



中国科学院深圳先进技术研究院  
SHENZHEN INSTITUTE OF ADVANCED TECHNOLOGY  
CHINESE ACADEMY OF SCIENCES

---

# 肾衰风险预警模型

——信息中心  
姓名：刘利明  
导师：蔡云鹏



## 背景及相关研究

高血压是严重危害全球公共健康的重大问题，其危害性不仅体现在其高发的频率，也是肾脏疾病的致病因素。据统计我国每年每100万人口中有90~100人进入肾功能衰竭阶段，目前有关肾功能衰竭的研究，主要集中在肾功能不全的疾病患者，传统上利用临床方法，肾小球滤过率（GFR）认为是评价肾功能的最好指标，及时准确的评估GFR对肾功能衰竭早期诊断及治疗有重要的意义，一般通过Tc—DTPA肾动态显像检查，测得GFR（rGFR）作为参考标准，而常见的估算肾小球滤过率（GFR），主要通过测量人体血清肌酐和血尿素氮指标，利用CDK—EPI公式，简化MDRD公式或者Mayo Quadratic公式以及中国改良简化MDRD公式预测GFR水平的适用性，分析估测的GFR与参考值（rGFR）的相关性。



## 病人来源

本课题研究数据来源于深圳市**68**家公立医院（含所有三甲医院）和**600**多家社康医院，总共覆盖全市**1450**万居民信息档案，历史数据一共涉及**1600**万份，总计超过**30**亿条医疗数据记录。

## 病人纳入

选取高血压患者作为研究对象，医疗诊断报告中高血压患者预后诊断出的高血压性肾衰竭、高血压性肾病、肾动脉硬化、肾小动脉硬化症、高血压性心脏病肾脏病伴心力衰竭、高血压性心脏病肾脏病伴肾功能衰竭、高血压性心脏病及肾脏病、急性肾衰竭、未特指的肾衰竭和2型糖尿病性肾病这**10**种病症中的**1**种或多种划分为肾功能衰竭患者。高血压患者但预后没有诊断出任何严重病症划分为高血压健康患者。病人选取肾功能衰竭患者及健康高血压患者作为研究人群。



## 病人选取标准

以高血压患者为研究对象，肾功能衰竭患者选取标准：选取高血压确诊日期唯一，且高血压确诊日期小于肾衰发病日期。高血压健康患者选取标准：选取高血压确诊日期唯一，且除去3年内健康高血压患者，即高血压患者最大随访日期减去确诊日期所得时间差小于等于3年的健康高血压患者。



## 特征字段

选取高血压随访表、高血压登记表、居民个人信息表、公共健康体检表、高血压ICD编码表，共涉及特征字段：性别、年龄、身高、体重、舒张压、收缩压、空腹血糖、心率、体质指数、随访方式、危险分层、高血压管理级别、呼吸困难、眼花耳鸣、肢体麻木、鼻出血、恶心呕吐、下肢水肿、头痛头晕、心悸、以及除去严重病症的其他诊断病症名称，共计**310**个特征。。

## 特征选取标准

以涉及人口信息、体检信息、诊断病例的特征为主，排除冗余、无意义，空缺值高于**70%**的特征。



## 人群基本情况

根据以上陈述，抽取出高血压肾功能衰竭患者**796**人，健康高血压患者**16908**人，由于数据不平衡性，对健康高血压患者依据高血压肾功能衰竭患者年龄和性别分层抽样出**808**人用于实验模型，其中高血压肾功能衰竭患者男性人数：**499**，女性人数：**297**，健康高血压患者男性人数：**475**，女性人数：**333**。



## 预处理

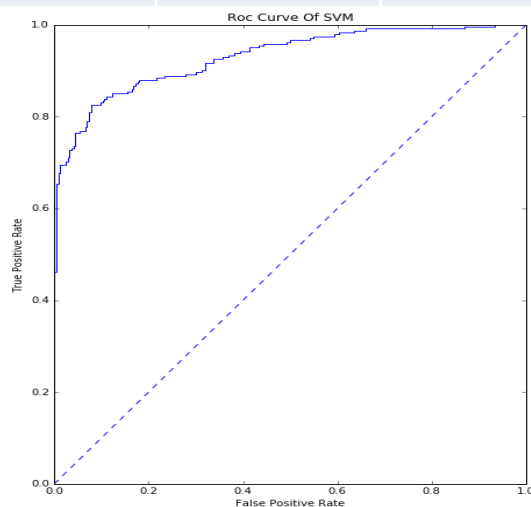
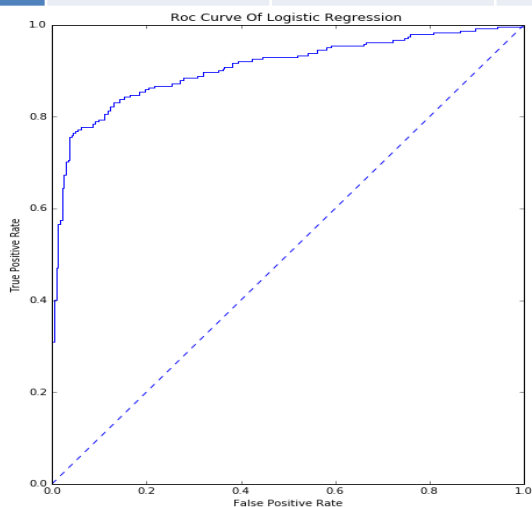
由于医学数据存在不一致，冗余、时间性、不完整和错误值等情况，针对不一致性，即一人多条诊断记录，对高血压肾功能衰竭患者我们选取随访日期小于肾功能发作日期内的记录特征的平均值，对健康高血压患者我们选取在高血压确诊日期小于随访日期小于高血压肾功能衰竭患者的最大时间段与确诊日期之和，这期间内的记录的特征平均值。错误值和不完整性都以不完整方法处理，通过k近邻方法填充， $k=10$ ，或者用众数、平均数填充。



## 建模

对模型所需数据的7份作为训练集，3份作为独立测试集，其中训练集人数1118人，包含实验组554人，对照组564人，独立测试集人数486人，包含实验组242人，对照组244人。分别使用logistics回归和svm算法，结果如下：

Model	accuracy	precision	specificity	sensitivity	auc	f-measure
Logistics	0.8457	0.8615	0.8689	0.8223	0.9078	0.8414
SVM	0.8621	0.9147	0.9262	0.7975	0.9309	0.8521







### 基线对比方法简述

就肾功能不全，临床上分为5个阶段，如图所示：其中第5阶段，为肾功能衰竭。临床上，经常监测血清肌酐（Scr）水平，同时行Tc—DTPA肾动态显像检查，测得GFR(rGFR)作为参考标准，用CKD—EPI公式、简化MDRD公式，Mayo Quadratic公式计算估测GFR，如果划分为第5阶段，说明该患者是肾功能衰竭患者。

CKD stage	GFR level (mL/min/1.73 m <sup>2</sup> )
Stage 1	≥ 90
Stage 2	60 - 89
Stage 3	30 - 59
Stage 4	15 - 29
Stage 5	< 15

CKD—EPI公式：

$$eGFR = 141 \times \min(Scr/k, 1)^a \times \max(Scr/k, 1)^{-1.209} \times 0.993^{Age} \times [1.018 \text{ if Female}] \times [1.159 \text{ if Black}]$$

其中Scr是血清肌酐，女性k=0.7、a=-0.329，男性k=0.9，a=-0.411  
简化MDRD公式：

$$eGFR = 186 \times \text{Serum Creatinine}^{-1.154} \times \text{Age}^{-0.203} \times [1.210 \text{ if Black}] \times [0.742 \text{ if Female}]$$

这里的血清肌酐是mg/dL为单位

MDRD公式：

$$eGFR = 170 \times \text{Serum Creatinine}^{-0.999} \times \text{Age}^{-0.176} \times [0.762 \text{ if Female}] \times [1.180 \text{ if Black}] \times \text{BUN}^{-0.170} \times \text{Albumin}^{+0.318}$$

包含了血清肌酐和血尿素氮（bun）

Mayo Quadratic公式：

$$eGFR = \exp(1.911 + 5.249/\text{Serum Creatinine} - 2.114/\text{Serum Creatinine}^2 - 0.00686 \times \text{Age} - [0.205 \text{ if Female}])^{[32]}$$



# 讨论

依据Logistics回归权重值情况，从大到小排列，查看重要风险因子。

特征	权重	特征情况
n18	2.0299	慢性肾病
Bit08	1.1154	缺铁性贫血、叶酸盐缺乏性贫血、后天性溶血性贫血.....
Bit35	0.9390	肺气肿、其他慢性阻塞肺病
Bit42	0.88226	急性肾炎综合征、慢性肾炎综合征、肾病综合征、未特指的肾炎综合征、急性肾小管-间质肾炎、慢性肾小管-间质肾炎 .....
I48_x01	0.7719	心房颤动
R60	0.74405	水肿,不可归类在他处者
R04	0.6957	呼吸道出血
E10_401	0.6252	1型糖尿病性周围神经病
E11_401	0.5828	2型糖尿病性周围神经病
I70_0	0.5474	主动脉的动脉粥样硬化
Bit16	0.4692	芳香氨基酸代谢紊乱、支链氨基酸代谢和脂肪酸代谢紊乱、糖蛋白代谢紊乱、操作后内分泌和代谢紊乱,不可归类在他处者.....
R55	0.4675	晕厥和虚脱
I95_900	0.4446	低血压
I25_901	0.4240	冠状动脉性心脏病
Bit30	0.4157	风湿热伴有心脏受累、风湿性二尖瓣疾病、风湿性三尖瓣疾病、多个心瓣膜疾病、其他风湿性心脏病
E16_200	0.3870	低血糖症
Bit29	0.3780	外耳炎、外耳的其他疾患、分类于他处的疾病引起的外耳疾患、咽鼓管炎和阻塞
I66	0.3624	大脑动脉的闭塞和狭窄,未造成脑梗死
I25_102	0.3621	冠状动脉粥样硬化
Bit49	0.3420	言语障碍,不可归类在他处者、诵读困难和其他符号识别功能障碍,不可归类在他处者



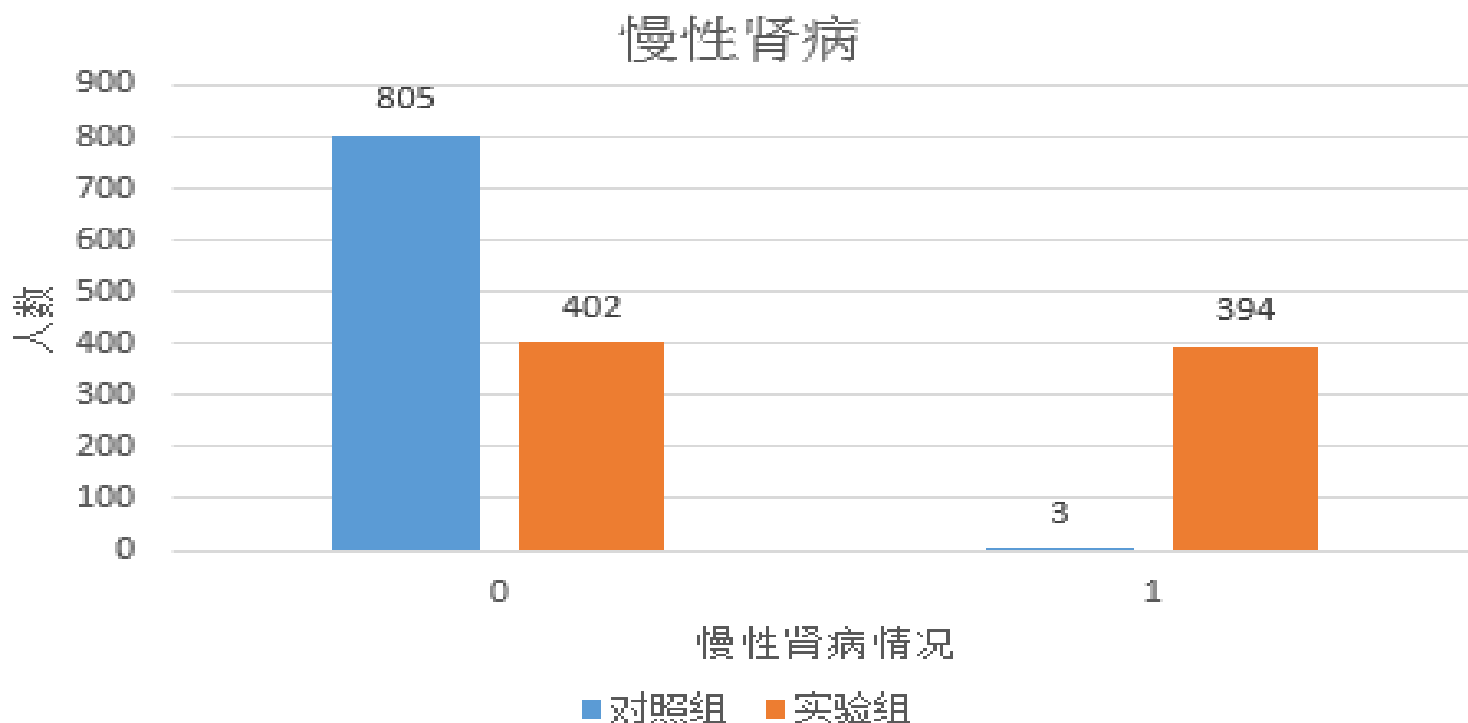
## 讨论

Bit33	0.3207	病毒性肺炎,不可归类在他处者、链球菌性肺炎、细菌性肺炎,不可归类在他处者、肺炎,病原体未特指
R11	0.3112	恶心和呕吐
E13_907	0.3099	继发性糖尿病
H35_004	0.2904	高血压性视网膜病变
.....	.....	.....
risk_stratify	-0.7022	危险分层
Sex	-0.3843	性别
palpitation	-0.3169	心悸
Bit03	-0.3358	沙眼、斑疹伤寒、急性脊髓灰质炎
Bit39	-0.3228	口腔、涎腺和颌疾病、埋伏牙和阻生牙、牙齿硬组织的其他疾病
Height	-0.3028	身高
R74	-0.2929	血清酶水平异常



## 特征分析

### 1、慢性肾病 (n18)





## 特征分析

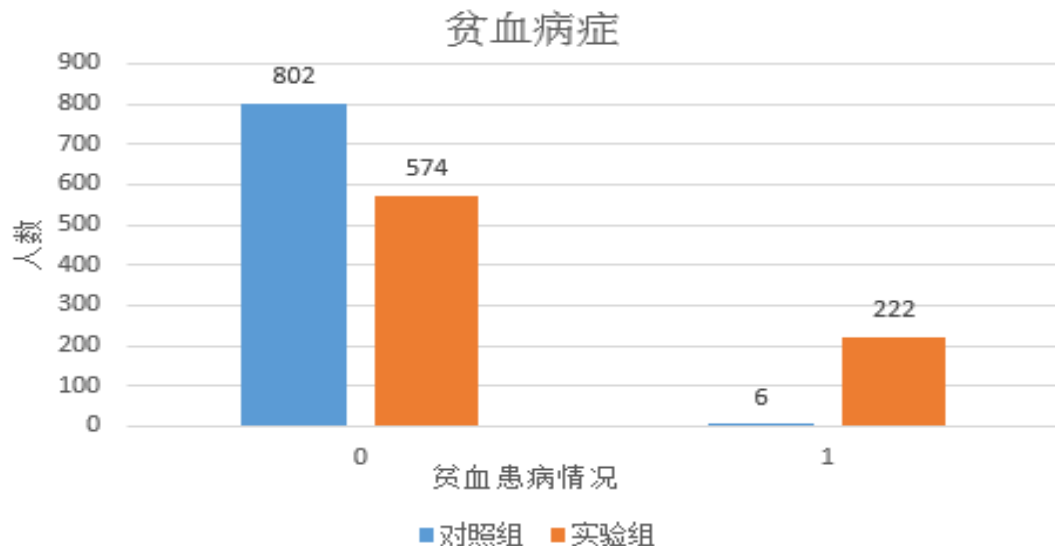
### 1、慢性肾病

从上图中可以看出对照组没有患慢性肾病病症即取值为0的人数805人，患有慢性肾病病症取值为1的人数为3人，实验组没有患慢性肾病病症取值为0的人数402人，患有慢性肾病病症取值为1的人数为394人，从而明显得到慢性肾衰病症能加剧肾衰的发生，临床上研究表明，在临床中，诊断为肾小球肾炎、隐匿性肾炎、肾盂肾炎、过敏性紫癜肾炎、红斑狼疮肾炎、痛风肾、iga肾病、肾病综合征、膜性肾病、肾病综合征、糖尿病肾病、高血压肾病、多囊肾，等等，当这些肾病的发病迁延难愈，时间超过三个月，病人尿液和相关的血液指标出现异常，肾脏病理学、影像学发现异常，或肾脏的肾小球有效滤过率低于60%，都可统称为“慢性肾病”。慢性肾病如未能及时有效救治，导致病情恶化进展，则随病程迁延，慢性肾病患者将发展成为慢性肾功能不全、肾衰竭



## 特征分析

### 2、贫血 (bit08)

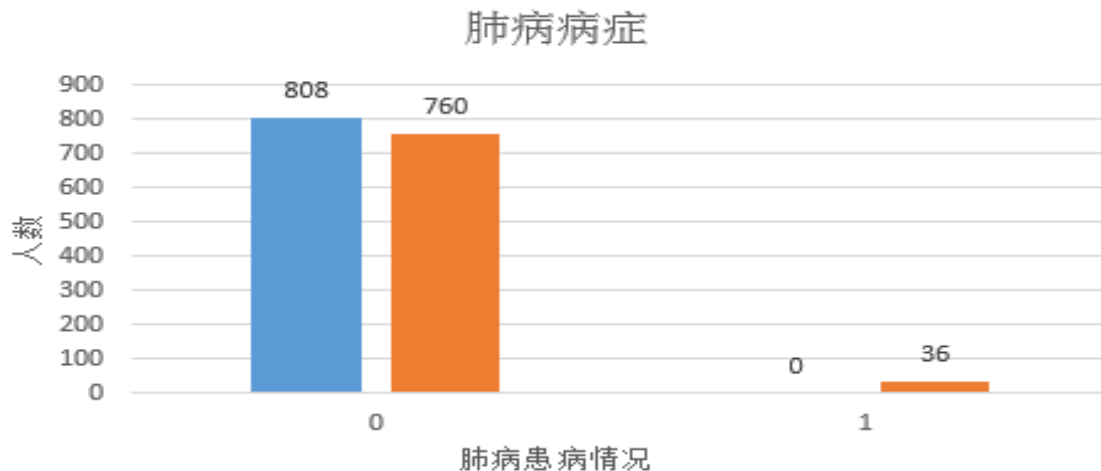


从图中可以看出对照组没有患贫血病症即取值为0的人数802人，患有慢性肾病病症取值为1的人数为6人，实验组没有患慢性肾病病症取值为0的人数572人，患有慢性肾病病症取值为1的人数为222人。从而明显得到患有贫血病症能加剧肾衰的发生。临床表明，贫血，是一种常见的疾病，然而它与肾脏也有着千丝万缕的关系，一般来说，贫血可发生在慢性肾功能衰竭的各个阶段。

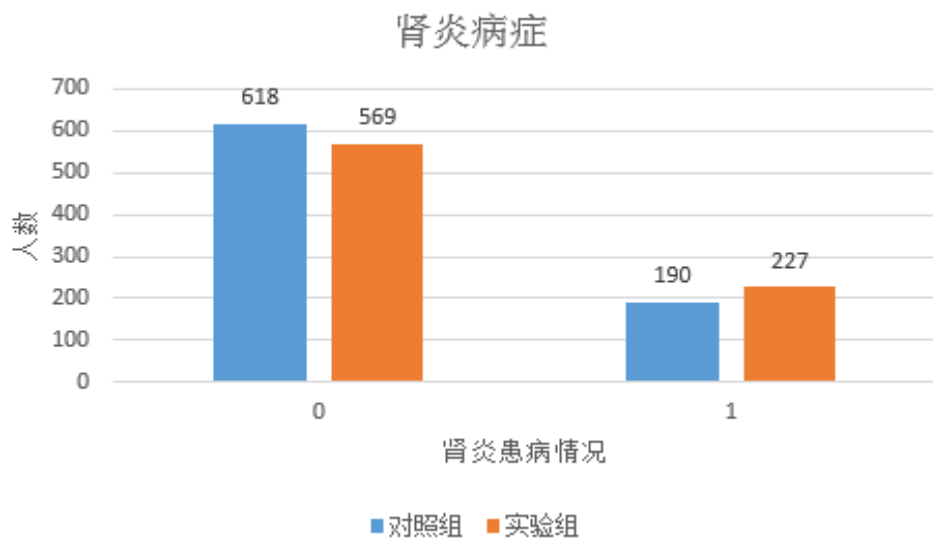


## 特征分析

### 3、肺病 (bit35)



### 4、肾炎病症 (bit42)





中国科学院深圳先进技术研究院  
SHENZHEN INSTITUTE OF ADVANCED TECHNOLOGY  
CHINESE ACADEMY OF SCIENCES

---

# Thank You

交流环节！！！！