· 专题报道 ·

慢性病趋势预测瓶颈剖析及优化模式构建分析*

刘蕊[1], 石建伟[2,3], 于德华[2], 庄淑卿[4], 董贞彬[4], 肖月[3], 刘娜娜[3], 王朝昕[1,3]

摘 要: 目的 基于对中国疾病预测研究的发展沿革、预测方法及研究瓶颈分析, 旨在通过优化慢性病趋势预测模式为中国慢性病防治提供一定的理论依据。**方法** 通过文献荟萃分析, 系统梳理中国慢性病预测发展现状及瓶颈, 分析优化预测模式。**结果** 中国慢性病预测重视度不足, 人群发病率预测较匮乏, 方法学应用仍停留在线性或多元回归层面。从人口、经济、社会 3 个范畴筛选出影响因素变量构建状态空间模型, 该优化模式比其他的时间序列自回归模型的拟合优度更高。**结论** 状态空间模型用于构建特定区域的慢性病趋势预测模型, 可大大提高长期预测的精度和灵敏性, 为循证决策提供强有力支撑。

关键词: 慢性病; 趋势预测; 瓶颈; 优化模式

中图分类号: R 181.2 **文献标志码:** A **文章编号:** 1001-0580(2017)11-1552-04 DOI: 10.11847/zgggws2017-33-11-03

Analysis on bottlenecks of chronic disease prevalence prediction and establishment of optimized model in China

LIU Rui*, SHI Jian-wei, YU De-hua, et al (* *Affiliated Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai* 200090, China)

Abstract: Objective To analyze the development, methodology, and bottleneck of research on prediction of chronic disease prevalence trend in China and to provide a theoretical basis by optimizing prevalence prediction model for prevention and control of chronic diseases. **Methods** We conducted a meta-analysis of relevant literatures, systemically reviewed development status and bottlenecks of chronic disease prevalence prediction in China, and optimized models for chronic disease prevalence prediction. **Results** There is a lack in concerns to the prediction for chronic disease prevalence trend and there are a few studies on prediction of chronic disease incidence. The application of methodology is still restricted to linear or multiple regression. We screened out influential variants in scopes of population, economy, and society and then constructed a state space model with those variants. The established model demonstrated a higher fitness than other time series autoregressive models. **Conclusion** The use of state space model in conducting prevalence prediction of chronic disease for a specific region could improve the precision and sensitivity of long term prediction and to provide strong evidences for evidence-based decision-making.

Key words: chronic disease; trend prediction; bottleneck; optimization model

至 2015 年年底, 中国 ≥60 岁人口超过 2.22 亿, 占总人口的比重进一步上升到 16.1%, 老龄化已成为我国社会必须面对的重大国计民生问题[1]。伴随老龄化的日趋严重, 慢性非传染性疾病(以下简称"慢性病")的比例也在逐渐升高, 在每年约 1 030 万各种因素导致的死亡中, 慢性病所占比例 >80%, 而慢性病的疾病负担也达到所有病种的 68.6%[2]。慢性病的有效防治离不开对现况问题的把握, 然而, 全国性的慢性病数据往往需要通过大规模的抽样调查采集, 而这些调查又面临着范围广、周期长等困难, 实施耗费大, 难以频繁进行。而高效准确的趋势预测可以极大简化对当前慢性病情况的了解过程, 慢性病的预测能够极大的节约资源, 为循证决策提供支撑。但是, 近年来有关慢性病的预测研究较为匮乏, 特别是预测方法和模式方面。针对这一空缺, 本研究逐一分析现有慢性病预测研究的瓶颈, 并提出适宜的慢性病趋势预测方法, 以期为我国慢性病防治提供一定的理论依据。

1 现有慢性病预测研究瓶颈分析

1.1 *疾病预测研究的发展沿革* 1993 年刘自远等[3]应用灰色动态模型对麻风病的患病率进行了预测, 直到 1997 年首例慢性病的预测模型才开始发表[4]。然而, 在 1997—2007 年, 极少有相关文章发表。相比国外, 仅仅科学引文索引(science citation index, SCI)和社会科学引文索引(social science cita-

* **基金项目:** 2016 年中央高校优秀青年教师计划(英才类); 中国博士后科学基金项目(2016M591719); 国家自然科学基金青年项目(71403185 & 71603182); 2016 年度上海市卫计委科研课题(20164Y0161)

作者单位: 1. 同济大学附属第十人民医院, 上海 200072; 2. 同济大学附属杨浦医院全科医学系; 3. 同济大学医学院; 4. 山东省日照市莒县疾病预防控制中心

作者简介: 刘蕊(1983 -), 女, 河南南阳人, 助理研究员, 硕士, 研究方向: 卫生事业管理。

通讯作者: 王朝昕, E-mail: supercell002@sina.com

数字出版日期: 2017 - 10 - 13 14:36

数字出版网址: http://kns.cnki.net/kcms/detail/21.1234.R.20171013.1436.014.html

中国公共卫生 CPH

tion index, SSCI)数据库在 2007 年前就发表慢性病患病率、发病率预测相关研究多达 3 000 篇以上。而接下来的 10 年,我国慢性病预测方面的研究仍然不如慢性病机制及临床、预防相关研究的增长速度,而是每年仅仅十余篇的数量保持稳定。从国内外慢性病预测相关研究的差异,不难发现我国其实仍处于由急性病临床诊疗到慢性病防治的过渡期,对于慢性病预测于政府卫生循证决策的重大作用仍未得到足够重视[5-6]。

1.2　慢性病预测研究的瓶颈剖析(表 1)　纵观近 20 年的慢性病预测相关文献,通过在中国知网、万方、维普三大中文数据库中,以“慢性病/慢性病”和“预测”为关键词进行检索,总共仅得出 89 篇相关文献,且对慢性病预测领域的相关问题进行检索分析发现,当前共识度最高的关键问题为“慢性病发病率患病率预测影响因素繁多,难以兼顾”、“慢性病预测方法单调,且不符合慢性病的数据特征”以及“当前研究多聚焦某种慢性病的个体发病几率,少人群研究”。不难发现,研究的核心瓶颈在于,慢性病预测重视程度不够,人群发病率的预测仍然较为匮乏,当前研究多聚焦个体的发病原因,方法学的应用也较为单调,停留在线性或多元回归层面。

表 1　慢性病预测研究的瓶颈分析($n=89$)

排序	问题阐述	提及数	提及率(%)
1	慢性病发病率患病率预测影响因素繁多,难以兼顾	42	46.85
2	慢性病预测方法单调,且不符合慢性病的数据特征	36	40.56
3	当前研究多聚焦某种慢性病的个体发病几率,少人群研究	34	38.46
4	我国卫生部门的慢性病防控决策缺乏趋势性的循证支撑	28	31.47
5	慢性病预测应用数据多为面上数据,对具体区域指导意义不大	21	23.08
6	慢性病预测的研究多数据陈旧	7	8.39
7	慢性病预测的病种单一,多聚焦糖尿病和心脑血管疾病	6	6.99
8	预测多缺乏大数据支撑	6	6.29

2　适宜的慢性病趋势预测方法分析(表 2)

当前慢性病趋势预测研究方法应用较单一,且基本均为线性回归,忽视了慢性病发病率的数据具有明显的时序性,具有非平稳的特性,基于最小二乘法的传统线性回归可能导致“伪回归”现象[7-8]。单位根检验、协整理论以及误差修正机制的提出突破了传统回归的局限,并消除了由于序列差分可能导致的研究变量间长期关系的信息损失[9-11]。因此,首先,在方法学的选择上,必须应用时间序列的预测方法,否则会掉入样本量小、数据不平稳的统计陷阱。

本研究将方法学的选取扩展到了趋势预测的广义领域,得到 130 篇文献,对当前应用最广泛的十二类方法进行了优缺点分析(表 2),从而确定状态空间模型为本次优化模式构建的核心预测方法。因为考虑慢性病数据的时序性,必须自回归,而普通自回归无法纳入影响因素,误差较大,状态空间模型在工科领域的精确应用彰显了其对影响因素纳入的强大功能,因此成为本次慢性病趋势预测的研究方法。

3　慢性病趋势预测的模式设想

3.1　广义范畴分析慢性病的相关影响因素(表 3)　在检索得到近 10 年 3 764 篇慢性病相关文献后,最终得到高度相关文献 1 260 篇,并根据经济、人口、医疗卫生和政策 4 个范畴来归类慢性病的相关影响因素。本文对四大范畴进行了细化,其中经济范畴仍然是国内生产总值(gross domestic product, GDP)影响最大(29.12%),人口则是≥65 岁人口比例(27.06%),卫生是每千人口卫生技术人员数(16.87%),而社会政策方面政府对卫生投入的多少(15.36%)至关重要。

3.2　状态空间模型对慢性病趋势预测的分析模式　从慢性病的病种分布、发病率情况及疾病顺位、门诊住院情况、用药行为、医疗费用、医保现况、就医意向及行为进行描述性分析。

状态空间模型预测慢性病趋势:模型在本研究中主要以 10 种慢性病优势病种的每月发病率作为时间序列的自回归变量,从人口、经济、社会 3 个范畴筛选出影响因素变量,构建状态空间模型,具体计算步骤如下:

记 $G\times1$ 向量 y_t 为 t 时刻的观测值,$t=1,2,\cdots,T$,而 $Z\times1$ 向量 z_t 是不可直接观测的,称为状态向量,则描述的动态行为的状态空间模型为:

$$\begin{aligned} z_t &= Fz_{t-1} + W_t \quad (1a) \\ y_t &= Ax_t + Hz_t + v_t \quad (1b) \end{aligned} \qquad (1)$$

其中 $K\times1$ 向量 x_t 为外生向量,矩阵 F,A 和 H 是系数矩阵,里面的元素可以是时变的,简单起见,我们先讨论系数矩阵为常数矩阵的情况。通常将方程(1)称为状态方程,或者叫转移方程,而将方程(2)称为信号方程或者观测方程。$Z\times1$ 阶向量 w_t 和 $G\times1$ 向量 v_t 都是向量白噪声过程,满足

$$E(w_t)=0E(w_tw_t')=\begin{cases} Wt=l \\ 0 \qquad t\neq l \end{cases}$$

表 2　卫生总费用预测方法提及率及各方法优缺点分析($n=130$)

序号	预测方法	优缺点简述	提及数	提及率(%)
1	主成分神经网络模型	优点:实现多变量时间序列预测,降低了预测误差,提高了预测精度。缺点:运算量大,更适合复杂现象的模拟预测。	1	0.77
2	时间序列模型	优点:突出时间因素在预测中的作用。缺点:较适合预测稳定的、时间方面延续的过程,缺乏考虑外界其他因素影响,预测误差大。	10	7.69
3	状态空间模型	优点:适用于多变量场合并能够识别带有趋势和季节成分的非平稳时间序列,能更好地诊断时间序列的动态性。缺点:模型的建立及预测精度受到数据类型、预测区间范围等的影响。	5	3.85
4	ARIMA 模型	优点:通过差分方法提取确定性信息,对确定性信息的提取比价充分。缺点:很难对模型作出直观解释。	13	10.00
5	协整回归	优点:经过 2 次广义差分法,消除了协整回归可能的自回归显现,提高了预测精度。缺点:依赖于另一同步趋势变量和平减指数的准确预测,不适合长期预测。	1	0.77
6	多因素回归模型	优点:考虑了卫生总费用的众多因素。缺点:忽略了卫生总费用的时间序列特征,预测精度不高。	39	30.00
7	计量经济学方法	优点:二维的数据(横截面和时间)构成,能控制 OLS 估计的偏差,模型参数更准确,降低了共线性,提高估计的有效性等,能更准确地反映经济变量的动态调整等。缺点:模型设定错误和数据收集不慎将引起较大偏差。	32	24.62
8	SimFin 模型	优点:结合实现卫生系统功能的财政需要与政府预算可能性模拟模型,评价宏观经济对卫生筹资的影响。缺点:受到基础模型中的参数的限制和影响。	2	1.54
9	GM(1,1)	优点:允许小样本预测,将随机性较强的原始数据处理变成数据列。缺点:预测精度受被预测对象的递变规律以及数据序列的光滑度影响。	9	6.92
10	DGM(1,1)	优点:允许小样本预测。缺点:对近似齐次指数序列的预测模拟效果较好,但对金斯非齐次指数序列的拟合精度差。	1	0.77
11	基于 Choquet 模糊积分的非线性组合	优点:同时实现较好的拟合精度和预测精度,能避免模型间的交互作用。缺点:较为复杂,较难实现。	1	0.77
12	系统动力学	优点:擅长处理高阶次、非线性、时变的复杂问题。缺点:缺乏严格的理论基础,精确度不高,仅被看做是一种系统建模的方法。	16	12.31

表 3　慢性病趋势影响因素范畴分析($n=1\ 260$)

类别	因素	提及数	提及率(%)
经济	GDP	367	29.12
	人均 GDP	231	18.30
	居民消费水平	163	12.97
	农村居民家庭人均纯收入	152	12.06
	城镇居民家庭人均可支配收入	152	12.09
	居民消费价格指数	126	9.97
	收入弹性系数	126	9.98
	相对价格等多	123	9.78
	卫生事业费	152	12.08
人口	≥65 岁人口比例	341	27.06
	大专及以上人口比例	124	9.82
	平均期望寿命	151	11.99
	年末总人口	190	15.06
	城镇人口比重	214	17.01
	年死亡人口数	156	12.35
医疗卫生	社会医疗保障经费占卫生总费用比例	149	11.86
	每千人口卫生技术人员数	213	16.87
	人均药品费用	126	9.97
	次均门诊费用	148	11.78
	诊疗人次数	165	13.07
	住院人次数	190	15.06
	床位数	165	13.08
	医疗保险参保人数	153	12.15
	新增医院设备、技术	166	13.17
	政府卫生事业费占财政预算支出比例	194	15.36
社会政策等	城镇化率	206	16.34
	医疗保障	171	13.57
	卫生管理体系	129	10.20
	政府卫生事业费占财政预算支出比例	168	13.35

中国公共卫生 CPH

$$E(v_t)=0E(v_t v'_t)=\begin{cases}Vt=l\\0 \qquad t\neq l\end{cases}$$

其中协方差矩阵 W 和 V 分别为 $Z\times Z$ 和 $G\times G$ 矩阵,简单起见,我们假设 w_t 和 v_t 不相关,即

$$E(w_t v'_1)=0\ \forall t,1$$

此外,假设干扰和初始状态 z_1 不相关,即

$$E(w_t z'_1)=0\ E(v_t z'_1)=0 \qquad t=1,2,\cdots,T$$

把状态空间模型(1)及其干扰的设定形式称为状态空间模型的基本形式。

描述 y_t 动态过程的状态空间模型的具体形式并不唯一:假设 T 为可逆矩阵,令

$$z_{*t}=Tz_t \tag{2}$$

注意到状态向量 z_t 与其线性变换 z_{*t} 在信息上是等价的,状态空间模型(1)可以等价表示成

$$z_{*t}=F_* z_{*t-1}+w_t$$

$$y_t=Ax_t+H_* z_{*t}+v_t$$

其中,

$$F_*=TFT^{-1}H_*=HT^{-1}w_{*t}=Tw_t$$

由此可以得出慢性病发病率的负根倒数矩阵,从而比其他的时间序列自回归模型的拟合优度更高,预测慢性病的未来发展趋势更加精确。

4　小　结

从 1997—2017 年这长达 20 年的慢性病预测相关文献研究中,出现过 1997—2007 年的空白期,当前慢性病防治逐渐得到重视但趋势预测研究却仍属忽视阶段,还存在方法应用不当的瓶颈。对比国外,形成了明显反差,尤以早早跨入老龄化社会的澳洲和欧洲为甚,其政府每年依据趋势预测报告进行医疗保险和卫生规划的制定[12],具有极大的循证决策意义,这也为我国未来的发展提供了启示。

本文从宏观层面对慢性病趋势预测进行了剖析,从现实困境到解决方案,提出了创新型的研究方法——状态空间模型。该模型适用于多变量场合并能识别带有趋势和季节成分的非平稳时间序列,能更好诊断时间序列的动态性特点,其广泛应用于精细化的工科领域[13-14],虽然医学领域应用极少[15],但其针对时序资料的精确拟合,却恰恰是慢性病趋势预测所需要的,且可以通过从经济、人口、医疗卫生、社会政策 4 个范畴筛选可得性指标,构建适合特定区域的慢性病趋势预测模型,相比自回归模型,将大大提高长期预测的精度和灵敏性。

参考文献

[1] 中华人民共和国民政部. 2015 年社会服务发展统计公报[R]. [2016-07-11]. http://www.mca.gov.cn/article/sj/tjgb/201607/20160700001136.shtml.

[2] 世界银行人类发展部. 创建健康和谐生活:遏制中国慢性病流行[EB/OL]. (2017-04-06)[2011-11-07]. http://ishare.iask.sina.com.cn/f/20721633.html.

[3] 刘自远,吴成建,易发超. 应用灰色动态模型预测麻风患病率[J]. 中国麻风杂志,1993,9(3):148-150. [4] 阎正民,毛正中. 城市慢性病费用灰色预测及分析[J]. 中国卫生事业管理,1997(2):93-95.

[5] 杜本峰,郭玉. 中国老年人健康差异时空变化及其影响因素分析[J]. 中国公共卫生,2015,31(7):870-877.

[6] 王萍,殷洪博,陈叶,等. 沈阳市人口老龄化趋势预测及社区居家养老护理模式的探讨[J]. 中国公共卫生管理,2013,29(3):312-314.

[7] 焦辛妮,王长义,汪东伟,等. 基于测量模型结构方程模型 GLS 与 WLS 比较[J]. 中国公共卫生,2015,31(1):104-108.

[8] 焦辛妮,汪东伟,王长义,等. 结构方程模型 GLS 与 WLS 性能比较[J]. 中国公共卫生,2015,31(9):1225-1229.

[9] 李国璋,江金荣,陈敏. 协整理论与误差修正模型在实证应用中几个问题的研究[J]. 统计与信息论坛,2010,25(4):21-24.

[10] 梅树江,周志峰,马汉武,等. 深圳市 ARIMA 在肾综合征出血热发病预测中应用[J]. 中国公共卫生,2015,31(7):936-938.

[11] 王瑞泽,周观君,张丽峰. 协整理论在中国的应用情况研究[J]. 统计与信息论坛,2007,22(4):21-25.

[12] Castillo MG, Ahmadi-Abhari S, Bandosz P, et al. Predicting future trends in disability individuals with cardiovascular disease: a modelling study to 2030[J]. Circulation, 2017, 135(s1):336-338.

[13] 李宏坤,何德鲁,张志新,等. 基于状态空间模型的可靠性评估方法[J]. 振动与冲击,2016,35(1):118-124.

[14] 张福利. 状态空间模型在季节性时间序列中的应用[D]. 大连:大连海事大学硕士学位论文,2011:1-65.

[15] 唐波,闫彬彬. 基于状态空间模型的我国卫生总费用分析[J]. 中国卫生经济,2014,33(3):50-51.

收稿日期: 2017-05-08　　(郑新编辑　郭薇校对)

中国公共卫生 CPH