

多水平Meta回归分析及其在流行病学研究中的应用

王安伟¹, 黄文丽²

(1.大理学院公共卫生学院, 云南大理 671000; 2.云南省地方病防治所, 云南大理 671000)

[摘要] 目的: 介绍多水平Meta回归分析方法及其在流行病学研究中的应用。方法: 检索并综合国内外Meta回归分析及流行病学资料Meta分析的相关文献。结果: 检索到相关文献资料数十篇。结论: Meta分析可看作是多水平分析的一个特例。合并流行病学资料时, 应进行Meta回归分析, 评估各项研究间的异质性大小及来源。

[关键词] 多水平模型; Meta回归; Meta分析

[中图分类号] R195.1 [文献标识码] A [文章编号] 1672-2345(2007)12-0084-03

Multilevel meta-regression analysis and its application to epidemiological research

WANG An-wei¹, HUANG Wen-li²

(1. School of Public Health, Dali University, Dali, Yunnan 671000, China;

2. Yunnan Institute of Endemic Diseases Control and Prevention, Dali, Yunnan 671000, China)

[Abstract] Objective: To introduce the methodological issues of multilevel meta-regression and its application to epidemiological studies. Methods: Collecting and synthesizing the papers concerning the method of multilevel meta-regression and its application to epidemiological research. Results: Literature from home and abroad was retrieved. Conclusions: Meta-analysis could be regarded as a special case of multilevel analysis. A meta-regression analysis should be conducted to explore and estimate the Heterogeneity in meta-analysis of data from epidemiologic studies.

[Key Words] multilevel model; meta-regression; meta-analysis

多水平Meta回归分析是多水平分析方法在Meta分析中的应用。对多水平Meta回归分析及其在流行病学研究中的应用进行介绍, 为流行病学资料的Meta分析提供参考。

1 Meta回归分析概述

1.1 Meta分析简介 Meta分析最早由英国教育心理学家Glass于1976年命名并将其定义为:“ The statistical analysis of large collection of analysis results from individual studies for the purpose of integrating the findings”。此后, 不少统计学家都对Meta分析下过定义, 但都倾向于“ Meta分析是对以往的研究结果进行系统定量综合的统计学方法”这一含义^[1-4]。

1.2 流行病学研究与Meta回归分析 流行病学研究方法通常分为四大类: 描述性研究、分析性研究、实验性研究和理论性研究。前两类均属观察性研

究, 是流行病学最常用的研究方法。观察性研究容易受到混杂偏倚和选择偏倚的影响, 各项研究的对象选择、研究方法等的不同都会导致研究间的异质性, 对异质性较大的资料进行传统的Meta分析可能会导致错误的结论, 从而误导读者。因此, 在对流行病学研究资料进行Meta分析时, 需分析各研究间的异质性, 并对异质性的来源进行评估^[5]。

Meta回归分析可评价研究间异质性的大小及来源。根据统计模型的不同, 可将Meta回归分析分为固定效应的Meta回归分析和随机效应的Meta回归分析两大类。基于固定效应模型的Meta回归分析假设多项研究具有一个共同的效应尺度, 各项研究效应尺度存在的差异主要是因为随机误差造成; 随机效应模型则假设各项研究不具有共同的效应尺度, 而是每项研究都有自己的效应尺度, 并将其定义为

[基金项目] 科技部科技攻关项目(2003BA712A11-01); 云南省科技厅攻关项目(2004NG13)

[收稿日期] 2007-09-25

[作者简介] 王安伟(1978-), 男, 云南昭通人, 硕士研究生, 医师, 主要从事地方病研究。

一随机变量, 该变量呈正态分布。固定效应模型要求资料满足正态性、观察值间的独立性和研究间的方差齐性, 如不满足这三个条件的资料进行固定效应的分析, 就有扩大犯I类错误概率的风险; 随机效应模型则只要求资料满足正态性, 其它两个条件则不作要求, 故扩大了适用范围。在此主要介绍基于随机效应模型的Meta回归分析(也称多水平Meta回归分析)^[6,7]。

2 多水平Meta回归分析

2.1 多水平统计模型简介 多水平模型又称“随机效应模型”、“随机系数模型”、“层次线性模型”或“混合效应模型”等, 是将II型方差分析理论与多元统计分析相结合的新技术, 是分析和处理具有层次结构特征数据的有力工具, 亦是当前国际上统计学研究中一个新兴而重要的领域^[8]。

多水平模型用于处理具有层次结构或嵌套结构的数据。如个人处于家庭中, 家庭处于社区中, 这就构成三个水平: 个人是基本水平(水平1), 家庭是中间水平(水平2); 社区是最高水平(水平3)。对于具有层次结构特征的数据, 多水平模型可将传统模型中单一的随机误差项分解到与数据层次结构相对应的水平上, 即分解出研究水平的变异, 并提供了进一步拟合研究水平上复杂误差结构的可能性^[9]。对于连续型反应变量, 两水平随机效应模型可表示为^[10]:

$$y_{ij} \sim N(X_{ij}\beta, \sigma^2_{\epsilon})$$

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{ij} \quad (1)$$

$$u_{0j} \sim N(0, \sigma^2_{u_0}) \quad e_{ij} \sim N(0, \sigma^2_{\epsilon})$$

式中, $i=1, 2, 3, \dots, n_j$, 示水平1单位, $j=1, 2, 3, \dots, m$, 示水平2单位; x_{ij} 为协变量; u_{0j} 代表水平2的残差, 表示第j个水平2单位的平均预测值与总均值之差, 即组间变异; e_{ij} 为水平1单位的残差, 表示方程残留的随机成分; $(\beta_0 + \beta_1 x_{ij})$ 为方程的固定部分, 回归系数 β_0 和 β_1 描述模型的固定效应; $(u_{0j} + e_{ij})$ 为方程的随机部分, 其方差 $\sigma^2_{u_0}$ 和 σ^2_{ϵ} 描述模型的随机效应。

以上模型称为方差成分多水平统计模型, 也是最简单的两水平模型, 它只包括随机参数 $\sigma^2_{u_0}$ 和 σ^2_{ϵ} , 即仅考虑截距的随机项, 故又称随机截距模型。

2.2 多水平Meta回归模型 Meta分析可被看成是多水平模型的一个特例, 资料具有至少两个层次结构特征, 即各项研究为高水平单位, 每项研究中的研究对象为低水平单位。在Meta分析中, 尽管各项

研究都针对同一问题或具有相同的研究假设进行研究, 但各项研究通常在研究设计或实施的诸多方面不同于另一项研究, 如纳入和排除标准、样本含量、质量控制等方面在不同的研究之间难免存在差异, 这就导致了“效应尺度”在不同研究之间的差异性。

Meta分析通常只能获取到水平2单位(研究水平)的聚集数据, 如各研究的均数及标准差、OR值、RR值等, 这就使Meta分析资料的多水平模型与一般的多水平模型有所区别。因此, 两水平Meta回归分析模型可改写为:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{ij} z_{ij} \quad (2)$$

式中, z_{ij} 是设计变量, 为第j项研究的权重因子, 如以各研究样本含量为权重因子, 则 $z_{ij} = 1/\sqrt{n_j}$; 若以各研究效应倒方差为权重因子, 则 $z_{ij} = 1/s_{ij}$ (标准误), 如不引入协变量 x_{ij} , 则该模型相当于D-L随机效应模型。水平1的误差 σ^2_{ϵ} 不可测, 分析时需设定水平1的方差 σ^2_{ϵ} 为1^[8,10]。

2.3 多水平Meta回归模型的参数估计方法 拟合多水平模型的方法主要包括基于经典的频率学派方法和贝叶斯方法。经典的频率学派方法主要有“极大似然法”、“迭代广义最小二乘法”、“限制性最大似然法”等, 对于多水平Meta回归分析, 限制性最大似然法被认为是最好的估计方法。当反应变量为二分类变量时, 所用的估计方法称为“拟似然法”^[11]。

另一种方法为基于贝叶斯(Bayes)的计算方法, Meta分析的Bayes方法近年来越来越受到重视。Bayes方法适合于任何参数后验分布的计算, 特别适合于处理高维问题和多余参数问题。BayesMeta分析模型充分考虑了模型参数的不确定性, 得出的是一种“光滑”估计^[12]。

随着计算机技术的发展, 上述各方法均可通过专门的统计软件实现。多水平分析的专用统计软件有MLwiN及HLM等, WinBUGS可进行复杂的Bayes分析。其它通用统计软件如SAS、Stata、SPSS、S-plus等也有相应的多水平统计模块或命令, 可执行不同程度的多水平统计分析。其中, Stata可通过metareg命令对聚集型数据进行两水平的Meta回归分析。由于各种软件采用的计算方法不同, 致使不同的软件的计算结果可能会出现微小差别。

3 多水平Meta回归分析在流行病学研究中的应用

下面根据纳入Meta分析的资料的不同类型, 以最常见的病例对照和对列研究资料的多水平Meta

回归分析方法来说明多水平Meta回归分析在流行病学研究中的应用。

3.1 能获取每项研究的个体水平资料 当通过各种渠道能够获取纳入Meta分析的每项研究的原始资料(包括个体研究对象特征变量等)时,多水平Meta回归分析相当于两水平的多水平分析,故可进行标准的多水平分析。对于病例对照和对列研究资料,可进行包含多个解释变量(可为水平1或水平2变量)的多水平Logistic回归分析。两水平Logistic回归模型可表示为:

$$\begin{aligned} y_{ij} &\sim \text{Binomial}(n_{ij}, p_{ij}) \\ \text{logit}(P_{ij}) &= \mu_0 + \mu_1 x_{ij} + u_{0j} \\ u_{0j} &\sim N(0, \sigma^2) \quad (3) \\ \text{var}(P_{ij}) &= (p_{ij}(1-p_{ij})/n_{ij}) \end{aligned}$$

当 σ^2 为0时,该模型演变为一般的logistic回归模型。 μ_0 为尺度参数,当反应变量确定服从二项分布时,尺度参数应该为1或接近1。

3.2 能获取每项研究的四格表资料 在进行Meta分析时,欲获取每项研究的个体水平原始资料一般是不可能的,通常我们只能收集到每项目研究的四格表资料。此时资料的分析可有多种选择:拟合两水平Logistic回归模型,可将研究水平(水平2)的变量引入模型,但水平1变量不能获取,故不能引入;根据四格表资料分别计算各研究的OR或RR值,以 $\ln OR$ 或 $\ln RR$ 为反应变量,按模型(2)进行两水平线性模型拟合。此时,也只能引入水平2的解释变量。对于对列研究资料,可计算各研究的分组发病率(病例组和对列组),以发病率的自然对数($\ln p_j$)为反应变量,以组别为解释变量(也可引入水平2的变量)进行两水平线性模型拟合。

3.3 只能获取每项研究的OR/RR值和样本量 此时,只能以 $\ln OR$ 或 $\ln RR$ 为反应变量,以水平2变量为解释变量,以样本量为权重系数进行两水平线性模型拟合,见模型(2)。

3.4 部分研究能获取个体水平的数据 在纳入Meta分析的部分研究可以获取到个体水平的数据,而其它研究只能获取到聚集水平的数据时,如果将个体水平数据转化为聚集水平,则会损失信息。多水平模型可将个体水平数据(individual data)和聚集水平数据直接组合进行Meta分析,充分利用现有数据信息,提高了参数估计的准确性^[13]。

4 多水平Meta回归分析的局限性

多水平Meta回归分析作为传统Meta分析的补充

和扩展,可判断纳入Meta分析的资料的异质性,并对其异质性的来源进行定量的评估,最后作出综合的判断。

然而,多水平Meta回归分析也是有条件的,对于连续型反应变量的多水平模型,要求反应变量呈正态分布,Logistic回归模型则要求反应变量呈二项分布。另外,如果是对聚集型数据进行多水平分析时,由于水平1的协变量数据不能获得,故只能对水平2的协变量进行分析,不能控制发生在水平1的混杂因素。由于流行病学研究很容易产生混杂偏倚,因此,在对流行病学资料进行Meta分析时,如有可能,应尽可能获取到每个研究个体水平的数据,进行标准的多水平分析。

[参考文献]

- [1] Anne Whitehead. Meta-Analysis of Controlled Clinical Trials (M). England: John Wiley & Sons Ltd, 2002. 1-10.
- [2] Matthias Egger, George Davey Smith, Douglas G Altman. Systematic Reviews in Health Care: Meta-Analysis in Context (M). 2nd edition London: BMJ Publishing Group, 2001. 3-8.
- [3] 刘伟远. Meta-分析的研究进展[J]. 新疆医科大学学报, 2003, 26(5): 512-514.
- [4] 方积乾, 陆盈. 现代医学统计学(M). 北京: 人民卫生出版社, 2002. 150-157.
- [5] Wolfgang Ahrens, Iris Pigeot. Handbook of Epidemiology (M). Springer, 2005. 829-857.
- [6] Simon G., Thompson, Julian P. T. Higgins. How should meta-regression analyses be undertaken and interpreted [J]. Statistics In Medicine, 2002, 21(15): 1559-1573.
- [7] 余松林. 混合线性模型[J]. 中国医院统计, 2006, 13(1): 70-75.
- [8] 杨珉, 李晓松. 医学和公共卫生研究常用多水平统计模型(M). 北京: 北京大学医学出版社, 2007. 126-165.
- [9] 李晓松, 刘巧兰, 倪宗瓚. 多水平统计模型在Meta分析中的应用研究[J]. 中国卫生统计, 1999, 16(3): 133-135.
- [10] Joop Hox. Multilevel Analysis Techniques and Applications (M). London: Lawrence Erlbaum Associates, 2002. 139-156.
- [11] Jon Rasbash, Fiona Steele, William Browne, et al. A User's Guide to MLwiN Version 2.0 (M). Bristol: Centre for Multilevel Modelling, University of Bristol, 2005. 12-34.
- [12] 郭秀娥, 刘立, 徐勇勇. 老年痴呆的多因素贝叶斯Meta分析[J]. 第四军医大学学报, 2001, 22(13): 1185-1186.
- [13] 朱彩蓉, 刘忠华, 孙克宏, 等. 组合聚集数据与个体数据的多水平Meta分析模型[J]. 四川大学学报(医学版), 2005, 36(6): 888-891.