

组合模型在肺癌死亡率预测中的构建和应用*

曲红梅¹, 白亚娜^{2△}

1. 西北民族大学医学院 预防医学教研室(兰州 730030); 2. 兰州大学公共卫生学院 流行病学与统计学教研室(兰州 730000)

【摘要】 目的 用3种预测方法对肺癌死亡率进行拟合,选择最优方法,建立组合模型并进行预测。方法 用动态数列、joinpoint 回归、指数平滑对2001~2013年金昌队列肺癌死亡率进行拟合,选择单项模型形成组合模型进行拟合并预测。组合模型权重系数的计算基于算术平均法、方差倒数法、均方误差倒数法、简单加权平均法。结果 单项预测模型以指数平滑法拟合精度最高,为79.67%,Joinpoint 线性回归拟合精度为74.27%。指数平滑与Joinpoint 线性回归进行组合,其中以算术平均法及均方误差倒数法预测效果最好,拟合精度分别为86.87%、85.80%。结论 组合模型优于单项预测法,可以用于肺癌死亡率预测。

【关键词】 肺癌死亡率 指数平滑 Joinpoint 线性回归 组合模型

Establishing a Combination Model in Predicting Mortality of Lung Cancer QU Hong-mei¹, BAI Ya-na^{2△}. 1. Preventive Medicine Teaching and Research Section, Medical College, Northwest University for Nationalities, Lanzhou 730030, China; 2. Teaching and Research Section of Epidemiology and Biostatistics, School of Public Health, Lanzhou University, Lanzhou 730000, China

△ Corresponding author, E-mail: baiyana@lzu.edu.cn

【Abstract】 Objective To identify a good combination model for predicting the mortality of lung cancer. **Methods** Mortality data of lung cancer from 2001-2013 were used to test three prediction model: dynamic series, exponential smoothing, and Joinpoint regression. Weight coefficients of the combination models were calculated using the arithmetic average method, the variance inverse method, the mean square error inverse method, and the simple weighted average method. **Results** The exponential smoothing model had the highest accuracy (79.67%) of prediction, followed by the Joinpoint linear model (74.27%). The combination of these two models resulted in better results. The arithmetic average method and the mean square error inverse method had the best prediction, with an accuracy of 86.87% and 85.80%, respectively. **Conclusion** The combined model has higher accuracy than the single models in predicting the mortality of lung cancer.

【Key words】 Lung cancer mortality Exponential smoothing Joinpoint linear regression
Combination model

在金昌队列癌症死因中,肺癌居于死因之首^[1]。为了解肺癌未来发展趋势,本研究利用3种单项预测方法对2001~2013年金昌队列肺癌死亡率进行拟合,遴选单项模型建立组合模型并对肺癌死亡率进行趋势预测。在以往疾病死亡预测中多采用单一预测方法,本研究应用组合模型对队列人群肺癌死亡率未来趋势做出预测,最终为队列人群制定针对性的肺癌防治策略提供依据和方向。

1 资料及方法

1.1 资料来源

应用金昌队列2001~2013年肺癌死亡率数据建立死亡率预测模型,利用模型对2014~2018年肺

癌死亡率进行预测。

1.2 预测方法与模型评价

1.2.1 单项预测模型的遴选及评价 通过动态数列、指数平滑、Joinpoint 回归模型进行肺癌死亡率的拟合并评价。模型评价指标为均方误差(MSE)、平均绝对误差(MAE)、平均绝对百分误差(MAPE)和拟合精度。

1.2.2 组合模型拟合 选择单项模型建立组合模型。组合模型4种权重系数的计算方法为基于算术平均法、方差倒数法、均方误差倒数法、简单加权平均法^[2]。

1.2.3 组合模型的遴选及预测 对不同方式组合模型进行拟合,按1.2.1方法进行模型的评价,遴选出最优组合模型并进行预测。

1.3 统计学方法

应用动态数列、灰色模型、线性回归、指数平滑、

* 西北民族大学引进人才项目(No. XBMUYJYC201626)和中央高校基金(No. 31920180024)资助

△ 通信作者, E-mail: baiyana@lzu.edu.cn

Joinpoint 回归、组合预测对肺癌死亡率进行预测分析, $\alpha=0.05$ 。

2 结果

2.1 预测方法的拟合

2.1.1 三种单项预测方法拟合 如表 1、图 1 所示,用 3 种单项预测方法拟合 2001~2013 年肺癌死亡率,可见指数平滑预测效果最好。如表 2 所示,3 种单项预测方法用于肺癌死亡率的拟合效果比较,指数平滑法拟合精度最高,为 79.67%,Joinpoint 线性回归精度为 74.27%,动态数列拟合精度最低为 63.03%。

2.1.2 模型优化组合及拟合 因指数平滑法和 Joinpoint 线性回归的拟合精度最高,故选择指数

表 1 三种单项预测方法拟合肺癌死亡率($\times 10^{-5}$)

Table 1 Lung cancer mortality predicted by the three predicting models ($\times 10^{-5}$)

Year	Measured value	Dynamic series	Exponential smoothing	Joinpoint linear regression
2001	67.44	—	67.58	41.86
2002	53.06	70.66	69.99	46.98
2003	64.42	73.87	62.01	52.10
2004	45.17	77.09	64.37	57.23
2005	46.91	80.31	53.59	62.35
2006	51.61	83.52	48.72	67.47
2007	42.39	86.74	49.13	72.6
2008	117.91	89.96	43.86	77.72
2009	117.11	93.18	87.91	82.84
2010	113.13	96.39	111.31	87.97
2011	92.55	99.61	120.56	93.09
2012	92.88	102.83	111.54	98.22
2013	118.02	106.04	105.62	103.34

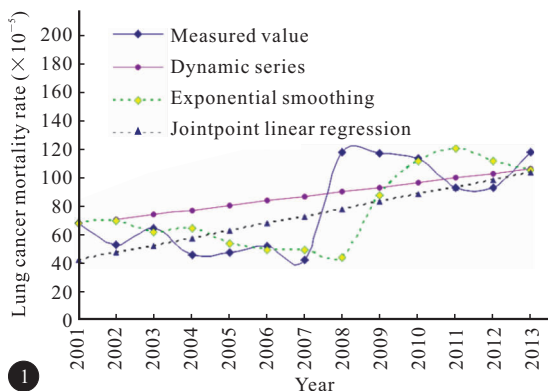


图 1 三种单项预测方法用于 2001~2013 年肺癌死亡率的拟合效果

Fig 1 Fitness of data (2001-2013) into the three predicting models on lung cancer mortality

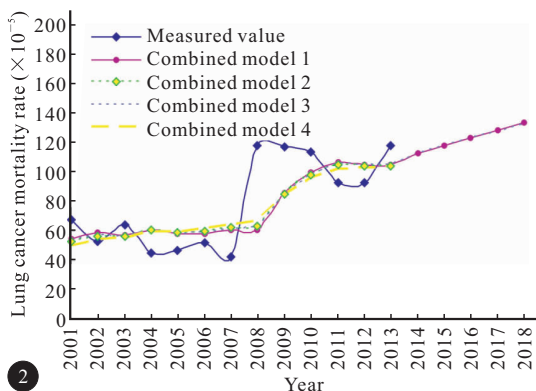


图 2 组合模型用于 2001~2013 年肺癌死亡率的拟合及对 2014~2018 年肺癌死亡率的预测

Fig 2 Fitness of data (2001-2013) into the combined model and forecasting of lung cancer mortality for 2014-2018

表 2 三种单项预测方法对肺癌死亡率拟合效果评价

Table 2 Precisions of the three predicting models on lung cancer mortality

Prediction method	MSE	MAE	MAPE	Precision/%
Dynamic series	620.46	22.19	36.97	63.03
Exponentialsmoothing	645.03	16.86	20.33	79.67
Joinpoint linear regression	466.01	18.29	25.73	74.27

MSE: Mean square error; MAE: Mean absolute error; MAPE: Mean absolute percentage error

平滑法与 Joinpoint 线性回归建立组合模型并进行拟合,基于算术平均法、方差倒数法、均方误差倒数法、简单加权平均法 4 种方法计算组合模型权重系数,组合预测模型 2001~2013 年肺癌死亡率拟合值见图 2、表 3。其中组合模型 1、组合模型 3 拟合效果较好,用于预测 2014~2018 年的肺癌死亡率。

2.2 模型评价

如表 4 所示,肺癌组合模型预测精度接近

表 3 组合模型拟合肺癌死亡率($\times 10^{-5}$)

Table 3 Lung cancer mortality ($\times 10^{-5}$) predicted by the combined model

Year	Measured value	Combined model 1	Combined model 2	Combined model 3	Combined model 4
2001	67.44	54.72	52.65	53.68	50.43
2002	53.06	58.49	56.63	57.55	54.65
2003	64.42	57.06	56.26	56.65	55.40
2004	45.17	60.80	60.22	60.51	59.61
2005	46.91	57.97	58.68	58.33	59.43
2006	51.61	58.10	59.61	58.86	61.22
2007	42.39	60.87	62.76	61.82	64.78
2008	117.91	60.79	63.52	62.16	66.43
2009	117.11	85.38	84.97	85.17	84.53
2010	113.13	99.64	97.76	98.69	95.75
2011	92.55	106.83	104.61	105.71	102.25
2012	92.88	104.88	103.81	104.34	102.66
2013	118.02	104.48	104.30	104.39	104.10

Weight coefficients of the combined models were calculated by the arithmetic average method, the variance inverse method, the mean square error inverse method, and the simple weighted average method, as is indicated in the combined model 1, 2, 3, and 4, respectively

90%，均高于上述任一单项预测模型。在 4 种权重系数的计算中，基于算数平均法及均方误差倒数法预测效果较好。

表 4 组合模型对肺癌死亡率拟合效果评价

Table 4 Precision of the combined models in predicting lung cancer mortality

Combined model	MSE	MAE	MAPE	Precision/%
1	95.64	8.54	13.13	86.87
2	128.95	9.92	15.25	84.75
3	111.79	9.24	14.20	85.80
4	170.03	11.39	17.51	82.49

MSE, MAE, MAPE; Denote the same as those in table 2

2.3 模型预测

肺癌死亡率组合预测中，指数平滑与 Joinpoint 线性回归的组合模型拟合效果最好，故用基于算数平均法、均方误差倒数法计算权重系数的组合模型预测 2014~2018 年金昌队列肺癌死亡率，结果如图 2、表 5 所示。

表 5 组合模型预测 2014~2018 年肺癌死亡率($\times 10^{-5}$)

Table 5 Forecasting of lung cancer mortality ($\times 10^{-5}$) in 2014-2018 using the combined model

Combined model	2014	2015	2016	2017	2018
1	112.80	117.87	122.95	128.02	133.09
3	112.45	117.52	122.60	127.68	132.76

3 讨论

在死亡率、发病率的预测中，多采用单项模型预测法，如动态数列平均增长速度、线性回归^[2]、灰色模型^[3]、指数平滑法^[4]、ARIMAR 模型^[5-6]，近几年也有少量用 Joinpoint 回归^[7-8]，较少应用组合模型^[2,9]。

在对肺癌死亡预测中，本研究采用指数平滑与 Joinpoint 线性回归进行组合，综合利用各单项预测方法所提供的信息，赋予适当权重系数得出组合预测模型。权重系数的计算是根据参与组合的单项模型的误差平方和，基于算术平均法、方差倒数法、均方误差倒数法、简单加权平均法^[10]计算而得。

在肺癌死亡率组合模型预测中，因各参与组合的单项模型的精度也影响组合模型的预测精度，故选择拟合精度较高的单项模型：指数平滑法与 Joinpoint 线性回归建立组合模型。其中，基于算数平均法及均方误差倒数法计算的权重系数的组合模型拟合效果均较好，肺癌死亡率拟合结果接近实际死亡率，与唐小我^[11]的研究一致，故用基于算数平

均法及均方误差倒数法计算的权重系数的组合模型(指数平滑法、Joinpoint 线性回归)预测肺癌死亡率，使预测结果更接近真实值。由于我们尚未查阅到 2014 年及以后肺癌死亡率的真实数据，故本研究建立的组合模型的预测效果有待进一步证实。

目前在恶性肿瘤预测中，组合模型应用较少^[2,9-10]。其它领域组合模型研究表明^[12]：参与组合的模型 2~5 个为宜，不同曲线组合精度更高。故本研究选取了两种不同类型单项模型进行组合——指数平滑法与 Joinpoint 回归。另外，组合模型权重系数的计算也很重要。本研究基于算数平均法与均方误差倒数法建立的组合模型误差较小，精度高，预测效果好。总之，组合模型预测精度大于单项模型，且预测误差较单项模型小。

到目前为止，不存在一种恶性肿瘤预测普遍适用的方法，不同数据要选择不同的模型。本研究遴选出精度较好的预测模型仅仅是 2001~2013 年这个时段比较合适的预测模型，因恶性肿瘤的发生和死亡受多种因素的影响，且有长期变化趋势，故随着年份的变化，要探讨不同的预测模型。本研究建立的组合模型预测精度较高，是预测效果较好的一种预测方法，可以用于其它疾病的预测。

参 考 文 献

- [1] QU HM, BAI YN, CHENG N, *et al.* Trend Analysis of Cancer Mortality in the Jinchang Cohort, China, 2001-2010. *Biomed Environ Sci*, 2015, 28(5):364-369.
- [2] 曲红梅, 白亚娜, 魁发瑞, 等. 组合模型对恶性肿瘤死亡率拟合度评价及预测方法的研究. *中华流行病学杂志*, 2017, 38(1):117-120.
- [3] 潘凤莲, 陈哲. GM(1,1)模型在子宫颈癌死亡率预测中的应用. *海南医学院学报*, 2013, 19(1):92-95.
- [4] 陈勇. 启东 1972-2001 年恶性肿瘤发病趋势分析及预测模型研究. 南京:南京医科大学, 2005 [2017-11-08]. http://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD0506&filename=2005089656.nh&uid=WEEvREcwSljHSldRa1FhcEE0RVZxbnd6eUNJRjRVpRHU0WGVLbERIZEtRdz0=MYM9A4hF_YAuvQ5obgVAqNKP CYeEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!&v=MTA3MDg3RGgxVDN5VHJXTTFGckNVUkxLZlplUnJGeURuVXJ6TVYxMjdHN093RjlmSnFaRjQSVi4ZVgxTHV4WV M=.
- [5] 朱慈华, 丁彦培, 叶波, 等. 运用 ARIMA 模型预测江岸区恶性肿瘤死亡率的变化趋势. *中华临床医师杂志(电子版)*, 2016, 10(7):41-42 [2017-11-08]. http://kns.cnki.net/KCMS/detail/11.9147.R.20161208.1011.014.html?uid=WEEvREcwSljHSldRa1FhcEE0RVZxbnd6eUNJRjRVpRHU0WGVLbERIZEtRdz0=MYM9A4hF_ (下转第 978 页)

- [3] TIDBURY LP, BLACK RH, O'CONNOR AR. Perceiving 3D in the absence of measurable stereo-acuity. *Br Ir Orthopt J*, 2014, 11: 34-38 [2017-11-23]. https://www.researchgate.net/publication/266717752_Perceiving_3D_in_the_absence_of_measurable_stereo-acuity. doi:10.22599/bioj.83.
- [4] TIDBURY LP, BROOKS KR, OCONNOR AR, *et al.* A systematic comparison of static and dynamic cues for depth perception. *Invest Ophthalmol Vis Sci*, 2016, 57(8): 3545-3553.
- [5] ALLEN B, HAUN AM, HANLEY T, *et al.* Optimal combination of the binocular cues to 3D motion. *Invest Ophthalmol Vis Sci*, 2015, 56(12): 7589-7596.
- [6] TIDBURY LP, O'CONNOR AR, WUERGER SM. Dynamic cues to binocular depth. *Br Ir Orthopt J*, 2016, 13(8): 29-34.
- [7] TAI Y, GOWRISANKARAN S, YANG S, *et al.* Depth perception from stationary and moving stereoscopic three-dimensional images. *Proceedings of SPIE*, 2013, 8648 [2017-11-23]. <https://www.mendeley.com/research-papers/depth-perception-stationary-moving-stereoscopic-three-dimensional-images/>. doi:10.1117/12.2005166.
- [8] HESS RF, MANSOURI B, THOMPSON B, *et al.* Latent stereopsis for motion in depth in strabismic amblyopia. *Invest Ophthalmol Vis Sci*, 2009, 50(10): 5006-5016.
- [9] MAEDA M, SATO M, OHMURA T, *et al.* Binocular depth-from-motion in infantile and late-onset esotropia patients with poor stereopsis. *Invest Ophthalmol Vis Sci*, 1999, 40(12): 3031-3036.
- [10] MOLLENHAUER KA, HAASE W. Preliminary report: dynamic stereopsis in patients with impaired binocular function. *Strabismus*, 2000, 8(4): 275-281.
- [11] WATANABE Y, KEZUKA T, HARASAWA K, *et al.* A new method for assessing motion-in-depth perception in strabismic patients. *Br J Ophthalmol*, 2008, 92(1): 47-50.
- [12] WARDLE SG, ALAIS D. Evidence for speed sensitivity to motion in depth from binocular cues. *J Vis*, 2013, 13(1). pii: 17 [2017-11-23]. <https://jov.arvojournals.org/article.aspx?articleid=2121215>. doi:10.1167/13.1.17.
- [13] READ JCA. Stereo vision and strabismus. *Eye*, 2015, 29(2): 214-224.
- [14] HORNG C, HSIEH Y, TSAI M, *et al.* Effects of horizontal acceleration on human visual acuity and stereopsis. *Int J Environ Res Public Health*, 2015, 12(1): 910-926.
- [15] WESTHEIMER G. The Ferrier Lecture, 1992. Seeing depth with two eyes: stereopsis. *Proc Biol Sci*, 1994, 257(1349): 205-214.

(2018-05-03 收稿, 2018-07-22 修回)

编辑 沈进

(上接第 962 页)

- Y AuvQ5obgVAqNK PCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!! &.v=MDU4MzBHNEg5Zk5ybzVNWk90bVl3OU16bVJuNm0N1QzZmxxV00wQ0xMN1I3cWViK1p0RkMzbFz3M01JbFk9UHIIU2Fy.
- [6] ZHANG X, ZHANG T, PEI J, *et al.* Time Series Modelling of Syphilis Incidence in China from 2005 to 2012. *PLoS One*, 2016, 11(2): e0149401 [2017-10-15]. <https://doi.org/10.1371/journal.pone.0149401>.
- [7] WONG MC, JIANG JY, GOGGINS WB, *et al.* International incidence and mortality trends of liver cancer: a global profile. *Sci Rep*, 2017, 7: 45846. doi:10.1038/srep45846.
- [8] WONG MCS, FUNG FDH, LEUNG C, *et al.* The global epidemiology of bladder cancer: joinpoint regression analysis of its incidence and mortality trends and projection. *Sci Rep*, 2018, 8(1): 1129-1140.
- [9] 陈亦新, 田辉. 应用组合模型分析预测恶性肿瘤死亡率. *黑龙江医学*, 2000(5): 24-25.
- [10] 戴华娟. 组合预测模型及其应用研究. 长沙: 中南大学, 2007 [2017-10-08]. http://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD2007&filename=2007171067.nh&uid=WEEvREcwSIJHSlDRa1FhcEE0RVZxbnd6eUNJRXPpRHU0WGVLbERIZEtRdz0=MYM9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!&.v=MDQzNzREbldyL0pWMTI3R2JLL0g5SEtxSkViUEISOGVYMUx1eFlTN0RoMVQzcvRyV00xRnJDVVJMS2ZaZVJyRnk=.
- [11] 唐小我. 最优组合预测方法及其应用. *数理统计与管理*, 1992(1): 31-35.
- [12] 姜明辉. 商业银行个人信用评估组合预测方法研究. 哈尔滨: 哈尔滨工业大学, 2006 [2017-11-04]. http://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CDFD&dbname=CDFD9908&filename=2007040225.nh&uid=WEEvREcwSIJHSlDRa1FhcEE0RVZxbnd6eUNJRXPpRHU0WGVLbERIZEtRdz0=MYM9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8gFw!!&.v=MDY3MTFyRnIEbldydlBWMTI3R2JPOEh0UE9xcEViUEISOGVYMUx1eFlTN0RoMVQzcvRyV00xRnJDVVJMS2ZaZVI=.

(2018-05-17 收稿, 2018-09-10 修回)

编辑 余琳