·基础理论与方法·

生命历程流行病学中的动态路径分析

田征文　曾广宇　吴诗蓝　黄麟婷　王贝子　谭红专
410008 长沙，中南大学湘雅公共卫生学院流行病与卫生统计学系
通信作者：谭红专，Email：tanhz99@qq.com
DOI：10.3760/cma.j.issn.0254-6450.2018.01.018

【摘要】 现代流行病学研究中，疾病或健康相关事件的发生常难以完全用短期的暴露状态来解释。生命历程流行病学着眼于生命早期阶段的暴露因素对个人整个生命历程中的健康或疾病状况所产生的长期影响，并逐渐得到重视。当对暴露因素的病因机制及其通过其他因素产生的作用大小进行分析时，由于时间因素的存在，传统统计分析方法难以满足生命历程流行病学中病因分析的需求。本文概述了能用于生命历程病因分析的动态路径分析方法，包括该模型的结构、意义及其在生命历程流行病学病因分析中的应用。同时说明了如何准备数据、进行病因机制分析，并证明动态路径分析模型可作为生命历程流行病学中有效的病因分析工具。

【关键词】 时依性变量；生存分析；有向无环图；生命历程流行病学

Dynamic path analysis on life course epidemiology

Tian Zhengwen, Zeng Guangyu, Wu Shilan, Huang Linting, Wang Beizi, Tan Hongzhuan
Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha 410008, China
Corresponding author: Tan Hongzhuan, Email: tanhz99@qq.com

【**Abstract**】 In the studies of modern epidemiology, exposure in a short term cannot fully elaborate the mechanism of the development of diseases or health-related events. Thus, lights have been shed on to life course epidemiology, which studies the exposures in early life time and their effects related to the development of chronic diseases. When exploring the mechanism leading from one exposure to an outcome and its effects through other factors, due to the existence of time-variant effects, conventional statistic methods could not meet the needs of etiological analysis in life course epidemiology. This paper summarizes the dynamic path analysis model, including the model structure and significance, and its application in life course epidemiology. Meanwhile, the procedure of data processing and etiology analyzing were introduced. In conclusion, dynamic path analysis is a useful tool which can be used to better elucidate the mechanisms that underlie the etiology of chronic diseases.

【**Key words**】 Time-dependent covariate; Survival analysis; Directed acyclic graphs; Life course epidemiology

现代流行病学研究发现，许多研究所关注的结局，其形成机制难以完全用短期的暴露状态来解释。生命历程流行病学关注的是事件发生发展过程中的长期机制，着眼于生命早期阶段的暴露因素对个人整个生命历程中的健康或疾病状况所产生的长期影响，在慢性病因分析方面有其优势[1-2]。如今已被广泛应用于疾病或健康状态长期机制的研究中。生命历程流行病学研究中的资料都是纵向随访资料，蕴涵有结局和时间两个方面的信息；暴露的协变量是在不同时间点反复测量的，且可能随时间变化；有些数据是不完整的，数据分布类型复杂。

对此类资料进行统计分析时，传统的方法有寿命表法、多元线性回归或Cox回归模型分析。但寿命表不能作多因素分析，多元线性回归往往不能处理截尾数据[3]。Cox比例风险回归模型虽能处理截尾数据，并对数据进行多因素分析，但模型中协变量要满足比例风险性，即假定所研究的人群在任何时间点上，发生事件的风险比例是恒定的（或者解释为某一个暴露在所有时间里，对发生事件的作用都是相同的），而实际研究中常难以满足该要求。且Cox回归在分析数据时，往往只纳入基线测量值，导致重复测量的数据未被充分利用[4-5]，当变量的效应不变，而测量值会随时间改变，即存在内在时依性变量时，其作用常常被低估[6]。此时，虽可使用带有时间

协变量的Cox非比例风险性模型，但是该模型没有考虑到各次观测间相关性的存在[7]，仍存在一定的局限性。本研究将介绍动态路径分析模型，该模型可以有效纳入重复测量的变量值，对暴露与结局之间的潜在关联进行分析，是生命历程流行病学中对时依性变量进行分析的一种有效工具[6]。

一、基本原理

动态路径分析是在Aalen加性模型的基础上[8-9]，由Fosen等[10]提出的将路径分析和Aalen模型相结合，对时依性变量之间的相互作用及其对结局变量的作用进行分析的一种方法。通过构建有向无环图和Aalen加性模型，可将暴露因素的作用分解成直接作用和间接作用并分别进行计算。直接作用主要通过Aalen模型的系数来体现，间接作用主要通过有向无环图的路径系数来体现。

1. 有向无环图：是一种变量之间有方向、不成环、可视化的因果关联路径图。可将各变量的关系用直观可视的图形表示，以梳理各变量间的关系，定性识别混杂因素，确认控制和消除混杂的充分调整集（如果调控其中的某一个或几个变量就足以控制所有混杂，那么由这些需要控制的混杂变量组成的集合称为充分调整集），常用于控制混杂因素[11-13]。鉴于有向无环图的线性及无环结构，可以通过最小二乘法使用回归方程计算每条路径的路径系数，从而可将路径分析视为普通多元回归的延伸[5]。此时，每条路径系数即为变量间的回归系数。

2. Aalen加性模型：可视为Cox模型的延伸，其主要特点是回归系数是随时间变化的函数[14]。该模型是一个半参数线性模型[8-10]，使用随机过程变量$N(t)$表示在时间间隔$(0,t)$内事件A出现的总次数。

在时点t时，对于结局变量$N(t)$的方程：

$$dN(t)=dB_0(t)+dB_1(t)X_1(t)+\cdots+dB_p(t)X_p(t)+dM(t)$$

其中，$dN(t)$是在时点t，所研究的对象发生结局事件A的风险大小。$dB_0(t)$是基线风险率。而$X_1(t)$，$X_2(t)$，…，$X_p(t)$是在时点t，时间因素对变量X_1，X_2，…，X_p产生的效应大小。$dB_1(t)$，…，$dB_p(t)$是回归系数，即时依性变量$X_p(t)$对结局事件的效应。$dM(t)$是鞅增量，指的是该时点t之前，上一次测量时，所研究个体发生结局事件A的整体概率[4]。

由于$dB_p(t)$的变异性较大，常用做法是计算累计回归系数$B_p(t)=\int_a^b dB_p(t)$。累积回归系数的含义为在特定研究时点$(a,b]$之间，当相关协变量的值变化1个单位，所研究的人群相比其他人群而言，出现结局事件的额外风险。在动态路径分析中，通常对$B_p(t)$计算其累积直接效应。同时，变量之间的效应（即有向无环图中每条路径的路径系数）为$\varphi_p(t)$，可通过一般路径分析计算回归系数。此时暴露因素通过时依性中介变量对结局事件的间接效应$dB_p(t)'$为$\varphi_p(t)dB_p(t)$。同样的，此处间接效应为累积间接效应。

二、动态路径分析的实施

以研究BMI与冠心病（CHD）的关系为例来说明动态路径分析。我们假设BMI对CHD有作用，这种作用可能包括直接作用和间接作用（可能通过影响SBP而影响CHD的发生）。我们对一个队列人群从生命早期开始追踪观察，在不同年龄阶段测量每个对象的BMI、SBP和CHD。由于个体的体重和血压在不同时间的测量值会有变化，因此均为时间依存变量。

1. 数据准备：若存在数据删失，则使用最近一次的测量值对数据进行填补。由于时依变量存在，把同一个研究对象的随访分成若干条记录，将两次随访之间的时间记为一个时间间隔，间隔以年为单位。为减少多次测量结果抽样误差的影响，将测量结果转换成标准化得分，得到z值。通过数据变换，剔除了量纲。

$$z=(x-\mu)/\sigma$$

2. 绘制有向无环图：根据先验知识，且为了简化研究内容，假定BMI、SBP及是否出现CHD之间存在如下动态路径关联（图1）。

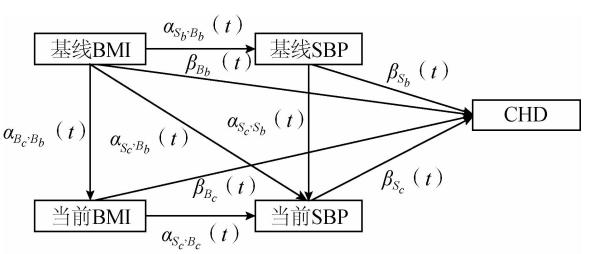

图1 BMI、SBP及CHD之间的动态路径（有向无环图）

3. 直接效应的估计：从路径图可知，直接效应包括从基线BMI到CHD$[\beta_{B_b}(t)]$，当前BMI到CHD$[\beta_{B_c}(t)]$，从基线SBP到CHD$[\beta_{S_b}(t)]$，及从当前SBP到CHD$[\beta_{S_c}(t)]$，都是各自对CHD的直接效应。

变量间存在如下关系：

$$\alpha(t)=\beta_0(t)+\beta_{B_b}(t)\ \mathrm{BMI}_b+\beta_{B_c}(t)\ \mathrm{SBP}_b+\beta_{S_b}(t)\ \mathrm{BMI}_c(t)+\beta_{S_c}(t)\ \mathrm{SBP}_c(t) \quad (1)$$

其中，$\alpha(t)$为个体在时点t出现CHD的风险，$\beta_0(t)$为基线风险率，即当余下所有变量取值为0时，研究对象出现CHD的风险大小。$\beta_{B_b}(t)$、$\beta_{B_c}(t)$、$\beta_{S_b}(t)$、$\beta_{S_c}(t)$分别为不同时点t时，协变量对结局变量的效应。这里的直接效应是通过Aalen加性模型估计，具体分析过程可使用R语言软件timereg程序包实现。由于Aalen模型中的效应为随时间变化的函数，因此，可以对其求累积风险系数，$B_p(t)=\int_a^b \mathrm{d}B_p(t)$。即在时点$a$到时点$b$这段时间内，当相关协变量的值变化1个单位，所研究的人群相比其他人群而言，发生心血管疾病的额外风险。此时所计算的效应为暴露因素对结局变量的直接效应。

4. 间接效应的估计：从路径图可知，从基线BMI到CHD的间接效应路径包括：①基线BMI通过当前BMI到CHD；②基线BMI通过基线SBP，到当前SBP，到CHD；③基线BMI通过基线SBP到CHD；④基线BMI，通过当前SBP到CHD；⑤基线BMI，通过当前BMI到当前SBP，再到CHD。上述5条间接路径中，还有5条路径系数是未知的，这些间接路径系数可通过一般通径分析模型进行拟合计算。

通过式2可估计基线BMI到基线SBP的效应$[\alpha_{S_bB_b}(t)]$：

$$\mathrm{SBP}_b=\alpha_{S_b}(t)+\alpha_{S_bB_b}(t)\ \mathrm{BMI}_b+\varepsilon_2(t) \quad (2)$$

通过式3可估计基线BMI到当前BMI的效应$[\alpha_{B_cB_b}(t)]$：

$$\mathrm{BMI}_c(t)=\alpha_{B_c}(t)+\alpha_{B_cB_b}(t)\ \mathrm{BMI}_b+\varepsilon_3(t) \quad (3)$$

通过式4可估计基线BMI到当前SBP的效应$[\alpha_{S_cB_b}(t)]$，基线SBP到当前SBP的效应$[\alpha_{S_cS_b}(t)]$及当前BMI到当前SBP的效应$[\alpha_{S_cB_c}(t)]$：

$$\mathrm{SBP}_c(t)=\alpha_{S_c}(t)+\alpha_{S_cB_b}(t)\ \mathrm{BMI}_b+\alpha_{S_cS_b}(t)\ \mathrm{SBP}_p+\alpha_{S_cB_c}(t)\ \mathrm{BMI}_c(t)+\varepsilon_4(t) \quad (4)$$

若一条间接效应路径包括两个或以上的路径效应（路径系数），通过将从暴露因素到结局变量的同一条路径上的各个系数相乘，可得到所研究的暴露因素通过其他因素对结局因素产生的影响大小。如间接路径1，基线BMI通过当前BMI到CHD，就包括从基线BMI到当前BMI的效应$[\alpha_{B_cB_b}(t)]$及从当前BMI到CHD的效应$[\beta_{B_c}(t)]$，该路径的总效应是这两个效应之积。同样，间接路径上的各系数也需要利用$B_p(t)=\int_a^b \mathrm{d}B_p(t)$，计算累积风险系数，即将不同时间段的风险累积。

5. 总效应的估计：所研究的暴露因素对结局变量的总效应即为直接效应和间接效应之和。在本例，可将基线BMI对是否发生CHD的总效应总结为：

直接效应（A）：$\beta_{B_b}(t)$

间接效应1（B）：$\alpha_{B_cB_b}(t)\beta_{B_c}(t)$

间接效应2（C）：$\alpha_{S_bB_b}(t)\alpha_{S_cS_b}(t)\beta_{S_c}(t)$

间接效应3（D）：$\alpha_{S_bB_b}(t)\beta_{S_b}(t)$

间接效应4（E）：$\alpha_{S_cB_b}(t)\beta_{S_c}(t)$

间接效应5（F）：$\alpha_{B_cB_b}(t)\alpha_{S_cB_c}(t)\beta_{S_c}(t)$

总效应＝A＋B＋C＋D＋E＋F

三、总结

生命历程理论关注人群健康的上游因素，扩大了关注面[1]。目前，生命历程理论研究中还存在挑战，如较难对动态变化的测量指标进行分析，以及传统统计分析方法的应用条件常难以满足等[1,15-16]。动态路径分析可用于分析随时间变化的暴露因素与疾病或者健康状态发生发展的联系。

相比Cox模型而言，Aalen模型为线性模型，可以将暴露因素对结局因素的影响分解为直接效应和间接效应，通过直接相加计算总效应，便于更好地了解病因形成机制。然而，模型还有待进一步的发展，一方面，需要使用样条函数检验其线性，增加模型的可信度；另一方面，模型还需更加完善，以便更好地纳入混杂因素及中介因素进行分析，减少未测量的混杂，从而得到尽可能接近真实的因果关联[5]。

利益冲突　无

参　考　文　献

[1] 郑媛，何电. 生命历程流行病学的发展与应用[J]. 中华疾病控制杂志，2015，19（2）：196-199. DOI：10.16462/j.cnki.zhjbkz.2015.02.024.
Zheng Y，He D. An overview of life course epidemiology[J]. Chin J Dis Control Prev，2015，19（2）：196-199. DOI：10.16462/j.cnki.zhjbkz.2015.02.024.

[2] 陆海霞，卢展民，王春美. 生命历程方法在口腔流行病学中的应用[J]. 国际口腔医学杂志，2013，40（3）：339-343. DOI：10.7518/gjkq.2013.03.017.
Lu HX，Lu ZM，Wang CM. Application of life course approach on oral epidemiology[J]. Int J Stomatol，2013，40（3）：339-343. DOI：10.7518/gjkq.2013.03.017.

[3] 郭孙伟，张照寰，许世瑾，等. Cox回归模型及其在医学中的应用[J]. 上海第一医学院学报，1985，12（4）：289-295.
Guo SW，Zhang ZH，Xu SJ，et al. Cox's regression model and its application in medicine[J]. Acta Academ Med Primae Shanghai，1985，12（4）：289-295.

[4] Therneau T，Crowson C，Clinic M. Using time dependent covariates

and time dependent coefficients in the cox model[A]. 2017: 1-25.

[5] Gamborg M, Jensen GB, Sørensen TIA, et al. Dynamic path analysis in life-course epidemiology[J]. Am J Epidemiol, 2011, 173(10): 1131-1139.

[6] Strohmaier S, Røysland K, Hoff R, et al. Dynamic path analysis-a useful tool to investigate mediation processes in clinical survival trials[J]. Stat Med, 2015, 34(29): 3866-3887. DOI: 10.1002/sim.6598.

[7] 郭丽娟. 寿命资料比例风险回归的多水平异质性模型[D]. 广州: 中山大学, 2007.
Guo LJ. Multi-variate heterogeneity model in the propotional risk regression in survival data[D]. Guangzhou: Sun Yat-sen University, 2007.

[8] Aalen O. A linear regression model for the analysis of life times[J]. Stat Med, 1989, 8(8): 907-925. DOI: 10.1002/sim.4780080803.

[9] Aalen O. A model for nonparametric regression analysis of counting processes[C]//Klonecki W, Kozek A, Rosiński J, eds. Mathematical Statistics and Probability Theory. New York, NY: Springer, 1980: 1-25.

[10] Fosen J, Ferkingstad E, Borganø, et al. Dynamic path analysis-a new approach to analyzing time-dependent covariates[J]. Lifet Data Analy, 2006, 12(2): 143-167. DOI: 10.1007/s10985-006-9004-2.

[11] Al-Jewair TS, Pandis N, Tu YK. Directed acyclic graphs: A tool to identify confounders in orthodontic research, Part Ⅰ[J]. Am J Orthodont Dentof Orthoped, 2017, 151(2): 419-422. DOI: 10.1016/j.ajodo.2016.11.008.

[12] Shrier I, Platt RW. Reducing bias through directed acyclic graphs[J]. BMC Med Res Methodol, 2008, 8: 70. DOI: 10.1186/1471-2288-8-70.

[13] 向韧, 戴文杰, 熊元, 等. 有向无环图在因果推断控制混杂因素中的应用[J]. 中华流行病学杂志, 2016, 37(7): 1035-1038. DOI: 10.3760/cma.j.issn.0254-6450.2016.07.025.
Xiang R, Dai WJ, Xiong Y, et al. Application of directed acyclic graphs in control of confounding[J]. Chin J Epidemiol, 2016, 37(7): 1035-1038. DOI: 10.3760/cma.j.issn.0254-6450.2016.07.025.

[14] 曹志强, 王杨, 李卫. Aalen模型在医学研究中的应用[J]. 中国卫生统计, 2015, 32(4): 335-337.
Cao ZQ, Wang Y, Li W. Application of Aalen model in medicine[J]. Chin J Health Stat, 2015, 32(4): 335-337.

[15] 李立明, 余灿清, 吕筠. 现代流行病学的发展与展望[J]. 中华疾病控制杂志, 2010, 14(1): 1-5.
Li LM, Yu CQ, Lv J. Modern epidemiology: the development and prospect[J]. Chinese Journal of Disease Control and Prevention, 2010, 14(1): 1-5.

[16] Shanahan MJ, Mortimer JT, Kirkpatrick Johnson M. Handbook of the Life Course[M]. New York, NY: Springer International Publishing, 2016.

（收稿日期：2017-06-18）
（本文编辑：王岚）

中华预防医学会流行病学分会第七届委员会名单

（按姓氏笔画排序）

主任委员 李立明

副主任委员 刘天锡 杨维中 吴 凡 何 耀 汪 华 胡永华 姜庆五 詹思延

常务委员 王 岚 叶冬青 余宏杰 汪 宁 沈洪兵 陆 林 陈 坤 周晓农 赵根明 段广才
贺 雄 唐金陵 曹务春 崔萱林

委 员 于雅琴 么鸿雁 王 岚 王 蓓 王开利 王文瑞 王定明 王素萍 王效俊 仇小强
叶冬青 冯子健 毕振强 吕 筠 庄贵华 刘天锡 刘殿武 闫永平 许汴利 严延生
杜建伟 李 丽 李 琦 李凡卡 李申龙 李立明 李亚斐 李俊华 李增德 杨维中
吴 凡 吴先萍 邱洪斌 何 耀 何剑峰 余宏杰 汪 宁 汪 华 沈洪兵 张 晋
张 颖 陆 林 陈 坤 陈可欣 陈维清 岳建宁 周宝森 周晓农 单广良 孟 蕾
项永兵 赵亚双 赵根明 胡东生 胡代玉 胡永华 胡志斌 胡国良 段广才 俞 敏
施 榕 施国庆 姜 晶 姜庆五 贺 雄 贾崇奇 夏洪波 栾荣生 唐金陵 曹广文
曹务春 崔萱林 董柏青 程锦泉 詹思延 蔡 琳 戴江红 魏文强

秘书长 王 岚

副秘书长 吕 筠