

doi:10.16055/j.issn.1672-058X.2020.0004.015

一种基于 Meta 分析结果的综合评价方法

马小艳

(重庆工商大学 数学与统计学院,重庆 400067)

摘要:Meta 分析是循证医学中最高级别证据合成的方法,建立一种循证的综合评价方法,其综合性评价研究比任何单项评价研究都能提供更精确、概括和客观的结果;因此,从评价方案、资料收集、文献质量评价、数据处理、统计分析、敏感性分析和评价结果分析与报告七个方面探讨了基于 Meta 分析的综合评价基本范式,并以精准扶贫满意度测评为例,综合评价了基层群众对扶贫政策满意度情况。Meta 分析的优越性是其决策过程基于更真实、更全面的证据,不仅整合了以前的工作,还阐明了未来研究的差距和重点;但其自身存在的方法学问题,以及如何将不同研究结果科学地放在同一测量尺度上,以避免“苹果”与“橘子”的混合仍是各领域面临的挑战。

关键词:Meta 分析;综合评价;循证医学;循证社会科学;科学决策;精准扶贫

中图分类号:C32 **文献标志码:**A **文章编号:**1672-058X(2020)04-0104-07

0 引言

Meta 分析是对研究结果进行科学、定量的综合,是循证医学中被广泛应用的一种证据合成方法,是循证医学中被广泛应用的一种证据合成方法。循证医学是指基于现有的最好临床证据,结合医疗条件、病人情况以及医生经验做出的医疗决策^[1]。证据是循证医学的核心基础,当前,系统评价和 Meta 分析被认为是循证医学中做循证决策的最高级别依据^[2],是针对某一具体问题研究证据的最佳手段,可以得到客观的综合评价结果。循证社会科学是应用循证医学的循证理念、思想和方法在社会实践中对管理、教育、经济等活动进行科学的综合评价,做出科学的决策,因此,证据依然是循证社会科学的核心,生产、评价和转化应用高质量的证据是循证社会科学的重要任务^[3]。无论是在医学研究领域,还是在社会科学研究领域,在这个信

息爆炸式增长,研究成果成百上千涌现的大数据时代,面对同一研究问题,其文献质量参差不齐,研究设计、研究对象也各有差别,单项研究成果可能不同甚至相互矛盾。如何去伪存真,把握总体水平或得到结论式答案,需要对这些研究成果进行科学的评价,得出最佳证据和结论。本文将循证医学被广泛应用的证据合成方法——Meta 分析引入社会综合测评方法中,提出基于 Meta 分析的综合评价基本范式,以期能合理应用 Meta 分析,在大量研究成果中解决看似矛盾的结果,使管理决策更科学,评价活动更符合现实。

1 基于 Meta 分析的综合评价基本范式

Meta 分析是一种基于评价结果再评价的综合评价方法,将文献评价和统计方法相结合,不仅可以整合科研设计,评价研究方案,找到研究差距与

收稿日期:2019-10-24;修回日期:2019-11-30.

作者简介:马小艳(1992—),女,重庆合川人,硕士研究生,从事综合评价研究.

重点,还可以量化评价研究结果,判断一致性,反映总体平均水平情况,并揭示差异来源,发现新的研究领域。将 Meta 分析应用于综合评价的基本范式如下:

(1) 提出要评价的问题,设计评价方案。确定 Meta 分析的研究目的,明确要评价的问题,还包括文献检索策略、文献纳入排除标准、数据处理方法以及统计分析方法。

(2) 收集文献资料。按纳入排除标准,如 PRISMA 声明^[2],系统检索相关文献,除了检索公开发表的期刊、图书等文献外,还应尽可能收集如会议论文、白皮书、调查报告等文献。检索手段包括计算机检索和手工检索,检索方法必须正确、全面,不能遗漏对结果评价有重要影响的文章。

(3) 纳入文献质量评价。Meta 分析是对原始研究的二次综合分析,若纳入文献质量低,且未对原始研究方法进行评价,其结果和结论可能是偏离的、不正确的。因此,必须严格对纳入的文献,根据其研究设计、实施过程、结果分析等过程进行质量评价。质量评价标准应根据该研究领域制定的条目^[2]与标准,组织专家讨论,评价结果应进行分类或排序,必要时,应根据分类再做亚组分析(即根据分类再分别做 Meta 分析)。好的质量评价结果应能够解释研究之间差异的原因。

(4) 数据提取与数据处理。确定纳入 Meta 分析的文献后,提取原始研究中的描述数据和效应值统计数据^[4],数据提取应尽可能全面、准确,避免偏倚、错误和重复劳动。描述数据主要包括文献来源、发表时间、研究方法、研究对象、评价指标、数据来源、结局指标等基本情况。效应统计数据是指将各研究结果放在同一尺度的统计指标,如 t 值、 P 值、回归系数、弹性系数、相关系数、标准误差、方差等。效应统计数据是将不同研究成果进行合并分析的重要数据,有的可直接在原文献中获得,有的需要间接计算或联系作者等方式取得。

(5) 统计处理与分析。主要有三方面的工作:同一尺度测量指标转化、异质性检验、整合分析。

① 同一尺度测量指标转化。因各原始文献中研究变量的定义、测评的方法、变量的赋值可能存

在差异,需要将各研究结果进行统计处理,转化同一尺度测量的效应量(Effect Size, ES)指标才能进行 Meta 整合分析,其一般转化形式为

$$ES_i = f_i(X_{i1}, X_{i2}, \dots, X_{ik})$$

ES_i 为第 i 个研究的效应统计指标, X_{ij} 为第 i 个研究中的第 j 个测评指标, f_i 的选择应根据评价目的和原始文献提供的数据确立最适合的效应量指标。

② 异质性检验。虽然按照严格标准纳入文献和质量评价,最大限度地减少了异质性来源,但仍可能存在不同质(即不相似)的情况,异质性的大小决定了模型的选择与是否有必要进行整合分析。因此,需要对各研究结果的评价指标即效应统计指标进行异质性检验。异质性检验有统计量法和图示法,最常用的统计量检验为 Q 检验^[4]。

若 Q 检验零假设 $H_0: ES_1 = ES_2 = \dots = ES_K$ (即各独立研究结果不存在异质性),备择假设 $H_1: ES_1 \neq ES_2 \neq \dots \neq ES_K$ (即各独立研究结果存在异质性),则 Q 检验统计量为

$$Q = \sum_{i=1}^K \omega_i (ES_i - \bar{ES})^2 \sim X^2(df) \quad (1)$$

式(1)中: $\omega_i = (se_i^2)^{-1}$, ES_i 为第 i 个研究中的效应量指标, \bar{ES} 为所有研究的平均效应量指标, se_i 为第 i 个研究的标准误差,自由度 $df = K - 1$, K 为研究总数。值得注意的是, Q 检验对纳入研究数量敏感,其值会因自由度的增加而增加,当纳入 Meta 分析文献数低于 30 时, Q 检验的检验效能较低。也就是说,当 Q 检验结果具有统计学显著性,各研究结果间存在异质性,但检验结果不具有显著性时,其不一定没有异质性。因此,为提高 Q 检验结果的稳定性,通常取 $\alpha = 0.1$ 作为显著性水平,即当 $p < \alpha = 0.1$ 时,研究间存在异质性^[5]。

③ 整合分析。Meta 整合分析的关键之处是平均效应量即合并效应量的计算,反映的是研究问题总体的平均效应情况,其一般形式为

$$ES_{\text{合并}} = \sum_{i=1}^K W_i ES_i$$

W_i 的确定决定了模型的不同,如固定效应模型

(Fixed effect model)中, $W_i^F = \frac{\omega_i^F}{\sum \omega_i^F}$, $\omega_i^F = \text{var}(ES_i)^{-1}$;

随机效应模型(Random effect model)中,对 ω_i^R 进行修正, 即 $W_i^R = \frac{\omega_i^R}{\sum \omega_i^R}$, $\omega_i^R = \left(G + \frac{1}{\omega_i}\right)^{-1}$, $G =$

$$\frac{Q - (K - 1)}{\left(\sum \omega_i - \frac{\sum \omega_i^2}{\sum \omega_i}\right)}, \omega_i = \text{var}(ES_i)^{-1} [5-6]。固定效应模$$

型和随机效应模型是 Meta 分析的两种基本模型, 当各研究间同质性较高时选择固定效应模型, 否则, 选择随机效应模型。

(6) 敏感性分析。敏感性分析是用于评价加入调节因素对 Meta 分析结果影响大小的方法。其目的是探索影响 Meta 分析研究结果的主要因素, 解决研究结果间的矛盾性, 解释产生不同结论的原因^[7]。敏感性分析可以使用文献质量、研究变量、数据类型、研究样本等作为调节变量, 探索调节变量对 Meta 分析研究结果的影响, 其基本模型为 Meta 回归模型, 估计方程一般为多元回归估计^[8]:

$$ES_i^{\text{effect}} = \alpha + \beta X_i + \gamma X_i^{\text{adjust}} + \varepsilon_i$$

ES_i^{effect} 为第 i 个独立研究结果的不同调节变量影响的合并结果, α 为常数项, X_i 为自变量向量, X_i^{adjust} 为调节变量向量, β 、 γ 为各变量系数, ε_i 为残差项。

(7) 评价结果分析与报告。Meta 分析是根据已有不同研究结果进行文献评价和统计分析相结合的综合评价。因此, Meta 分析最终的评价结果与

报告应包括文献质量评价、Meta 整合分析、Meta 回归分析以及 Meta 分析结果的解释。更具体地, 一是根据 Meta 整合分析结果, 把握总体平均水平, 得出一个更为合理、概括和有据可查的结论; 二是根据文献质量评价、Meta 回归分析等结果解释影响这个研究主题结论差异的来源, 分析调节变量影响效应量的大小; 三是根据综合评价结果, 总结目前该研究领域存在的问题, 分析未来的研究方向应该如何开展。若在综合研究过程中发现了新的研究假说, 应该明确指出来。

2 实例分析

以精准扶贫满意度测评为例, 综合评价全国精准扶贫满意度总体情况, 并分析差异来源。

2.1 文献资料

检索策略: 输入“扶贫”、“精准扶贫”、“扶贫绩效”、“扶贫满意度”等关键词, 全面检索知网、政府网等网站。

纳入排除标准: 纳入文献须为实证研究, 排除理论性、叙述性研究; 纳入文献须报告总体满意度或各评价指标满意度等结局指标; 排除无法用李克特量表对评价结果进行同尺度转化的研究。

筛选结果: 纳入本次 Meta 分析的文献共计 37 篇, 其中核心期刊 8 篇, 一般期刊 14 篇, 学位论文 15 篇(表 1)。文献质量评价如表 2 所示。

表 1 入选文献基本信息(部分)

Table 1 Basic information of selected literatures (part)

研究编号	作者	发表年份	文献来源	研究地区	研究对象	抽样方法	样本量/个	结局指标
1	刘琪	2015	D	成都龙泉驿区	农村居民	分层抽样	527	1*2*3*4*5*
2	孙璐	2015	D	四川大小凉山	建档立卡贫困户	分层判断与随机抽样	966	1*2*3*4*5*
3	刘昌昊	2015	J	武陵山区	农村居民	随机抽样	382	1*2*3*4*
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
36	梁传波	2019	J	苏北五市	建档立卡贫困户	-	278	1*2*3*4*
37	张亚南	2019	J	崇礼、张北县	建档立卡贫困户	-	230	1*3*5*

注: 文献来源“J”=“期刊”, “D”=“学位论文”; 结局指标: “1*”=“总体满意度”, “2*”=“总体标准差”, “3*”=“各评价指标满意度”, “4*”=“各评价指标标准差”, “5*”=“各评价指标满意度占比”。

表 2 纳入研究的文献质量评价(部分)

Table 2 Evaluation of literatures quality in included studies(part)

评价条目	说 明	NO. 1	NO. 2	NO. 3	NO. 4	...	NO. 36	NO. 37
研究目的	清楚描述了研究目的或研究假设	Y	Y	Y	Y	...	Y	Y
研究设置	清楚描述了研究地点及相关资料	Y	Y	Y	Y	...	Y	Y
问卷设计	咨询了相关专家或实地进行了 调研或参照了有关部门制定的问卷	N	N	Y	Y	...	N	N
调查方式	清楚描述了调查的主要方式	Y	Y	Y	Y	...	Y	Y
调查对象	清楚描述了调查对象并在调查前进行了识别	Y	Y	Y	Y	...	Y	Y
样本大小	清楚描述了样本的选择方法	Y	Y	Y	Y	...	N	N
评价指标	提供了评价指标体系或清楚描述了评价指标	Y	Y	Y	Y	...	Y	Y
调查过程	对调查员进行了事前培训或充分考虑了 调查对象的文化差异或采取措施弥补了问卷上的不足	Y	N	N	Y	...	Y	N
调查结果	提供了信效度检验	N	N	N	N	...	N	N
	提供了调查样本基本特征	Y	N	Y	Y	...	Y	N

注:“NO. *i*”=“第 *i* 个研究”;“Y”=“有”,“N”=“未提及或不清楚”。

2.2 统计分析

本次纳入的文献满意度均采用李克特量表进行量化,即分值 5(非常满意)、4(比较满意)、3(一般)、2(不满意)、1(非常不满意),总体满意度具有可比性,效应量指标、标准误差、权重计算公式分别为

$$ES_i = \bar{X}_i = \frac{\sum x_i}{n_i}$$

$$SE_i = \frac{s_i}{\sqrt{n_i}}$$

$$w_i = \frac{1}{SE_i^2} = \frac{n_i}{s_i^2}$$

运用软件 CMA2.0 对数据进行运算,得到研究样本总体效应值及异质性检验结果(表 3)。

表 3 样本总体效应检验结果

Table 3 Total effect test results of samples

模 型		95% 的置信区间		渐进值		文献数
		下限	上限	Z 值	P 值	
固定效应	3. 316	3. 307	3. 325	715. 960	0. 000	37
随机效应	3. 445	3. 265	3. 624	37. 573	0. 000	

注:异质性检验结果: $Q=13\ 593.914, p=0.000, df=36, I^2=99.735$ 。

通过表 3,本次纳入文献异质性检验结果 $Q=13\ 593.914, p=0.000$,且 $I^2>50\%$,因此采用随机效应模型进行合并,获得总体效应量为 3.445,处于比

较满意和一般之间,总体满意度不是特别高。为进一步查看各调节变量对扶贫政策满意度的测评是否有影响,进行 Meta 二元异质性检验,结果如表 4。

表 4 Meta 二元异质性检验结果

Table 4 Meta binary heterogeneity test results

调节变量	文献 数量/个	样本 容量/个	ES	95% 的置信区间		Q 值	P 值
				下限	上限		
发表时间							
2015	3	1875	3.494	2.327	4.661	937.300	0.000
2017	8	3584	3.569	3.267	3.872	576.541	0.000
2018	22	9578	3.374	3.314	3.615	9 970.556	0.000
2019	4	988	3.550	3.216	3.885	367.181	0.000
研究地区							
东北	2	356	2.617	0.976	4.259	795.668	0.000
华北	5	3 140	3.425	2.997	3.853	2362.902	0.000
华东	5	1 677	3.624	3.359	3.889	167.253	0.000
华南	3	922	3.318	3.068	3.568	76.234	0.000
华中	3	303	3.290	3.080	3.499	22.411	0.000
西北	6	4 299	3.537	2.934	4.140	4 100.769	0.000
西南	11	4 460	3.532	3.124	3.941	3 702.954	0.000
其他	2	868	3.563	3.240	3.887	56.894	0.000
研究对象							
建档立卡贫困户	25	10 136	3.535	3.301	3.769	12 815.360	0.000
农村居民	12	5 889	3.257	3.066	3.447	711.915	0.000
抽样方法							
随机抽样	17	5 569	3.376	3.098	3.654	8 403.551	0.000
多阶段抽样	2	2 234	2.688	1.940	3.436	309.157	0.000
分层抽样	5	3 634	3.662	3.477	3.847	102.809	0.000
全面调查	2	381	3.763	3.160	4.365	85.052	0.000
其他	11	4 207	3.535	3.239	3.830	2 164.084	0.000

根据表 4,从年份来看,基层群众对扶贫政策的满意度没有太大差异;从地区来看,东北地区效应量 $ES=2.617$,处于不满意水平,各地区表现了不同满意度水平;从研究对象看,建档立卡贫困户的评价要高于一般农村居民;而抽样方法中,全面调查的满意度水平最高($ES=3.673$)。发表时间、研究地区、研究对象、抽样方法的 Meta 回归影响系数分别为 0.022、0.054、0.101、0.029 (p 值均小于 0.01),亦发现研究对象对单项研究结果的差异影

响最大,研究地区其次,说明了对调查对象的识别与各区域的实施情况对扶贫政策的测评均有较大的影响。

2.3 评价结果

本次纳入 Meta 分析的单项研究总数 $K=37$,总样本量 $N=16 025$,覆盖了中国七大地理区域,时间包含了 2015—2019 年(2016 除外),其综合评价结果如下:

(1) 从文献质量评价来看,大部分单项测评研

究都进行了实地调研,咨询了相关专家,参照了扶贫绩效考核标准,并对调查对象进行了精准识别,使评价研究建立在更加符合事实基础之上。

(2) 从单项测评研究结果看,有的评价结果处于非常满意和比较满意之间,而有的结果处于非常不满意与不满意之间,表现了较大的差异性。

(3) 通过 Meta 整合结果,总的来看,全国群众对精准扶贫政策的满意水平为 3.445,处于比较满意与一般之间,距离满意有一定的差距。

(4) 根据二元异质检验和 Meta 回归分析结果,从政策提出到实施,随着时间的推移,满意程度没有太大的影响;建档立卡贫困户的满意度最高,而在东北的测评结果最低,处于不满意水平,研究对象的识别和不同研究地区对单项研究结果造成了一定的差异性。

(5) 更进一步,还可以建立在 Meta 分析基础上对扶贫政策绩效指标进行分析,并对满意度影响因素进行更具体地分析,本文仅从总体满意度以及影响单项测评结果差异因素进行了分析与评价。

3 结 语

以精准扶贫满意度测评为例,综合评价了全国群众对扶贫政策满意度情况,与单个满意度测评研究相比,基于 Meta 分析的扶贫政策满意度测评扩大了样本量,从总体上把握了基层群众对精准扶贫政策的评价情况;利于在诸如研究时间、研究对象、研究方法、研究区域等多种维度进行对比分析,分析各单项评价不一致原因;建立在大量同类研究成果之上的研究,提高了效应的评估力度,结果更令人信服。这是一个信息时代,随着计算机科学和大数据的发展,各个领域的研究数量巨大,并在不断增长,Meta 分析能够以最大限度从大量可用的研究中挖掘有效的信息。单项的初步研究可能被看作是对证据积累的贡献,而不是对科学问题结论性答案的揭示,将在循证医学广泛应用的 Meta 分析用于社会综合测评,能够消化和吸收从某一特定领域的大量出版物中获得的信息,其综合性评价研究比任何

单项评价研究都能提供更真实、更全面的证据。但是,确定最有效的决策证据并不容易,Meta 分析亦有其自身的局限性,其方法论的合理性与将不同研究成果放在同一尺度测量的有效性依赖于 Meta 分析的方法与技术的发展。当前,跨学科的合作、大型数据库的建立以及更多综合研究方法应用标准的出版有望推动 Meta 分析在各个领域得到发展与应用。

参考文献 (References):

- [1] 刘露,崔馨月,张宇晶,等. 系统综述、循证医学与医疗器械上市前临床评价的探讨[J]. 中国医疗设备, 2019,34(9):157—160,167
LIU L, CUI X Y, ZHANG Y J, et al. Study of Systematic Review, Evidence-Based Medicine and Pre-Marketing Clinical Evaluation of Medical Devices[J]. China Medical Devices, 2019, 34(9): 157—160, 167 (in Chinese)
- [2] 杨克虎,李秀霞,拜争刚. 循证社会科学研究方法:系统评价与 Meta 分析[M]. 兰州:兰州大学出版社,2018
YANG K H, LI X X, BAI Z G. Evidence-Based Social Science Research Methods: Systematic Evaluation and Meta-Analysis[M]. Lanzhou: Lanzhou University Press, 2018 (in Chinese)
- [3] 杨克虎. 循证社会科学的生产、发展与未来[J]. 图书与情报,2018(3):1—10
YANG K H. Evidence-Based Social Science: The Origin, Development and Prospects[J]. Library & Information, 2018(3): 1—10 (in Chinese)
- [4] 徐双丽,吕昭河. Meta 分析方法在经济学中的发展及应用[J]. 统计与信息论坛,2019(8):34—41
XU S L, LV Z H. Study on the Application Paradigm of Meta-Analysis in Economics[J]. Statistics & Information Forum, 2019(8): 34—41 (in Chinese)
- [5] 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
LUO J, LENG W D. Theory & Practice of Systematic Review/Meta-Analysis [M]. Beijing: Military Medical Science Press, 2013 (in Chinese)
- [6] TRIKALINOS T A, SALANTI G, ZINTZARAS E, et al. Meta-Analysis Methods[J]. Advances in Genetics, 2008,

60;311—334

- [7] 黄爱君,詹思延. 系统综述和 Meta 分析[J]. 中国药物应用与监测,2009,6(4):257—259
- HUANG A J, ZHAN S Y. Systematic Review and Meta-Analysis[J]. Chinese Journal of Drug Application and Monitoring,2009,6(4):257—259(in Chinese)

- [8] 彭俞超,顾雷雷. 经济学中的 Meta 回归分析[J]. 经济学动态,2014(2):126—131
- PENG Y C, GU L L. Meta Regression Analysis in Economics[J]. Economic Perspectives,2014(2):126—131(in Chinese)

A Comprehensive Evaluation Method Based on Meta-analysis Results

MA Xiao-yan

(School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, China)

Abstract: Meta-analysis is the highest level of evidence synthesis method in evidence-based medicine. It is applied in social comprehensive evaluation to establish an evidence-based comprehensive evaluation method. Its comprehensive evaluation research can provide more accurate, general and objective results than any single evaluation research. Therefore, the basic paradigm of comprehensive evaluation based on meta-analysis was discussed from seven aspects including evaluation program, data collection, literature quality evaluation, data processing, statistical analysis, sensitivity analysis and evaluation result analysis and report. Taking the accurate poverty alleviation satisfaction assessment as an example, the satisfaction of grassroots people with Poverty Alleviation Policies is evaluated comprehensively. The advantage of meta-analysis is that its decision-making process is based on more authentic and comprehensive evidence, which not only integrates the previous work, but also clarifies the gap and focus of future research. However, its own methodological problems and how to scientifically put different research results on the same measurement scale to avoid the mixing of "apple" and "orange" are still the challenges for all fields.

Key words: meta-analysis; comprehensive evaluation; evidence-based medicine; evidence-based social sciences; scientific decision; targeted poverty alleviation

责任编辑:李翠薇

引用本文/Cite this paper:

马小艳. 一种基于 Meta 分析结果的综合评价方法[J]. 重庆工商大学学报(自然科学版),2020,37(4):104—110

MA X Y. A Comprehensive Evaluation Method Based on Meta-analysis Results[J]. Journal of Chongqing Technology and Business University (Natural Science Edition),2020,37(4):104—110