· 流行病学研究 ·

丙肝疫情 3 种不同疾病预测预警方法比较*

姜超[1],刘文东[2],胡建利[2],朱叶飞[2],鲍昌俊[2],汤奋扬[2],彭志行[1],陈峰[1]

摘　要:目的　分析 Markov 链模型、Holt-Winters 指数平滑模型、控制图预警模型 3 种预测预警方法对江苏省丙型肝炎(丙肝)疫情的预测效果,为丙肝的预防控制提供科学依据。**方法**　收集江苏省传染病疫情管理系统 1997—2009 年法定报告传染病疫情中丙肝的发病资料,分别建立 Markov 链模型、Holt-Winters 指数平滑模型以及控制图预警模型,比较 3 种方法对丙肝预测的效果。**结果**　Markov 链模型预测江苏省 2009 年 9—12 月丙肝的发病数均 >120 例,与实际发病例数的 164、125、145、138 例相符;Holt-Winters 季节迭加模型($\alpha = 0.46, \beta = 0.001, \gamma = 0.001$)预测效果较好,RMSE 和 MAPE 分别为 12.629 85 和 19.18%;控制图 5 年预警模型可以用于丙肝疫情的预警,将 P_{70} 作为预警界值,灵敏度、特异度均 >90%。**结论**　Markov 链模型、Holt-Winters 指数平滑模型是丙肝预测的良好方法,控制图预警模型可用于丙肝疫情的预警。

关键词:预测;丙型肝炎(丙肝);Markov 链模型;Holt-Winters 指数平滑模型;控制预警模型
中图分类号:R 512.6+3　**文献标志码:**A　**文章编号:**1001-0580(2015)04-0390-04　DOI:10.11847/zgggws2015-31-04-02

Prediction and early warning for HCV: comparison of three methods

JIANG Chao*, LIU Wen-dong, HU Jian-li, et al (*Department of Epidemiology and Health Statistics, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu Province 211166, China)

Abstract: Objective　To apply and compare different models in prediction of the trend of hepatitis C virus (HCV) epidemic in Jiangsu province and to provide bases for the control of HCV epidemic. **Methods**　The data on infectious diseases report in Jiangsu province form 1997 to 2009 were collected. Markov chain model and Hole-Winters exponential smoothing forecasting model were used to predict epidemics and control chart was used to detect outbreaks or epidemics of HCV infection. **Results**　Markov chain model predicted that the incidence of HCV from September to December in 2009 would be more than 120 cases and the prediction was consistent with the case number reported. Hole-Winters-additive model ($\alpha = 0.46, \beta = 0.001, \gamma = 0.001$) had good performance in predicting HCV epidemic, with the root mean square error (RMSE) and mean absolute percentage error (MAPE) of 12.6 and 19.1%, respectively. A five-year control chart was chosen as an essential method in early warning system. The early warning threshold for HCV epidemic was 70 percentile and at the threshold level, the sensitivity and specificity for HCV epidemic prediction were over 90%. **Conclusion**　Markov chain model and Hole-Winters exponential smoothing forecasting model are good method for prediction of HCV epidemic and control chart can be used in the early warning.

Key words: prediction; HCV; Markov chain; exponential smoothing; control chart

丙型肝炎(丙肝)是由丙型肝炎病毒(hepatitis C virus, HCV)感染所致的一种慢性肝脏疾病。HCV 是一种折叠的单链核糖核酸病毒,属肝炎病毒黄病毒科家族[1]。自 1989 年被发现以来,就在世界范围内被认为是导致肝纤维化、肝硬化以及肝细胞癌的主要原因[2-5]。据报道,全球大约 >3%(约 1.7 亿)的人口感染了 HCV[6]。在中国也存在着丙肝疫情暴发的各种隐患。为此,尽早发现传染病暴发的迹象,及时采取有效控制措施,可有效降低丙肝造成的发病和死亡。为分析不同预测预警方法对江苏省丙肝疫情的预测效果,为丙肝的预防控制提供科学依据,本研究收集江苏省传染病疫情管理系统 1997—2009 年法定报告传染病疫情中丙肝的发病资料,分别建立 Markov 链模型、Holt-Winters 指数平滑模型以及控制图预警模型,比较 3 种方法对丙肝预测的效果。结果报告如下。

1　资料与方法

1.1　资料来源　资料来源于江苏省传染病疫情管理系统 1997 年 1 月—2009 年 12 月法定报告传染病疫情中丙肝的发病资料。

1.2　方法

1.2.1　Markov 随机过程方法　当随机过程在时刻 t_{n-1} 状态已知的条件下,其在时刻 t_n 所处的状态仅

* **基金项目:** 江苏省自然科学基金(BK2010079);国家科技重大专项(2011ZX10004-902);江苏省优势学科建设项目;江苏省科教兴卫工程(ZX201109)
作者单位: 1. 南京医科大学公共卫生学院流行病与卫生统计学系,江苏 南京 211166;2. 江苏省疾病预防控制中心
作者简介: 姜超(1989-),男,江苏南京人,硕士在读,研究方向:传染病预测预警。
通讯作者: 陈峰,E-mail:fengchen@njmu.edu.cn
数字出版日期:2015-3-9 14:33
数字出版网址:http://www.cnki.net/kcms/detail/21.1234.R.20150309.1433.005.html

与时刻 t_{n-1} 的状态有关，这样的过程称为 Markov 过程。有限个 Markov 过程的整体称为 Markov 链，它是由状态转移概率联系起来的一个个状态所组成的“链条”，通过研究链内有限 Markov 过程的状态及相互关系进而预测链内的未来状况，据此作出决策。疾病的发生的走势属于随机过程，根据疾病资料的取值范围可以划分状态，通过计算状态的频数可以求得各个状态的初始概率，然后再求各状态相互转移出现的频数，确定一阶转移概率矩阵。按照式 $p_n = p_1^n$，可求 n 阶转移概率矩阵，即是预测值状态间的各阶概率矩阵，按照最大转移概率原理即可作出预测。

1.2.2　指数平滑法　指数平滑预测方法通过给近期的观察值以较大的权重，给远期的实际值以较小的权数，使预测值既反映最新的信息，又能反映历史资料的信息，从而使预测结果更符合实际。指数平滑法可以分为一次指数平滑、二次指数平滑以及多参数指数平滑[7]。本研究采用的为 Holt-Winter 季节迭加模型，该方法用于序列具有线性趋势和加法季节变化的序列。y_t 平滑后的序列 $\hat{y}_t$ 由下式计算：$\hat{y}_{t+k} = a_t + b_t k + S_{t+k}, t = s+1, s+2, \cdots, T$，其中 a_i 表示截距，b_t 表示斜率，$a_t + b_t k$ 表示趋势，S_t 为加法模型的季节因子，s 表示季节周期长度，月度数据 $s = 12$，季度数据 $s = 4$。这 3 个系数有下面的递归式定义：$a_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(a_{t-1} + b_{t-1})$；$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{i-1}$；$S_t = \gamma(y_t - a_t) + (1-\gamma)S_{t-s}$。其中 $k > 0$，α、β、γ 在 0 ~ 1，预测值由下式计算：$y_{T+k} = a_T + b_T k + S_{T+k-s}$，其中 S_{T+k-s} 用样本数据最后 1 年的季节因子。其他模型具体公式参照文献[8]。

1.2.3　控制图法　按照控制图法预警模型原理，参考国内外传染病预警的一些方法建立控制图法 3 年模型和控制图法 5 年模型 2 种丙肝流行的预警模型。以研究疾病既往 3 或 5 年该月的月发病数作为基线数据，计算指定的百分位数（P_{50}、P_{60}、P_{70}、P_{80}、P_{95}）作为候选预警界值，建立预警模型，计算不同预警界值时预警模型的敏感性和特异性，并借助 ROC 曲线优选出最合适的预警模型及其预警界值。

1.3　统计分析　采用 Excel 2003 进行数据录入与整理；应用 SAS 9.2 求转移概率矩阵，采用 Eviews 5 进行时间序列的分析，应用 SPSS 13.0 进行统计分析并绘制 ROC 曲线。

2　结　果

2.1　Markov 链模型丙肝疫情预测（图 1、表 1）　根据 1997—2008 年江苏省丙肝发病的时间顺序及发病数大小，按照所规定的发病数取值范围（状态 1：$X \leqslant 25$；状态 2：$25 < X \leqslant 50$；状态 3：$50 < X \leqslant 120$；状态 4：$X > 120$）确定各年度月发病数所处的状态。各年度月发病数及其趋势见图 1。按照各月发病数所处的状态，计算出各状态相互转移出现的次数，建立一阶转移概率矩阵。应用 SAS 9.2 求出各阶转移概率矩阵：

$$P_1 = \begin{pmatrix} 0.67 & 0.25 & 0.08 & 0.00 \\ 0.29 & 0.67 & 0.05 & 0.00 \\ 0.00 & 0.12 & 0.62 & 0.26 \\ 0.00 & 0.00 & 0.19 & 0.81 \end{pmatrix}$$

$$P_2 = \begin{pmatrix} 0.52 & 0.34 & 0.12 & 0.02 \\ 0.38 & 0.52 & 0.08 & 0.01 \\ 0.03 & 0.15 & 0.44 & 0.38 \\ 0.00 & 0.02 & 0.27 & 0.71 \end{pmatrix}$$

$$P_3 = \begin{pmatrix} 0.44 & 0.37 & 0.14 & 0.05 \\ 0.40 & 0.45 & 0.11 & 0.03 \\ 0.07 & 0.16 & 0.35 & 0.42 \\ 0.01 & 0.05 & 0.30 & 0.65 \end{pmatrix}$$

$$P_4 = \begin{pmatrix} 0.40 & 0.37 & 0.15 & 0.08 \\ 0.40 & 0.42 & 0.13 & 0.06 \\ 0.09 & 0.16 & 0.31 & 0.44 \\ 0.02 & 0.07 & 0.31 & 0.61 \end{pmatrix}$$

2009 年 8 月的发病数为 156 例，处于状态 4，在一阶转移概率矩阵中处于第 4 行，该行以第 4 列的转移概率（$P = 0.81$）最大，表示状态 4 向状态 4 转移，故可认为 2009 年 9 月的丙肝发病数处于第 4 状态，其发病数 > 120 例，$P_2 \sim P_4$ 阶转移概率矩阵中，最大概率均在第 4 列，分别为 0.71、0.65、0.61。据此可推断未来 4 个月丙肝的发病数均超 > 120 例，与 2009 年 9—12 月实际月发病数的 164、125、145、138 例相一致。

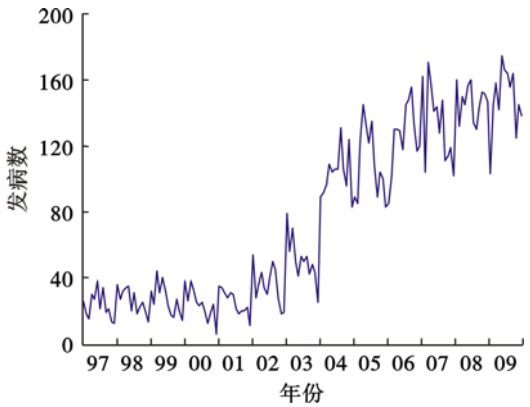

图 1　江苏省 1997—2009 年丙肝月发病趋势图

表 1　1997—2009 年丙肝各月状态转移次数

状态	1	2	3	4
1	24	9	3	0
2	12	28	2	0
3	0	4	21	9
4	0	0	8	35

2.2　转移概率矩阵(表 2、3，图 2)　江苏省 1997—2009 年丙肝发生病例数总体呈现上升趋势，并且伴随有季节周期性波动，因此选择 Holt-Winter 季节模型进行拟合和预测。在建模过程中参数值的确定是通过不同参数值的组合进行建模，在参数的选择过程中，要使 RMSE 最小的模型为最优模型，然后利用最优的参数值组合，通过公式递推出时间序列各时间点的预测值，使拟合和预测达到最佳效果。除此之外 MAPE 也作为模型拟合效果的评价参考标准。参数设置和最优参数设置如表 2 所示，可见 Holt-Winter 季节迭加模型的 RMSE 和 MAPE 2 个指标均低于其他模型，拟合效果较好，因此选择 Holt-Winter 季节迭加模型对数据进行拟合，将 1997—2008 年做为训练数据，拟合江苏省实际丙肝发病数，并对 2009 年各月丙肝发病数做出预测，拟合及预测效果见表 3。Holt-Winter 季节迭加模型对 2009 年各月丙肝发病数的预测相对误差整体较小，除 1 月的误差略大外，其余各月误差均在较低水平。从宏观分析，该模型对于第 2、3 季度的预测相对误差较小，效果较好，对于第 1、4 季度的预测误差较大、效果略差。从图 2 可见，Holt-Winter 季节迭加模型预测的丙肝发病的趋势与原序列十分接近。

表 2　指数平滑法参数设置和拟合评估

方法	α	β	γ	RMSE	MAPE
一次指数平滑	0.48			16.340 39	24.13%
二次指数平滑	0.14			16.943 21	26.62%
Holt-Winter 非季节模型	0.45	0.001		16.234 87	25.74%
Holt-Winter 季节乘积模型	0.14	0.001	0.001	19.839 59	25.35%
Holt-Winter 季节迭加模型	0.46	0.001	0.001	12.629 85	19.18%

表 3　Holt-Winter 季节迭加模型拟合江苏省 2009 年丙肝月发病数

月份	发病数(例)	预测数(例)	绝对误差	相对预测误差
1	103	173	70	40.46%
2	145	159	14	8.81%
3	158	177	19	10.73%
4	142	177	35	19.77%
5	175	173	2	1.16%
6	166	172	6	3.49%
7	164	170	6	3.53%
8	156	172	16	9.30%
9	164	164	0	0.00%
10	125	164	39	23.78%
11	145	163	18	11.04%
12	138	152	14	9.21%

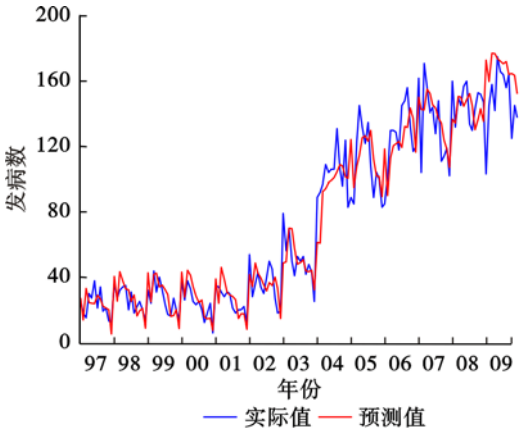

图 2　Holt-Winter 季节相加模型拟合序列与原序列比较

2.3　江苏省 2009 年丙肝月预警界值计算(图 3)　传染病“流行”系指该疾病的发病水平超过历年一般水平，实际应用时需要对其进行量化。本研究以近 5 年非流行年发病数为基础，计算丙肝月平均发病数作为基线值，并计算各平均数的 s，以作为预警界值判断疾病的流行。将计算所得的基线值代入模型，用各候选预警界值(P_{50}、P_{60}、P_{70}、P_{80}、P_{95})对 2009 年江苏省丙肝传染病疫情分月进行预警，并借助 ROC 曲线优选出最合适的预警模型和预警界值。由图 3 可见，选用 P_{70}作为预警界值时预警模型的灵敏度和特异度均较好。

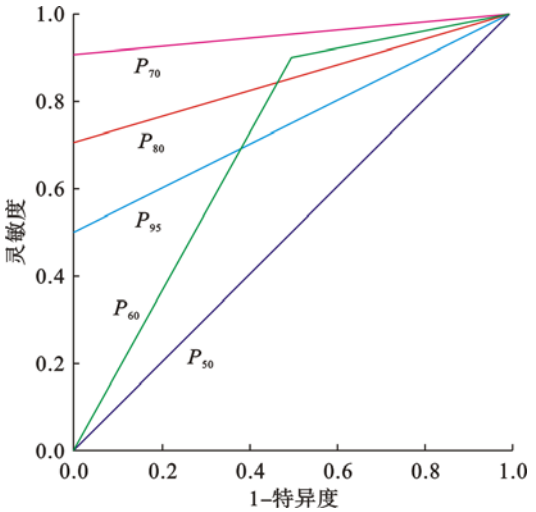

图 3　不同预警界值下的 ROC 曲线

3　讨　论

丙肝发病情况的预测预报研究是一项意义深远的工作，也是流行病学研究的重要课题之一。世界卫生组织 2010 年 5 月在一次专家共识研讨会上将 HCV 的慢性感染作为头等重要的公共卫生问题，所有的医疗当局都被邀请加入到努力预防、识别和快速治疗感染的行动中来[9]。本研究将丙肝的月发

病数划分为 4 个状态,利用 Markov 链方法,根据所分状态,预测 2009 年 9—12 月丙肝的发病数均 > 120 例,与实际发病数的 164、125、145、138 例相符。Markov 链预测是区间预测,虽然降低了预测的精确度,但却提高了预测的准确度,在含有未知因素和随机性的情况下预测发病数实际值的范围,对防病更具有现实意义。为跟踪数据的发展,以便于直观疾病的发展趋势,本研究同时采用了指数平滑法进行预测,通过观察数据可以发现江苏省丙肝发生病数总体呈现上升趋势,且伴有季节周期性波动,因此选择 Holt-Winter 季节模型进行拟合和预测。通过比较发现,Holt-Winter 季节迭加模型($\alpha = 0.46, \beta = 0.001, \gamma = 0.001$)预测效果较好,RMSE 和 MAPE 分别为 12.629 85 和 19.18%。指数平滑法在进行预测时可以得到每个月具体的预测发病数,提高了预测的精度。在对疾病进行预警时,本文采用了 5 年控制图法,通过 ROC 曲线筛选,发现将 P_{70} 作为预警界值时预警模型的灵敏度和特异度均 >90%。该方法简单易行,可避免不断对参数进行调整的弊端。

由于在发展中国家 HCV 感染的流行病学并不能很好的描述其特性。即便在建立很好疾病监测系统的国家,急性病报告系统也会低估 HCV 感染的发病率[10]。虽然直接测量 HCV 感染的发病率十分困难,但是仍然可以依赖数学模型推断发病的趋势[7]。本研究直接用发病数的预测来了解未来丙肝的发病情况,将区间预测与点预测结合,对疾病的发展趋势有更全面的了解,为防控决策提高更可靠更准确的依据。

参考文献

[1] 许可,汤奋杨. 血液透析及丙肝病毒感染者 F 蛋白抗体分析[J]. 中国公共卫生,2012,28(3):333 – 335.

[2] Thomas HC, Lemon SM, Zuckerman AJ. Viral hepatitis[M]. 3rd ed, Hoboken: Blackwell Publishing, 2005: 824 – 840.

[3] Knipe DM, Howley PM. Field virology[M]. 5th ed, Philadelphia: Lippincott Williams and Wilkins Immunology, 2007: 3 – 4.

[4] Kanwal F, Hoang T, Kramer JR, et al. Increasing prevalence of HCC and cirrhosis in patients with chronic hepatitis C virus infection[J]. Gastroenterology, 2011, 140(4): 1182 – 1188.

[5] 孔晶,储春丽. 丙肝病毒 F 基因对 c-myc 转录调控作用[J]. 中国公共卫生,2010,26(12):1508 – 1510.

[6] Hajarizadeh B, Grebely J, DoreG J. Epidemiology and natural history of HCV infection[J]. Nat Rev Gastroenterol and Heptology, 2013, 10(1038): 553 – 562.

[7] Shepard CW. Global epidemiology of hepatitis C virus infection[J]. The Lancet Infectious Disease, 2005, 5(9): 558 – 567.

[8] 孙敬水. 计量经济学[M]. 清华大学出版社,2004: 365 – 370.

[9] World Health Organization. Global alert and response: hepatitis C[EB/OL]. [2013 – 01 – 14]. http://www.who.int/csr/disease/hepatitis/whocdscsrlyo2003/en/index1.html.

[10] Hagan H, Snyder N, Hough E, et al. Case-reporting of acute hepatitis B and C among injection drug users[J]. J Urban Health, 2002, 79(4): 579 – 585.

收稿日期: 2014-02-24　（郭薇编校）

· 流行病学研究 ·

医生工作付出 – 回报失衡现状及影响因素*

孙葵,尹文强,黄冬梅,于倩倩,赵延奎,郭洪伟,胡金伟

摘　要: 目的　对不同人口学特征的医生工作付出 – 回报失衡感进行比较,以发现工作付出 – 回报失衡重点人群,并探讨医生工作付出 – 回报失衡的影响因素。**方法**　采用多阶段分层抽样方法抽取山东省济南、潍坊、德州 3 个市的 1 910 名医生,利用课题组编制的公立医院医生工作感受量表进行调查,并对数据进行描述性分析、方差分析和单因素非条件 logistic 回归分析。**结果**　55.7% 的医生自感处于工作付出 – 回报中高度失衡状态;学历、职称、是否为临床教师和月均收入对医生工作付出 – 回报失衡感均有影响($P < 0.05$);其中,学历方面:学历越高失衡感越强,失衡感最强的是博士学历者[(1.79 ± 0.75)分];职称方面:失衡感最低的是未定职称者[(1.40 ± 0.73)分],失衡感最强的是住院医师[(1.57 ± 0.79)分]和主治医师[(1.57 ± 0.74)分];临床教师[(1.66 ± 0.77)分]失衡感高于非临床教师[(1.52 ± 0.74)分];收入方面:随着收入水平升高,医生工作付出 – 回报失衡感递减,其中以月均收入≤2 000 元者[(1.61 ± 0.87)分]失衡感最强,月均收入≥5 001 元者[(1.40 ± 0.56)分]失衡感最低;对医生工作付出 – 回报失衡影响最大的前 3 位因素为职业风险($\beta = 0.22$)、学习压力($\beta = 0.18$)和精神付出($\beta = 0.17$)。**结论**　应重点关注高学历、低职称、低收入等特征医生的工作付出 – 回报失衡现象。

关键词:人口学变量;医生;付出 – 回报失衡;影响因素

中图分类号:R 192　**文献标志码**:A　**文章编号**:1001-0580(2015)04-0393-04　DOI:10.11847/zgggws2015-31-04-03

* **基金项目**: 国家自然科学基金(71103129);教育部人文社会科学研究项目(09YJA630114);山东省教育厅资助项目(J11WG24)

作者单位: 潍坊医学院管理学院,山东 潍坊 261053

作者简介: 孙葵(1982 –),女,山东寿光人,讲师,硕士,研究方向:卫生人力资源管理。

通讯作者: 尹文强,E-mail: yinwq@wfmc.edu.cn

数字出版日期:2015 – 3 – 9 14:39

数字出版网址:http://www.cnki.net/kcms/detail/21.1234.R.20150309.1439.009.html