

• 综述 •

空间回归方法在慢性病研究中的应用进展*

叶子容^{1,2} 陈佳^{1,2} 周霁^{1,2} 方亚^{1,2,Δ}

随着社会经济的发展和人们生活方式的转变,慢性病发病率不断攀升,已成为严重困扰个人、家庭和社会的焦点问题。在世界范围内,2010 年估计有 3450 万人死于慢性病,较 20 年前有明显的增加^[1]。目前我国慢性病患者已超过 2 亿人,占总人口的 20% 以上,其中 2012 年我国 18 岁及以上成人高血压、糖尿病、慢性阻塞性肺病患病率分别为 25.3%、9.7%、9.9%,各类慢性病患者率居高不下^[2]。更为严重的是,慢性病的病程长,难治愈,严重影响居民的生活质量。除了慢性病本身,其并发症的防制也耗费大量的资源,给个人、家庭及社会带来的经济负担已成为一个突出问题^[3]。为了能够更好地预防慢性病,该领域的相关研究一直倍受关注。近年来,有研究表明慢性病的发生往往呈现空间聚集性,其发病存在明显的地区差异,说明疾病的发生发展与地理环境因素紧密相关^[4]。而在慢性病危险因素的研究中,传统回归方法忽略了与疾病发生息息相关的地理信息^[5-6],同时,传统方法往往基于研究样本相互独立的假定,缺乏分析具有空间自相关性数据的能力。空间回归方法则可以很好地解决上述问题,其目的在于探索慢性病的发生与空间相关变量的关系,并与 GIS (geographic information system) 结合,将结果以可视化的形式呈现。另外,空间回归可以有效控制空间混杂因素,是慢性病发病因果关系研究的一种推进。

空间回归方法自 20 世纪 90 年代被广泛应用于健康相关问题的研究,尤其是研究传染病的影响因素^[7-9]。随着慢性病成为人类健康的第一大威胁,空间回归方法也逐步被应用于慢性病的研究,这些方法可以依据其参数估计的原理归为两类,即基于频率估计的空间回归模型和基于贝叶斯估计的空间回归模型,前者主要包括多水平模型、空间自回归模型、地理加权回归模型,后者包括 BYM 模型 (Besag York and Mollié Model) 和广义结构可加回归模型。本文旨在梳理空间回归方法在慢性病研究中的应用进展,并比较各方法的优缺点。

* 基金项目:厦门市科技计划项目(350Z20143006)

1. 厦门大学公共卫生学院(361102)

2. 福建省高校卫生技术评估重点实验室

Δ 通信作者:方亚, E-mail: fangya@xmu.edu.cn.

基于频率估计的空间回归模型

1. 多水平模型

多水平模型适用于具有层次结构特征的数据,其将传统回归方法原来的单一随机误差分解到与数据层次结构相对应的各水平上,使得该方法比一般线性回归模型放松了样本相互独立的前提假定。因此,多水平模型广泛应用于慢性病危险因素的空间研究,它同时考虑了个体水平和区域水平的危险因素对慢性病发病的影响,具有识别区域水平影响因素效应的能力^[10]。其中,以多水平 logistic 回归模型应用最为广泛^[11-14],如 Yu Xu 等人基于中国非传染病监测系统开展的糖尿病与居住地、居住地经济水平、社会环境等区域影响因素的关系研究^[12]。同时,在研究单位时间、单位空间内某种慢性病发病的影响因素时,由于事件发生概率较小,这类数据往往服从 Poisson 分布,需要进一步采用多水平 Poisson 回归模型,如 A. Bocquier 等人开展的根据小区域尺度的地理因素对糖尿病治疗率影响的研究^[15]。

然而,基于纵向数据的慢性病危险因素研究在因果推断的研究上更具有说服力,为实现这一目的,Maigeng Zhou 等人应用多水平负二项回归模型,基于全国代表性数据,首次同时研究时间和空间变量对糖尿病死亡率的影响^[16]。该研究将以往局限于截面数据的多水平回归模型拓展到了基于纵向数据的应用,实现了基于时空二维数据的危险因素探索。

2. 空间自回归模型

在分析具有区域与个体两个层次的数据时,多水平模型可以很好地解决样本间互不独立的问题,但是,很多流行病学研究,地理区域划分清晰、互相接壤,并且互相影响,存在较强空间自相关性,需要空间回归的研究方法将地理空间当成互有联系的整体,以充分描述和解释空间变量对慢性病的影响。其中,空间自回归模型(simultaneous autoregressive models SAR)就是一种解决空间自相关性的回归模型,它包括含因变量空间滞后项的空间滞后模型(spatial lag model, SLM)、含误差项空间滞后项的空间误差模型(spatial error model, SEM)及同时含有空间滞后项和误差项空间滞后项的空间杜宾模型(spatial Durbin model,

SDM)^[17]。实际研究中,可根据残差是否独立、拉格朗日乘数和稳健拉格朗日乘数是否有统计学意义、 R^2 大小、信息准则量(AIC、SC)大小等对模型进行评价及筛选,最终将模型设定为SLM、SEM或SDM^[18]。SAR模型因此也被用于解决慢性病空间数据存在空间自相关性的问题^[19-21],如基于小尺度区域,Adam Drewnowski等人应用SAR模型研究糖尿病发病率与社会经济水平和饮食环境变量的关系,该研究很好地克服了空间自相关的问题,最终为制定以社区为基础的糖尿病预防控制措施提供依据^[19]。

3. 地理加权回归模型

SAR模型考虑了邻近区域因变量互相之间潜在的影响,但是其回归系数不随空间位置的变化而变化,因此其本质是一个全局回归模型。而慢性病的发生往往存在外生性相互作用的效应,即某研究单元的因变量受邻近单元自变量影响^[22]。为了解决这一问题,Brunsdon等人于1998年提出了地理加权回归模型(geographically weighted regression,GWR)^[23]。GWR的原理是,基于空间研究局部光滑的思想,将数据的空间结构嵌入回归模型中,使得模型回归参数变为观察点的地理位置函数^[24]。与OLS回归全局估计相比,该模型的优势是,它是一种局部估计模型,可以充分展现影响因素对疾病发生发展的影响随空间变化的趋势,因此,GWR模型在慢性病的研究中常用于空间异质性的诊断,以及探索慢性病的发病和社会人口学、社会经济学、环境相关变量及食物可及性等因素的关系^[25-28]。如Carlos Siordia等人利用GWR模型,研究社会经济学变量与糖尿病的关系,最终发现某些区域贫穷和糖尿病之间没有相关关系,推翻了传统研究中对于贫穷与糖尿病关系的认识^[25]。可见,GWR模型正是利用局部回归系数解释了自变量和因变量之间关系的空间异质性。Aletta Dijkstra等人及Chan等人更是分别从预测准确率和模型拟合优度两个角度对GWR模型与OLS回归进行比较,发现GWR模型均优于OLS回归^[27-29]。

但是GWR也存在不足,其一,模型源于一般线性模型,所以不适用于事件发生数较小的慢性病研究;其次,GWR模型假设所有研究自变量与因变量之间均存在空间异质性,而这与实际情况相背离。针对上述问题,Nakaya等人对GWR模型进行了延伸,提出地理加权泊松回归(geographically weighted Poisson regression,GWPR),以及对GWPR模型的延伸模型即半参数地理加权泊松回归(semi-parametric GWPR,sGWPR),该模型同时考虑了区域可变回归系数和全局稳定回归系数两部分^[30]。

需要注意的是,无论是GWR还是GWPR,在建模过程中都面临两个重要而困难的问题:空间权重函数的

设定和最优宽带的选择。其中,较为普遍的空间权重函数有三种^[23],见表1。

表1 空间权重函数比较

名称	函数公式
距离阈值法	$W_{ij} = \begin{cases} 1 & d_{ij} < \theta \\ 0 & d_{ij} > \theta \end{cases}$
高斯函数法	$W_{ij} = \exp(-d_{ij}^2/\theta^2)$
bi-square 函数法	$W_{ij} = \begin{cases} (1 - d_{ij}^2/\theta^2)^2 & d_{ij} < \theta \\ 0 & d_{ij} > \theta \end{cases}$

d_{ij} 是点*i*和*j*在空间上的距离,而 θ 是核宽带。距离阈值法是最简单的空间权重函数,但是其空间权重在距离 θ 点出现突变,与空间的平滑性和渐变性相悖,较少被使用。而bi-square函数法因其计算消耗较小,往往在大样本分析中优于高斯函数。事实上,相比于空间权重函数的选择,不同宽带对于模型最后的估计结果影响更大。宽带决定了核区域的大小,所以,较大的宽带所估计的参数具有较小的空间差异,而较小的宽带则会带来较大的局部变异,使得参数估计结果的置信区间变大。对于宽带的选择,目前缺乏统一的标准,研究者常以最小赤池信息准则(AIC)来选择最优宽带^[31]。

基于贝叶斯估计的空间回归模型

由于每一个区域所特有的信息可以当作先验信息,因此贝叶斯模型具备研究空间问题的能力。同时,贝叶斯模型可以检验出疾病与危险因素之间是否具有空间相关效应,进而有选择地估计变量的空间效应,从而优于前述频率估计方法的空间回归模型。此外,基于充分利用空间相关的先验信息,贝叶斯模型对于粗糙的及缺失的数据有更稳定的参数估计^[32]。

1. BYM 模型及其拓展

1991年,Besag York和Mollié等对前人的模型进行了拓展,将空间结构化与非结构化的异质性放入模型中,新衍生的模型被Mollie称为卷积模型(convolution model)。后人为了纪念两位科学家的贡献,将该模型命名为BYM模型^[33]。早在2007年,BYM模型就已运用于慢性病危险因素及发病率估计的研究^[34-35]。近年来,在慢性病的空间研究中,为了同时考虑区域水平和个体水平危险因素的空间相关性,BYM分层模型同时纳入固定效应和随机效应来估计疾病危险因素的影响^[36-38]。

进一步地,人类多发的几种慢性病往往具有相似危险因素、相似空间分布规律,这要求研究者在开展慢性病空间研究时,能够同时纳入多种疾病进行研究,探索疾病间的相互关系。而上述研究多以一种疾病作为目标疾病,为弥补这一不足,有研究者基于BYM框

架,应用多疾病自回归模型(multivariate conditional autoregressive regression, MCAR)或共同成分模型(shared component model, SCM),达到分析两种或多种疾病空间模式的目的^[39-42]。值得一提的是,在近年来的慢性病空间研究中,SCM 受到了越来越多的关注。该模型的基本思想是,许多慢性病有着共同的危险因素,从而潜在的风险变异可以被分解成共同的和特定的风险变异两部分,其中,共同成分反映多种疾病的共同风险因子,特有成分反映每种疾病特有的随空间变化的风险因子。Feltbower 等将 SCM 用于研究儿童急性粒细胞白血病和 I 型糖尿病空间分布的相似性^[39],Havulinna 等将 SCM 用于缺血性中风与急性心力衰竭的空间相似性研究^[40]。不仅如此,相比于生态学回归以及 MCAR,SCM 估计结果更精确,拟合效果更优^[43]。近年来,对于 SCM 模型的应用有了进一步发展,该方法逐步聚焦于慢性病发病的性别差异研究^[43-44]。总体而言,SCM 的结果不仅直观展现了多种疾病共有的相对风险,还展示每一种疾病特有的相对风险,该方法也因此成为病因探索性分析的重要工具^[45]。

但是,BYM 和 SCM 在参数估计过程中均涉及高维计算,所以在实际应用中常采用马尔科夫链蒙特卡罗模拟(MCMC)的方法进行参数估计。进一步地,有研究者认为参数先验分布的选择是关键,若选择不恰当的先验信息,可能导致不合理的后验分布结果^[46],因此,往往需要进行敏感性分析,以确保参数估计结果的稳定性。常用的先验分布包括,高斯分布、均匀分布、伽马分布等,近年来研究发现,当研究区域或研究的组数较少时,非中心 t 分布或半柯西分布是更好的选择^[47]。

2. 广义结构可加回归

在慢性病的研究领域,无论是基于频率估计的空间回归模型还是 BYM 及其拓展的回归方法都很好地解决了空间自相关和空间异质性问题。然而,慢性病的发生发展除了存在不同空间分布的特点外,在时间维度上的流行趋势也值得关注。因此,空间研究方法需要同时具备控制区域自相关性和时间自相关性的能力。其次,在慢性病发病率预测及危险因素研究中,仅仅控制空间自相关性还远远不够,研究者期望在控制空间自相关性时,能够量化这种空间差异,从而揭示疾病发病风险的空间模式。为了实现上述两个目标,在广义可加模型(generalized additive models, GAM)的基础上利用马尔科夫链蒙特卡罗模拟进行贝叶斯推断的回归方法得到发展,该回归模型被命名为广义结构可加回归(generalized structured additive regression, STAR)^[48]。STAR 可以同时纳入具有非线性随机效应的时间协变量以及空间协变量,并利用惩罚曲线、马

尔科夫随机场、高斯随机场等方法来拟合协变量的随机效应^[49]。Chien 等人多次将 STAR 模型运用于慢性病的研究中,于 2013-2015 年基于美国郡县水平先后研究乳腺癌病死率的时空特征^[50],美国哮喘发病的时空差异^[51],PM2.5 对糖尿病的影响^[52]。在上述研究中,研究者分别运用二阶段样条的惩罚曲线来估计时间协变量的随机效应,以及马尔科夫随机场来拟合模型的空间效应,在控制时间相关和空间相关的同时,也量化了疾病发生风险的时间差异和空间差异。

小 结

近年来,空间回归方法被逐步运用于慢性病发病率的预测、危险因素探讨、区域分布模式探索等研究。在不断探索的过程中,空间回归方法也得到长足的发展。始于多水平模型,放松了经典 OLS 模型样本独立的假设,使得用于空间研究的数据得到更准确及稳定的估计;之后,以 SAR 模型及 GWR 模型为代表的生态学回归则充分考虑了空间自相关性和相邻区域的相互作用,是空间回归研究方法的一大进步。随后,空间回归方法的发展从基于频率估计转向基于贝叶斯估计的回归模型的发展。基于贝叶斯估计的方法,从 BYM 到 SCM 等,都很好地克服了空间数据有缺失、较粗糙等问题,保证了估计结果的稳定。STAR 模型同时考虑时间效应和空间效应,很好地将时间空间结合在一起,既控制了时间、空间的相关性对模型估计结果的影响,也量化了疾病发生发展在时空维度上的差异。

总体而言,随着空间统计分析软件的革新、数据采集空间定位的精细化、空间统计学研究人员水平的提升,空间回归方法在慢性病研究中的应用也不断发展,主要趋势有:第一,加强获取时空维度的数据,开展时间空间相结合的研究,以便深入理解疾病的病因及个体与环境的作用机制;第二,结合健康及流行病学的理论,进一步探索空间分析中生成空间权重矩阵的方法,并证实其合理性;第三,提高数据搜集的精细化,进一步细化空间分析尺度,基于空间层面开展探索性研究;最后,不断积累空间数据、时空数据,继续推动空间大数据、时空大数据的发展。

参 考 文 献

- [1] Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380(9859): 2095-2128.
- [2] 中华人民共和国国家卫生和计划生育委员会.《中国居民营养与慢性病状况报告(2015)》新闻发布会文字实录. *中国实用乡村医生杂志* 2015 (15): 1-5.
- [3] 滕海英, 许丁才, 熊林平, 等. 西安市社区老年人慢性病医疗需求与负担调查分析. *中国卫生统计* 2013; 30(2): 259-260.

- [4] Caswell JM. Prevalence of reported high blood pressure in Canada: investigation of demographic and spatial trends. *Journal of Public Health* 2016; 1-11.
- [5] 刘彩萍, 张嫣平, 胡皓, 等. 宁夏银川市某城区流动人口慢性病及影响因素调查研究. *中国卫生统计* 2015 03: 477-478.
- [6] Yang W, Lu J, Weng J et al. Prevalence of Diabetes among Men and Women in China. *New England Journal of Medicine*, 2010, 362 (12): 1090-1101.
- [7] Jia Z, Jia X, Liu Y et al. Spatial analysis of tuberculosis in migrants and residents, Beijing, 2000-2006. *Emerg Infect Dis*, 2008, 14 (9): 1413-1419.
- [8] Maciel ELN, Pan W, Dietze R et al. Spatial patterns of pulmonary tuberculosis incidence and their relationship to socio-economic status in Vitoria, Brazil. *The International Journal of Tuberculosis and Lung Disease* 2010, 14 (11): 1395-1402.
- [9] Xia J, Cai S, Zhang H et al. Spatial, temporal, and spatiotemporal analysis of malaria in Hubei Province, China from 2004-2011. *Malar J* 2015, 14 (1): 145.
- [10] Baker J, White N, Mengersen K. Spatial modelling of type II diabetes outcomes: a systematic review of approaches used. *Royal Society open science* 2015 2(6): 140460.
- [11] Zhou M, Astell-Burt T, Bi Y et al. Geographical variation in diabetes prevalence and detection in china: multilevel spatial analysis of 98,058 adults. *Diabetes Care* 2015 38(1): 72-81.
- [12] Xu Y, Wang L, He J, et al. Prevalence and Control of Diabetes in Chinese Adults. *Jama* 2013 310(9): 948-958.
- [13] Astell-Burt T, Feng X, Kolt GS et al. Understanding geographical inequities in diabetes: multilevel evidence from 114,755 adults in Sydney, Australia. *Diabetes Res Clin Pract* 2014, 106(3): e68-e73.
- [14] Gonzalez G, Wilson SMW, Thorpe RJ. Examining Place As a Social Determinant of Health Association Between Diabetes and US Geographic Region Among Non-Hispanic Whites and a Diverse Group of Hispanic/Latino Men. *Family & Community Health* 2015 38(4SI): 319-331.
- [15] Bocquier A, Cortaredona S, Nauleau S et al. Prevalence of treated diabetes: Geographical variations at the small-area level and their association with area-level characteristics. A multilevel analysis in South-eastern France. *Diabetes Metab* 2011 37(1): 39-46.
- [16] Zhou M, Astell-Burt T, Yin P et al. Spatiotemporal variation in diabetes mortality in China: multilevel evidence from 2006 and 2012. *BMC Public Health* 2015, 15(1): 633.
- [17] 黄秋兰, 唐咸艳, 周红霞, 等. 四种空间回归模型在疾病空间数据影响因素筛选中的比较研究. *中国卫生统计*, 2013, 30(3): 334-338.
- [18] Anselin L. Exploring spatial data with GeoDaTM: a workbook. *Center for Spatially Integrated Social Science* 2005: 198-199.
- [19] Drewnowski A, Rehm CD, Moudon AV et al. The Geography of Diabetes by Census Tract in a Large Sample of Insured Adults in King County, Washington, 2005-2006. *Preventing Chronic Disease*, 2014, 11(1): E125.
- [20] Green C, Hoppa RD, Young TK et al. Geographic analysis of diabetes prevalence in an urban area. *Social Science & Medicine* 2003 57(3): 551-560.
- [21] 热伊拉·吾斯曼. 应用地理信息系统技术分析新疆糖尿病的空间分布. *新疆医科大学* 2015.
- [22] Yang T, Shoff C, Noah AJ. Spatialising health research: what we know and where we are heading. *Geospatial Health* 2013 7(2): 161-168.
- [23] Brunson C, Fotheringham S, Charlton M. Geographically weighted regression - modelling spatial non-stationarity. *The Statistician*, 1998, 47(3): 431-443.
- [24] 覃文忠. 地理加权回归基本理论与应用研究. 同济大学 2007.
- [25] Siordia C, Saenz J, Tom SE. An Introduction to Macro-Level Spatial Nonstationarity: a Geographically Weighted Regression Analysis of Diabetes and Poverty. *Human geographies* 2012 6(2): 5-13.
- [26] Margolis D J, Hoffstad O, Nafash J, et al. Location, Location, Location: Geographic Clustering of Lower-Extremity Amputation Among Medicare Beneficiaries With Diabetes. *Diabetes Care* 2011 34(11): 2363-2367.
- [27] Chan TC, Chiang PH, Su MD et al. Geographic disparity in chronic obstructive pulmonary disease (COPD) mortality rates among the Taiwan population. *PLoS One* 2014 9(5): e98170.
- [28] Park J, Yoon S, Na MH et al. The effects of air pollution on mortality in South Korea. *Procedia Environmental Sciences*, 2015, 26: 62-65.
- [29] Dijkstra A, Janssen F, De Bakker M, et al. Using Spatial Analysis to Predict Health Care Use at the Local Level: A Case Study of Type 2 Diabetes Medication Use and Its Association with Demographic Change and Socioeconomic Status. *Plos One* 2013 8(8): e72730.
- [30] Nakaya T, Fotheringham AS, Brunson C, et al. Geographically weighted Poisson regression for disease association mapping. *Statistics in Medicine* 2005 24(17): 2695-2717.
- [31] Siordia C, Saenz J, Tom SE. An Introduction to Macro-Level Spatial Nonstationarity: a Geographically Weighted Regression Analysis of Diabetes and Poverty. *Human geographies* 2012 6(2): 5-13.
- [32] Baker J, White N, Mengersen K. Spatial modelling of type II diabetes outcomes: a systematic review of approaches used. *R Soc Open Sci*, 2015 2(6): 140460.
- [33] Besag J, York J, Molli A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 1991 43(1): 1-20.
- [34] Prieto RR, Garcia-Perez J, Pollan M, et al. Modelling of municipal mortality due to haematological neoplasias in Spain. *Journal of Epidemiology and Community Health* 2007 61(2): 165-171.
- [35] Dominguez-Berjon MF, Gandarillas A, Segura Del Pozo J et al. Census tract socioeconomic and physical environment and cardiovascular mortality in the Region of Madrid (Spain) . *Journal of Epidemiology and Community Health* 2010 64(12): 1086-1093.
- [36] Hao Y, Balluz L, Strosnider H et al. Ozone, Fine Particulate Matter, and Chronic Lower Respiratory Disease Mortality in the United States. *American Journal of Respiratory and Critical Care Medicine*, 2015, 192(3): 337-341.
- [37] Hajna S, Ross NA, Joseph L, et al. Neighbourhood Walkability and Daily Steps in Adults with Type 2 Diabetes. *Plos One* 2016, 11(3): e151544.
- [38] Goungounga JA, Gaudart J, Colonna M, et al. Impact of socioeconomic inequalities on geographic disparities in cancer incidence: comparison of methods for spatial disease mapping. *BMC Medical Research Methodology* 2016, 16(1): 136.
- [39] Feltbower RG, Manda S, Gilthorpe MS, et al. Detecting small-area similarities in the epidemiology of childhood acute lymphoblastic leukemia and diabetes mellitus, type 1: A Bayesian approach. *American*

- Journal of Epidemiology 2005 ,161(12) : 1168-1180.
- [40] Havulinna AS ,Paakkonen R ,Karvonen M ,et al. Geographic patterns of incidence of ischemic stroke and acute myocardial infarction in Finland during 1991-2003. *Annals of Epidemiology* ,2008 ,18(3) : 206-213.
- [41] Manda SOM ,Feltbower RG ,Giltorpe MS. Investigating spatio-temporal similarities in the epidemiology of childhood leukaemia and diabetes. *European Journal of Epidemiology* 2009 24(12) : 743-752.
- [42] Tassone EC ,Waller LA ,Casper ML. Small-area racial disparity in stroke mortality: an application of Bayesian spatial hierarchical modeling. *Epidemiology (Cambridge ,Mass.)* 2009 20(2) : 234.
- [43] Ibanez-Beroiz B ,Librero-Lopez J ,Peiro-Moreno S ,et al. Shared component modelling as an alternative to assess geographical variations in medical practice: gender inequalities in hospital admissions for chronic diseases. *BMC Medical Research Methodology* 2011 ,11(1) : 1.
- [44] Xu L ,Lai D ,Fang Y. Spatial analysis of gender variation in the prevalence of hypertension among the middle-aged and elderly population in Zhejiang Province ,China. *BMC Public Health* 2016 ,16(1) : 447.
- [45] Onicescu G ,Hill EG ,Lawson AB ,et al. Joint disease mapping of cervical and male oropharyngeal cancer incidence in blacks and whites in South Carolina. *Spat Spatiotemporal Epidemiol* ,2010 ,1(2-3) : 133-141.
- [46] Gustafson P ,Hossain S ,MacNab YC. Conservative prior distributions for variance parameters in hierarchical models. *Canadian Journal of Statistics-revue Canadienne de Statistique* 2006 34(3) : 377-390.
- [47] Nathoo FS ,Ghosh P. Skew-elliptical spatial random effect modeling for areal data with application to mapping health utilization rates. *Stat Med* 2013 32(2) : 290-306.
- [48] Fahrmeir L ,Lang S. Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* ,2001 ,50(2) : 201-220.
- [49] Kneib T ,Fahrmeir L. Structured additive regression for categorical space-time data: A mixed model approach. *Biometrics* 2006 62(1) : 109-118.
- [50] Chien L ,Yu H ,Schootman M. Efficient mapping and geographic disparities in breast cancer mortality at the county-level by race and age in the US. *Spatial and Spatio-temporal Epidemiology* 2013 5: 27-37.
- [51] Chien L ,Alamgir H. Geographic disparities of asthma prevalence in south-western United States of America. *Geospatial Health* ,2014 ,9(1) : 97-108.
- [52] Chien L ,Alamgir H ,Yu H. Spatial vulnerability of fine particulate matter relative to the prevalence of diabetes in the United States. *Science of the Total Environment* 2015 508: 136-144.

(责任编辑: 郭海强)

(上接第 136 页)

在这个过程中不可避免地会有可观测的和不可观测的混杂因素同处理因素共同对结局变量产生影响。可观测的混杂因素产生的偏倚已经有较多的手段去处理。而不可观测的混杂因素 往往是个体异质性 人们通常对其分布状况是没有先验信息 实验设计和观察期间也没有办法对其进行有效的控制。在统计过程中直接估计出来的难度非常大 因此在处理效应的估计中目前主流的做法是将不可观测的混杂在估计的过程中去掉。本文采用的处理效应方法在假设个体异质性的离散程度较大的前提下 对带有不可观测混杂的两值结局变量的处理效应进行估计具有较好的估计性质 比较简便可行。

参 考 文 献

- [1] Rubin DB. Inference and missing data (with discussion) . *Biometrika* ,1976(63) : 581-592.
- [2] Rubin DB. Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics* ,1978(6) : 34-58.
- [3] Wooldridge JM. *Econometric analysis of cross section and panel data*. The MIT press 2002.
- [4] Rosenbaum PR ,Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* ,1983(70) : 41-55.
- [5] Luo Z ,Gardiner JC ,Bradley CJ. Applying Propensity Score Methods in Medical Research: Pitfalls and Prospects. *Medical Care Research and Review: formerly Medical Care Review* 2010 67(5) : 528.
- [6] Vansteelandt S ,Daniel RM. On regression adjustment for the propensity score. *Statistics in Medicine* 2014(33) : 4053-4072 ,
- [7] 张亮 李婵娟 夏结来 等. 倾向得分区间匹配法用于非随机对照试验的探索与研究. *中国卫生统计* 2012 29(1) : 53-57.
- [8] 别玉坤 杨成林 曹卫 等. 倾向得分匹配法对两种方法治疗胆总管结石的疗效再评价. *中国普通外科杂志* ,2014 ,23(8) : 1059-1062.
- [9] KD Song ,HK Lim ,H Rhim 等. 肝切除术复发肝癌反复肝切除术与射频消融术的比较: 倾向得分匹配研究. *国际医学放射学杂志* 2015 38(4) : 378.
- [10] Takuma Y ,Takabatake H ,Morimoto Y ,et al. 采用倾向得分匹配法比较符合米兰标准肝癌病人的肝动脉化栓栓塞联合射频消融术与手术切除. *国际医学放射学杂志* 2014 (1) : 82.
- [11] Gao W ,Bergsma W ,Yao Q. Estimation for Dynamic and Static Panel Probit Models with Large Individual Effects. *Journal of Time* 2016.
- [12] Arellano M. Discrete choices with panel data . *Investigaciones Economicas* 2001 27: 423-458
- [13] Bartolucci F ,Farcomeni A. A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association* 2009 , 104(486) : 816-831.
- [14] Honoré BE ,Kyriazidou E. Panel data discrete choice models with lagged dependent variables. *Econometrica* 2000 68(4) : 839-874.
- [15] Lancaster T. The incidental parameter problem since 1948. *Journal of econometrics* 2000 95(2) : 391-413.

(责任编辑: 郭海强)