

癌症负担预测方法学的研究进展

杨之洵, 陈万青

(国家癌症中心/国家肿瘤临床医学研究中心/中国医学科学院北京协和医学院肿瘤医院, 北京 100021)

摘要:目前,癌症已成为我国居民的首要死因,且随着人口的老齡化,癌症负担日益严重。根据现有数据对今后癌症负担进行准确的预测是制定中长期癌症防控策略的基础。利用长期、高质量的癌症发病或死亡数据构建适宜的模型,可分析癌症负担的变化趋势并对负担进行预测。本文归纳了国内外针对癌症负担预测的研究,对当下应用最广泛的几种预测模型进行了介绍。

关键词:癌症;发病;死亡;预测;模型

中图分类号:R73 文献标识码:A 文章编号:1004-0242(2019)04-0241-05

doi:10.11735/j.issn.1004-0242.2019.04.A001

Progress of Research on the Methodology of Prediction of Cancer Burden

YANG Zhi-xun, CHEN Wan-qing

(National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China)

Abstract: Cancer has become the leading cause of death in China and cancer burden is getting heavier along with the ageing of population. Precise prediction of cancer burden based on current data is the prerequisite for the making of mid- and long-term cancer prevention policies. Appropriate models fitted with long-term and high-quality data on cancer incidence and mortality could be used to analyze the change of trend of cancer burden and predict the burden in the future. Herein we summarize recent researches about the prediction of cancer burden and make a brief introduction of the widely used methods.

Key words: cancer; incidence; mortality; prediction; model

目前,全球范围内尤其是发展中国家的癌症负担呈现加重的趋势。2012年全球57%的癌症新发病例和65%的死亡病例发生在发展中国家^[1]。作为最大的发展中国家,我国的癌症负担同样不容乐观。据估计,2015年我国约有429万癌症新发病例和281万死亡病例,癌症已成为我国首要的致死因素^[2]。癌症患者住院费用总额高达1771亿人民币,相比于2011年增长了84.1%^[3]。除了人口的增长和老齡化以外,吸烟等与患癌有关的危险因素流行情况的变化也是癌症发病和死亡增加的主要原因。

为了解日益加重的癌症负担,方便决策者更有效地将有限的医疗资源进行分配,也为了向科研人员提供估计干预措施或筛查效果的依据,对未来癌

症发病和死亡的预测至关重要。理论上,预测癌症发病率的模型中自变量的选择应该与各致癌因素有关,但目前对于大多数癌症的病因研究尚不充分,对潜在危险因素的认识还未完善到可以用来预测负担的程度。因此,癌症负担预测中多以时间为自变量,用其代表人群对各种已知和未知危险因素的暴露情况^[4]。

目前针对癌症负担预测最常用的假设是:若排除外部干预的影响,近期癌症负担的变化趋势可以在未来得到延续^[5,6]。基于既往的发病率或死亡率的发展趋势推算未来的率,再乘以估计的未来人口就可以对发病数或死亡数进行预测。因此,长期、高质量的以人群为基础的癌症发病或死亡数据是精确估计癌症负担的重要前提。同样重要的是可靠人口数量的预测^[7]。

本文归纳了国内外有关癌症发病或死亡预测的

收稿日期:2018-10-11;修回日期:2018-12-10

基金项目:中国医学科学院医学与健康科技创新工程(2016-12M-2-004);
科技部基础工作专项(2014FY121100)

通信作者:陈万青, E-mail: chenwq@cicams.ac.cn

研究,并对主要的预测方法和模型进行了介绍。

1 年龄-时期-队列模型

年龄-时期-队列模型(age-period-cohort model, APC 模型)是一种特殊的泊松对数线性模型,它同时考虑了年龄、时期、出生队列三个因素对癌症发病率或死亡率的影响。随着年龄的增长,身体暴露于各种致癌因素的时间会变长,积累的能导致细胞不规律增殖的基因突变也会增多,因此几乎所有癌症的发病率都随着年龄的增长而上升。时期因素是在特定时间点影响所有年龄组的发病或死亡的环境、社会和经济因素,如战争、饥荒或经济危机。队列因素是与一整代人有关的因素,其差异往往导致两代人各年龄组之间发病率的不同,如对吸烟的接受度和生育习惯等生活方式的改变等^[7,8]。

拟合 APC 模型前,需要将数据分性别以年龄分组,一般每 5 岁一组,分为 18 组(0-4、5-9、10-14、15-19……80-84、85+),再将发病时期每 5 年分为一组。基于既往的趋势可以在未来得到延续的假设,APC 模型多利用最近三个或四个 5 年的观察数据预测未来癌症发病率或死亡率的变化。考虑到当下的趋势可能会逐渐趋于平缓,Moller 等^[9]假定在预测的第三、第四个 5 年仅 50%的线性趋势得以维持。Arnold 等^[10]在此基础上做了优化,假定该趋势在预测的第二、第三和第四个 5 年将分别减少 25%、50%和 75%。利用 R 语言的 Nordpred、Epi、apc 和 BAPC 包均可实现 APC 模型的拟合。

Chen 等^[11]利用 APC 模型分析了 1998~2007 年我国肺癌的发病数据,估计出 2020 年全国肺癌新发病例数可达近 70 万人,其中人口老龄化为主要因素,时期和队列因素影响较小。Wong 等^[12]用 1976~2010 年中国香港的乳腺癌数据预测到 2021~2025 年发病率将从 2011~2015 年间的 56.7/10 万上升到 62.5/10 万,死亡率则将从 9.3/10 万下降到 8.6/10 万。此外,APC 模型还被广泛应用于英国^[13,14]、西班牙^[5]、巴西^[15]、韩国^[16]和印度^[17]等国,以及国际食管癌^[10]、肝癌^[18]和睾丸癌^[19]负担的预测研究。在实际研究中,几个低年龄组的癌症发病或死亡例数通常较少,其随时间的变化趋势不仅无法反映实际情况,还会对预测结果的准确性产生影响。在模型拟合时

可将这些低年龄组进行合并^[15,20,21]或直接从研究中剔除^[12,16,17]。

APC 模型可以从流行病学的角度量化并比较年龄、时期和出生队列因素对癌症负担的影响,要求数据精确到各年龄组,是目前分析癌症负担的变化趋势及预测未来负担的研究中被应用最广泛的方法。虽然模型包含的参数越多拟合效果往往越好,但更多的参数有时会使结果的可信区间变宽,不如参数少的模型精确。因此在拟合 APC 模型的同时也可以用年龄-时期模型(age-period model, AP 模型)和年龄-队列模型(age-cohort model, AC 模型)的结果作验证。

2 Joinpoint 回归模型

Joinpoint 回归模型 (Joinpoint regression model) 是以时间(年)为自变量、癌症发病率或死亡率为因变量的基于泊松分布的对数线性模型,通过一系列连接点(joinpoint)将数个对数线性片段拼接起来,以此描述癌症发病或死亡的时序变化趋势。模型初始默认连接点数量为 0,即趋势为单调递增或递减,通过蒙特卡洛置换检验方法判断连接点的个数、每个连接点的位置及相应的 *P* 值,最终根据贝叶斯信息准则(bayesian information criterion, BIC)选出最优模型。Joinpoint 回归模型既可以识别趋势发生改变的时间点,又可以通过计算年度变化百分比(annual percentage change, APC)及其 95%可信区间来量化趋势的幅度并检验这种变化是否具有统计学意义^[22-24]。该模型也适用于其他慢性疾病发病率或死亡率的分析^[23]。

目前实现 Joinpoint 回归模型最常用的方法是依靠美国国家癌症研究所的 Joinpoint 软件。我国学者利用 Joinpoint 软件对我国 1989~2008 年的癌症死亡趋势进行了分析,发现男性死亡标化率在 1989~2001 年的年度变化百分比为-2.0%,在 2001~2008 年下降速度减慢,年度变化百分比为-0.1%。女性的标化率也呈现出下降幅度先快后慢的趋势,但时间节点不同。1989~1997 年和 1997~2008 年间女性癌症死亡标化率的年度变化百分比分别为-2.3%和-0.1%^[25]。

除年度变化百分比外,Joinpoint 软件还可以计

算其加权平均值,即平均年度变化百分比(average annual percentage change, AAPC)来综合反映整个研究时间段内率的变化趋势。陈万青等^[26]发现 1989~2008 年间我国男性和女性癌症发病粗率的 AAPC 分别为 2.0%和 3.0%,而同期标化发病率的 AAPC 分别为-0.6%和 0.8%,提示了我国导致癌症负担不断增加的主要因素是人口老龄化。和 APC 模型一样,基于现有癌症发病和死亡的趋势可以在未来得以保持的假设, AAPC 成为了利用 Joinpoint 软件预测癌症负担的关键指标^[6,27]。

Quante 等^[6]通过 1999~2011 年间德国的癌症发病数据和 1998~2012 年的死亡数据计算了负担最严重的 15 个癌种各年龄组的发病率和死亡率,分别计算了十余年间发病率与死亡率的 AAPC。因为某些年龄组的癌症发病数或死亡数过小,对其计算 AAPC 将使结果的不稳定性增加,所以研究只计算了全人群的 AAPC,未细分至各年龄组。将各个癌种 2011 年在各年龄组中的发病率与德国联邦统计局对 2020 年和 2030 年德国各年龄组人口的预测值相乘,再用 Joinpoint 软件计算出的 AAPC 值进行调整, Quante 等推算出了 2020 年和 2030 年各癌种的发病数,并同理利用 2012 年的死亡数据推算出了 2020 年和 2030 年的各癌种死亡人数。

Joinpoint 回归模型既可以分析分年龄组率的时序趋势,又适用于全人群的数据。相比于 APC 模型,它对数据的要求较低且步骤更简单,是癌症趋势分析及预测研究中另一种常见的方法。

3 差分自回归移动平均模型

差分自回归移动平均模型 (autoregressive integrated moving average model, ARIMA 模型)是一种重要的时间序列分析预测模型,分为季节性 ARIMA 模型和非季节性 ARIMA 模型。前者多用于传染性疾病负担的时序分析和预测,后者多用于癌症负担的分析和预测。ARIMA 模型将预测对象随时间推移而形成的时间序列视为一个随机序列,即除去个别的偶然因素引起的观测值外,时间序列是一组依赖于时间的随机变量。这组随机变量所具有的依存关系或自相关性暗示了预测对象发展的延续性,而这种自相关性一旦被相应的数学模型描述,即可用时

间序列的过去和现在观测值预测未来值。

癌症的发病率和死亡率随时间的变化大多呈现出不平稳 (non-stationary),需要进行差分。非季节性 ARIMA 可以用 ARIMA(p,d,q)来表示,其中 p 为自回归模型的阶数,d 为时间序列成为平稳时序所做的差分次数,q 为移动平均模型的阶数。利用 R 语言中的 forecast 包可以构建 ARIMA 模型,并选择赤池信息量准则 (Akaike's information criterion, AIC)和 BIC 值最小的模型为相对最优。Xu 等^[28]尝试了多种参数组合后发现 ARIMA(1,0,0)最适合拟合中国鼻咽癌 1991~2013 年的死亡趋势。Tsoi 等^[29]利用 ARIMA 模型预测出英国、瑞典 2030 年结直肠癌的发病率相比于 1988 年将分别上升 4.8%和 4.7%,而克罗地亚和哥斯达黎加将分别上升 47.0%和 18.5%。

和 Joinpoint 回归模型一样,在构建 ARIMA 模型时不需要分年龄组的发病或死亡数据,只需要从数理的角度考虑时间和总的率之间的关系。相比于前两种方法,ARIMA 模型在癌症负担预测的研究中应用较少,更多出现在传染性疾病和社会经济因素的时序分析和预测研究中,如增加烟草税、加强控烟法案等措施对吸烟率的影响^[30]、细菌性病疾^[31]和结核病^[32]的发病等。

4 其他模型

“灰色系统”这一概念最早由我国学者邓聚龙在 1982 年提出,指的是一部分信息已知、另一部分信息未知且各因素间有不确定关系的系统。GM(1,1)模型是灰色系统重要的组成部分之一,多次被我国学者应用于癌症负担的预测^[33-35],其中(1,1)表示一个变量的一阶微分方程。一般的数学模型是对原始数列建模,且要求有足够的样本量和特定的分布。GM(1,1)模型不受这些条件的约束,具有所需样本量小、不需要典型的概率分布、计算简便的特点。模型通过对原始数据进行累加,消除其随机性及波动性,得到规律性更强的新时间序列后拟合其中的规律并进行外推,最后形成微分方程模型。GM(1,1)模型适合对符合一定特征的数据进行预测,不适合具有极大波动性和随机性的数据。

指数平滑法 (exponential smoothing)采用“厚近薄远”的原则对数据进行加权平均和修匀,对近期观

察值赋予较大的权重,对远期观察值赋予较小的权重,可以去除时间序列中的随机波动并减弱异常数据的影响,使历史规律得以更好地体现,对未来发展趋势做出合理的预测。指数平滑法根据参数个数的不同可以分为单指数平滑法(只有水平项)、双指数平滑法(有水平项和趋势项)和三指数平滑法(有水平项、趋势项和季节效应)。Yasmeen 等^[36]用指数平滑法预测出了美国女性乳腺癌死亡率的下降趋势,但在某些年龄组内,黑人女性的下降幅度不如白人女性明显。

5 总 结

目前,我国处在经济飞速发展的阶段,社会经济因素、生活习惯和医疗水平等都在不断发生变化。这些都会对癌症负担产生影响,例如近年来我国肝癌标化发病率下降趋势明显^[37],而甲状腺癌和前列腺癌等癌种的发病率增长迅速^[2]。因此,若干年后肝癌可能不再是防治的重点,甲状腺癌和前列腺癌则有可能成为最常见的癌种之一。癌症防治策略的制定需要前瞻性。基于既往发病或死亡数据对今后的癌症负担进行较为准确的预测,不仅可以为医学研究提供依据,还可以为癌症预防工作提供目标和参考,从而优化癌症防治策略的选择,将医疗资源更高效地分配。

参考文献:

[1] Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012[J]. *CA Cancer J Clin*, 2015, 65(2): 87-108.

[2] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015[J]. *CA Cancer J Clin*, 2016, 66(2): 115-132.

[3] Cai Y, Xue M, Chen W, et al. Expenditure of hospital care on cancer in China, from 2011 to 2015 [J]. *Chin J Cancer Res*, 2017, 29(3): 253-262.

[4] Dyba T, Hakulinen T, Paivarinta L. A simple non-linear model in incidence prediction[J]. *Stat Med*, 1997, 16(20): 2297-2309.

[5] Cleries R, Martinez JM, Moreno V, et al. Predicting the change in breast cancer deaths in Spain by 2019: a Bayesian approach[J]. *Epidemiology*, 2013, 24(3): 454-460.

[6] Quante AS, Ming C, Rottmann M, et al. Projections of cancer incidence and cancer-related deaths in Germany by 2020 and 2030[J]. *Cancer Med*, 2016, 5(9): 2649-2656.

[7] Bray F, Moller B. Predicting the future burden of cancer [J]. *Nat Rev Cancer*, 2006, 6(1): 63-74.

[8] Clayton D, Schifflers E. Models for temporal variation in cancer rates. I: age-period and age-cohort models [J]. *Stat Med*, 1987, 6(4): 449-467.

[9] Moller B, Fekjaer H, Hakulinen T, et al. Prediction of cancer incidence in the Nordic countries up to the year 2020 [J]. *Eur J Cancer Prev*, 2002, 11 Suppl 1: S1-S96.

[10] Arnold M, Laversanne M, Brown LM, et al. Predicting the future burden of esophageal cancer by histological subtype: international trends in incidence up to 2030[J]. *Am J Gastroenterol*, 2017, 112(8): 1247-1255.

[11] Chen WQ, Zheng RS, Zeng HM. Bayesian age-period-cohort prediction of lung cancer incidence in China[J]. *Thorac Cancer*, 2011, 2(4): 149-155.

[12] Wong IO, Schooling CM, Cowling BJ, et al. Breast cancer incidence and mortality in a transitioning Chinese population: current and future trends [J]. *Br J Cancer*, 2015, 112(1): 167-170.

[13] Smittenaar CR, Petersen KA, Stewart K, et al. Cancer incidence and mortality projections in the UK until 2035[J]. *Br J Cancer*, 2016, 115(9): 1147-1155.

[14] Pesola F, Ferlay J, Sasieni P. Cancer incidence in English children, adolescents and young people: past trends and projections to 2030[J]. *Br J Cancer*, 2017, 117(12): 1865-1873.

[15] Souza DL, Jerez-Roig J, Cabral FJ, et al. Colorectal cancer mortality in Brazil: predictions until the year 2025 and cancer control implications[J]. *Dis Colon Rectum*, 2014, 57(9): 1082-1089.

[16] Lee JE, Lee SA, Kim TH, et al. Projection of breast cancer burden due to reproductive/lifestyle changes in Korean women (2013-2030) using an age-period-cohort model[J]. *Cancer Res Treat*, 2018, 50(4): 1388-1395.

[17] Dikshit RP, Yeole BB, Nagrani R, et al. Increase in breast cancer incidence among older women in Mumbai: 30-year trends and predictions to 2025[J]. *Cancer Epidemiol*, 2012, 36(4): e215-e220.

[18] Valery PC, Laversanne M, Clark PJ, et al. Projections of primary liver cancer to 2030 in 30 countries worldwide[J]. *Hepatology*, 2018, 67(2): 600-611.

[19] Le Cornet C, Lortet-Tieulent J, Forman D, et al. Testicular cancer incidence to rise by 25% by 2025 in Europe? Model-based predictions in 40 countries using population-based registry data[J]. *Eur J Cancer*, 2014, 50(4): 831-839.

[20] Yang L, Parkin DM, Li LD, et al. Estimation and projec-

- tion of the national profile of cancer mortality in China: 1991–2005[J]. *Br J Cancer*, 2004, 90(11):2157–2166.
- [21] Moller B, Fekjaer H, Hakulinen T, et al. Prediction of cancer incidence in the Nordic countries: empirical comparison of different approaches[J]. *Stat Med*, 2003, 22(17): 2751–2766.
- [22] Qiu D, Katanoda K, Marugame T, et al. A Joinpoint regression analysis of long-term trends in cancer mortality in Japan(1958–2004)[J]. *Int J Cancer*, 2009, 124(2):443–448.
- [23] Lopez-Campos JL, Ruiz-Ramos M, Soriano JB. Mortality trends in chronic obstructive pulmonary disease in Europe, 1994–2010: a joinpoint regression analysis[J]. *Lancet Respir Med*, 2014, 2(1):54–62.
- [24] Kim HJ, Fay MP, Feuer EJ, et al. Permutation tests for joinpoint regression with applications to cancer rates[J]. *Stat Med*, 2000, 19(3):335–351.
- [25] Zeng HM, Zheng RS, Zhang SW, et al. Trend analysis of cancer mortality in China between 1989 and 2008[J]. *Chinese Journal of Oncology*, 2012, 7(1):525–531.[曾红梅, 郑荣寿, 张思维, 等. 1989–2008年中国恶性肿瘤死亡趋势分析[J]. *中华肿瘤杂志*, 2012, 7(1):525–531.]
- [26] Chen WQ, Zheng RS, Zeng HM, et al. Trend analysis and projection of cancer incidence in China between 1989 and 2008[J]. *Chinese Journal of Oncology*, 2012, 7(1):517–524.[陈万青, 郑荣寿, 曾红梅, 等. 1989–2008年中国恶性肿瘤发病趋势分析[J]. *中华肿瘤杂志*, 2012, 7(1):517–524.]
- [27] Rahib L, Smith BD, Aizenberg R, et al. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States [J]. *Cancer Res*, 2014, 74(11):2913–2921.
- [28] Xu ZX, Lin ZX, Fang JY, et al. Mortality characteristic and prediction of nasopharyngeal carcinoma in China from 1991 to 2013 [J]. *Asian Pac J Cancer Prev*, 2015, 16(15): 6729–6734.
- [29] Tsoi KKF, Hirai HW, Chan FCH, et al. Predicted increases in incidence of colorectal cancer in developed and developing regions, in association with ageing populations[J]. *Clin Gastroenterol Hepatol*, 2017, 15(6):892–900.e4.
- [30] Wakefield MA, Coomber K, Durkin SJ, et al. Time series analysis of the impact of tobacco control policies on smoking prevalence among Australian adults, 2001–2011 [J]. *Bull World Health Organ*, 2014, 92(6):413–422.
- [31] Wang K, Song W, Li J, et al. The use of an autoregressive integrated moving average model for prediction of the incidence of dysentery in Jiangsu, China[J]. *Asia Pac J Public Health*, 2016, 28(4):336–346.
- [32] Mao Q, Zhang K, Yan W, et al. Forecasting the incidence of tuberculosis in China using the seasonal auto-regressive integrated moving average (SARIMA) model [J]. *J Infect Public Health*, 2018, 11(5):707–712.
- [33] Li DJ. Analysis of incidence and mortality rates of malignant tumors in Hebei Provincial cancer registry areas in 2013 and prediction of incidence of malignant tumors in high-risk area[D]. Hebei Medical University, 2016.[李道娟. 河北省肿瘤登记地区 2013 年恶性肿瘤发病死亡分析及高发区恶性肿瘤发病预测[D]. 河北医科大学, 2016.]
- [34] Xiao J, Huang JP, Zhang M, et al. Trends and forecast of hepatocellular carcinoma in Nantong, China: mortality rates from 1999 to 2011 [J]. *Chinese Journal of Hepatology*, 2015, 9(2):663–668.[肖静, 黄建萍, 张敏, 等. 南通地区 1999 至 2011 年肝癌死亡率的变动趋势和今后 5 年发展趋势的预测[J]. *中华肝脏病杂志*, 2015, 9(2):663–668.]
- [35] Lin YL, Wu XQ, Lin TQ. Analysis and forecasting of population mortality and life lost trend due to lung cancer among Xiamen residents [J]. *Chinese Journal of Lung Cancer*, 2016, 2(5):77–81.[林艺兰, 伍啸青, 林田泉. 厦门市居民肺癌死亡与减寿趋势分析及预测 [J]. *中国肺癌杂志*, 2016, 2(5):77–81.]
- [36] Yasmeeen F, Hyndman RJ, Erbas B. Forecasting age-related changes in breast cancer mortality among white and black US women: a functional data approach [J]. *Cancer Epidemiol*, 2010, 34(5):542–549.
- [37] Zuo TT, Zheng RS, Zeng HM, et al. Analysis of liver cancer incidence and trend in China [J]. *Chinese Journal of Oncology*, 2015, 9(8):691–696.[左婷婷, 郑荣寿, 曾红梅, 等. 中国肝癌发病状况与趋势分析 [J]. *中华肿瘤杂志*, 2015, 9(8):691–696.]