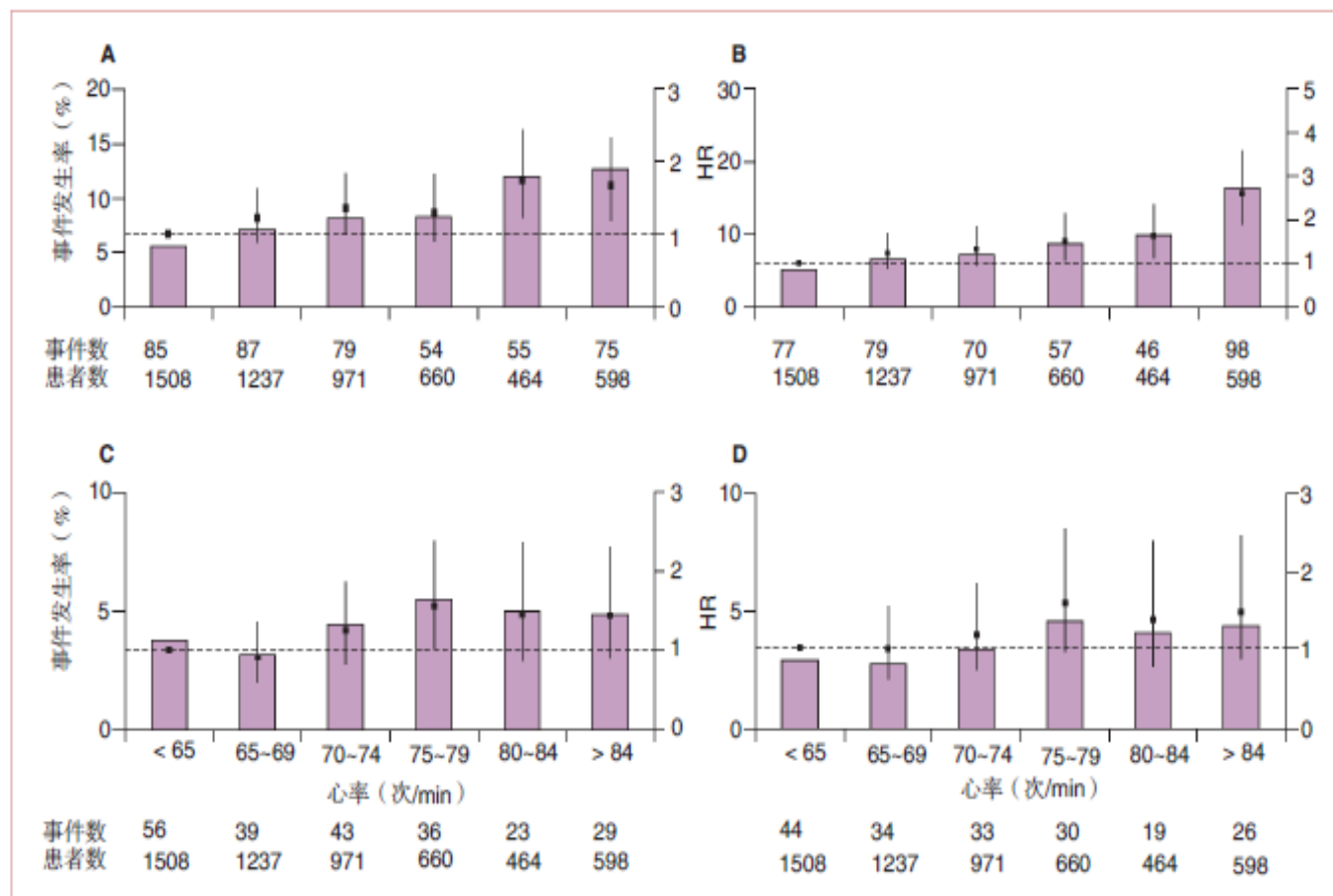
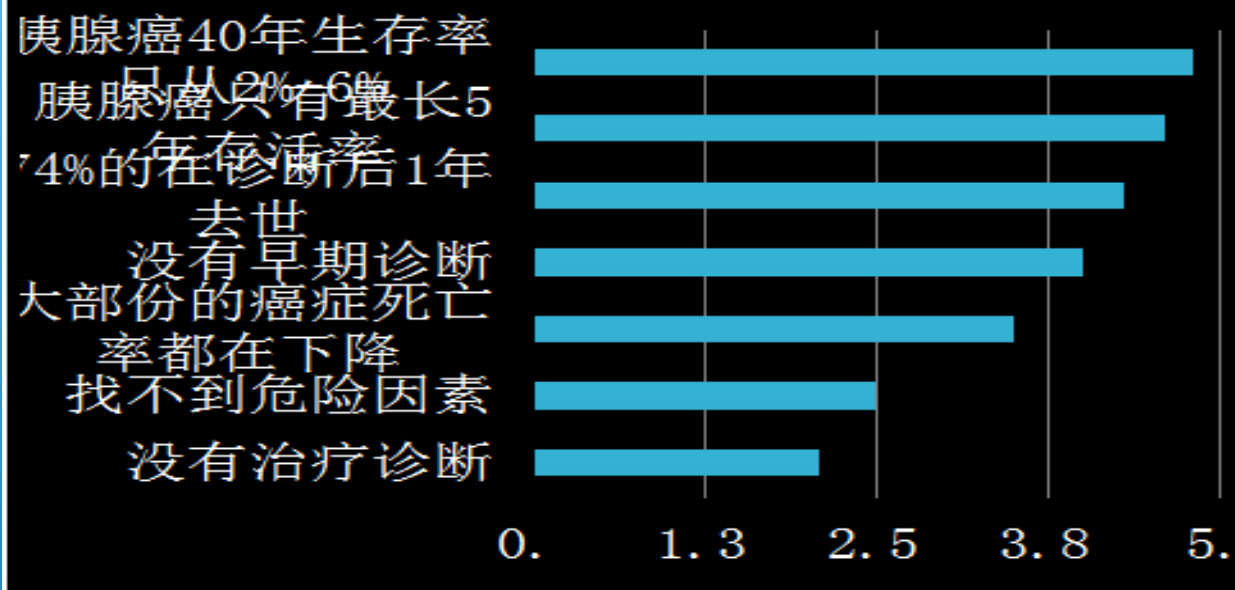


图2 按照心率分组的事件粗发生率及HR

按照心率分组的事件粗发生率（条；左侧标尺）以及HR（95%CI，右侧标尺），与心率<65次/min有关的（A）心血管死亡、（B）心力衰竭入院；（C）心肌梗死入院；（D）冠状动脉血运重建术。水平线为HR = 1。表1中以基线时差异较小的变量进行校正的HR（ $P < 0.05$ ）

- 结 论 在冠心病和左室收缩功能障碍的患
- 者中，心率升高（ ≥ 70 次/min）意味着心血管疾病
- 结局风险升高，同时对心力衰竭相关性结局与冠
- 状动脉事件相关性结局有着不同的影响

胰腺癌特点



案例4：胰腺癌与二型糖尿病

| 变量名称 | 变量指标 |
|---------------|---|
| 血糖控制和胰岛素敏感性指标 | OGTT 中的空腹血糖水平 (FPG) (mmol /L) |
| | OGTT 中的空腹胰岛素水平 (FINS) (pmol /L) |
| | OGTT 中的餐后2 h 血糖水平 (PBG) (mmol /L) |
| | OGTT 中的餐后2 h 胰岛素水平 (FINS) (pmol /L) |
| | 糖化血红蛋白 (HbA1c) (%) |
| 糖尿病相关的脂代谢指标 | 第一次就诊时的总胆固醇水平 (TC) (mmol /L) |
| | 第一次就诊时的甘油三酯水平 (TG) (mmol /L) |
| | 第一次就诊时的低密度脂蛋白胆固醇水平 (LDL - C) (mmol /L) |
| | 体重指数 (BMI) |
| 糖尿病相关的肝、肾功能指 | 总胆红素水平 (TBIL) (μ mol /L) |
| | γ - 谷氨酰转肽酶水平 (γ - GT) (u / L) |
| | 胱抑素C (Cystatin C) (mg /L) |
| | 肌酐 (酶法) (CR) (μ mol /L) |

胰腺癌与糖尿病的指标差曲线

