



荔枝人寿  
Lychee Life  
Insurance

# 重大疾病保险的“水晶球”

报告团队：世纪战舰队



01

报告背景



### 重大疾病保险

- **重疾险**作为本公司健康险中的**主力产品线**，对公司整体规模和利润提升具有显著影响
- 近两年中公司一款**主力重大疾病保险产品**赔付情况较差
- 重疾保障是终身责任，可能存在较高的**长期赔付风险**
- **新冠疫情**对国民日常生活和消费观念等产生严重影响，对重疾险的赔付情况也产生了一定干扰

# 研究过程

0

对现有数据进行分析处理，剔除小样本数据和异常数据；通过外部公开渠道进行数据拓展，并使用合理方法填补缺失数据

0

使用Logit回归模型就相关风险要素指标对重疾发生率的影响进行验证和分析，并检验模型的准确性

0

使用精算方法计算粗重疾发生率，据此观察各风险要素对重疾发生率的影响规律，进一步验证回归模型结果

0

建立风险预测逻辑，计算各风险要素相对风险评分并赋予其影响权重，形成风险预测工具并窗口化

0

基于风险要素及其对重疾发生率的影响模式，对现有保单样本进行统计分析，探讨重疾险产品赔付情况恶化的背后成因

0

总结已有研究，提出可行后续研究途径，并为本公司降低重疾险赔付风险提供可能的商业建议

# 研究及应用价值

额外扩展了地区数据和水资源数据，丰富了风险要素



建立了各风险要素的相对风险评分线性计算方法并用规划求解方法赋予权重



## 两个创新

## 四种价值



验证了可能的风险要素及其影响模式，为后续进一步探索风险要素、优化风险预测工具打下基础



有效预测新投保保单的风险等级，可以为承保端提供商业建议



对本公司进行赔付恶化成因分析，多维探索了近两年赔付恶化的可能原因



根据研究结论给出针对性建议，降低赔付风险，改善公司经营情况

# 02 数据处理及 指标选取

# 数据处理

1

## 现有数据样本

选取本公司主力重疾险产品2018-2020年的相关承保和理赔数据保单样本共100万件，其中出险保单9248件。数据包含的基本保单信息有：被保险人性别、基本保额、所在省市、城市线、缴费年限、投保年龄、是否理赔等。

2

## 城市线指标缺失

根据《2021城市商业魅力排行榜》，将襄樊市（即襄阳市）评为三线城市，将巢湖市认定为六线城市。

3

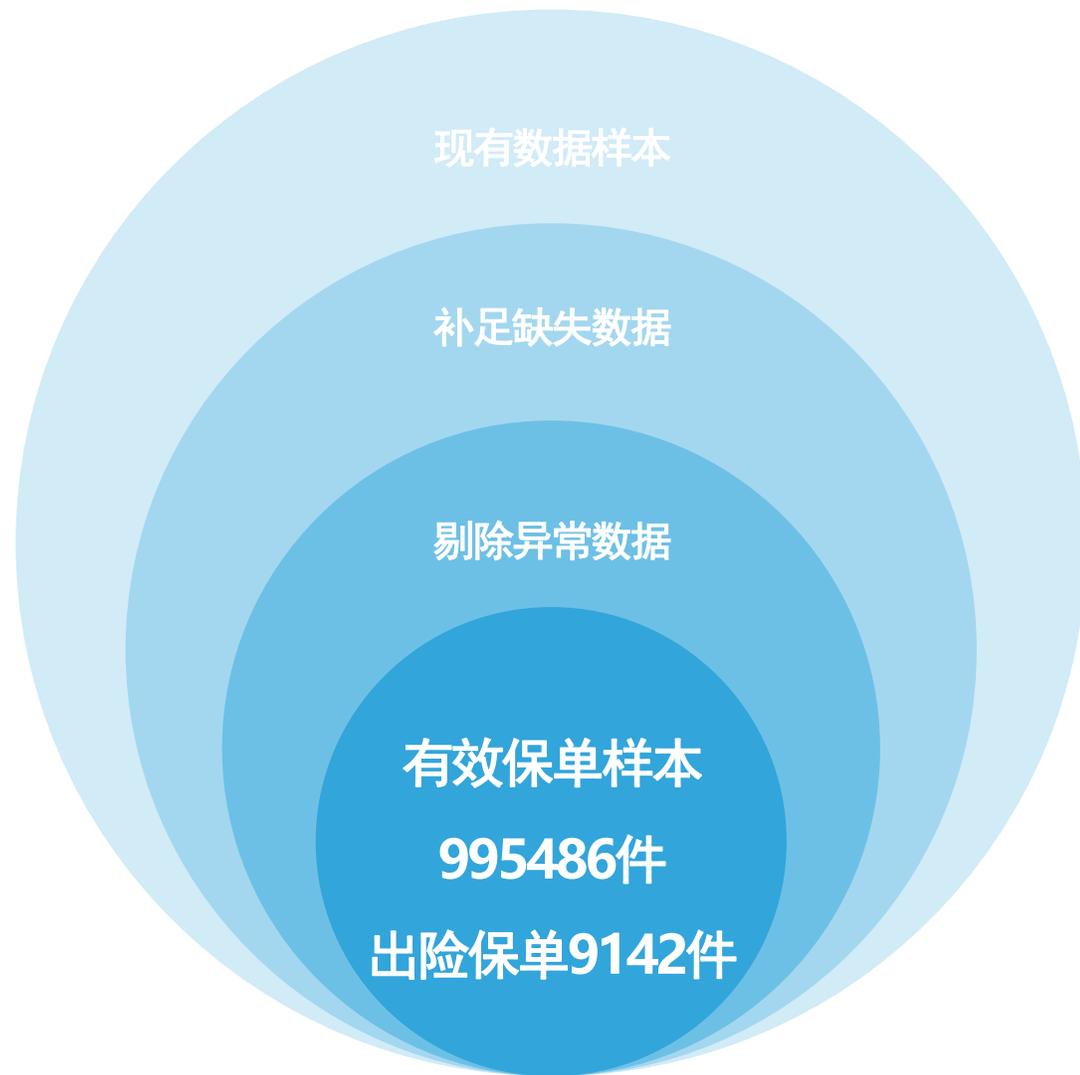
## 剔除异常数据

剔除出险日期和理赔日期间隔时间过长的出险保单样本共105件。

4

## 剔除56-60岁保单样本

该年龄段保单样本在本次研究数据中属于大龄样本，且保单数量过少，较难产生可靠估计结果。



# 指标选取

1 性别

2 年龄

3 城市线

城市线在一定程度上代表了城市的经济水平和医疗水平，可能对重疾发生率产生多种影响

4 缴费年限

缴费年限可能包含逆选择的影响

5 保险金额

有研究表明重疾险的保险金额选择中存在逆选择现象

6 地区

城市所处的地理位置、饮食习惯、气候、水文均会对重疾发生率产生影响。例如，东北地区甲状腺瘤发生率高达全国平均水平的172%

7 水资源

水资源稀缺程度会影响重疾发生率

性别

城市线

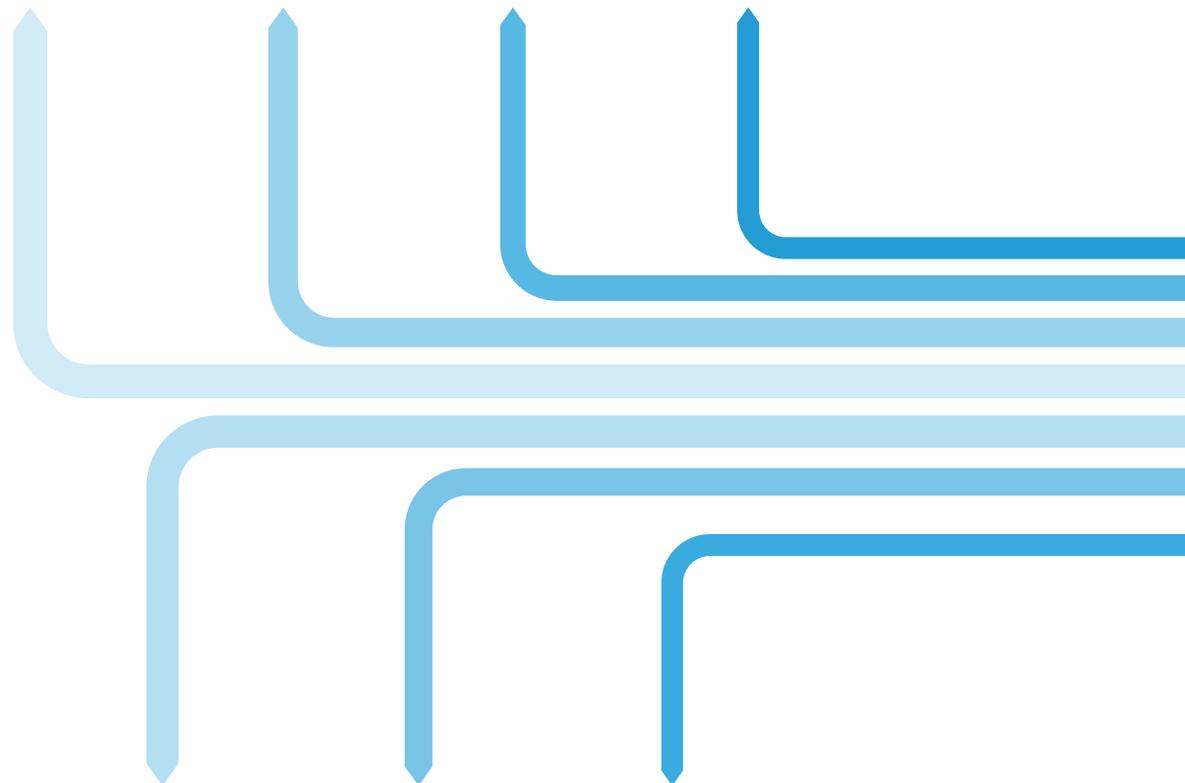
保险金额

地区

年龄

缴费年限

水资源



# 03 回归分析

# 回归分析

变量符号	说明
claim	是否理赔：理赔为 1，未理赔为 0
gender	性别：男性为 1，女性为 0
age	年龄：取年龄段中值
age2	年龄二次方项
db	东北地区 0-1 变量：属于东北地区即为 1
hb	华北地区 0-1 变量：属于华北地区即为 1
hz	华中地区 0-1 变量：属于华中地区即为 1
hd	华东地区 0-1 变量：属于华东地区即为 1
hn	华南地区 0-1 变量：属于华南地区即为 1
xb	西北地区 0-1 变量：属于西北地区即为 1
xn	西南地区 0-1 变量：属于西南地区即为 1
city_level	城市线：按城市线由高到低赋予 1-7 值
contri_year	缴费年限
water	人均水资源：按用水紧张程度递增赋予 1-5 值
water2	人均水资源二次方项
SI	保险金额：取保险金额段中值；100 万以上赋值 100 万
SI2	保险金额二次方项

- 使用Logit模型进行回归，首先假设解释变量与被解释变量之间呈线性关系：

$$\begin{aligned} \text{logit}\left(\frac{\text{claim}_i}{1 - \text{claim}_i}\right) = & \alpha \cdot \text{gender}_i + \beta \cdot \text{age}_i + \gamma \cdot \text{db}_i + \delta \cdot \text{hb}_i + \varepsilon \cdot \text{hz}_i + \epsilon \cdot \text{hd}_i \\ & + \theta \cdot \text{hn}_i + \vartheta \cdot \text{xb}_i + \mu \cdot \text{xn}_i + \pi \cdot \text{city}_{\text{level},i} + \rho \cdot \text{contri}_{\text{year},i} \\ & + \sigma \cdot \text{water}_i + \tau \cdot \text{SI}_i \end{aligned}$$

- 通过比较各个解释变量在不同模型设定下的结果显著性，发现年龄、水资源、保险金额对被解释变量的影响呈非线性，应在回归方程中加入这三个解释变量的二次项，变为：

$$\begin{aligned} \text{logit}\left(\frac{\text{claim}_i}{1 - \text{claim}_i}\right) = & \alpha \cdot \text{gender}_i + \beta_1 \cdot \text{age}_i + \beta_2 \cdot \text{age}_i^2 \\ & + \gamma \cdot \text{db}_i + \delta \cdot \text{hb}_i + \varepsilon \cdot \text{hz}_i + \epsilon \cdot \text{hd}_i + \theta \cdot \text{hn}_i + \vartheta \cdot \text{xb}_i \\ & + \mu \cdot \text{xn}_i + \pi \cdot \text{city}_{\text{level},i} + \rho \cdot \text{contri}_{\text{year},i} \\ & + \sigma_1 \cdot \text{water}_i + \sigma_2 \cdot \text{water}_i^2 + \tau_1 \cdot \text{SI}_i + \tau_2 \cdot \text{SI}_i^2 \end{aligned}$$

# 回归分析

	(1) 总回归模型
gender	-0.373*** (0.0218)
age	0.135*** (0.0110)
age2	-0.000794*** (0.000154)
db	0.447*** (0.0516)
hb	0.0645 (0.0567)
hz	0.221*** (0.0493)
hd	0.230*** (0.0471)
hn	0.0685 (0.0501)
xb	-0.0306 (0.0720)
xn	0 (.)
city_level	-0.0182 (0.00944)
contri_year	0.00766** (0.00236)
water	0.120 (0.0795)
water2	-0.0221* (0.0112)
SI	0.0174*** (0.00343)
SI2	-0.000157** (0.0000510)

Standard errors in parentheses  
\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

	(1) 性别	(2) 年龄	(3) 地区	(4) 缴费年限	(5) 保险金额	(6) 水资源	(7) 城市线
gender	-0.371*** (0.0217)						
age		0.143*** (0.0107)					
age2		-0.000948*** (0.000148)					
db			0.418*** (0.0491)				
hb			-0.0283 (0.0501)				
hz			0.0960* (0.0471)				
hd			0.192*** (0.0442)				
hn			-0.0413 (0.0466)				
xb			-0.122 (0.0668)				
xn			0 (.)				
contri_year				-0.0339*** (0.00201)			
SI					-0.00928** (0.00297)		
SI2					0.000124** (0.0000423)		
water						0.380*** (0.0771)	
water2						-0.0605*** (0.0108)	
city_level							-0.00145 (0.00710)

Standard errors in parentheses  
\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

- 女性的重疾发生概率高于男性
- 重疾发生概率随着年龄增长呈先增后减的倒U形规律
- 不同地理位置对重疾发生概率影响不同，整体上对其存在显著影响

# 回归分析

	(1) 年龄&缴费年限	(2) 年龄&保险金额	(3) 城市线&水资源
age	0.136*** (0.0109)	0.141*** (0.0107)	
age2	-0.000826*** (0.000153)	-0.000903*** (0.000148)	
contri_year	0.00709** (0.00234)		
SI		0.0140*** (0.00331)	
SI2		-0.000136** (0.0000505)	
city level			-0.0279*** (0.00816)
water			0.379*** (0.0773)
water2			-0.0626*** (0.0109)

Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

- 仅考虑缴费年限对重疾发生率的影响时，发现重疾发生概率随缴费年限增加而降低。控制年龄后，发现重疾发生概率随缴费年限增加而增加，与总回归模型结果一致
- 保险金额对重疾发生率的影响同样需要控制年龄，否则形成影响方向相反的伪回归。不同的是，保险金额对重疾发生概率的影响呈先增后减的倒U形规律
- 城市线变量与水资源变量存在相对较高的相关性（0.47），因此探究城市线的影响需要控制水资源变量以避免内生性，城市越发达，重疾发生率越高

# 04 重疾发生率估计

# 粗重疾发生率精算方法

分地区、性别、年龄、缴费年限、  
保险金额、城市线、水资源的粗重疾  
发生率 $CI$

$CI$

$factor$

$factor$ 表示风险因素

$$CI_{factor} = \frac{\theta_{factor}}{E_{factor}}$$

$E_{factor}$ 表示某风险因素下的  
风险暴露数，为精确计算出的保单  
累计生效时长，单位为年，按保额  
加权计算

$E_{factor}$

$\theta_{factor}$

$\theta_{factor}$ 表示某风险因素下的重疾  
数，按保额加权计算

# 结果分析

## 01 东北地区发生率最高

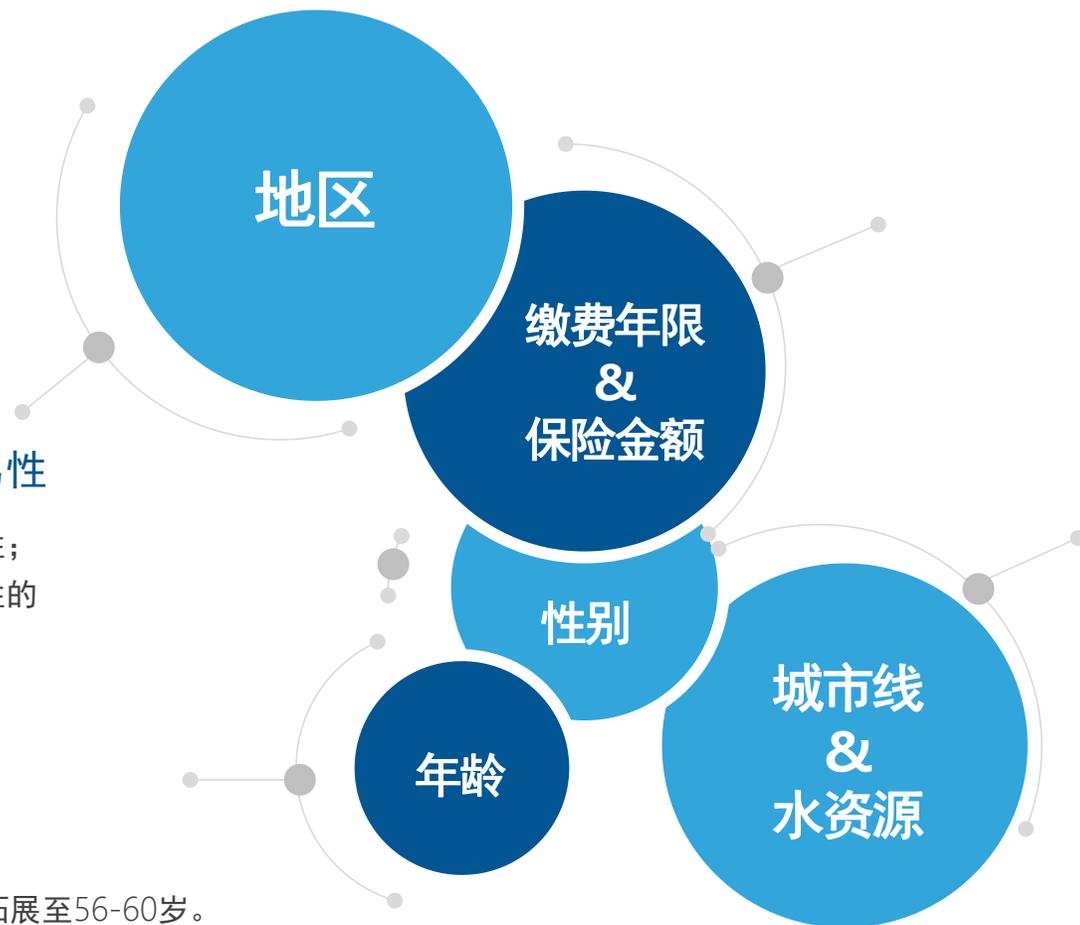
恶性肿瘤在东北地区发生率最高，为全国平均值的132%；该现象与饮食习惯和气候有关。北方人日常饮食偏厚重，盐分、油脂大，饮酒、抽烟多。

## 02 女性的重疾发生率高于男性

许多重大疾病在女性中的发生率高于男性；甲状腺癌的发病率居恶性肿瘤首位，女性的甲状腺癌发病率高于男性。

## 03 重疾发生率随年龄增长呈倒U形规律

通过绘图拟合得到二次曲线，再将年龄拓展至56-60岁。



## 04 缴费年限&保险金额

缴费年限和保险金额的影响需要控制年龄变量。使用缴费年限和保险金额对重疾发生率的边际影响来计算其风险等级。缴费年限对重疾发生率的影响呈线性，保险金额对重疾发生率的影响呈倒U形。

## 05 城市线&水资源

水资源与城市线具有一定相关性，计算各城市线下各水资源稀缺程度等级的粗重疾发生率。水资源稀缺程度越高，粗重疾发生率先增大后减少；城市越发达，粗重疾发生率越高。

# 05 风险预测工具

## 相对风险评分

基于各风险要素指标粗重疾发生率的测算结果，建立如下各风险要素指标的相对风险评分计算方法：

$$\text{相对风险评分} = \frac{\text{粗重疾发生率} - \min(\text{粗重疾发生率})}{\max(\text{粗重疾发生率}) - \min(\text{粗重疾发生率})}$$

缴费年限和保险金额两个指标对重疾发生率的影响需要控制年龄段前提，使用平均边际影响方法计算相对风险评分。

例如，缴费年限对重疾发生率的影响呈线性，平均边际影响为0.0000573，则其相对风险评分为

$$\frac{\text{缴费年限} - 10}{20}$$

保险金额对重疾发生率的影响呈倒U形，对各保额段取中间值计算对重疾发生率的总影响，基于将粗重疾发生率替换为对重疾发生率总影响的相对风险评分公式，估计各保额段的相对风险评分。

## 相对风险评分

对各风险要素指标赋予权重，使得权重总和为1，各指标相对风险评分加权可得单张保单的相对风险评分，且范围为0至1。若保单的相对风险评分计算方法完全拟合实际，相对风险评分处于前9142位的保单样本应当出险。

$$\begin{aligned} \min \quad & \sum_{i=1}^{995846} |Predict_i - Actual_i| \\ \text{s.t.} \quad & 0.1 < Weight_{gender} < 0.3 \\ & 0.1 < Weight_{region} < 0.3 \\ & 0.1 < Weight_{age} < 0.3 \\ & 0.05 < Weight_{ppp} < 0.15 \\ & 0.05 < Weight_{SI} < 0.15 \\ & 0.1 < Weight_{city\&water} < 0.3 \end{aligned}$$

指标	权重
性别	0.1103503
地区	0.2433333
年龄	0.1508800
缴费年限	0.1321599
保险金额	0.1055975
城市线&水资源	0.2576790

基于权重结果可计算得到单张保单的相对风险评分，即风险程度等级。

对所有保单样本（995846件）计算相对风险评分，可得整体人群的相对风险评分分布。单张保单在整体人群中的相对风险水平，本团队将其量化为指标：该保单的风险低于xx比例的总保单。

# 风险预测工具

## Step 1

打开“风险预测工具”，输入保单基本信息



Risk Prediction Tool-Input

输入保单基本信息

Information

年龄 Age 16-20岁

性别 Gender F

省份 Province 北京市

城市 City 北京市

缴费年限 Pay Premiums Period 11

保险金额 Sum Insured 00-05万不含

Cancel Risk Results

## Step 3

得到风险预测结果：  
个人的风险等级及风险相对水平



Risk Prediction Tool-Input

Information

Age 16-20岁

Gender F

Province 北京市

City 北京市

Pay Premiums Period 11

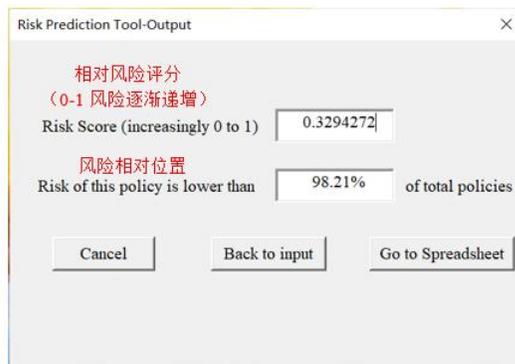
Sum Insured 00-05万不含

Cancel Risk Results

转到输出窗口

## Step 2

点击“Risk Results”，读取风险预测结果



Risk Prediction Tool-Output

相对风险评分  
(0-1 风险逐渐递增)

Risk Score (increasingly 0 to 1) 0.3294272

风险相对位置  
Risk of this policy is lower than 98.21% of total policies

Cancel Back to input Go to Spreadsheet

## Step 4

结束本次预测，或重新输入信息，或查看风险预测源数据。



Risk Prediction Tool-Output

Risk Score (increasingly 0 to 1) 0.3294272

Risk of this policy is lower than 98.21% of total policies

Cancel Back to input Go to Spreadsheet

删除并退出 返回输入界面 关闭工具并打开表格

## 风险预测工具操作流程

# 06 赔付恶化成因分析

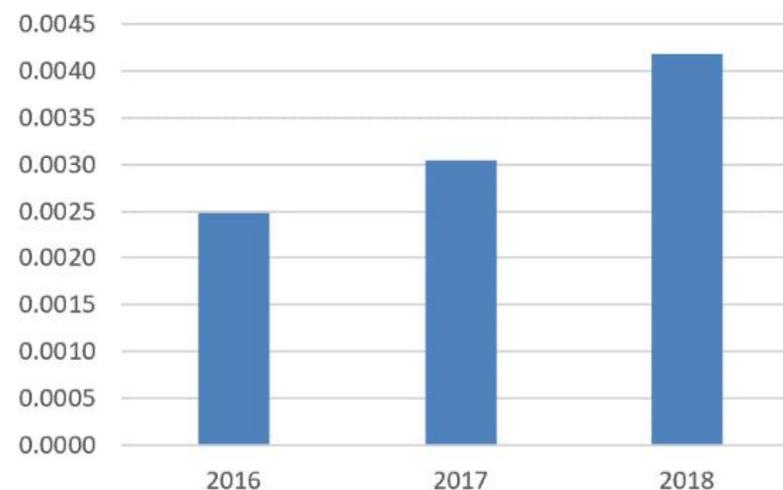
# 赔付恶化成因分析

- 将重疾发生率估计结果与《中国人身保险业重大疾病发生率表（2020）》进行对比，发现基于本公司重疾险产品赔付情况计算得到的重疾发生率高于行业重疾发生率，且差距随年龄增加递减
- 公司重疾险产品出险速度逐年上升

表 12 重疾发生率对比

年龄	行业重疾表		本公司
	男	女	
16-20 岁	[0.313,0.380]	[0.256,0.328]	0.4319
20-25 岁	[0.380,0.551]	[0.328,0.500]	1.0048
25-30 岁	[0.592,0.812]	[0.552,0.882]	1.5685
30-35 岁	[0.812,1.364]	[0.882,1.580]	2.2027
35-40 岁	[1.364,2.299]	[1.580,2.758]	3.0986
40-45 岁	[2.299,3.778]	[2.758,4.385]	4.3854
45-50 岁	[3.778,6.101]	[4.385,5.587]	5.974
50-55 岁	[6.101,9.698]	[5.587,7.205]	6.533

2016-2018投保年两年内出险率



# 赔付恶化成因分析

## 2016-2018投保年出险率年龄分布



## 重疾险出险年轻化

- 16-30岁青壮年的重疾险出险率逐年提升，26-40岁投保人的出险人数约占总出险人数的65%
- 本公司重疾发生率明显高于行业重疾发生率，但差距随年龄增加递减

## 2016-2018投保年地区业务份额分布

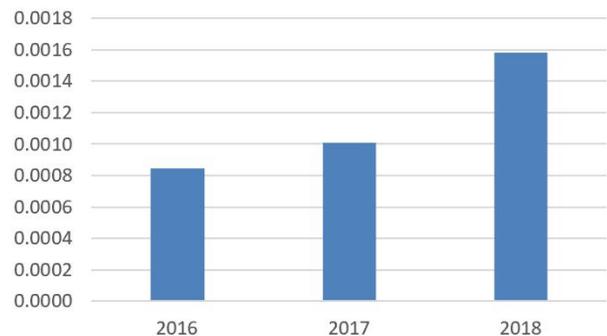


## 地区业务结构改变

- 华南地区业务占比大幅减小，其他地区除华东地区外，业务占比均体现出逐年上升的趋势，导致整体逆选择影响程度逐渐上升

# 甲状腺癌影响显著

2016-2018投保年甲状腺癌两年内出险率分布



## 出险速度提升

- 超三分之一的出险保单因甲状腺癌出险
- 甲状腺癌出险速度大幅提升，可能由于近两年科技进步等原因提升了甲状腺癌检出率
- 甲状腺疾病过度治疗和逆选择现象严重，进一步恶化检出率提升对赔付的影响

## 甲状腺癌的性别差异

- 女性投保人数量高于男性、女性投保人的甲状腺癌出险率和出险速度都高于男性这两点因素共同导致本公司重疾险产品赔付情况恶化

2016-2018投保年两年内甲状腺癌重疾出险率分性别分布



2016-2018投保年各地区两年内甲状腺癌出险率分布



## 地区业务结构改变

- 甲状腺癌筛查率和出险率低的华南地区业务份额减小、筛查率和出险率高的东北等地区业务份额增加，提升投保人整体甲状腺癌筛查率，导致本公司重疾险赔付情况恶化

# 07 结论与建议

# 结论



## 风险要素

证明性别、年龄、地区、缴费年限、保险金额、城市线、水资源七个风险要素对重疾发生率有显著影响

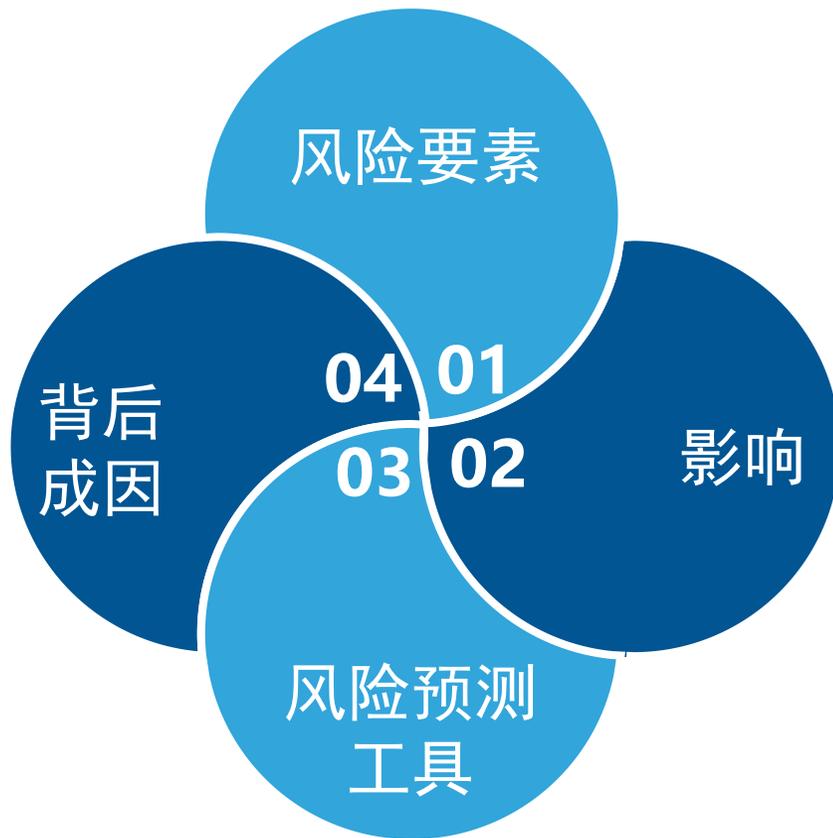


## 赔付情况恶化背后成因

主要从年龄、地区、性别、逆选择、甲状腺癌等角度出发，探讨本公司主力重疾险产品赔付情况恶化的背后成因

- 重疾险出险年轻化
- 业务结构改变，逆选择程度加深
- 甲状腺癌出险占比超三分之一且出险速度提升，可能由于科技进步、业务结构改变等原因提升了检出率，同时因其自身特点逆选择程度较高，加剧赔付情况恶化

.....



## 各风险要素的影响

女性的重疾发生率高于男性；  
 年龄对重疾发生率的影响呈先增后减的倒U形规律；  
 地区对重疾发生率影响显著，东北地区的重疾发生率最高、西北地区的重疾发生率最低；  
 控制年龄后，缴费年限对重疾发生率影响正向显著、保险金额对重疾发生率影响呈先增后减的倒U形规律；  
 控制水资源后，城市越发达，重疾发生率越高；  
 控制城市线后，水资源稀缺程度对重疾发生率的影响呈先增后减的倒U形规律



## 风险预测工具

计算相对风险评分，形成风险预测工具并窗口化，可以得到投保保单的个人风险等级和相对风险等级，预测误差率为1.66%

# 建议

## 严格承保

对投保人所在城市信息进行多维风险分析，并针对拥有风险要素中风险较高指标的投保人进行严格承保，如东北地区、水资源稀缺地区等；对青中年人加强承保阶段的健康状况审核力度；提高承保阶段对甲状腺疾病的筛查要求

## 合理调整业务结构

在地区业务结构调整时将不同地区重疾筛查率不同和逆选择程度不同的影响纳入考虑，修正地区业务结构调整策略或及时调整因业务结构改变而产生的经营假设变化

## 保单跟踪

定价和承保流程中引入现有发现的风险要素影响模式，例如风险要素的影响规律和出险年轻化趋势等；随时间变化持续追踪有效保单的赔付情况，以最新的风险预测研究结果作为经营依据，避免因出险速度上升等变化造成公司经营情况恶化

## 优化风险预测工具

例如：婚姻状况、是否有子女、血型、区分轻症和重症、对重疾发生率进行修匀和波动性调整、优化风险要素权重等





荔枝人寿  
Lychee Life  
Insurance

感谢观看！

报告团队：世纪战舰队