

I^2 检验在Meta分析异质性评价中的误用

王世琦, 蒋青青, 黄 申, 谢雨霖, 曹世义

华中科技大学同济医学院公共卫生学院 (武汉 430030)



【摘要】 Meta 分析异质性是指 Meta 分析纳入的各单项原始研究间存在的差异。选用恰当的方法或指标识别和测量异质性是评价 Meta 分析结果可靠性的重要步骤。尽管 I^2 检验是目前使用极为广泛的 Meta 分析异质性检验方法, 但从 I^2 统计量定义可知, I^2 检验并不能真实展现异质性大小, 当前 Meta 分析异质性评价方法误用情况严重。本研究系统梳理 Meta 分析常用异质性评价方法, 通过两个实例展示 I^2 统计量在 Meta 分析异质性检验中的误用情况, 并介绍能够合理展现 Meta 分析异质性的统计指标, 以期提高 Meta 分析异质性评价的可靠性和规范性。

【关键词】 Meta 分析; I^2 检验; I^2 统计量; 异质性

The misuse of the I^2 test in the evaluation of heterogeneity of Meta-analysis

Shi-Qi WANG, Qing-Qing JIANG, Shen HUANG, Yu-Lin XIE, Shi-Yi CAO

School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

Corresponding author: Shi-Yi CAO, Email: caoshiyi@hust.edu.cn

【Abstract】 The heterogeneity of Meta-analysis refers to the differences among the individual studies included in the Meta-analysis. Selecting appropriate methods to identify and measure heterogeneity is an important step to evaluate the reliability of Meta-analysis results. The I^2 -squared (I^2) test is the commonly used heterogeneity test of Meta-analysis, but according to the definition of I^2 statistic, I^2 test cannot truly show the heterogeneity, and the current evaluation methods of heterogeneity of Meta-analysis are seriously misused. This study aimed to systematically review the commonly used heterogeneity evaluation methods of Meta-analysis, introduce the misuse of I^2 statistic in the test of heterogeneity of Meta-analysis through two examples, and propose statistical indicators that can truly show the heterogeneity of Meta-analysis, in order to provide reference for improving the reliability and normativity of Meta-analysis researches.

【Keywords】 Meta-analysis; I^2 test; I^2 statistic; Heterogeneity

在 Meta 分析中, 异质性是指纳入的单项研究间存在的差异, 包括由受试者、干预措施和结局指标等差异引起的临床异质性, 由研究设计和偏倚引起的方法学异质性, 以及统计学异质性^[1]。如果异质性较大, Meta 合并结果的可靠性就会大

大降低, 甚至在异质性过大时应当放弃 Meta 分析而仅进行系统性综述。因此, 在进行 Meta 分析时, 选用恰当的异质性评价手段来识别和测量异质性至关重要。

常用的 Meta 分析异质性检验方法以 Q 检验

DOI: 10.12173/j.issn.1004-4337.202304137

基金项目: 国家卫生健康委员会卫生技术评估重点实验室 (复旦大学) 开放基金项目 (FHTA2023-01)

通信作者: 曹世义, 博士, 副教授, 博士研究生导师, Email: caoshiyi@hust.edu.cn

<https://slyyx.whuzhmedj.com/>

为主,并在此基础上演变出 H 检验和 I^2 检验。近年来发表的 Meta 分析研究中,研究者常用 I^2 取值来划定研究的异质性程度,以 25%、50% 和 75% 区分为低、中、高度异质性,当 $I^2 > 50%$ 时认为存在明显异质性,采用随机效应模型, $I^2 \leq 50%$ 时则采用固定效应模型。然而,这种方式并不符合 I^2 统计量的定义,也无法真实地展示研究异质性大小^[2-3]。本文梳理了目前常用的 Meta 分析异质性评价方法,解读 I^2 统计量误用情况,并介绍能够反映 Meta 分析异质性的统计指标。

1 Meta分析常用的异质性评价方法

Cochrane 手册认为,纳入同一个 Meta 分析的所有研究都不可避免地存在差异,假如一项 Meta 分析中得出的合并效应量为 1.2,其纳入的单项研究效应量分布在 1.1 到 1.3 之间或 0.5 到 1.9 之间,两者的临床价值将大不相同,因此任何 Meta 分析都必须进行异质性评价,常用异质性评价方法有图示法和统计学检验两类。

1.1 图示法

图示法以森林图为主,可显示每项研究及合并效应量的置信区间,如果各单项研究的置信区间很少发生重叠或几乎不重叠,则提示 Meta 合并结果的异质性较大。除森林图外,拉贝图、Galbraith 星状图、漏斗图等也可应用于异质性评价,代表各单项研究的点偏离代表合并效应量的线条越远则提示该研究造成 Meta 分析异质性的可能性越大。

图示法可以简单直观地体现各单项研究偏离合并值的情况,但是存在明显局限性,其主观解读空间较大,且无法定量估计异质性大小,所以在实际研究中往往采用统计学检验来定量评估 Meta 分析的异质性。

1.2 统计检验法

常用的 Meta 分析异质性统计学指标有 Q 统计量、 I^2 统计量、 H 统计量等^[4]。 Q 统计量是标准化尺度上所有单项研究关于平均效应量的离差平方和,展示 Meta 分析可观察到的所有差异,其计算公式如下^[5]:

$$Q = \sum_{i=1}^k \left(\frac{X_i - M}{SE_{X_i}} \right)^2$$

X_i 为第 i 项研究的效应量, M 为固定效应加权的平均效应量, SE_{X_i} 为第 i 项研究的标准误。 Q 统计量服从自由度为 $k-1$ 的卡方分布,当 $P > 0.05$

时可认为研究间的差异由抽样误差引起,即 Meta 分析不存在显著异质性,反之,则说明研究间差异无法完全由抽样误差解释,则必须考虑 Meta 分析的异质性。 Q 统计量在异质性检验中应用非常广泛,但容易受到纳入研究数量的影响,如果纳入研究较多, Q 检验易出现假阳性,如果纳入研究较少,易出现假阴性。

H 和 I^2 统计量在 Q 统计量的基础上,通过校正自由度来降低研究数量对异质性检验结果的影响,相对稳定可靠,其计算公式如下:

$$H = \sqrt{\frac{Q}{df}}$$

$$I^2 = \frac{Q - df}{Q} \times 100\%$$

df 为自由度, I^2 统计量的分子反映了除去抽样误差以外各研究间的离差平方和,故 I^2 统计量反映了研究间异质性在合并效应量总差异中所占的比例^[5]。当 $Q \leq df$ 时, I^2 取值为 0,故 I^2 的取值范围在 0%~100% 之间。需要注意的是, I^2 和 H 统计量只能反映研究间异质性和总差异之间的关系,而不能用来直接判断 Meta 分析异质性大小。

2 I^2 统计量在Meta分析异质性评价中的误用

在部分 Meta 分析中,研究者用 I^2 统计量简单地描述异质性大小,以 I^2 取 25%、50% 和 75% 这三个值来分别划定低、中、高水平的异质性,但 I^2 统计量与异质性大小并无直接对应关系,简单使用 I^2 统计量表达异质性可能会存在 I^2 很大而实际异质性较小,或 I^2 很小而实际异质性较大的谬误。本文将通过两个 Meta 分析实例来展示这一误用。

2.1 I^2 较大而实际异质性较小的研究实例

在一项关于非工作时间入院对急性缺血性脑卒中患者死亡率影响的 Meta 分析中,与工作时间入院相比,非工作时间入院的急性缺血性脑卒中患者发生院内死亡的比值比为 1.08,95% 置信区间为 (1.03, 1.12), I^2 统计量为 69%^[6]。按照目前常用的 Meta 分析结果解释方式,该研究提示非工作时间入院可以显著增加急性缺血性脑卒中患者的院内死亡风险,但效应量异质性较大。然而该 Meta 分析的单项研究效应值分布情况如图 1-A 所示,研究间实际离散程度较小。

2.2 I^2 较小而实际异质性较大的研究实例

在一项关于膳食中摄入 α -亚麻酸对心血管死亡风险影响的 Meta 分析中,与摄入少量 α -亚麻酸的人群相比,摄入大量 α -亚麻酸者死于心血管疾病的比值为 0.89, 95% 置信区间为 (0.81, 0.97), I^2 统计量为 5.6%^[7]。按照目前常用的 Meta 分析结果解释方式,该研究提示膳食中摄入大量 α -亚麻酸是心血管疾病的保护因素,且研究异质性极小。然而该 Meta 分析的单项研究效应值分布情况如图 1-B 所示,研究间实际离散程度较大。

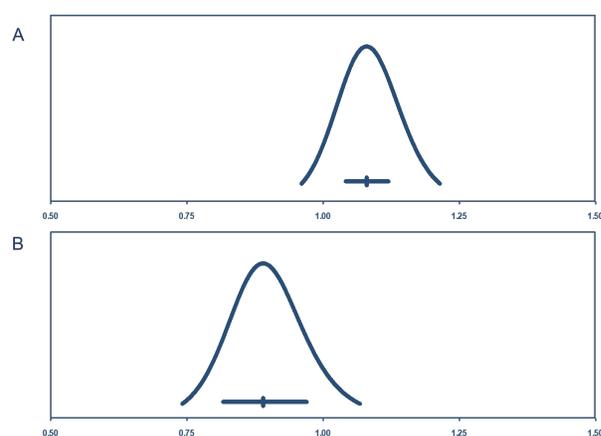


图1 两项Meta分析实例中纳入研究效应值的分布情况

Figure 1. Distribution of the effect sizes of included studies in the two Meta-analysis examples

注: A.该研究的效应量及95%置信区间为1.08 (1.03, 1.12), 95%预测区间为 (0.96, 1.21); B.该研究的效应量及95%置信区间为 0.89 (0.81, 0.97), 95%预测区间为 (0.74, 1.07)。

2.3 I^2 统计量的解读

上述两个实例中,实例 1 的 I^2 明显大于实例 2, 但后者的研究间实际离散程度却更大,说明 I^2 统计量并不能直接判断研究异质性大小。

在 Meta 分析中,需要区分观察到的效应量和真实的效应量。观察到的效应量是对总体人群效应量的估计,但由于抽样误差,观察到的效应量总是与该人群中的真实效应量不同,其方差比真实方差更大。如果每个单项研究的效应值完全相等,那么观察到效应量的预期方差 (V_{OBS}) 将等于各单项研究的误差方差 (V_{ERR}), 即:

$$V_{OBS} = V_{ERR}$$

当每个单项研究的效应值不相等时,这个等式将引入真实效应的方差 (T^2), 即:

$$V_{OBS} = V_{ERR} + T^2$$

根据 I^2 统计量的定义, I^2 计算公式可以改写为:

$$I^2 = \frac{T^2}{V_{OBS}}$$

I^2 统计量反映了研究间异质性在合并效应量总差异中所占的比例,故实例 1 中的 I^2 为 69% 仅表明观察到的差异有 69% 是由异质性而非抽样误差引起,实例 2 中 I^2 为 5.6% 仅表明观察到的差异有 5.6% 是由异质性而非抽样误差引起。对于较小的总差异来说,较大的 I^2 依旧会得出较小的异质性,而相应地,对于较大的总差异来说,较小的 I^2 依旧会得出较大的异质性。

3 真实展现Meta分析异质性的统计指标

总体效应量的 95% 置信区间反映的是总体效应估计值的精确程度,即样本总体的平均水平会落在这个范围内,而不是样本中 95% 的个别值会落在这个范围内。当研究者应用 Meta 分析为临床实践提供指导时,希望结果能够适用于绝大部分患者,此时更需要总体效应值的预测区间。

预测区间是均值两侧两个标准差的范围,如果样本符合正态分布假设,则可以预期总体内 95% 的个体真实效应将落在这个区间内。预测区间直观而清晰地提供了关于绝对离散程度的信息,在 Meta 分析当中应当汇报预测区间来体现研究结果的异质性^[8]。

在实例 1 中,汇报研究异质性时不应只汇报 $I^2=69%$, 而应表明绝大多数非工作时间入院的急性缺血性脑卒中患者的院内死亡风险将落在 0.96 到 1.21 之间。在实例 2 中,汇报研究异质性时不应仅汇报 $I^2=5.6%$, 而应表明绝大多数摄入大量 α -亚麻酸者的心血管疾病死亡风险将落在 0.74 到 1.07 之间。

3.1 预测区间的计算

Meta 分析的预测区间和原始研究类似,需要获得真实效应量的标准差。可使用 Q 和 df 来计算真实效应方差 T^2 的估计值,其分子是反映真实效应变化的离差平方和,分母 C 是基于标准化偏差的权重因子, W_i 是研究的倒方差加权^[8]。

$$T^2 = \frac{Q - df}{C}$$

$$C = \sum_{i=1}^k W_i - \frac{\sum_{i=1}^k W_i^2}{\sum_{i=1}^k W_i}$$

3.2 预测区间的软件实现

Meta 真实效应量标准差的实际计算过程较为复杂, 在常用 Meta 分析工具中已嵌入计算 T^2 的命令, 输出结果常用 τ^2 表示。预测区间的计算往往也可以使用软件进行, 如在 R 软件的“meta”程序包中包含“prediction”逻辑判定, 在 meta 主命令中加入“prediction = TRUE”即可直接输出效应值的预测区间。在网站 <https://meta-analysis-books.com> 中, 可通过免费提供的电子表单绘制纳入研究效应量的正态分布曲线并计算总体效应量的预测区间, 从而直观展现效应量的置信区间和预测区间范围, 相应计算公式由 Borenstein 等提供^[9]。

参考文献

- 1 陈维, 赵守盈, 罗杰, 等. 元分析中三种统计异质性估计方法的比较[J]. 西南师范大学学报(自然科学版), 2015, 40(4): 112–116. [Chen W, Zhao SY, Luo J, et al. Comparison of three estimators of statistical heterogeneity in meta-analysis[J]. Journal of Southwest China Normal University(Natural Science Edition), 2015, 40(4): 112–116.] DOI: 10.13718/j.cnki.xsxb.2015.04.022.
- 2 Borenstein M. In a meta-analysis, the I^2 -squared statistic does not tell us how much the effect size varies[J]. J Clin Epidemiol, 2022(152): 281–284. DOI: 10.1016/j.jclinepi.2022.10.003.
- 3 Migliavaca CB, Stein C, Colpani V, et al. Meta-analysis of prevalence: I^2 statistic and how to deal with heterogeneity[J]. Res Synth Methods, 2022, 13(3): 363–367. DOI: 10.1002/jrsm.1547.
- 4 王若琦, 秦超英. Meta 分析中异质性检验方法的改进[J]. 科学技术与工程, 2012, 12(10): 2256–2259. [Wang RQ, Qin CY. The improvement of testing methods for heterogeneity in meta-analysis[J]. Science Technology and Engineering, 2012, 12(10): 2256–2259.] DOI: 10.3969/j.issn.1671-1815.2012.10.002.
- 5 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis[J]. Stat Med, 2002, 21(11): 1539–1558. DOI: 10.1002/sim.1186.
- 6 Sorita A, Ahmed A, Starr SR, et al. Off-hour presentation and outcomes in patients with acute ischemic stroke: a systematic review and meta-analysis[J]. Eur J Intern Med, 2014, 25(4): 394–400. DOI: 10.1016/j.ejim.2014.03.012.
- 7 Naghshi S, Aune D, Beyene J, et al. Dietary intake and biomarkers of alpha linolenic acid and risk of all cause, cardiovascular, and cancer mortality: systematic review and dose-response meta-analysis of cohort studies[J]. BMJ, 2021(375): n2213. DOI: 10.1136/bmj.n2213.
- 8 Borenstein M, Higgins JP, Hedges LV, et al. Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity[J]. Res Synth Methods, 2017, 8(1): 5–18. DOI: 10.1002/jrsm.1230.
- 9 Borenstein M. Research note: in a meta-analysis, the I^2 index does not tell us how much the effect size varies across studies[J]. J Physiother, 2020, 66(2): 135–139. DOI: 10.1016/j.jphys.2020.02.011.

收稿日期: 2023 年 04 月 19 日 修回日期: 2023 年 05 月 13 日
本文编辑: 李 阳 黄 笛

引用本文: 王世琦, 蒋青青, 黄申, 等. I^2 检验在 Meta 分析异质性评价中的误用[J]. 数理医药学杂志, 2023, 36(8): 561–564. DOI: 10.12173/j.issn.1004-4337.202304137
Wang SQ, Jiang QQ, Huang S, et al. The misuse of the I^2 test in the evaluation of heterogeneity of Meta-analysis[J]. Journal of Mathematical Medicine, 2023, 36(8): 561–564. DOI: 10.12173/j.issn.1004-4337.202304137