

系列问答。

流 行 病 学 知 识 问 答

A部分 吴尊友¹ B部分 徐兆炜² 组稿与总审 曾光¹

A 多因素分析中数学模型及变量的选择

A₀₃ 选择数学模型会面临哪些问题?

答:随着计算机的飞速发展和统计软件的不开发,多因素数学模型在流行病学资料分析中的应用越来越广泛。同时,滥用数学模型的现象也越来越多。

当研究者决定使用多因素数学模型进行数据分析,假设估计暴露因素 X 的效应时,他需要控制其他 6 个潜在混杂因子的混杂作用。在这种情况下,有多少种模型可供选择呢?假设研究者决定用 logistic 回归模型,6 个被控变量中有 4 个要进入模型,考虑到被控变量间的交互作用和被控变量与暴露因子 X 之间的交互作用,再进一步考虑数据的形式转换,模型的可选择种类会错综复杂。

从上述可能的模型中选出最佳的模型是一个复杂的过程。不论怎么选,其结果都会或多或少有混杂,只是程度不同而已。因此,我们希望尽可能地挑选出残余混杂最小的模型来,而最后挑选出来的模型是否真正是残余混杂最小的,研究者本人也心中无数。于是,多数研究者都是依靠计算机去挑选“最佳模型”。常用的方法是:选用最方便的计算机软件,连续性变量不经转换直接输入,按计算机软件设计的挑选变量机制来确定模型。而实际情况是计算机软件只考虑纯粹的数字运算,而没有考虑变量本身和变量间的流行病学含意及其生物学的意义。

A₀₄ 选择数学模型时应考虑哪些问题?

答:尽管数学模型在描述和概括数据方面比分层分析法更为有效,但这一优势却是以潜在的更大偏倚为代价。数学模型的前提假设比分层分析要严格的多。如果其中的一个或多个假设条件不能满足,根据模型得出的估计值往往失真。

为减少数学模型估计值的失真,以下几点可以

考虑:①用现有的数据和既往的研究结果来验证所选择数学模型的流行病学假设条件和统计学的假设条件是否满足;②用多种模型对数据进行分析,看是否得出一致的结果;③把数学模型结果与分层分析比较,看是否吻合。

每个人都希望能够方便地使用多因素数学模型。对于病例对照研究,几乎无一例外地应用 logistic 模型。而对于队列研究,常用的数学模型包括 logistic 模型、Cox 模型和指数模型。

在多数的流行病学研究中,logistic 模型、Cox 模型和指数模型都有两个重要的流行病学假设:

①疾病比值比与暴露因素 X 之间呈指数线性关系;

②暴露因素与混杂因子间呈相乘关系。比如说,暴露因素每增加一个单位,发病率增加 2 倍;混杂因子每增加一个单位,发病率增加 3 倍;如果暴露因素与混杂因子同时各增加一个单位,那么发病率增加 6 倍(2×3 倍)。

而在实际的流行病学应用中,以上两个条件很难满足,至于这两个条件不能满足时,模型计算的估计值与真实值偏差多少,无从知晓。应该认识到,流行病学研究都是尽最大可能寻找出最接近真实值的估计值。因此,在应用这些模型时要注意几点。

①如果模型不正确,根据模型计算的回归系数估计值有误差。

②即使模型本身是正确的,如果变量进入模型的形式不当,暴露因素与疾病间剂量反应关系也会失真。

③在多数的流行病学应用中,如果模型不正确,在异常暴露水平上的估计值偏差更大。

④如果模型本身不适合,简单地数据进行数据形式转换或在模型中加入交互作用因子,并不足以纠正模型不正确所造成的偏倚。

当模型的应用条件不符合时,如果仍坚持使用,不仅估计值会严重偏倚,而且还有可能出现人为的假象。比如说,暴露因素与疾病发病率间存在 U 型

1 中国预防医学科学院 北京 100050

2 中国预防医学科学院流行病学微生物学研究所

的剂量反应关系,即使暴露因素与疾病之间的这种剂量反应关系很强,如果使用上述线性模型对这一数据进行分析,也不会发现其间的关联。

A₅₅ 如何判断所选定的数学模型是否合理?

答:如果能够选择正确或者相对正确的模型固然好。可是,并没有一种最完美的挑选模型的方法,每一种选择方法都或多或少存在误差。目前广泛采用的检验方法是模型拟合度检验,但这并不足以确认模型是否合理,尚需要通过多种其他方法,如回归诊断,来帮助判断模型的条件假设是否满足。

有很多种作图法可检验数据是否满足模型的假设条件,比如把实际观察值与模型估计值作成点图,如果观察值与估计值成为(或近似)一条直线,则说明模型应用条件满足,可以使用。

有时,模型的假设条件不满足,而采用常规的检验方法或回归诊断方法又检查不出来,可以考虑选用其他数学模型进行分析,看是否能得出相同或类似的结果。

A₅₆ 如何选择合理的变量进入数学模型?

答:筛选变量的策略一般都是从剔除不适宜变量开始。研究某种暴露与疾病的关系时,如果被控混杂因素是受暴露的影响或受疾病的影响,那么,这种因素就不易作为潜在的混杂因素进入模型。因此,挑选变量的第一步就是把这一类型的变量从所有候选混杂因素中剔除。

经过第一步筛选后,可能仍有很多潜在的混杂因子需要控制。如果把这些变量都列入模型,特别是当变量之间高度相关时,模型很不稳定,估计值就会严重偏倚。因此,要尽可能地减少混杂因子的数量,而决定减少到哪几个变量为止,又会遇到难题。不管采用什么方法挑选,被剔去的变量应该被假设为不是混杂因子。而在实际应用中,判断这些剔下来的变量是否真正为非混杂因子,也并不那么简单。

在卫生学研究中常见的两种变量选择方法是逐步回归法和估计值改变法。

最常用的逐步回归法挑选变量的过程是这样:首先把所有变量分为两大类,一类是必须保持在模型中的变量,如暴露因子,这一变量不会在选择过程中被剔去,其余变量在回归过程中或被选入或被剔除。其基本原理是:

选入:在所有未进入模型的变量中,寻找进入模型后其回归系数最大的那个变量,如果其回归系数达到预先确定的显著性水平(如 $P < 0.05$),则被选入模型。

剔除:在所有已进入模型的变量中,寻找回归系数最小的那个变量,如果其回归系数达不到预先设定的显著性水平,则从模型中剔除。

选入和剔除一直进行到所有达到一定水平的变量被选入或所有达不到一定水平的变量被剔除为止。

逐步回归自动筛选法又分为逐步递增筛选法和逐步递减筛选法。逐步递增筛选法是先必须保持在模型中的变量(如暴露因素)放入模型,然后一个一个地逐步把可能的混杂因子加入到模型中,直到没有变量可达到显著性水平进入为止。逐步递减法是把所有的变量放入模型,然后一个一个地剔除,直到所有达不到显著性水平的变量被剔除为止。无论是递增法还是递减法,由于在变量的选择过程中,数学模型并没有考虑混杂因素与暴露因子之间的关系,因此,最后的模型并不一定是最合理的,有时,可能是很不合理的。

估计值改变法筛选变量是根据被选变量对暴露因素估计值的改变大小来决定。

选入:在所有未选入模型的变量中,寻找进入模型后对暴露因素估计值改变最大的那个变量,如果暴露因素估计值的改变超过预先设定的水平(如 10%),则选入模型。

剔除:在所有已选入模型的变量中,寻找剔出模型后对暴露因素估计值改变最小的那个变量,如果改变值小于预先设定的水平,则剔出模型。

同样,这种筛选法也不能保证最后的模型合理。

自动变量筛选法并没有考虑既往的研究结果。在变量筛选过程中,应该充分利用先前研究结果所提供的各种信息,比如说,在研究某暴露因素与肺癌的关系时,吸烟应该考虑为混杂因素,如果采用自动筛选法,可能会出现吸烟这个变量选不进最后的模型,在这种情况下,应该把吸烟保持在模型中。

综上所述,在进行多因素分析时,无论是选择模型,还是选择变量,不能完全依赖根据纯统计理论而设计的计算机软件来作决定,而应该综合考虑既往研究结果和变量及变量间关系的流行病学和生物学含义来作判断。

B 克雅氏病和疯牛病流行病学进展

B₇₈ 什么是克雅氏病(CJD)?

答:克雅氏病是一种罕见的、既有传染性、又有遗传性的人类传染性海绵状脑病(TSE),以进行性痴呆、肌阵挛、特殊的脑电图(EEG)表现和神经病理学病变为特征的亚急性中枢神经系统紊乱疾病。

1920年 Creutzfeldt和 1921年 Jakob先后报告 7例这种中枢神经系统变性病例, 1966年以该二氏命名该病,简称 CJD或克雅氏病。此病目前尚无有效治疗方法,病死率几乎为 100%。潜伏期 10~ 15年(可长达 30多年)。病人发病年龄为 55至 75岁(平均 60岁左右),病人发病后 1~ 2年内死亡(根据 185例散发型 CJD病例统计,病程平均约为 4个月)病人确诊均取决于死后脑组织镜检结果。近来有的学者报告生前可采取扁桃腺进行活体检查和脑脊液中 14-3-3蛋白质的测定来诊断,但一般认为仅能做出疑似 CJD的诊断。全世界各国发病率基本相似,每年每 100万人口 1人(0.5~ 2人),随年龄增长而增高。性别和职业对发病率无显著差异。此病可分为三型:①散发型 CJD,约占 80%,如何被传染目前不详;②传染型或医源型 CJD,可通过被此病致病因子污染材料或其加工的产品而传播感染,如生物制品的药物(如生长激素注射)、角膜移植、脑手术时污染的手术器械引起的医院感染等,能否通过血液传染(输血)还有争议,约占 5%;③家族型或遗传突变型 CJD,病人有明显的家族史,约占 10%~ 15%。由于 CJD发病率低,且生前诊断困难,所以各国文献报告仅 1 435例,但其中确诊的为 588例。近年来各大洲均有散发的病例报告。英国 BSE病牛数虽很高,但 CJD年发病率与其他国家相似,约百万分之一,年发病人数在 50例左右。

B₇ 什么是新型克雅氏病(nvCJD)?

答:新型或突变型 CJD(new variant CJD,简称 nvCJD或 vCJD),是英国政府根据英国国家 CJD监测中心(R. G. Will等)提供的资料,于 1996年 3月 20日宣布并命名的,同年 4月初 WHO在日内瓦召开的专家研讨会上亦正式采用了这个命名。简单地说,它是一组与典型散发型 CJD表现不一、可能与 BSE密切相关的、较年轻的 CJD病例。当时仅 10例,男 4例,女 6例,发病年龄均在 40岁以下(16~ 39岁,平均 27.5岁),发病年份为 1994~ 1995年,已死亡 8例的病程为 7.5~ 22.5个月(平均 12个月),存活 2例病程分别已超过 6和 22个月(女性,16岁)。但到 1997年 10月,英国已报告 nvCJD病例 21例,法国报告 1例,尚未见到其他国家报道,所以目前全球总数为 22例。nvCJD具有 CJD的一般特征外,还有以下特点:①发病年龄较轻(40岁以下,平均 27.5岁);②临床表现不同(除有共济失调、痴呆、肌阵挛等典型症状外,9例早期出现性格改变症状,1例在整个 7.5个月病程内有足疼症状);③没有

CJD典型的 EEG图谱。此外,这 10例病人病史没有发现其他可能的致病因素(医源性或家族性),而有吃食牛肉史,所以一般认为系经口感染。

B₈ 什么是疯牛病? 流行情况如何?

答:疯牛病就是牛海绵状脑病(Bovine Spongiform Encephalopathy,简称 BSE)的俗称,是牛的一种慢性致死性中枢神经系统疾病。该病在英国发现,经组织病理学检查于 1986年 11月为英国 John W. Wilesmith等确诊,但是于 1988年方予报道。潜伏期 2~ 8年(在 2岁以下的牛中曾观察到疯牛症状)。临床症状主要有行为异常(惊恐、烦躁不安、好斗),姿势和运动异常(四肢伸展过度、后肢运动失调、震颤、跌倒、麻痹)以及感觉异常(对外界声音和触摸敏感、搔痒)等表现,食欲和体温正常,但奶产量降低,体重减轻,几周后病情恶化,卧地不起。主要靠死后脑组织的检查结果确诊,生前可根据早期症状做出初步诊断或疑似诊断。发病后或屠宰或病死,无治疗方法,病死率 100%,病程为 2周至 6个月。目前英国共报告 BSE确诊病牛约 160 000多头,占全球总数的 99%以上,大多为 3~ 5岁的奶牛。目前报告有 BSE的国家为 12个。最近报告,BSE流行已被控制。

B₈ 克雅氏病和疯牛病的致病因子是什么? 克雅氏病与疯牛病是否有关联?

答:近几年 Prion做为致病因子的学说虽仍有争议,但已被较普遍地接受。Prion是一种无核酸的感染性蛋白质或蛋白质感染因子,译名还有朊病毒、毒蛋白、普里安等,系美国神经病学家、生物化学家斯坦利·普鲁西纳(Stanley B. Prusiner)教授于 1982年首先发现和命名的。由于他在细菌、病毒、真菌、寄生虫等以外提出了一种新病原体,对今后深入研究及防治 CJD及 BSE等传染性海绵状脑病将会起到不可估量的作用,所以于 1997年 10月荣获 1997年度诺贝尔医学奖。普鲁西纳认为,Prion是人及动物大脑组织中一种正常结构蛋白质,即存在所有神经细胞表面的糖蛋白,正常情况下与神经传导有关,不会致病,但由于目前还不清楚的原因,导致编码它的基因 PRNP突变,即促使这种对蛋白水解酶敏感的 Prion(PrPsen)突变为对酶有抵抗力的 Prion(PrPres),在形态上后一种蛋白质会扭曲变形,如果将这种 PrPres与 PrPsen一起孵育,后者亦将变形,神经元将死亡,致使大脑成海绵状而蜕变,从而引起 CJD、BSE等海绵状脑病的症状,因而在生物学性状上获得传染性。有的学者认为,Prion复制如同酵母

菌,因为酵母菌不仅有 Pion,而且它可出芽分生,但是许多学者认为这不适用于人类和高等生物;有的学者认为是依靠正常组织中的 PrP复制,即蛋白质在逆转录方式下指导蛋白质自身复制。它对理化因素抵抗力较强,耐高温,134℃湿热 30~60分钟才能达到消毒目的,这种特性对医源性 CJD的传播和牛骨加工的动物饲料感染 BSE,甚至争议中的 nv CJD传播颇为重要。近有人提出某种细菌与 CJD致病有关的推论,正在进一步验证中。最近报告的两个实验动物试验结果证明,BSE的致病因子与 nv CJD的致病因子是相同的,但与散发型、家族型、医源型的显著不同。因此,有的学者认为 nv CJD是 BSE在人类中的复制品,而有“人型 BSE”或“牛型 CJD”的提法。

B₂: “疯牛病风波(危机)”是怎么回事?

答:“疯牛病风波”是由于 1996年 3月 20日英国政府宣布和承认 10名 nv CJD病人与疯牛病有关联后引起的一场对疯牛病的大恐慌,不仅影响英伦三岛而且迅速蔓延到欧洲各国,甚至全世界。由于无法排除这 10例病人与疯牛病的关联,欧盟于 1996年 3月 27日宣布禁止英国牛肉出口,英国要在 4年内屠宰 300~600万头牛来防止 BSE流行及传播,要求英国宣布近期内屠宰 250万头 30个月龄以上牛的计划。这使英国养牛业和牛肉业受到致命性打击,经济损失惨重(高达 200亿英镑以上)。

B₃: 我国克-雅氏病和疯牛病流行情况如何?已采取什么预防措施?

答:总的说来,我国克-雅氏病病例很少,目前仅报告了 20多例,散发在北京、上海、湖南、青岛、大连、长春等地,主要来自临床报告资料及论著报道,大多病例均经病理学证实。但是 1996年台湾曾报道 1976~1995年的 14例确诊的 CJD病例,病程均在一年以内,近两年台湾又报道了 2例女性 CJD病例,但均不同于 nv CJD。目前我国没有发现 nv CJD,而 CJD年发病率亦远远低于全球平均年发病率。

我国对控制疯牛病已做了大量的工作。1990年 6月农业部等单位即发出“关于严防牛海绵状脑病传入我国的通知”,明确规定在英国没有消灭 BSE以前,严禁从英国进口牛、牛精液和牛胚胎以及其他牛的产品,对已进口的应立即停止使用,并开展检疫,及时通报疫情。1992年 6月 BSE被我国列入进境动物检疫一类传染病之列。1996年 3月 20日后我国农业部又发出紧急通知,要求全国加强防范 BSE。国家动植物检疫局一直依法严格检疫,加强执法力度,严防疯牛病传入我国,更在首都等主要城市的饭店、肉联厂等生产、销售网点以及各口岸加强检疫工作,防止疯牛病传入我国。迄今为止,我国已成功地将疯牛病拒于国门之外,但仍不能掉以轻心。