

整合科研设计与统计分析的方法： 广义综合评价

郭春雪¹ 沈 宁¹ 胡良平^{1 2*}

(1. 军事医学科学院生物医学统计学咨询中心, 北京 100850;

2. 世界中医药学会联合会临床科研统计学专业委员会, 北京 100029

* 通信作者: 胡良平, E-mail: lphu812@sina.com)

【摘要】 本文目的是介绍一种能将科研设计与统计分析整合在一起的方法, 即广义综合评价。通过指出传统综合评价方法的局限性, 并陈述了在实践中人们对很多事情从设计、实施到完成的质量评价的广泛需求, 提出有必要将传统综合评价拓展为广义综合评价这一观点。结合基本常识和有关专业知识, 笔者提纲挈领地呈现了如何对科研设计的质量、课题实施的质量、系统评价的质量和统计分析的质量进行综合评价的思路和要领, 言简意赅地阐释了广义综合评价的概念、功能和基本做法。

【关键词】 科研设计; 统计分析; 系统评价; 传统综合评价; 广义综合评价

中图分类号: R195.1

文献标识码: A

doi: 10.11886/j.issn.1007-3256.2017.01.004

Integrate the study design with the statistical analysis: generalizing comprehensive evaluation

Guo Chunxue¹, Shen Ning¹, Hu Liangping^{1 2*}

(1. Consulting Center of Biomedical Statistics, Academy of Military Medical Sciences, Beijing 100850, China;

2. Specialty Committee of Clinical Scientific Research Statistics of World Federation of Chinese Medicine Societies, Beijing 100029, China

* Corresponding author: Hu Liangping, E-mail: lphu812@sina.com)

【Abstract】 This article aimed to introduce a method, entitled generalizing comprehensive evaluation, which integrated the study design with the statistical analysis. The necessity of being aware that the traditional comprehensive evaluation should be expanded to a more generalized level, which takes the following aspects into consideration: some limitations were found out in the traditional comprehensive evaluation method. While application and practice, an increase in demand was seen ranging from study designing, research implementing to quality evaluation of data analysis. Combining common sense with solid specialty background, the authors exemplified how to use this method in study design, implementation, systematic reviews, and the statistical analysis, while the essence of the method was also presented. Moreover, the authors briefly interpreted the concept of the method of generalizing comprehensive evaluation, and the functions and the procedures for practice of the method were further introduced.

【Keywords】 Study design; Statistical analysis; Systematical review; Traditional comprehensive evaluation; Generalizing comprehensive evaluation

1 概 述

1.1 问题的提出

众所周知, 科研设计与统计分析是统计学全部内容中最重要的一部分^[1]。然而, 综合评价与这两部分内容之间是什么关系呢? 翻阅统计学教科书, 似乎可以感觉到: 综合评价方法^[2-3]只是众多统计分析方法中所占篇幅很少的一种, 它往往被排除在“常用统计分析方法”之外。显然, “综合评价”的概念和功能显得过分“狭窄”, 似乎与这个“词”本身根本不相符。由基本常识可知, “综合评价”可以被用

于很多方面, 它包括“科研设计、课题实施和课题验收”全过程中所涉及到的各种内容, 甚至包括用于对除“科学研究”以外的很多其他事物或现象的评价。

人们不禁要问“系统评价”又是什么呢? 文献[4]对系统评价的定义为: 针对某一具体临床问题, 系统、全面地收集全世界所有已发表或未发表的临床研究, 采用临床流行病学的原则和方法对研究进行严格的评价, 筛选出符合纳入标准的研究, 进行定性或定量(Meta分析)合成, 从而得出可靠的结论。从此定义不难得出这样的结论“系统评价”完全可以被收藏入“综合评价”的锦囊之中。

简而言之, 现今统计学教科书中的“综合评价”

项目基金: 国家高技术研究发展计划课题资助(2015AA020102)

的概念与功能需要拓展,使其更加名正言顺、实至名归。

1.2 传统综合评价的实质是什么

在统计学教科书上所介绍的“综合评价”,其有以下两种定义:①对一个复杂系统的多个指标进行总评价的特殊方法^[2];②针对诸多评价对象的多指标信息,应用一定的数学原理与方法(包括数理统计方法),对数据进行适当的加工和提炼,以求得评价对象优劣等级或顺序的过程^[3]。在讲述“综合评价的一般步骤”时,文献[2]声明:对某事件进行多因素综合评价的过程,实质上就是一个科学研究与决策的过程,原则上应当包括设计、收集资料、整理资料和分析资料几个基本阶段。这里来自文献[2]的两种关于“综合评价”的概念和功能的“界定”是大相径庭的,第1种说法^[2-3]似乎强调:综合评价仅仅是基于“多个指标”汇总后求得一个综合指标,进而实现对无序样品进行排序与分档的一类统计分析方法;而第2种说法^[2]似乎强调:综合评价是囊括对“科学研究与决策过程”进行评价的全方位统计分析方法。

事实上,统计学教科书上所介绍的具体“综合评价方法(例如:综合评分法、综合指数法、层次分析法、Topsis法、秩和比法等)”是针对“多个指标”求取一个综合指标的“综合评价方法”,故可以被称为“传统综合评价”^[2-3];而真正意义上针对“科学研究与决策过程”的综合评价方法,统计学教科书上好像没有涉足,它似乎应该被称为“广义综合评价”。

然而,从传统综合评价方法表面上看,似乎确实是针对“多个指标”的;但从其计算方法和得出的结论来看,实际上,应属于将“无序样品转化成有序样品,进而将它们划分入优劣等级中去”的一类统计分析方法。具体地说,应属于“基于多个指标在 n 个个体(被评价对象)身上的取值,通过数学和统计分析技术得出一个“综合指标”,进而,计算出每个个体在该综合指标上的得分值。然后,将得分值由小到大(或由大到小)排序后再进行分档,从而得出:在 n 个个体中,哪些个体应归类于优、良、中、差(具体需要分几档,由研究者自行确定)不同等级的一种简单的评价结论”。

2 传统综合评价的概念与功能需要拓展

2.1 传统综合评价的局限性

从前面“1.2节”的最后一个自然段可以清楚地看出:传统综合评价的概念与功能具有很大的局限性。因为客观世界中的事物和现象都是很复杂的,哪怕是一个很小的问题,要想真正全面、客观、科学、严谨地去研究它并彻底把握它,期望得出的结果和结论经得起时间和实践的检验,仅靠“传统综合评价方法”是无能为力的。

2.2 应将传统综合评价拓展为广义综合评价

为了使“综合评价”实至名归,有必要将“传统综合评价”拓展为“广义综合评价”。笔者建议可从以下三个方面进行拓展:

其一,从概念上拓展——广义综合评价不是一个或一类具体的方法,而是关于如何提出、解决和评价某实际问题的一个系统工程的评价。

其二,从内容上拓展——广义综合评价的内容包括:所提出的问题是否真正属于一个科学的、有研究价值的问题;与研究问题对应的研究目标是否明确、具体;为实现研究目标而制订的研究设计方案是否科学完善、系统全面、精准高效、经济可靠且具有可操作性;研究项目实施过程中是否具有可操作性的标准操作规程和实时精准的质量控制策略以及具体落实情况;研究项目验收阶段是否获得了值得分析的研究资料以及对研究资料的分析和挖掘是否全面彻底。

其三,从功能上拓展——广义综合评价应包括科学研究以及非科学研究领域,特别是在科学研究领域内,应包括科研设计(含课题框架设计和课题技术设计)、项目实施(含质量控制、数据收集与管理)、项目验收(含研究过程的监管、数据分析的核准和结果及结论的验证)和成果推广应用(含社会效益与经济效益的评估、投入与产出比例的核算)。

3 广义综合评价的核心与要领

3.1 广义综合评价的概述

由前述“2.2节”的内容可知,广义综合评价涵盖面非常宽泛。事实上,除了前述提及的“与科研有关的内容”外,还可以包括很多其他内容,例如:对人类开发太空和海洋的利弊进行综合评价、对一

个国家的某项大政方针实施效果的综合评价,诸如如此的可以用上综合评价的事物或现象不胜枚举。本文仅把注意力局限于“对科研项目从科研设计到统计分析的质量进行综合评价”,谈一点笔者的肤浅认识,以期达到抛砖引玉之目的。

3.2 对科研设计的质量进行综合评价

3.2.1 对科研设计质量进行综合评价的概述

科研设计的质量体现在“科研设计方案”上,其应由两部分组成,即“课题框架设计方案”与“课题统计设计方案”^[5]。前者为科研项目提供“物质基础”,后者(说明:宜将“课题统计设计方案”修改为“课题技术设计方案”,因为统计知识只有紧密结合基本常识和各科专业知识,才能如虎添翼)为科研项目提供“技术支撑”。因此,科研设计的质量高低完全取决于前述提及的两种设计方案的质量。

3.2.2 课题框架设计方案的核心内容

在课题框架设计方案中,应将“课题概况”“课题承担情况”“课题技术问题”“课题进度”和“课题经费预算”五个方面的核心内容陈述清楚、到位。从这些内容中可以清楚地体察到:课题研究单位和课题研究者在“人力、物力、财力、时间、硬件与软件、技术”等方面具备坚实和雄厚的“物质基础”,是对课题框架设计方案质量进行综合评价的关键点。

3.2.3 课题技术设计方案的核心内容

在课题技术设计方案中,应将“基本常识”“研究问题涉及的各科专业知识”和“统计设计知识”三个方面的核心内容陈述清楚、到位。从这些内容中可以清楚地体察到:课题研究单位和课题研究者“在“伦理道德、研究类型、三要素的把握、四原则的遵守、调查或试验设计类型的选定、比较类型的确定”等方面具备坚实和牢固的“技术支撑”,是对课题技术设计方案质量进行综合评价的关键点。

3.3 对课题实施的质量进行综合评价

3.3.1 对课题实施质量进行综合评价的概述

无论一个课题设计方案制订得多么科学完善,它毕竟只是写在纸上或电子文档上的东西,必须通过研究人员将其落实到具体的课题实施之中去,才能发挥其应有的价值。要想对课题实施的质量进行

综合评价,应抓住以下四个重要方面:第一,落实研究设计方案的程度;第二,执行标准操作规程的效果;第三,对实施过程的质量控制力度;第四,研究结果的记录质量。

3.3.2 课题实施中的一盏指路明灯

在课题实施过程中,研究者和所有参与者必须在“课题设计方案”这盏指路明灯的指引下,有条不紊地开展课题研究。不应随便偏移或改变课题设计方案,万一因客观原因或随机效应产生了无法抗拒的偏移,应将有关的原因和结果记录在案,以便将来采取合适的方式进行校正。可以这样说,认真仔细检查课题设计方案在课题实施中的具体落实情况,就是对课题实施质量进行综合评价的第一个关键点。

3.3.3 课题实施中的一个重要依据

课题实施的全过程通常需要被分割成多个不同的阶段,在每个阶段上,往往是由一位或多位研究者和参与者同时发出一系列“动作”并产生相应的“半成品”或“局部结果”;将产生于多个阶段的“半成品”或“局部结果”串起来,就形成了课题的最终结果。在这一漫长且由多位实施者参与的过程中,每个阶段上的“动作”是否“标准”必将影响最终结果的准确性。因此,课题实施中的一个重要依据就是“标准操作规程”。显而易见,考查“有无标准操作规程以及是否严格按其实施”,就是对课题实施质量进行综合评价的第二个关键点。

3.3.4 课题实施中的一个有力举措

由于不同的研究者或操作者的技术水平、熟练程度和责任心不尽相同,加之环境和条件的变化,即使有标准操作规程作为“动作”的依据,也很难完全避免自始至终不会发生任何偏差的情况。在课题实施的任何时间节点上,如何能及时且精准地发现已经出现的“偏差”并及时采取有效措施予以纠正,就是保证研究结果精准可靠的有力举措,其实质就是在课题实施过程中能够实时精准进行质量控制。控制的源头可能来自下面的诸多方面:研究者、受试者(特别是受试者的心理因素)、环境与条件(仪器、试剂)等。因此,考查“有无质量控制以及质量控制的成效如何”,就是对课题实施质量进行综合评价的

第三个关键点。

3.3.5 课题实施中的一个对应关系

以数据说话的试验性或调查性研究课题,得出结论的有力证据是科研数据。因此,科研数据的质量是结论正确与否的重要根基。体现科研数据质量的要素有以下几点:原始性、实时性和精准性。所谓原始性,就是所记录下的科研数据未经任何修正或改动;所谓实时性,就是所获得的科研数据自其产生那一刻就被记录在案了;所谓精准性,就是采用了正确的测定方法在恰当的时间节点上客观精准地度量了被测事物或现象的特定指标的数量大小。前述提及的所有要素都需要通过“试验记录”全面、系统、完整地反映出来。“试验记录”通常包括以下几种具体情况:“试验记录本”或“病例报告表(简称 CRF)”或“某调查研究项目的调查表”或相应的“电子记录系统”。有一个简单的方法可用来评价某个研究课题所获得的科研数据是否具有前述提及的三个基本要素,就是看“课题实施中是否存在一个对应关系”,即“试验记录”中所记载的内容和数据是否都能从与其对应的设计方案中找到“缘由和依据”。因此,找到“试验记录”与“设计方案”之间的很好对应关系,就是对课题实施质量进行综合评价的第四个关键点。

3.4 对系统评价的质量进行综合评价

3.4.1 对系统评价质量进行综合评价的概述

可能有相当多的人把“Meta 分析”当作“系统评价”的代名词,这似乎是以点代面的真实写照。不言而喻,“Meta 分析”就如同传统综合评价方法一样,只是众多统计分析方法中的很小的一类,它常被排除在“常用统计分析方法”之外。而“系统评价”则不同,它是“在全面收集所有有关研究的基础上,对所有纳入的研究逐个进行严格的评价,并联合所有研究结果进行综合分析和评价,必要时进行 Meta 分析,以得出尽量客观、审慎的结论^[4]。”

由此可知,要想对系统评价的质量进行综合评价,应抓住以下五个重要方面:第一,检索文献所来自数据库的种类和质量;第二,检索文献语种的数量、检索范围和检索策略;第三,涉及到的所有同类研究本身的科研设计质量;第四,涉及到的所有同类研究本身的课题实施和数据分析质量;第五,作为

“系统评价项目”本身涉及到的“条目、一览表或清单、量表”和“相应的质量评估报告”的质量。

3.4.2 用于系统评价的文献来源种类与质量

系统评价通常是基于众多同类研究的“文献研究”,而不是基于系统评价者亲自开展的试验或调查或理论推导研究。不难想象,系统评价的质量与文献来源的种类与质量是密切相关的。迄今为止,世界上的数据库数目繁多,各数据库所涉猎的内容范围不尽相同,数据库本身的质量也存在差异。所以,对一个特定的系统评价研究项目而言,其所采用的数据库种类多少与质量高低,是对该系统评价项目质量的综合评价的第一个关键点。

3.4.3 用于系统评价的具体文献被提取的依据

在指定的数据库中,用于进行系统评价的具体文献是如何被提取出来的呢?这个问题取决于检索文献的语种、检索范围和检索策略。显然,所提取的将被用于系统评价项目的文献来自于“语种多、范围广和检索策略的‘智商高’的数据库”是一个极其重要的前提条件。因此,前述的这些方面就是对该系统评价项目质量的综合评价的第二个关键点。

3.4.4 用于系统评价的研究项目本身的设计质量

不难想象:从所有被检索出来的文献中提取信息的质量高低,在很大程度上取决于系统评价者为该系统评价项目所制订的“设计方案”的质量。这个“设计方案”不同于为一个具体的试验或调查研究项目所制订的“设计方案”,前者的主要内容是如何从所检索到的文献中提取出能系统全面深入反映源研究项目与“设计、实施、分析和结论”有关的重要信息或证据;而后的主要内容是关于如何更好地实现试验或调查目的所做的一切考虑和安排。所以,评估用于系统评价的研究项目本身的设计质量(它必须能全面客观真实地反映源文献的科研设计质量)就是对该系统评价项目质量的综合评价的第三个关键点。

3.4.5 用于系统评价的研究项目本身的实施和分析质量

在此标题下需要陈述的很多内容与本文“3.4.4 节”类似,篇幅所限,不再赘述。所以,评估用于系统评价的研究项目本身的实施与分析质量

(同时,它还必须能全面客观真实地反映源文献的实施与分析质量),就是对该系统评价项目质量的综合评价的第四个关键点。

3.4.6 作为系统评价项目本身在核心要素上的质量

作为一个系统评价研究项目,不可避免地会用到条目(由单个条目组成,可能是影响研究结果的偏倚因素,如分配方案的隐藏、盲法、随访结果及失访病例的处理方法)、清单或量表(由多个评价研究质量和偏倚风险的条目组成,给每个条目都评分,能定量地估计整个研究的质量)。那么,条目或清单或量表在内容上、范围上和核心要素上考虑周到、做到“准确、全面、简练”就是对该系统评价项目质量的综合评价的第五个关键点。

3.5 如何对统计分析的质量进行综合评价

3.5.1 对统计分析质量进行综合评价的概述

一位有着一辈子丰富经验和深厚统计功底的统计分析师与一个初出茅庐且仅学了一点统计学皮毛的统计分析者都可能会承担十分艰巨和复杂的统计分析任务。那么问题在于:人们如何去评价一个研究项目中的统计分析质量的好坏呢?

由一个研究项目所产生的科研资料,通常,其内容是十分丰富的。简单地说,它一定是多因素多指标的产物。要想对其进行全面、深入的统计分析,关键要把握好以下几点:第一,应有充分的证据表明,待分析的科研资料是值得分析的;第二,应能全面、系统地诊断出科研资料所具备的前提条件或存在的问题;第三,应依据科研资料所对应的以下诸多方面,合理地选择一系列对应的具体统计分析方法。前面提及的诸多方面概括如下:①每项具体统计分析的目的是什么;②科研资料来自的研究类型、设计类型和比较类型分别是什么;③原因变量与结果变量的性质分别是什么;④同时参与某项具体统计分析的原因变量与结果变量的个数分别是多少;⑤希望统计分析结果的精确度是多高;⑥拟采用的统计分析软件的可信度如何;⑦若采用自编统计软件,其所依赖的算法本身是否科学严谨、软件本身是否可靠。

3.5.2 必须有证据表明待分析的科研资料是值得分析的

一些人可能会认为:科研资料只要经过了统计分析,其结论就是可信的。显然,这是过分夸大了统计分析的作用。事实上,正确合理地使用统计分析方法,只能将已经存在于科研资料中的规律呈现出来,而它绝不可能在根本不存在规律的科研资料中创造出规律来。换句话说,只有确实包含了“客观规律”的科研资料,才是值得进行统计分析的,问题是事先如何进行判断呢?

判断一个科研项目所产生出的科研资料是否值得进行统计分析的方法是考察在以下几个方面是否存在严重的“毛病或硬伤”:①科研设计方案中;②课题实施过程中;③产生和获取科研资料过程中。在前述的三个方面,若存在严重的“毛病或硬伤”,其科研资料就是不值得分析的。最常见的不值得分析的科研资料有以下几种情形^[6]:①人为编造的科研资料;②在统计设计的三要素与四原则方面存在严重瑕疵的科研资料;③产生于质量控制不严的科研资料;④经过错误方法加工整理后的科研资料;⑤不符合特定统计分析方法要求的科研资料;⑥缺失数据过多的科研资料。

3.5.3 必须有证据表明待分析的科研资料满足所选用的统计分析方法的要求

若某研究项目在以下三个方面做得很好,有望保证所获得的科研资料是值得分析的^[6]:①制订出了科学完善的科研设计方案;②按标准操作规程实施并严格控制了课题实施过程中的质量;③有实时精准记录科研资料的规格化表格或方法。

尽管如此,还要看待分析的科研资料是否满足所选用的统计分析方法的要求,因为很多统计分析方法都是在特定的假设之下推导出来的。在这些假设之中,有些是在统计学教科书上明确写出来的,还有一些是隐含的。例如,对于单因素两水平设计一元定量资料而言,人们最常选用的差异性统计分析方法是 t 检验。然而, t 检验的书面前提条件分别为:独立性、正态性和方差齐性;隐含的两个前提条件分别为:①两组受试对象在一切非试验因素(包括研究者已观测的和未观测的)方面均衡可比;②所考查的试验因素与未考查的非试验因素之间的交互作用效应不存在或可以忽略不计。事实上,在很多人所完成的 t 检验中,能满足前述提及的两个隐含前提条件的并不多。

3.5.4 必须有证据表明所做的各种统计分析都是无懈可击的

通常,一个研究项目对应的全部科研资料(可被称为全数据集)都是相当复杂的,需要根据不同的分析目的将其进行分解形成多个不同的子数据集,再根据各子数据集所对应的具体情况,选择不同的统计分析方法对其进行处理。因此,就不可避免地会涉及到两个方面的问题:其一,对全数据集的拆分是否科学合理;其二,所做的全部统计分析是否系统全面。必须有证据表明所做的各种统计分析都是无懈可击的,才表明对该研究项目的统计分析质量是令人满意的。

4 结 语

综上所述,广义综合评价包括的内容确实丰富多彩。从宏观上来看,可以涵盖对人类所做过的任何事或自然界中发生过的任何事的评价;从课题研究角度来看,可以囊括对科研设计、课题实施和课题验收等环节的质量评价;仅从统计分析角度来看,综合评价除了包含传统综合评价方法之外,几乎可以包括其他所有的统计分析方法(如差异性分析、相

关与回归分析、聚类与分类分析等)。

事实表明,将传统综合评价拓展为广义综合评价,还“综合评价”的真面目,至少可以把与统计学有关的所有内容整合在一起,使其发挥“共鸣”“联动”和“整体”效应,对人们客观地认识和评价事物或现象具有很大的现实意义。

参考文献

- [1] 胡良平. 科研设计与统计分析[M]. 北京: 军事医学科学出版社, 2012: 129-730.
- [2] 孙振球. 医学统计学[M]. 北京: 人民卫生出版社, 2002: 373-396.
- [3] 孙颀龄. 中国医学统计百科全书: 统计管理与健康统计分册[M]. 北京: 人民卫生出版社, 2004: 30-94.
- [4] 罗杰, 冷卫东. 系统评价/Meta 分析理论与实践[M]. 北京: 军事医学科学出版社, 2013: 1-22.
- [5] 胡良平. 课题设计与数据分析—关键技术与标准模板[M]. 北京: 军事医学科学出版社, 2014: 1-51.
- [6] 胡良平, 胡纯严, 鲍晓蕾. 应用数理统计[M]. 北京: 电子工业出版社, 2015: 1-36.

(收稿日期: 2017-02-11)

(本文编辑: 吴俊林)

统计学培训通知

北京华斯泰生物医学科技有限公司拟于 2017 年 4 月 21 日-23 日在北京冠京饭店(北京市丰台区丰台北路 79 号)举办“基于 SAS 和 R 软件的广义差异性分析实战训练培训班”,将邀请具有三十多年统计学教学经验的著名统计学专家全程授课。与此次统计学培训班有关的详细情况和报名

方式请通过下面的网址或邮箱或直接电话咨询等方式获取。
网址: www.Huasitai.com; E-mail: Training@Huasitai.com;
电话: 010-57699287, 010-57699286; 联系人: 李老师 13366403928, 章老师 13121195178。