

/ 第 1 章 /

元分析概述

在学术研究的浩瀚星海中，每一个研究虽然如同星辰般璀璨，却可能因视角的局限而显得孤立无援。当研究者们面对纷繁复杂的数据与结论时，如何拨开迷雾，探寻到更为普遍、更为可靠的真理？元分析便是为解决这一问题应运而生的一盏明灯。它不仅是一种数据统计方法，更是连接过去与未来、个体与整体的桥梁。

元分析，这一术语由 Glass 在 1976 年首次提出，其内涵随着学术研究的深入而不断丰富。它不仅仅是对已有研究成果的简单汇总，更是一种科学的、系统的综合分析方法。通过严谨的文献筛选、数据提取与统计分析，元分析能够揭示隐藏在众多独立研究背后的普遍规律与趋势，为理论构建与实践指导提供坚实的证据基础。

在本章，我们将深入探讨元分析的基本概念与分类。首先，我们将明确元分析的定义与内涵，揭示其作为一种综合分析方法的独特价值。其次，我们将介绍元分析的两大主要类型——定量元分析与质性元分析，阐述它们各自的研究方法、适用条件及在不同学科领域中的应用情况。最后，我们还将探讨元分析与传统文献综述之间的区别与联系。

通过本章的学习，读者将能够掌握元分析的基本框架与操作流程，为后续章节的深入学习奠定坚实的基础。

1.1 相 关 概 念

1.1.1 什么是元分析

元分析的英文是 meta-analysis, 其中 meta 是英文中的前缀, 有“更加全面或超常规的综合”的意思。国内许多学者将 meta-analysis 译为元分析、荟萃分析、整合分析、汇总分析、二次分析、集成分析、再分析、汇后分析等, 还有许多学者来用“meta 分析”。元分析还有许多同义词, 如总观评述 (overview)、定量评论 (quantitative review)、定量综合 (quantitative synthesis)、资料汇总 (data pooling)、资料综合 (data synthesis) 等。

第一次使用“元分析”这一概念的是美国学者 Glass, 他在 1976 年美国教育研究联合会 (American Education Research Association) 的发言致辞中首次提出元分析概念。

- Glass 对元分析提出的定义: 以综合已有的发现为目的, 对单个研究结果的集合的统计学分析方法。
- Brown 等认为元分析是指“整合相关的单个研究的发现, 并汇集成大样本资料进行统计分析”, 它有助于识别企业单个独立研究难以判断的变量关系以及权变效应。之后的许多统计学家也做了类似定义, 即把元分析仅看作是对以往研究结果进行定量合并的统计分析方法。
- Sacks 等提出的定义: 对以往的研究结果进行统计学的合并和严谨的综述的方法。
- 罗湘认为元分析是依靠搜集已有或未发表的具有某一可比特性的文献, 应用特定的设计和统计学方法进行分析与综合评价, 使之有可能对具有不同设计方法及不同样本数的研究结果进行综合比较的方法。
- Milos 把前阶段的元分析称作典型或定量元分析 (classic meta-analysis or

quantitative meta-analysis)。随着元分析方法应用范围的扩大和其自身的发展,学者们认识到元分析中应该既有定量分析也有定性分析。因此,他们不再简单地将元分析看作一种统计分析,而是将其看作一种通过综合多个同类研究的结果,对研究效应进行定量合并的分析研究过程和系统方法。因此,简单地说,元分析就是一种应用特定的设计和统计学方法对以往的研究结果进行整体的和系统的定性与定量分析的方法。

- Stanley 在讨论了定量文献综述的研究方法后,认为元分析因其在总结、评价和分析实证研究方面具有的明显的科学性优势和广泛的影响而成为这方面最重要的统计学方法。Stanley 的这一观点已经为不同学科的研究者所广泛认同。同样,管理学研究也存在定量文献综述的问题,因此,运用元分析进行定量文献综述,成为管理学研究的必然趋势。

综合上述学者的观点,元分析是以综合已有的发现为目的,对单个研究结果进行综合的统计学分析方法,也称为“典型或定量元分析”。换句话说,元分析是一种数据分析方法,它通过收集、整理以往专家学者的实证研究,并运用一些测量和统计分析技术,对关于同一个问题的多项独立的定量研究结果进行再分析。通过系统地提炼和整合来自许多研究的数据,找出研究问题或者所关注变量之间的关系,进而就给定的研究问题得出更具普适性、更可靠的结论,从而弥补传统文献综述的不足。因此,元分析也被视作跨研究评判结果的一种有力工具。元分析已经有半个多世纪的历史,最初主要应用于医学领域,现在越来越广泛地应用于社会科学领域的各个学科中。

从数据统计方法的角度而言,元分析是对众多现有实证文献的再次统计,利用相应的统计公式对相关文献中的统计指标进行再一次的统计分析,从而可以根据获得的统计显著性等来分析两个变量间真实的相关关系。元分析要求对每个观察到的相关系数进行研究样本大小的权重处理,从而产生经过权重处理后总体相关性的评价估计值。这个估计值的误差包括总体样本的真实误差、样本误差,以及测量误差。因此为了获得精确的总体相关性及其误差,需要对样

本误差和测量误差等进行修正，找出“调节变量（moderating variable 或 moderator）”分组研究。另外，元分析对使用的数据进行了一定的限制要求。如“一个变量在不同的研究中有多种衡量指标”出现时，首先得将这种“异质性”进行处理以保证数据来源及统计方式的一致性。

1.1.2 元分析的分类

元分析的分类主要基于其研究方法和研究对象的不同，可以大致分为以下几种类型：定量元分析和质性元分析。定量元分析是对实证文章进行总结和分析的方法，而质性元分析则是对于质性研究进行证据整合的方法。

定量元分析是一种统计方法，它通过对多个独立研究的结果进行综合分析，以评估某一效应或关系的总体效应量。这种方法可以提高统计功效，增加研究结论的普适性和可信度。定量元分析广泛应用于医学、心理学、教育学、社会科学等多个领域，它为研究者提供了一种强有力的工具，以整合现有证据，指导实践和制定政策。然而，这种方法也有其局限性，如依赖于研究质量的一致性、可能的发表偏差以及统计模型的选择等。

质性元分析，又叫作定性或质化元分析、元综合等，作为一种新兴的质性研究整合方法，由 Stern 等在 1985 年首次提出，是一种兼具探索、归纳和整合性的研究工具，主要是将以往零散的单案例或多案例研究、扎根理论研究、民族志研究等质性研究的结论进行整合，以寻求对某一领域进行系统性回顾和全新解释，并构建普适性的理论框架。质性元分析最早应用于医疗保健和护理方面的研究，初衷是顺应“循证医学”的研究热潮。该方法逐渐被社会、教育、心理等研究领域所借鉴，成为对零散的质性研究进行元分析的有效工具。2013 年，Hoon 借鉴了教育学质性元分析、循证管理以及管理案例研究等领域开创学者的思想，首次系统地将质性元分析方法应用到管理案例研究的整合中，并给出了详细的使用步骤和操作指南。自此，质性元分析方法在组织管理领域逐渐应用开来。2014 年，Rauch 等认为 Hoon 提出的质性元分析难以真正实现管

理研究结论的普遍性，因此提出了另一种范式的质性元分析，该范式通过增大研究数量来有力预测变量之间的因果效应，但由于其发表案例中的公开数据信息要求较高，因此这种范式的质性元分析未被管理学者们广泛使用。

无论是定量元分析还是质性元分析，它们都提供了一种系统化的方法来综合现有研究，通过统计合成或主题综合来评估和解释研究结果。与文献综述相比，元分析依赖于严格的纳入标准和透明的方法论，旨在提供更客观、可量化的结论，而文献综述则更侧重于叙述性分析和理论发展，可能包含主观判断，如表 1-1 所示。

表 1-1 定量元分析、质性元分析和文献综述的区别

维 度	定量元分析	质性元分析	文 献 综 述
方法属性 (What)	定量方法，通过公开发表的定量研究准确预测因果效应	质性方法，通过公开发表的质性研究更科学地解释现象	质性方法，主观评估公开发表的研究领域的现状
方法作用 (Why)	提供变量之间确定性的因果关系，验证和发展理论	逼近现实地解释现象，并建构出理论	系统呈现研究现状，并提出未来展望，具有一定理论贡献
使用时机 (When)	定量研究领域的成果较为丰富，但存在较多矛盾结论	质性研究领域的成果较为丰富，但各个结论之间缺少连接和对比	研究领域的成果较为丰富，并且研究方向正在发生转变
研究对象 (Where)	定量研究，适用大样本	质性研究，包括访谈、民族志、案例研究等，适用于小样本	定量、质性或混合研究
实施流程 (How)	通过系统的文献检索程序，归类相似含义的概念，整合效应值验证假设	通过系统的文献检索程序，归类相似含义的概念，涉及重新概念化	根据研究需要检索文献，归类相似含义的概念，并以一定的框架总结原有结论，追求逻辑合理化
可复制性 (Whether)	可复制性强	可复制性较强	可复制性弱

1.1.3 元分析中的常见概念

在进行元分析的学术探究之前，对一系列专业术语的精确理解至关重要。

这些术语构成了元分析研究方法论的基础，并为后续的深入分析提供了必要的理论支撑。元分析主要涉及以下 5 个关键概念。

1. 效应量（effect size）

效应量是一个用来反映效果重要性的值或者说是反映两个变量之间关系的强度值，它是元分析中的项目单位。效应量大致可分为三组：相关系数（ r ）组、标准平均数（ d ）组和概率比/风险比（risk ratio）（OR）组，详见表 1-2。

表 1-2 效应量分类

r	d	OR
相关系数	标准平均数	概率比/风险比
用于回归分析	用于实验法	用于二分法的数据
例子：工作满意度与工作绩效之间的相关性系数	例子：一个人的业绩表现在事件发生前后的差异	例子：使用药物和不使用药物组生存或者死亡的概率

选择这些效应量指标的原因在于它们的通用性和适用性，而其他指标存在一定的局限性。例如：① t 检验和 z 分数与样本量有关，这可能导致在小样本研究中得出的结论被过度放大；②卡方值不提供效应大小的直接度量；③ B 值与其他的自变量有关，可能受到多重共线性的影响。

2. 抽样误差（sampling error）

抽样误差是抽样统计量与相应的总体参数之间的差值。它是一种随机误差，但是在元分析中，它可以被预测。抽样误差的决定因素有两个，分别是样本量大小与总体估计。当样本量无限大时，抽样误差为零，其抽样估计实际可被看作是总体参数。所以数据量大的样本所占权重应更高，因为其更接近真相。真相越显著，抽样误差也会越小。比如，打喷嚏时头会移动这种现象，这只需要调查几个人就能得到结果。但是除了以上原因外，还有一些人为因素可能会影响误差结果，如测量误差、范围限制、不完善的结构效度、转录误差等。

3. 固定效应模型 (fixed effects model) 和随机效应模型 (random effects model)

固定效应模型假设所有效应大小之间的变化都是由于抽样误差造成的, 换句话说, 效果大小的不稳定性仅仅是由于主题级别的“噪声”造成的。随机效应模型假设效应大小的变化来源于抽样误差和效应在总体中的实际变异。换句话说, 效应大小的不稳定性既可能由研究对象的随机“噪声”引起, 也可能源于不同研究之间真实且未被测量的差异 (即每项研究可能在估计略有不同的总体效应大小)。

通俗地说, 当真相只有一个时, 比如一个人的身高具体是多少, 就用固定效应模型; 反之则使用随机效应模型。实际上, 在管理学的研究中较多使用的是随机效应模型。

4. 发表偏差 (publication bias)

发表偏差是指在学术研究中, 具有统计学显著性研究结果的研究更有可能被报告和发表, 而无显著性意义或无效的结果则不太可能被报告和发表。这种偏差可能源于研究者、审稿人或编辑在选择论文发表时对研究结果的方向和强度的选择, 从而使出版的过程不再是随机事件, 进而压制了某些研究的发表。常用于识别发表偏差的方法有漏斗图 (funnel plot) (经典漏斗图和附加轮廓线漏斗图) 法、失安全系数法、剪补法、森林图法及累积森林图法。

5. 异质性 (heterogeneity)

在元分析中, 真实效应值的变化程度称之为异质性。对于实证研究论文, 通常在讨论因果关系时由于回归样本的局限, 即研究得出的因果关系可能仅是使用样本得出的结论, 要尽力说明样本的一般性和代表性, 使得样本更好地符合总体情况从而得出一般化的结论。现实情况是这种一般化的结论并不一定是稳定的, 可能由于样本本身存在的一些特质或者是受外部的一些客观因素的干扰, 从而产生与主要结论不一样的因果关系, 这便是进行异质性讨论的原因,

即一般的因果关系中可能包含了异质性结果，因此需要将这些异质性找出来讨论，使得研究结论更加符合现实情况。

异质性的检验是元分析中的一个关键步骤，常用的检验方法包括 Q 统计量检验和 I^2 统计量检验，以此来定量评估异质性的。 Q 统计量的计算如下：

$$Q = \sum_{i=1}^k W_i (Y_i - M)^2$$

式中， W_i 为第 i 个研究的权重值； Y_i 为第 i 个纳入研究的效应量； M 为纳入研究的平均效应量； k 为相应的研究数。也就是说， Q 实际上是效应量的加权离均差平方和，叫作加权平方和（WSS）。

Higgins 等提出的 I^2 统计量的计算如下：

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

式中， df 为自由度； I^2 随 Q 的变化而变化。如果发现显著的异质性，研究者可能需要进行亚组分析或元回归（meta-regression）来探究异质性的原因，或者选择不同的元分析模型（如固定效应或随机效应模型）来处理研究间变异（between-study variance）。

如上所述，元分析作为一种系统性的研究方法，其有效性建立在对一系列专业术语的精确理解上。这些术语不仅构成了元分析方法论的基础，而且为深入分析提供了理论支撑。元分析的关键概念包括效应量、抽样误差、固定与随机效应模型、发表偏差和异质性。效应量作为元分析的基本单位，通过不同的统计指标反映研究效果的重要性。抽样误差是随机的，但在元分析中可预测，并影响数据的权重分配。固定效应模型适用于假设研究间无本质差异的情况，而随机效应模型则考虑了研究间的变异性。发表偏差是指显著性结果更可能被发表，可能导致研究结论的偏差。异质性则涉及对研究结果变异的评估，对泛化结论的稳定性至关重要。

1.1.4 元分析的注意事项

1. 界定清晰的概念

与其他的研究方法类似，在元分析中，研究者需要梳理研究问题中所涉及

的概念并进行清晰的界定。任何社会科学中涉及的变量都以两种方式定义，即概念性定义和操作性定义。具体而言，研究者在元分析中应该注意以下两点。

第一，界定清楚概念性定义及其广度。概念性定义描述了变量的质量，这些质量独立于时间和空间，但可以用来区分与概念相关或无关的可观测的事件，如“成绩”一词的概念性定义为“一个人在学术领域的知识水平”。概念性定义的广度不同，它所涉及事件的数量也不同，如果“成绩”被定义为“通过努力或费尽心力获得的东西”，那么这个概念就比仅与学术有关的定义宽泛得多。不同的概念性定义和广度可能会得出不同的结果，因此，研究者在元分析中需要定义清楚研究问题中所涉及的不同概念及其广度。

第二，确保操作性定义准确地反映了概念性定义。清晰概念的最终参照点必须依托于可直接观测的事实和清晰的属性，尽管管理研究中相当多的概念不能依托于直接观测而是靠间接的推测性观测，但不管怎样的推测性观测，最终要归结到某些清晰的测度指标。例如，“企业创新能力”的概念总是要落实到“研发经费”“专利数”或“高级科技人员比例”等这样一些可测度的指标上来。在元分析中，为了将概念与可观测的事件联系起来，需要定义操作性变量并确保其反映了概念性定义，以契合研究者想要研究的问题。

2. 界定研究假设和研究模型

在梳理和界定了研究的相关概念后，元分析研究者需要界定研究假设和研究模型，从而将研究问题转化为可以分析量化的关系。

假设的提出需要考虑一些要求，如假设以现有理论为基础，假设必须以一定的客观事实为依据；假设需要明确表述变量之间的关系，包括清晰地界定自变量、因变量，明确因变量与自变量之间的关系；假设必须是可以验证的。一个“好”的假设应该具备以下三个条件：第一，能够表现理论或命题中预期的变量关系；第二，不应过于空泛，它必须可以进行验证；第三，尽可能简单扼要。在元分析研究提出的假设中，还需要考虑变量之间的关系是线性的还是曲线的？潜在的中介变量（mediator）、调节变量是什么？并以此构建完整的研究模型。

1.2 元分析的特点和优势

1.2.1 客观性

元分析作为一种系统性的研究方法，通过定量的统计手段，有效克服了传统定性文献综述中存在的主观性和偏见问题。它通过制定严格的文献纳入标准和系统化的操作流程，系统性地收集、整合和分析现有研究的数据，以客观的统计方法合并和分析多项独立实证研究的结果，从而避免了以往定性文献综述的主观性和研究者的个人偏见，研究结论更加客观与可靠。

元分析的客观性特点，使其在多个学科领域得到了广泛应用，包括医学、心理学、教育学、管理学等。它不仅为研究者提供了一种整合现有证据的有效工具，也为政策制定者和实践者提供了基于实证的决策支持。

1.2.2 高效性

元分析将已有的实证研究进行统计整合，不用再进行大样本的实证检验，节约了成本，具有高效性的优势。这种方法允许研究者在现有研究的基础上，进一步探索和验证假设，提高了研究的效率和经济性。元分析的高效性体现在几个方面：第一，它能够利用已有的研究成果，减少了对原始数据收集的需求。第二，元分析所用的主要是合并研究结果和对研究结果进行齐性检验的统计方法，这提高了统计功效，使得研究者能够检测到较小的效应量，这在单个研究中可能难以实现。第三，元分析还能够揭示研究结果的一致性和异质性，为研究领域提供更全面的视角。

元分析的高效性并不意味着它是一个简单的过程。研究者需要进行系统性的文献检索、严格纳入和排除的标准制定、数据质量评估以及复杂的统计分析。

1.2.3 精确性

通常情况下，单一实证研究的样本有限，很难代表整个总体，因此可能导致结论的可靠性不高。虽然增加研究的样本量可以提高可复制性，但它依然难以提供高度准确的结论，而且需要更多的研究经费支持。相比之下，元分析方法通过整合已有的实证研究数据，可以有效地提高统计结果的精确性，而不需要额外的昂贵资源支出。

元分析通过对大量研究数据的整合和分析以及规范的统计方法的应用，克服了小样本的局限性，精确性大大提高。与单个实证研究相比，基于大样本的元分析具有更高的统计效力，这使得研究者能够更准确地描述变量之间的关系。

综上所述，元分析作为一种高效的研究方法，通过整合现有研究数据，不仅节约了研究成本，而且提高了研究结论的可靠性和精确性。这种方法在学术研究中的重要性日益凸显，为研究者提供了一种强有力的工具，以便他们更全面、更深入地理解研究主题。

1.3 基本流程

元分析是一种系统性地整合和分析多项独立研究结果的统计方法。元分析的基本流程如图 1-1 所示。

1. 选题

什么样的题材适合作元分析？选题是元分析重要环节之一，从一定程度上来说选题直接决定了该元分析的价值。在该步骤需要自问：“我想研究的概念或干预手段是什么？”“怎样测量相关概念和结果变量？”通过回答这些问题，确定研究证据是否与自己感兴趣的问题或假设相关。同样，在问题形成过程中，需要决定感兴趣的是对变量进行简述，还是研究两个或多个变量之间的关系，以及此变量之间的关系在本质上是相关的还是存在因果关系的。总之，一个好的选题应立足于专业的实际情况并具有一定的指导价值。

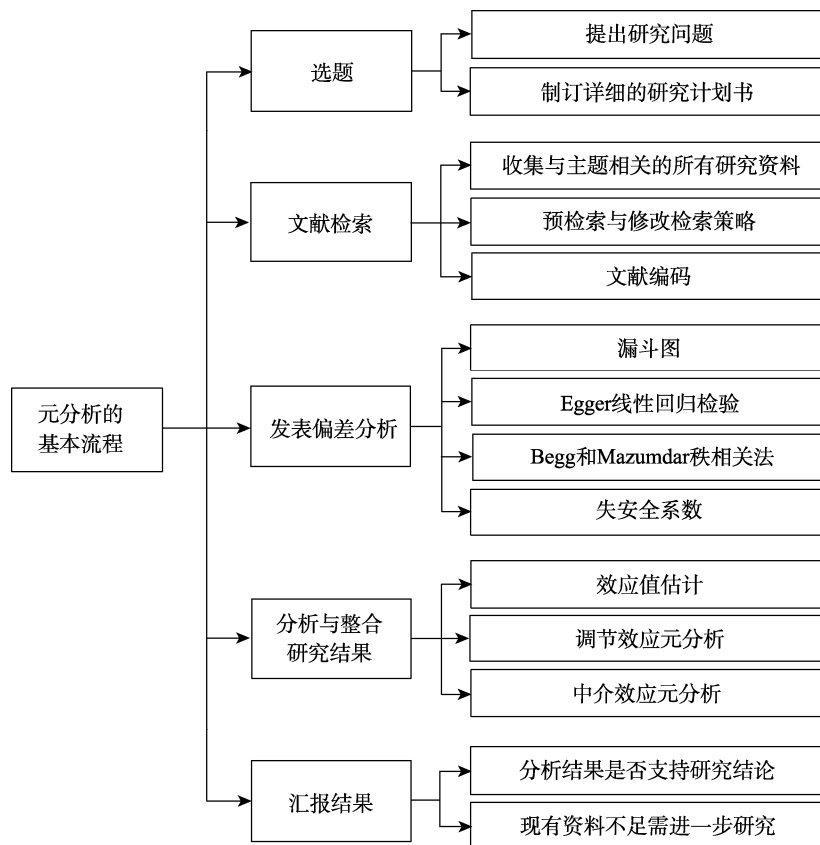


图 1-1 元分析的基本流程

第2章将介绍在选题阶段将会遇到的决策点。其中涉及的就是相关概念的定义范围以及所采用的操作测量方式，还涉及一手研究的设计类型及其如何与所期望的研究推论相联系。

2. 文献检索

由于元分析是对某一主题已有的研究进行综合分析，所以尽可能全面、系统地收集相关文献是进行文献检索的基本原则。文献检索阶段包括对研究目标元素总体的选择。在一手社会科学研究中，研究目标通常包括个体或群体。在研究综述中，由于要对两个目标进行推论，确定目标群体则相对复杂。首先，累积的结果要能够反映之前关于该研究主题的所有研究结果。其次，所收集的

研究文献能一般化到该研究主题领域的个体或群体。

之后,需要对文献进行编码。研究编码阶段要求研究者考虑他们想从每个研究中收集什么信息,数据收集工具包括问卷调查、行为观察和(或)生理测量。在研究综合中,数据收集工具则涉及所确定的、与研究问题相关的信息。这些信息不仅包括与理论或实际问题相关的研究特征,即自变量和因变量的性质,还包括研究进行的方式,研究的设计、实施和研究的统计结果。在这个阶段,研究者除了要明确收集何种信息并给出该信息清晰的定义外,还需要开发一个培训信息收集人员的标准流程,确保他们能以可靠、易解释的方式收集信息。

第3章将探讨文献的定位检索策略与方法、文献检索步骤,以及社会科学家可获得的文献资料源清单,讨论如何获取和使用关键的信息源,以及每个资料源中包含的信息可能存在哪些偏差。同时,第3章还将提供一些信息收集的具体建议,帮助研究者从与研究主题相关的实证研究中收集必要的信息,并介绍所需采取的步骤以更好地培训研究编码人员。

3. 发表偏差分析

发表偏差是元分析最常见的系统误差。因为阳性结果比阴性结果更容易发表,所以形成了为数不少的“抽屉文件”,此时再根据发表文献所做的综合分析就有可能歪曲真实效应。因此,在收集信息后,研究者需要对所收集数据的“质量”或其与研究问题的相关性做出关键的判断。而研究综合者需要对被整合的文章的研究方法进行讨论,确定其收集数据的方式是否适用于解决手头的研究问题。

第2章将讨论研究设计(例如,关联或因果)如何对应研究中出现的不同问题,第4章则将讨论如何分析存在的发表偏差,并就发表偏差校正与预防提出一些方法。

4. 分析与整合研究成果

在数据分析过程中,研究者将收集到的各个独立数据点进行汇总,整合成一个整体。数据分析要求研究者将系统的数据模式与“噪声”(或随机波动)

区分开来。在一手研究和研究综合中，这一过程通常与统计程序的应用有关，其中可能会涉及效应值的估计、调节效应与中介效应等统计问题。

第 5~7 章将阐释一些整合独立研究结果的方法及元分析方法。此外，还将展示调节效应与中介效应的操作方法。

5. 汇报结果

研究者需要解释累积的证据，并确定从数据中得出的结论。这些结论包括相关关系是否得到数据的支持，以及这种关系的确定性程度。此外，结论还涉及研究发现对不同类型的样本、处理方法、结果和情境的普适性或特殊性。最后，撰写一篇描述这些结果的论文是研究工作的最后任务。

第 8 章将列举一些对研究综合做出断定时应遵循的决策规则；而且，还将提供关于如何进行研究综合的其他 6 个阶段的说明以及报告所需信息的具体准则。