维普资讯 http://www.cqvip.com

论 著 文章编号:0253-3626(2007)10-1065-05

乳腺癌患者住院医疗费用的时间序列分析应用研究

邓 丹,王润华,周燕荣

(重庆医科大学公共卫生学院卫生统计与信息管理教研室,重庆 400016)

【摘 要】目的:制定乳腺癌患者住院医疗费用的最优预测模型。方法:针对乳腺癌患者的病例资料,分别拟合回归预测模型与时间序列预测模型。结果:①住院医疗费用中比例最大的为药品费,其次为材料费,并呈逐年上升趋势;②住院天数、有无癌灶转移和治疗措施是影响乳腺癌患者住院医疗费用的 3 个主要因素;③无论一元或多元,时序预测模型效果均优于回归预测模型。结论:对于预测单病种住院医疗费用,时间序列预测模型更具优势。

【关键词】乳腺癌;住院医疗费用;时间序列分析

【中国图书分类法分类号】R195.1 【文献标识码】A 【收稿日期】2007-04-30

The application study of time series analysis in breast cancer patients' hospitalized costs

DENG Dan, et al

(*Department of Health Statistics and Information Management, College of Public Health, Chongqing Medical University*)

【Abstract】*Objective*: We establish the most suitable predict model in breast cancer hospitalized costs. *Methods*: We applied a series of regression predict models and time series predict models to establish the most suitable predict model. *Results*: ①The breast cancer hospitalized costs are higher than before, the biggest proportion of which is drug costs and material costs; ②The three major influenced factors of breast cancer hospitalized costs are hospitalized time, finding cancer tissues or not and treatment measures; ③Time series predict models are better than regression predict models. *Conclusion*: The time series analysis will be more advantage to predict the hospitalized costs of single disease.

【Key Words】Breast cancer; Hospitalized costs; Time series analysis

时间序列分析[1-4](Time Series Analysis),是运用历史的观点,根据系统有限长度的运行记录,并通过量的手段建立能较精确地反映时序中所包含动态依存关系的数学模型的一种统计分析方法。运用时序分析不仅从数量上揭示某一现象的发展变化规律或从动态的角度刻画某一现象与其它现象之间的内在数量关系及其变化规律性;而且能够运用时序模型预测与控制现象的未来行为。本文旨在对重庆市近 3 年乳腺癌住院患者的病案统计资料,以科学的理论方法为指导,运用相应数理统计工具进行多种分析处理(以时间序列分析为主),探讨更为合理的统计分析方法,拟制定适合现行社会经济实际情况的单病种住院医疗费用控制模型,为卫生行政部门制定正确的卫生经济政策与科学的医疗费用管理方法,探索符合我国国情的合理的收费标准,同时也为促进基本医疗保险制度的顺利实施提供科学依据。

作者介绍:邓 丹(1977-),女,讲师,博士生,
主要研究方向:多元统计在医学中的应用。

1 资料来源与处理

1.1 资料来源

重庆市某三甲医院 1999 年 1 月 1 日至 2001 年 12 月 31 日出院结算的主要诊断为乳腺癌 (ICD-9 编码:174.9)的病例,共 350 例。收集 350 例患者的住院病历和住院医疗费用报表,并核查、补充,以确保数据的准确。

1.2 指 标

包括社会学特征(年龄、婚姻情况、生育情况、费别)、疾病特征(住院日、确诊天数、有无并发症、有无癌灶转移、治疗措施)、医院诊疗质量(诊断符合率、治疗效果)与住院费用(总费用、药品费、住院费、手术费、治疗费、检查费、材料费、其他费)19 项指标。

维普资讯 http://www.cqvip.com

1.3 资料处理

运用 EPI2.0 进行数据录入,并逻辑查错。运用 SAS8.0 统计软件包分析数据。

1.4 时序分析模型

趋势外推法、指数平滑法、Markov 模型、灰色数列模型与状态空间模型。

2 结果与分析

2.1 费用构成情况比较

可见:(1)对人均住院费用构成排序,药品费第一,其次是材料费,再次为治疗费与检查费,构成最少为住院费;(2) 药品费与材料费有明显逐年上升倾向,其中药品费总体趋势略有减少,但始终占有相当大比重(40%以上);材料费从 1999 年的 17.1% 迅速上涨到 2001 年的 27.8%,达 10 个百分点;(3) 其余费用 3 年变化不大。

2.2 线性回归分析

本次费用数据弱近似正态分布,经对数转换并未呈现更理想的线性趋势,故用原始数据分析。

2.2.1 以住院费用为结果变量,住院日为原因变量,作一元回归分析(α=0.05)

$Y=7256.45+226.47X4$ $F=76.42$

$R^2=0.1801$ $P<0.0001$

虽然回归方程成立,但 R2 显示住院日对费用的线性贡献并不大,即回归效果差。

2.2.2 以住院费用为结果变量,其他为原因变量,作多元逐步回归分析(α=0.05)

$Y=1400.24+218.21X4-987.45X7+2337.63X8$

$F=33.79$ $R^2=0.2266$ $P<0.0001$

可见,本研究中 3 个反映疾病特征的指标-住院日长短(X4)、有无癌转(X7)与采取何种治疗措施(X8)是影响乳腺癌患者住院医疗费用的主要因素。其中,随着住院天数的延长,医疗费用会大幅度上升;若患者发生癌转,医疗费用将有不同程度地上涨;采取不同的治疗措施,单纯手术最低廉,而(手术+化疗+放疗)三者合一的措施自然费用最为昂贵。同时,回归方程虽然成立,但多元线性因果模型效果也不理想。

表 1 1999~2001 3 年人均住院医疗费用(元)及相应的构成比情况

年份	例数	人均住院日	人均住院医疗费用(元)与人均费用构成(%)						
			总费用	药品费	住院费	手术费	治疗费	检查费	材料费	其他费
1999	114	21.66	8275.25	3570.41	181.96	929.75	875.33	1125.49	1415.07	177.99
			100.00	43.15	2.20	11.23	10.57	13.60	17.10	2.15
2000	147	18.61	13279.81	6022.90	200.61	1362.13	1394.28	1331.85	2775.48	189.29
			100.00	45.35	1.52	10.27	10.50	10.03	20.90	1.43
2001	89	18.21	13374.87	5961.19	196.20	844.11	1293.31	1172.80	3718.21	178.19
			100.00	44.58	1.48	6.33	9.68	8.78	27.81	1.34

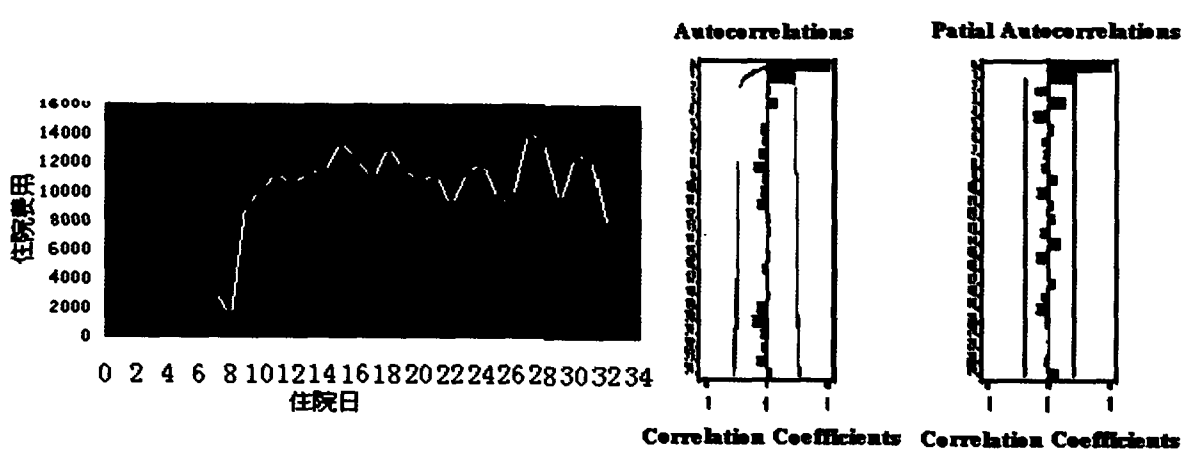


图 1 原始序列的散点图及对应的 ACF 图与 PCF 图

维普资讯 http://www.cqvip.com

2.3 时间序列分析

2.3.1 数据整理 针对时序分析对时间变量唯一性的要求,作者提出以下手段处理序列中的重复数据:(1)"同期平均值法":采用将相同住院日的患者归为一类的方法,在每一类中分别寻找欲分析变量(住院医疗费用)的数据分布与相应特征值(均数或中位数),并以此作为该时点的分析数据。(2)"邻期校正平均值法":对于患者数极少的个别时点,寻找其特征值较为困难。即可取前一时点费用和后一时点费用的平均值 $\bar{Y}_i=(Y_i-1+Y_{i+1})/2$,则该时点的替代值 $Y_i'=\bar{Y}_i+(\bar{Y}_i-Y_i)$,$Y_i$ 为该时点的费用实际值。

2.3.2 通过随机性和平稳性检验,可以认为原始序列是一个随机、基本平稳的时间序列。

2.3.3 拟合时序模型,并进行短期外推预测

多项式曲线模型:

$Y=-0.48X^4+40.46X^3-1234.40X^2+16048.00X-63264.00$ $R^2=0.7101$ X 为住院日

模型拟合效果良好;但住院 32 天的患者费用呈下降趋势,故向前预测效果与实际出入较大。

简单指数平滑模型:

$F_{t+1}=0.63678X_t+(1-0.63678)F_t$ 初值为前 3 期实际值的平均值 $\alpha=0.63678$

模型拟合效果较好,并能较理想地预测短期内患者住院医疗费用的变化。

灰色数列 GM(1,1)模型:

$Y'(t+1)=56657832.26e^{0.008t}-56654720.82$

t 为住院日

回代检验发现后验差值 C 远远大于 0.65(模型为优的标准),故拟合的灰色模型不适于外推预测。这可能因为原始序列本身存在一定的波动性,以致于该法不太容易实施。

Markov 链预测模型:

由一阶概率随机矩阵经过四步递推计算后,得到最终的转移概率矩阵:

$$P_{(1)}=\begin{bmatrix}0.1700 & 0.2780 & 0.3945 & 0.1566\\ 0.1560 & 0.2793 & 0.4015 & 0.1632\\ 0.1604 & 0.2795 & 0.3992 & 0.1609\\ 0.1388 & 0.2350 & 0.3379 & 0.1355\end{bmatrix}\xrightarrow{P_{(1)}P_{(2)}P_{(3)}P_{(4)}P_{(5)}}=$$

$$\begin{bmatrix}0.1700 & 0.2780 & 0.3954 & 0.1566\\ 0.1560 & 0.2793 & 0.4015 & 0.1632\\ 0.1604 & 0.2795 & 0.3992 & 0.1609\\ 0.1388 & 0.2350 & 0.3379 & 0.1355\end{bmatrix}$$

可见,最终的预测区间为第三状态,意味着住院日再延长 4 天的患者其平均住院医疗费用将在 11000~13000 元之间波动。同时进行回代检验,拟合效果比较满意。

2.3.4 多元时序分析-状态空间预测模型 引入影响住院医疗费用的 3 个主要因素——X4(住院日)、X7(癌灶转移)与 X8(治疗措施),和总费用 Y 共同拟合状态空间模型,并进行适用性检验和短期预测(Kalman 滤波技术):

$$\begin{bmatrix}Y_1(t+1)\\ Y_2(t+1)\\ Y_3(t+1)\\ Y_4(t+1)\end{bmatrix}=\begin{bmatrix}0.809886 & -1.920150 & 3.591007 & 0.000064\\ -0.014530 & 0.225934 & -0.128340 & 0.000022\\ 0.002108 & 0.508032 & -0.017000 & 0.000058\\ 58.884570 & 3841.703 & 998.6417 & 0.219685\end{bmatrix}$$

$$\begin{bmatrix}Y_1(t+1)\\ Y_2(t+1)\\ Y_3(t+1)\\ Y_4(t+1)\end{bmatrix}+I\begin{bmatrix}X_1(t+1)\\ X_2(t+1)\\ X_3(t+1)\\ X_4(t+1)\end{bmatrix}$$

$$\begin{bmatrix}Y'_1(t+1)\\ Y'_2(t+1)\\ Y'_3(t+1)\\ Y'_4(t+1)\end{bmatrix}=I\begin{bmatrix}Y_1t\\ Y_2t\\ Y_3t\\ Y_4t\end{bmatrix}$$

其中 $X_1(t+1)$、$X_2(t+1)$、$X_3(t+1)$和 $X_4(t+1)$分别代表 X4、X7、X8 与 Y 在输入中发挥效应。状态方程模型和输出方程模型的参数经过统计检验,大部分具有显著性意义($P<0.05$);进行回代预测,残差近似白噪声,即认为模型有效;同时作短期外推预测,效果较为理想。

2.4 线性回归分析与时间序列分析的比较

首先采用 F 检验:F=较差法的剩余均方/较优法的剩余均方 υ 为相应分析的自由度 $\alpha=0.05$

再用平均百分绝对误差(MAPE Mean Absolute Percentage Error)进行描述:

$$MPAE=\frac{\sum\left|\frac{Y_i-\hat{Y}_i}{Y_i}\times 100\right|}{n}$$

Y_i 为实测值,$\hat{Y}_i$ 为预测值,n 为样本例数

2.4.1 一元直线回归分析与一元时序分析的比较

作 F 检验发现:多项式曲线模型优于一元线性回归($F=3.75$),而指数平滑模型与一元线性回归差别不明显($F=1.45$)。但从 MAPE 值可见,一元时序分析的模型精度(分别为 22.13 和 23.07)和预测结果均好于一元线性回归分析的模型精度(44.12)和预测结果。

维普资讯 http://www.cqvip.com

2.4.2 多元逐步回归分析与多元时序分析的比较 同样地,作 F 检验发现:状态空间模型与多元逐步回归差别不大(F=1.93),但 MAPE 值仍说明,多元时间序列分析(状态空间模型)的模型精度(18.24)和预测结果好于多元逐步回归分析的模型精度(40.96)和预测结果。

表 2 一元分析结果的比较

出院日	实际值	一元线性回归 预测值	一元线性回归 残差	多项式曲线模型 预测值	多项式曲线模型 残差	指数平滑模型 预测值	指数平滑模型 残差
7	3111.44	8941.74	-5730.30	1313.60	1797.75	3553.00	-441.56
8	1394.10	9068.21	-7674.11	4871.01	-3476.91	3272.00	-1877.90
9	8917.88	9294.69	-376.81	7532.47	1385.41	2076.00	6841.88
10	10077.24	9521.16	556.08	9443.00	634.24	6433.00	3644.24
11	11223.23	9747.63	1475.60	10736.03	487.20	8754.00	2469.23
12	10431.91	9974.10	457.81	11533.48	-1101.57	10326.00	105.91
13	11306.24	10200.57	1105.67	11945.76	-639.52	10394.00	912.24
14	11214.00	10427.04	786.96	12071.76	-857.76	10975.00	239.00
15	13389.48	10653.51	2735.97	11998.88	1390.61	11127.00	2262.48
16	12189.12	10879.98	1309.14	11802.98	386.14	12568.00	-378.88
17	10788.55	11106.45	-317.90	11548.45	-759.90	12327.00	-1538.45
18	13050.01	11332.92	1717.09	11288.13	1761.88	11347.00	1703.01
19	11465.59	11559.40	-93.81	11063.37	402.22	12432.00	-966.41
20	10651.25	11785.87	-1134.62	10904.00	-252.75	11816.00	-1164.75
21	11237.54	12012.34	-774.80	10828.36	409.18	11074.00	163.54
22	8962.72	12238.81	-3276.09	10843.25	-1880.53	11178.00	-2215.28
23	11371.20	12465.28	-1094.08	10943.98	427.22	9767.00	1604.20
24	11914.86	12691.75	-776.89	11114.34	800.52	10789.00	1125.86
25	9403.60	12918.22	-3514.62	11326.63	-1923.02	11506.00	-2102.40
26	9534.16	13144.69	-3610.53	11541.60	-2007.44	10167.00	-632.84
27	13905.67	13371.16	534.51	11708.53	2197.15	9764.00	4141.67
28	13522.13	13597.63	-75.50	11765.16	1756.97	12401.00	1121.13
29	9267.33	13824.11	-4556.78	11637.74	-2370.41	13115.00	-3847.67
30	12428.30	14050.58	-1622.28	11241.00	1187.30	10665.00	1763.30
31	12347.66	14277.05	-1929.39	10478.16	1869.50	11788.00	559.66
32	7535.89	14503.52	-6967.63	9240.93	-1705.04	12144.00	-4608.11
33	*	14729.99		-		9209.65	
34	12653.21	14956.46		-		11138.67	
35	*20707.94	15182.93		-		12494.25	
36	12185.86	15409.40		-		13570.80	
37	12872.46	15635.87		-		14480.00	
MAPE			44.12		22.13		23.07

表 3 多元分析结果的比较

出院日	实际值	多元逐步回归模型 预测值	多元逐步回归模型 残差	状态空间模型 预测值	状态空间模型 残差
7	3111.44	8288.06	-5166.62	指为初始值	
8	1394.10	8496.27	-7102.17	3093.01	-1698.91
9	8917.88	8714.48	203.40	6616.33	2301.55
10	10077.24	8932.69	1144.55	8328.07	1749.17
11	11223.23	9150.90	2072.33	9640.29	1582.94
12	10431.91	9369.11	1062.80	9950.93	480.98
13	11306.24	9587.32	1718.92	9835.98	1470.26
14	11214.00	9805.53	1408.47	10086.94	1127.06
15	13389.48	10023.74	3366.01	10125.56	3263.92
16	12189.12	10241.95	1947.17	10662.36	1526.76
17	10788.55	10460.16	328.39	10457.55	331.00
18	13050.01	10678.37	2371.64	10208.75	2841.26
19	11465.59	10896.58	569.01	10764.44	701.15
20	10651.25	11114.79	-463.54	10475.25	176.00
21	11237.54	11333.00	-95.46	9356.60	1880.94
22	8962.72	11551.21	-2588.49	10542.92	-1580.20
23	11371.20	11769.42	-398.22	10102.06	1269.14
24	11914.86	11987.63	-72.77	10690.06	1224.80
25	9403.60	12205.84	-2802.24	10868.37	-1461.77
26	9534.16	12424.05	-2889.89	10375.57	-841.41
27	13905.67	12642.26	1263.41	10463.14	3442.53
28	13522.13	12860.47	661.66	11482.38	2039.75
29	9267.33	13078.68	-3811.35	11457.01	-2189.68
30	12428.30	13296.89	-868.59	10581.17	1847.13
31	12347.66	13515.10	-1167.44	7492.77	4854.89
32	7535.89	13733.31	-6197.42	7533.94	1.95
33	*	13951.52		9378.81	
34	12653.21	14169.73		10341.88	
35	*20707.94	14387.94		10612.34	
36	12185.86	14606.15		10730.64	
37	12872.46	14824.36		10815.51	
MAPE			40.96		18.24

3 讨 论

3.1 统计分析方法的比较

徐国祥[5]在“体制转轨时期宏观经济分析和预测的方法研究”报告中曾指出——如果研究者选择预测方法的标准是追求预测精度极大化,则最好选择时间序列预测模型,至少短期预测可以如此。作者同时运用一元线性回归预测法、多元逐步线性回归预测法、多项式曲线预测模型、指数平滑预测模型、灰色数列预测模型、Markov 预测模型和状态空间预测模型七种方法对费用时序进行研究,得到了相似的结论:不论针对一元变量还是多元指标进行分析,时间序列分析拟合的模型无论是其适用性检验,还是短期外推预测,都比相应的线性回归分析结果具有更高的精确度。这是由于当费用的影响因素之间存在错综复杂的相关时,线性回归这种结构式的因果模型就无法很好地体现其优越性。而时序分析则可根据所观察的住院医疗费用序列在住院日为 7~32 天有限长度内所记录下的动态数据,定量地挖掘出该段时间内费用的演变规律,并预测费用短期未来的发展趋势。明显地,对费用序列建立并检验合格的时间序列模型具有更为实际的可行性。

但是,每一种时序分析都有各自的优缺点,在运用时应予以考虑。其中,状态空间模型的精度(MAPE 为 18.24)远远高于其他任何一种分析。可见,状态空间模型能较完整地反映费用时序系统的动态性、相关性、系统内部状态以及内部状态与外部的输入输出变量的联系,而且还能将变量时序转变为向量时序,处理多输入多输出变量情况下的建模。这些都是其他模型不能比拟的。提示实际工作中,应尝试将该模型应用于各种各样的时间序列,特别是费用时序,以拟合适合于不同状态下的模型,从而更好地起到对卫生医疗管理工作,尤其是

维普资讯 http://www.cqvip.com

医疗费用的指导、管理和制约作用。需要注意的是，建立状态空间模型的前提是动态系统符合 Markov 特性，若观察时序不具备该性质，则无法有效地应用此法。

3.2 资料问题与处理

针对原始时序出现的重复现象，作者提出“同期平均值法”与“邻期校正平均值法”处理数据并用于分析。发现模型拟合效果和预测精度与预想的相差不大，故不失为两种可以尝试并改进的数据预处理方法。

3.3 影响乳腺癌患者住院医疗费用的主要因素

住院天数、治疗措施和有无癌转是影响乳腺癌患者住院医疗费用的主要因素。其中，住院天数是影响医疗费用的首要因素。明显地，住院日与住院医疗费用有直接的正比关系，缩短住院日正是控制费用的有效措施，而且大有作为；同时，缩短住院日会提高有效住院天数，还对加强医院的医疗质量管理和在新形式下参与医疗市场竞争有着至关重要的作用。

3.4 管理对策

住院医疗费用的有效降低和合理控制是符合广大人民共同利益的，也是医疗卫生服务领域的大势所趋。随着中国加入 WTO 以及医疗保险制度的逐步出台，在社会统筹资金不很充足的条件下，医院要不断加强自身建设，以提高服务质量和服务水平为目的，以加强医护人员的业务能力和职业道德为手段，最终使医院走上可持续发展的道路，实现经济效益与社会效益的同步增长；医院要切实做到办院理念的转变，切实做到以病人为本，时时处处为病人着想，让病人花最少的钱，得到最优质的医疗服务。各级卫生行政管理部门，应在调查分析的基础上制定单病种的费用标准及相关配套政策，以便更有效地对医院进行引导、管理和规范。

4 有待深入研究与解决的问题

本文仅仅运用时序分析对病种住院医疗费用进行了一个初步的分析与探讨；无论从纵深面，还是从水平面，均有待继续研究。

参 考 文 献

[1] Box G.E.P,Jenkins G.M,Reinsel G.C. Time Series Analysis: Forecasting and Control[M]. 3rd Edition,Prentice-Hall,Inc,1994.

[2] Box G.E.P,Jenkins G.M,Reinsel G.C.时间序列分析、预测与控制顾岚主译[M]. 中国统计出版社,1994.

[3] 王振龙,顾 岚. 时间序列分析[M]. 中国统计出版社,2000.

[4] 方积乾,陆 盈. 现代医学统计学[M]. 第一版,人民卫生出版社,2002.5.

[5] 徐国祥.体制转轨时期宏观经济分析和预测的方法研究[M]. 上海财经大学出版社,1999.2.

(上接第 1015 页)

其原因可能是乳腺癌的多种抑癌基因突变产生了许多含 PTC 的 mRNA，从而加强了 NMD 的降解，使其执行其功能的抑癌蛋白减少，最终导致肿瘤的发生。Y14 和 Upf1 直接反映 NMD 的功能状况。提示 NMD 机制的作用强弱可能与人乳腺细胞的恶性行为、侵袭能力等有关。

参 考 文 献

[1] Weischenfeldt-J,Lykke-Andersen-J,Porse-B. Messenger RNA surveillance:neutralizing natural nonsense[J]. Curr-Biol,2005;15(14):R559-562.

[2] Michael R,Culbertson,Eric Neeno-Eckwall.Transcript selection and the recruitment of mRNA decay factors for NMD in Saccharomyces cerevisiae[J]. RNA,2005;11:1333-1339.

[3] 冉 亮,任国胜. 无意义密码子介导的 mRNA 降解(NMD)机制的研究进展[J]. 重庆医学,2005;34(12):1895-1897.

[4] Gehring-N-H,Kunz-J-B,Neu-Yilik-G,et al. Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements[J]. Mol-Cell,2005;20(1):65-75.

[5] Kashima-I,Yamashita-A,Izumi-N,et al. Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay[J]. Genes-Dev,2006;20(3):355-367.

[6] Kunz-J-B,Neu-Yilik-G,Hentze-M-W,et al. Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation[J]. RNA,2006;12(6):1015-1022.

[7] Sheila V,GRAHAM. Nonsense-mediated decay breaks the circle [J]? Biochem J,2003;373.

[8] V.Narry Kim,Jeongsik Yong.The Y14 protein communicates to the cytoplasm the position of exon-exon junctions[J]. EMBO,2001;20:2062-2068.