

# 元分析研究法

姜铠丰

胡佳

俄亥俄州立大学

## 本章大纲

- 一、引言
- 二、元分析的作用
- 三、开展元分析的步骤
- 四、元分析的基本分析模型与方法
- 五、元分析中调节变量的检验
- 六、元分析与结构方程模型的结合
- 七、元分析中的其他技术细节
- 八、常见的元分析工具的介绍
- 九、结语

## 一、引言

元分析在英文中对应的称呼是 meta-analysis。在中文里，曾经有学者将其翻译为荟萃分析，聚合分析，综合分析等。但随着这一统计技术在中国研究人员中的普及，人们基本上把“元分析”认定为 meta-analysis 的中文译名（毛良斌，郑全全，2005；张翼，樊耘，赵菁，2009；魏江，赵立龙，冯军政，2012）。元分析在很多研究领域都有着广泛的应用，比如说，医学，教育学，心理学等。在组织和管理科学的研究中，元分析近年来也得到了越来越广泛的应用。Aguinis 等人（2011）曾经对发表在五本最优秀的英文管理期刊中的研究进行了统计，发现从上个世纪 80 年代以来，这五本期刊一共发表了近 200 篇元分析，其中涉及到对 5500 多个关系进行的总结归纳。我们对中文管理期刊中的元分析也进行了简单的搜索，发现从 2008 到 2017，研究人员已经发表了 110 多篇元分析，并且文章数量有逐年递增的趋势。由此可见，元分析已经成为了对组织和管理科学进行研究的重要方法，并且受到了国内外学术期刊的普遍重视。

为了帮助更多的中国学者了解和使用这一方法，我们将在本章节中对元分析的基本原理和实际应用进行介绍。我们本身并不是对元分析的理论和技术进行研究的学者，但是曾经有过一些在英文期刊发表和审阅元分析研究的经验。鉴于我们的背景，在本章节的撰写中，我们侧重对于元分析的流程和实际操作的介绍。这一思路跟 Rosenthal（1991）的 *Meta-Analytic Procedures for Social Research* 和 Lipsey 和 Wilson（2001）的 *Practical Meta-Analysis* 一书有些类似，更有助于初学者对这一方法有深入浅出的了解。同时，我们也把自己和同行在发表元分析研究的过程中遇到的一些技术细节拿出来跟大家分享，希望对一些已经有经验的同行也有所帮助。另外，我们想强调的是，本章节中介绍的很多内容（比如公式等）并非我们原创，我们会尽最大可能对重要内容的出处给出标注，但受章节篇幅和个人精力所限，遗漏之处请原作者和读者理解。

## 二、元分析的背景和作用

元分析是一种对以往的实证研究结果进行归纳和总结的统计方法。元分析首先在医学领域得到应用，然后在心理学和教育学等领域得到进一步发展。1904 年 Karl Pearson 第一次采用元分析的思路收集了发表在医学期刊上的关于伤寒接种问题多篇文章进行了汇总。Pratt 和 Rhine 等人在 1940 年第一次采用元分析的方法对 145 篇关于超感知觉（Extrasensory Perception）的实验研究进行了归纳总结。Gene V. Glass 在 20 世纪 70 年代最早明确提出了元分析的概念，并对元分析的基本方法做出了奠基性的贡献，之后这一方法在多个领域得到了广泛应用。在组织和管理科学中，Larry V. Hedges, John E. Hunter, Jacob Cohen, Robert Rosenthal, 和 Frank L. Schmidt 等人对这一统计方法的推动起到了重要作用。

在对这一方法进行具体介绍之前，我们首先想谈一下元分析这种方法对我们的研究有什么样的帮助。Hunter 和 Schmidt（2004）在他们经典的元分析的书中用工作满意度和

组织承诺之间的关系作为例子对这一问题进行了讨论。假如说，我们想通过以往的实证研究结果来了解工作满意度到底跟组织承诺有什么样的关系。通过文献搜索，我们找到了30篇对这一关系进行研究的文章，这些研究有的是基于比较大的样本，有的是基于比较小的样本，有的样本来自于大公司，有的则来自小公司，有的样本中年长的人较多，有的则年轻的人较多，有的样本中女性比例较大，有的则男性比例较大，等等。通过仔细观察，我们可能发现，尽管这30篇研究都是对同一关系进行的研究，但是他们得到的结论并不一致，甚至有的差距很大。比如有的研究发现，这两个变量的线性关系是0.20，在统计上是显著的，而有的研究则发现二者的线性关系是0.40，但是在统计上是不显著的。还有的研究发现，这两者的线性关系是负数-0.10，在统计上是不显著的。针对这些并不完全一致的结果，我们很难明确地知道究竟工作满意度和组织承诺之间的关系是多大，在统计上是不是显著的。如果没有元分析的话，我们在进行文献综述的时候很可能会得出一个模棱两可的结论。而元分析这一方法能够通过统计技术对以往基于不同特点的样本得出的结论进行定量分析，从而对两个变量之间的真实关系有着更准确的估计。

除了对两个变量的相关关系有更准确的估计之外，元分析还能告诉我们这二者之间的关系在不同的研究中存在怎样的差异。由于样本选择的随机性，来自不同样本的研究结果不可避免的存在一定程度上的差异。元分析可以帮助我们了解两个变量的相关关系在不同研究中的基本分布情况，同时还能够一些统计量来告诉我们这些差异在多大程度上受样本误差和其他一些常见的误差（比如测量误差）的影响，是否还受到这些误差以外的其他因素的影响。在此基础之上，元分析能够进一步地帮助我们判定两个变量之间的相关关系是否会受到某些变量的调节。比如说，Meyer等人（2002）汇报了组织承诺和满意度之间相关关系的元分析结果，他们发现情感承诺和薪酬满意度之间的关系要显著地低于情感承诺和工作满意度之间的关系，在这里满意度的类型就成为了一个能够对组织承诺—满意度关系在不同研究中的差异进行解释的调节变量。

此外，一般的实证研究通常只针对某一变量和少数几个其他变量之间的关系进行研究，因此我们不太可能从一个单一的实证研究中获知这一变量跟很多其他变量之间的整体关系是什么样的。但是，我们可以借助元分析同时对多个变量之间的相关关系进行估计，并将相关分析的结果跟其他统计方法（比如结构方程模型）结合来对变量之间的整体关系或者中介过程进行研究。举例来说，Colquitt等人（2000）对培训动机的前因和后果变量进行了研究，通过将元分析和结构方程模型的结合，他们检验了一个包含16个变量的中介模型。而这对于一个一般的实证研究来说是很难实现的。

综上所述，我们把元分析的基本作用总结为三点：

- 第一，针对某一具体的相关关系，对来自不同研究样本的结果进行整合，从而得出对这一关系更为接近样本总体的估计。
- 第二，对某一相关关系在不同研究样本间的差异进行分析，进而找出能对这些差异进行解释的调节变量。

第三，将元分析结果和其他统计方法结合，对某一变量与其他多个变量直接的整体关系进行分析。

尽管不同的学者对于元分析的用途有着不同的看法，但是上述这三点是我们自己的研究中以及跟同行的交流中体会到的最重要的三点。这三点往往也成为研究者着手对某一研究开展元分析的出发点。接下来，我们将对这三点的具体操作进行详细介绍。

### 三、开展元分析的步骤

在跟一些发表过元分析的同行的交流中我们发现，很多人在一开始都会觉得掌握元分析的统计技术是最重要，但是却忽略在数据分析之前的步骤。经过不断摸索之后才逐渐明白，数据分析之前的准备工作实际上在很大程度上决定一项元分析的理论贡献和质量。因此在这一部分，我们简要介绍一下开展一项元分析的基本步骤。总体来说，我们将元分析的基本步骤分为四步：选题，文献搜索，数据编录，和数据分析。

#### 1. 选题

我们对于元分析选题的一个基本建议是：不要为了做元分析而开展一项元分析的研究。元分析只是众多研究方法中的一种，只有当这种方法比其他的研究方法能够更好地回答我们的研究问题的时候，才考虑采用这种方法。换言之，在一篇元分析的文章里，我们不要过分强调元分析作为一种研究方法的优点，而要强调研究题目和元分析的特点之间的契合。举例来说，在我们关于高绩效工作系统的元分析中（Jiang, Lepak, Hu, & Baer, 2012），我们检验了一个关于人力资源管理系统的三个维度如何影响公司绩效的中介模型。在最初的投稿中，我们只是检验了一个基本的从人力资源管理系统到员工结果再到公司绩效的中介模型，我们的出发点是，这一模型还没有被以前的元分析研究过，所以这项元分析是有价值的。然而，这一研究模型是战略人力资源管理的一个基本模型，很多基于问卷和一手数据的研究已经从某些方面检验了这个模型，所以即使我们的元分析结果可能比单一研究的结果更接近样本总体的情况，但从理论上并没有太多新意，整个研究是一种“为赋新词强说愁”的感觉。在接下来的修改中，我们重新从理论出发去考虑有什么问题是以前的理论模型没有完全解释同时很难被单一的实证研究解决的。在评审的提示下，我们开始对人力资源管理系统的成分进行细分，并将与员工相关的结果进行了细分，在此基础上提出了不同的管理措施对不同的员工结果有不一样的影响。经过这样的修改，元分析成为了解答这一问题的最好的选择，它不仅仅让研究在实证上有优势，同时在理论上也有了创新。

针对上面提到的元分析的用途，元分析的选题一般有三个主要的角度。首先，如果能够找到一个研究变量从来没有发表过任何元分析的研究，那么可以考虑一下是否有必要用元分析的方法对现有的研究结果进行一下总结。在2000年以前英文期刊中发表过的元分析，有很多都是采用这种角度来提出研究问题的。然而，随着元分析研究方法的普及，在组织和管理科学里已经越来越难找到一个还没有过元分析的研究变量了。即使有这样的

变量，简单地总结归纳这个变量与其他变量之间的相关关系也很难有重要的理论贡献。第二，我们可以考虑一下以往关于某个关系的研究是否存在不一致甚至是相互矛盾的结论，然后从理论出发去探讨是否有存在一些调节变量能够解释为什么这一关系在不同研究间存在差异。如果能够找到有意义的调节变量，通常能够对这个关系的理解有理论上的贡献。举例来说，曾经有三个不同的研究团队同时对离职率和组织绩效的关系进行了元分析

(Hancock, Allen, Bosco, McDaniel, & Pierce, 2013; Heavey, Holwerda, & Hausknecht, 2013; Park & Shaw, 2013)，在各自的研究中，他们都找到了一些能够对这一关系的变异进行解释的调节变量，比如离职的类型，绩效的类型，行业的类型，组织的大小，样本所在的区域等等。第三，我们可以考虑通过元分析来构建一个结构方程模型来检验与某一变量相关的中介机制。举例来说，Liu 等人 (Liu, Jiang, Shalley, Keem, & Zhou, 2016) 对个体创造力的作用机制进行了元分析发现个人特征和情境因素会通过不同的作用机制对个体创造力产生影响，同时他们也发现了一个在以往文献中并未强调过的新的中介变量。近期发表的元分析的文章多从后两个角度来入手，有些文章还同时采用这两种角度来选取研究题目。

## 2. 文献搜索

在确定研究选题之后，我们可以开始选取一些关键词来搜索相关的文献。因为元分析旨在对所有与研究选题相关的实证研究进行归纳总结，所以我们要在这个步骤要尽可能地搜索以往发表过的和以及已完成但是还未发表的文章。搜索已发表的文章可以通过以下途径来完成：1) 通过搜索引擎或文献数据库进行搜索；2) 通过主要期刊进行手动搜索；3) 参考以前发表过的关于该研究选题的综述文章的文献列表。搜索已完成但还未发表的文章可以通过以下途径来完成：1) 搜索近期的会议文章；2) 搜索硕博论文数据库；3) 询问经常在该领域发表文章的学者是否有还未发表的研究；4) 通过专业领域的邮件群发列表来征求未发表的研究。在这里我们想强调的是，搜索未发表的研究对于一项元分析来说是很重要的。如果在一项元分析中我们只包含了发表过的研究，那么元分析的结果可能不会真实地反映样本总体的情况。这是因为有的学者发现 (比如 Dickersin, Min, & Meinert, 1992; Hedges, 1984)，显著的结果比不显著的结果更容易获得发表的机会。因此，如果一项元分析没有包含这些潜在的未发表的研究的话，会有可能受到所谓的发表偏差 (publication bias) 的影响 (Banks, Kepes, & McDaniel, 2012)。

通过初步的文献搜索，我们通常会得到很多不相关或者无法直接用于元分析的文章。因此，我们需要设定一些录用标准对初步搜索的结果进行筛选。一些基本的标准是：1) 必须是实证研究；2) 必须汇报了样本大小；3) 必须汇报了相关系数或者均值差 (针对实验研究而言)，或者其他能够转换成相关系数或均值差的统计量 (比如  $t$  值,  $z$  值,  $F$  值, 和卡方值；有些元分析还会采用回归系数或者偏相关系数作为汇总的对象)。这些基本标准决定了被筛选进来的文章能够提供元分析所必需的数据。另外，研究者也可以根据选题设定一些其他的标准，比如有的研究只关注组织层面的关系而排除个体层面或者跨层次的研究，有的研究只研究工作场所的关系而排除来自学生样本的研究，还有的研究为避

免共同方法偏差对结果的影响而排除有数据同源问题的研究，等等。另外，如果我们能够明显判断出，两个或多篇文章使用了同一个数据库，那么我们只考虑那个提供了最多数据信息的研究。而如果一篇文章包含了多个独立的样本，我们则把它们当作独立的研究来对待。

### 3. 数据编录

Lipsey 和 Wilson (2001) 对如何对数据进行编码和录入有着详细的介绍。不同的学者可能会采用不同的录入方式，但包含的基本信息可以归结为以下几类：1) 研究的基本信息（比如作者姓名，年份，期刊名称等）；2) 录入者的编码（为了保证数据录入的准确性，通常建议有至少两名录入者）；3) 基本数据信息（比如样本大小，相关系数，量表的信度等）；4) 可能的调节变量（比如国家，行业，样本类型，变量类型，研究设计等）。有的元分析还可能设计到对一些含义相似的概念进行整合，比如在前面提到的人力资源管理的元分析 (Jiang et al., 2012) 就将不同的管理措施分成三个大的维度，在录入的时候就需要研究者对这些类别进行判定和标记。

在确定了初步的录入模板之后，研究者可以随机选取 5-10 篇文章进行录入，发现在制定模板的时候没有考虑到的一些问题。比如在录入的时候有可能会发现新的研究变量，就需要对模板进行扩展。再比如有些录入的情况（例如变量合并）需要依据录入者的主观评价，就需要明确录入标准是明确而统一的。对录入模板进一步修订之后，可以由两名或以上的录入者独立完成剩余的数据录入，并检查录入结果的一致性。在实际操作中，我们经常发现，对于客观数据（比如相关系数，样本大小等）的录入一般不太容易出现录入偏差。然而，有些需要录入者主观判断的内容需要额外注意。比如，有些时候需要录入者判定是否研究样本来自于某一特定类型的组织（比如，非营利性组织或营利性组织），样本中被调查对象的工作性质（比如，复杂的工作或简单的工作）等。这一类信息需要在录入之前有详细的说明，以便录入者在登录信息是有所参考。一旦在录入中出现了之前说明中没有包含的情况，要及时对录入说明进行调整，并根据新的标准来完成接下来的录入工作。其最终的目的是为了确保所有录入的信息是准确无误的。为了方便他人验证分析的结果，研究者们通常要整理好录入的数据并作为研究的附录。

### 4. 数据分析

在完成了基本信息的收集之后，我们就可以开始对数据进行分析。我们把具体分析的细节放在接下来的章节介绍。在这里，我们先强调和区分一下元分析的主要分析对象。元分析主要是通过通过对来自各个实证研究的 effect size 的分析来对以往的实证研究进行汇总。Effect size 这个词在中文里曾经被翻译成“效果量”（魏江等，2012）或者“效应值”（王永贵和张言彩，2012），其基本含义是用一个数据统计量来表示某一种现象的强度。为了方便起见，我们在接下来的文中用效应值来代表 effect size。在组织和管理科学的元分析中，比较常见的效应值有两种：第一种是针对两个连续变量之间的相关系数，也就是

我们通常所说的  $r$ ，第二种是实验研究中用来比较两组实验对象的均值差距的  $d$  值。 $r$  值和  $d$  值和其他常见的假设检验统计量（比如  $t$  值， $z$  值， $F$  值，和卡方值）之间存在相互转换的关系。因此，如果一篇实证研究没有直接汇报  $r$  值和  $d$  值的话，我们可以通过转换来得到这些信息。*Practical Meta-Analysis* 一书的作者之一 David B. Wilson 在其网站上 (<http://cebcp.org/practical-meta-analysis-effect-size-calculator/>) 提供了一系列的可以对不同效应值进行转换的工具，在这里我们推荐给大家作为参考。当然，元分析可以处理的效应值还包括其他类型，比如回归方程系数等（王永贵和张言彩，2012），在这里我们就不做详细介绍。

#### 四、元分析的基本分析模型与方法

元分析的主要目的是通过对不同实证研究的效应值进行整合来得出对样本总体的估计。简单来说，对样本总体的估计是对各个实证研究的效应值进行加权平均（weighted mean）。之所以采取加权平均数的形式是因为统计学家们通常认为大样本比小样本更能够准确的反映样本总体的情况，所以来自大样本的效应值在元分析中应该被赋予更多的权重。虽然这个基本原理不难理解，但是在元分析的发展过程中，有很多学者提出了不同的模型和方法来实现加权平均的想法（比如 Hedges & Vevea, 1998; Hunter & Schmidt, 2004; Rosenthal & DiMatteo, 2001 等）。在这里，我们无意于讨论各种方法的利弊，而打算具体介绍一下在组织和管理科学中经常被用到的假设和方法。

##### 1. 元分析的两种估计模型

元分析对于样本总体的效应值估计有两种不同的估计模型：固定效应模型（fixed-effect model）和随机效应模型（random-effect model）。固定效应模型假设元分析中包含的研究都来自同一个样本总体，由于这个总体的平均效应值是固定的，那么来自这个总体的不同研究所得出的效应值在理论上是同质（homogeneous）的（Hunter & Schmidt, 2000）。与此相反，随机效应模型假定元分析中的研究反映不同的样本总体，这些不同的样本总体可以被看成是一个更大的总体的样本，由于这些样本总体的平均效应值是不同的，那么不同实证研究的效应值在理论上就是异质（heterogeneous）的（Hedges, 1992）。这两种估计模型的主要区别在于它们对于平均效应值误差来源的看法。固定效应模型认为误差主要是由来自同一总体的抽样误差造成的；而随机效应模型除了包含抽样误差项之外还考虑了样本总体均值间的差异。

根据 Aguinis 等人（2011）对 196 篇发表在五份英文管理期刊中的元分析的统计结果，87.5%的元分析采用了随机效应模型。这反映了大多数的组织和管理学的研究者们都假定来自实证研究的效应值反映了多个不同的总体。Hunter 和 Schmidt（2004）更是直接指出，随机效应模型反映了社会科学（包含组织管理科学在内）的常态，因此他们提出的元分析的方法只针对随机效应模型。然而有些学者认为（比如 Field & Gillett, 2010），固定效应模型也有它适用的情况。比如说，如果我们把元分析的题目限定为北京市三甲医院

的护士工作强度和工作压力之间关系，那么来自不同实证研究的结果就很有可能反映了同一个样本总体。在这种情况下采用固定效应模型进行分析也可能是有道理的。针对这种情况，Hedges 等人提出的方法 (Hedges, 1984, 1992; Hedges & Olkin, 1985; Hedges & Vevea, 1998) 同时考虑了固定效应和随机效应模型。但随之而来的问题是，固定效应模型得出的结论就只能适用于这个特定的样本总体，而不应该被适用于更广泛的总体（比如全国的护士）。鉴于这两种模型的特点，我们建议大家认真地考虑选题的对象和范围来决定相对应的估计模型，并以此为依据选择具体的分析方法。

## 2. 元分析的两种常见方法

在组织和管理科学的研究中，两种最常见的元分析的方法是 Hunter 和 Schmidt (2004)的方法和 Hedges 等人的方法 (Hedges, 1984, 1992; Hedges & Olkin, 1985; Hedges & Vevea, 1998)。根据 Aguinis 等人 (2011) 的统计，Hunter 和 Schmidt 的方法在组织管理研究中的应用更普遍（有 83.5% 的研究使用了这一方法），而仅有 3.2% 人使用了 Hedges 等人的方法。在这里，我们用相关系数  $r$  为例子对两种方法的基本分析方法进行介绍。如果有对这两种方法的比较感兴趣的同行，可以参考以下这些文章 (Field, 2001, 2005; Hafdahl & Williams, 2009 等)。

### 1) Hunter 和 Schmidt 的方法

首先我们看一下，Hunter 和 Schmidt 的方法是如何计算加权平均后的平均效应值的。我们用  $r_i$  来代表来自第  $i$  个研究里的效应值，用  $n_i$  来代表第  $i$  个研究的样本大小，那么加权平均的效应值  $\bar{r}$  为：

$$\bar{r} = \frac{\sum_{i=1}^k n_i r_i}{\sum_{i=1}^k n_i} \quad (1)$$

相对应的，来自各个样本的效应值方差 (variance of sample)  $\hat{\sigma}_r^2$  为公式 (2) 所示，开平方的结果就是效应值的平方差 (standard deviation)  $SD_r$ 。

$$\hat{\sigma}_r^2 = \frac{\sum_{i=1}^k n_i (r_i - \bar{r})^2}{\sum_{i=1}^k n_i} \quad (2)$$

根据这两个公式以及元分析所包含的研究数量  $k$ ，我们可以进一步计算出加权平均效应值  $\bar{r}$  的置信区间 (confidence interval)。通过对置信区间的检验，我们能够得知是否平均效应值在某个统计水平是显著不等于 0 的。公式 (3) 给出的是  $\bar{r}$  的 95% 的置信区

间。其中  $\sqrt{\frac{\hat{\sigma}_r^2}{k}}$  代表的是  $\bar{r}$  的标准误 (standard error)。

$$\begin{aligned} \text{置信区间上界} &= \bar{r} + 1.96 \sqrt{\frac{\hat{\sigma}_r^2}{k}} \\ \text{置信区间下界} &= \bar{r} - 1.96 \sqrt{\frac{\hat{\sigma}_r^2}{k}} \end{aligned} \quad (3)$$

前文中我们提到了，Hunter 和 Schmidt 的方法采用的是随机效应模型的假设。在随机效应模型里，样本效应值的方差有两个来源：抽样误差方差（sampling error variance,  $\hat{\sigma}_e^2$ ）和效应值的样本总体方差（variance in population,  $\hat{\sigma}_\rho^2$ ）。抽样误差方差的公式是：

$$\hat{\sigma}_e^2 = \frac{(1 - \bar{r}^2)^2}{N-1} \quad (4)$$

在这里， $\bar{N}$ 指的是包含在元分析里的平均样本大小。如果我们把抽样误差方差从样本效应值方差 $\hat{\sigma}_r^2$ 减掉之后，得到的就是效应值的样本总体方差 $\hat{\sigma}_\rho^2$ 。

$$\hat{\sigma}_\rho^2 = \hat{\sigma}_r^2 - \hat{\sigma}_e^2 \quad (5)$$

使用效应值的样本总体方差，Hunter 和 Schmidt 推荐计算加权平均效应值的信用区间（credibility interval），与置信区间用来判定平均效应值的显著性水平不同，信用区间代表的是总体均值的分布情况。公式（6）给出了 80%的信用区间的计算公式，其中 $\sqrt{\hat{\sigma}_\rho^2}$ 是总体效应均值的标准差（standard deviation 或者 SD）。80%的信用区间指的是有 80%样本效应值都分布在这个区域。

$$\begin{aligned} \text{信用区间上限} &= \bar{r} + 1.28 \sqrt{\hat{\sigma}_\rho^2} \\ \text{信用区间下限} &= \bar{r} - 1.28 \sqrt{\hat{\sigma}_\rho^2} \end{aligned} \quad (6)$$

以上是 Hunter 和 Schmidt 关于如何计算样本总体加权平均效应值及其相关统计量的基本计算公式。这些公式给出的结果只校正了抽样误差（sampling error）的影响，通常被成为基本校正整合分析（bare bones meta-analysis）的结果。然而，Hunter 和 Schmidt 强调实证研究中观测到的效应值会受到多种误差的影响，除了抽样误差以外，效应值还可能受到测量误差（measurement error），二元化（dichotomization）和全距变异误差（range variation）等共 11 类误差的影响。因此，他们还提供了对各种误差进行校正的计算公式。因为测量误差是在组织管理科学中最常见到的一种误差，所以我们在这里只介绍如何在元分析中校正测量误差。

测量误差是一种由于测量结果无法完全反映测量对象而导致观测到的效应值比真实效应值偏小的一种误差。如公式（8）所示，通常来说，要用量表的信度系数来对观测到的效应值进行校正。这里的  $r$  代表观测到的效应值， $\rho$  代表校正后真实的效应值（在 Hunter 和 Schmidt 的书中他们也用  $r_c$  来表示）， $r_{xx}$  和  $r_{yy}$  分别代表自变量  $x$  和因变量  $y$  的信度水平。这里需要指明的是，对于一般的变量而言，我们可以用内部一致性信度（Cronbach 的 alpha 系数）来作为变量的信度水平。但是对于在个体层面测量但是最终加总到更高层次的变量来说，我们可以考虑用组间信度（interrater reliability 或者 ICC2）来

对测量误差进行修正。例如，Combs, Liu, Hall, 和 Ketchen (2006) 在对高绩效工作系统的元分析中就采用了这种信度对测量误差进行修正。

$$r = \sqrt{r_{xx}}\sqrt{r_{yy}}\rho \quad (7)$$

Hunter 和 Schmidt 把 $\sqrt{r_{xx}}\sqrt{r_{yy}}$ 的乘积看作是一个削弱因子 (attenuation factor)  $A$ ，如果我们需要考虑其他误差的影响，也可以将这些误差的影响以乘积的形式整合到削弱因子  $A$  里。接下来，我们用这个削弱因子  $A$  来对前面提到的公式进行校正。首先，对于校正后的效应值，我们不能直接采取加权平均数的方式来进行计算，而是需要先计算一个权重值  $w_i$ ，通过这个权重值，我们可以对校正后的效应值进行加权平均。在这里我们用 $\bar{\rho}$ 来代表校正后的加权平均效应值，有的发表的文章中用 $\bar{r}_c$ 来指代。 $\rho_i$ 指的是第  $i$  个研究中校正后的效应值。

$$w_i = (n_i - 1)A_i^2 \quad (8)$$

$$\bar{\rho} = \frac{\sum_{i=1}^k w_i \rho_i}{\sum_{i=1}^k w_i} \quad (9)$$

相对应的，校正后的效应值方差 (variance of sample)  $\hat{\sigma}_o^2$  为：

$$\hat{\sigma}_o^2 = \frac{\sum_{i=1}^k w_i (\rho_i - \bar{\rho})^2}{\sum_{i=1}^k w_i} \quad (10)$$

校正后的抽样误差方差的公式是：

$$\hat{\sigma}_e^2 = \frac{\sum_{i=1}^k w_i \sigma_{\rho}^2 (e)_i}{\sum_{i=1}^k w_i} \quad (11)$$

公式 (12) 中的 $\sigma_{\rho}^2 (e)_i$ 是第  $i$  个研究中修正后的抽样误差的方差，其计算公式为：

$$\sigma_{\rho}^2 (e)_i = \frac{(1 - \bar{r}^2)^2}{A_i^2} \quad (12)$$

如果我们把抽样误差方差从样本效应值方差 $\hat{\sigma}_o^2$ 减掉之后，得到的就是校正后效应值的样本总体方差 $\hat{\sigma}_{\rho}^2$ 。

$$\hat{\sigma}_{\rho}^2 = \hat{\sigma}_o^2 - \hat{\sigma}_e^2 \quad (13)$$

95%置信区间和 80%信用区间的公式分别为：

$$\begin{aligned} \text{置信区间上界} &= \bar{\rho} + 1.96 \sqrt{\frac{\hat{\sigma}_o^2}{k}} \\ \text{置信区间下界} &= \bar{\rho} - 1.96 \sqrt{\frac{\hat{\sigma}_o^2}{k}} \end{aligned} \quad (14)$$

$$\begin{aligned} \text{信用区间上限} &= \bar{\rho} + 1.28 \sqrt{\hat{\sigma}_{\rho}^2} \\ \text{信用区间下限} &= \bar{\rho} - 1.28 \sqrt{\hat{\sigma}_{\rho}^2} \end{aligned} \quad (15)$$

## 2) Hedges 等人的方法

Hedges 等人的方法 (Hedges, 1992; Hedges & Olkin, 1985; Hedges & Vevea, 1998) 跟 Hunter 和 Schmidt 的方法主要有三个方面的区别：(1) Hedges 等人的方法考虑了固定效应模型和随机效应模型，而 Hunter 和 Schmidt 的方法只是针对随机效应模型；(2) Hedges 等人的方法使用了费舍尔的 Z 转换 (Fisher, 1921)，首先将效应值转化成 Z，而 Hunter 和 Schmidt 的方法不需要转换；(3) 两种方法采用了不同的方式来决定权重的大小。接下来，我们开始先介绍 Hedges 等人对于固定效应模型的加权平均效应值的计算。

Hedges 等人的方法首先需要通过费舍尔的公式将效应值从相关系数  $r$  转换成 Z。如果用  $r_i$  来代表第  $i$  个研究中的效应值的话，其转换公式为公式 (16)。经过转换以后的效应值是一个有着均值为  $\bar{Z}_r$ ，方差为  $1/(n_i - 3)$  的正态分布 (Field, 2005)。Z 值可以通过公式 (17) 再转换为相关系数  $r$ 。

$$Z_{r_i} = \frac{1}{2} \text{Log}_e \left( \frac{1+r_i}{1-r_i} \right) \quad (16)$$

$$r_i = \frac{e^{(2Z_i)} - 1}{e^{(2Z_i)} + 1} \quad (17)$$

转换后的 Z 值可以通过公式 (18) 来计算加权平均效应值  $\bar{Z}_r$ ，不同与 Hunter 和 Schmidt 的方法，Hedges 和 Olkin (1985) 建议在固定效应模型中使用的权重值是研究内方差 (within-study variance, 即  $1/(n_i - 3)$ ) 的倒数。 $\bar{Z}_r$  的具体计算公式为：

$$\bar{Z}_r = \frac{\sum_{i=1}^k (n_i - 3) Z_{r_i}}{\sum_{i=1}^k (n_i - 3)} \quad (18)$$

Hedges 等人还介绍了一个  $Q$  统计量对效应值的同质性进行检验 (Hedges & Olkin, 1985)， $Q$  统计量遵循的是以  $k-1$  为自由度的卡方分布，其具体公式如下，在后文中我们还将介绍如何通过  $Q$  来判定是否存在调节变量的情况。

$$Q = \sum_{i=1}^k (n_i - 3) (Z_{r_i} - \bar{Z}_r)^2 \quad (19)$$

在对随机效应模型的平均效应值进行估计的时候，Hedges 等人建议要在权重值中同时考虑研究内 (within-study variance) 和研究间 (between-study variance) 的方差。在前面我们提到了，研究内的方差是  $1/(n_i - 3)$ 。在这里，我们用  $\tau^2$  来代表研究间的方差，其估计值  $\hat{\tau}^2$  的计算公式为：

$$\hat{t}^2 = \frac{Q-(k-1)}{\sum_{i=1}^k(n_i-3) - \frac{\sum_{i=1}^k(n_i-3)^2}{\sum_{i=1}^k(n_i-3)}} \quad (20)$$

我们把转换效应值的研究内方差  $1/(n_i-3)$  和研究间方差的估计值  $\hat{t}^2$  相加之后取其和的倒数，就能够得到在随机效应模型中使用的权重值。并通过权重和转换后的效应值计算随机效应模型中的加权平均转换效应值。在实际的计算过程中， $\hat{t}^2$  的结果有可能是一个负数，在这样的情况下，Hedges 等人建议将其作为 0 来处理。

$$w_i^* = \left(\frac{1}{n_i-3} + \hat{t}^2\right)^{-1} \quad (21)$$

$$\bar{Z}_r^* = \frac{\sum_{i=1}^k w_i^* Z_{ri}}{\sum_{i=1}^k w_i^*} \quad (22)$$

最后，Hedges 等人还提供了计算置信区间所需要的均值标准误（standard error）的公式：

$$SE(\bar{Z}_r^*) = \sqrt{\frac{1}{\sum_{i=1}^k w_i^*}} \quad (23)$$

相对应的，我们就可以得到置信区间的公式，以 95% 的置信区间为例：

$$\text{置信区间上界} = \bar{Z}_r^* + 1.96 SE(\bar{Z}_r^*)$$

$$\text{置信区间下界} = \bar{Z}_r^* - 1.96 SE(\bar{Z}_r^*) \quad (24)$$

接下来，再使用公式 (17) 对公式 (22) 和 (24) 的结果进行转换得到相关系数的结果。

以上就是我们对最常见的两种元分析的基本分析方法的介绍。其中涉及到很多计算公式虽然看起来有些繁琐，但能够帮助我们理解每一种元分析的基本思路。在实际操作中，我们不一定需要根据这些公式自己动手计算，在后面我们会介绍一些已有的针对元分析的统计软件和程序。作为应用元分析这种方法的研究者来说，只需明白在元分析中需要关注哪些统计结果以及它们各自代表的意思即可。

## 五、元分析中调节变量的检验

除了对样本均值进行估计以外，元分析的一个重要作用就是对可能影响效应值的调节变量进行判定和检验。首先，我们需要做的是根据初步分析的结果来判定调节变量是否存在。有些学者 (Sagie & Koslowsky, 1993; Steel & Kammeyer-Mueller, 2002) 曾经对不同的调节变量的判定方法进行了比较。在这里我们主要介绍三种比较常见的做法。

第一，前文提到的  $Q$  统计量（见公式 19）是对效应值的同质性进行检验的统计量，该统计量如果显著，则预示着效应值可能受到潜在的调节变量影响（Hedges & Olkin, 1985）。

第二，Hunter 和 Schmidt（1990）提出一个 75% 的法则。具体来说，我们需要计算校正后的效应值方差有多少比例是可以被抽样误差（sampling error）或者其他误差（比如测量误差，measurement error）等解释的。如果这个比值少于 75%，那么就认定有潜在的调节变量存在。公式（11）和公式（10）的结果的比率就反映了这个数值。

第三，我们还可以通过信用区间来判定是否有潜在的调节变量存在。如果信用区间很宽，包含了 0，那么也预示着潜在的调节变量的存在（Hunter & Schmidt, 2004）。

在我们得出调节变量有可能会影响效应值大小的结论之后，接下来我们就开始考虑如何对具体的调节作用进行检验。通常来说，研究者们应该考虑一些有理论依据的变量（比如，行业类型有可能影响人力资源管理系统和公司绩效的关系）和一些跟实证研究特点有关的变量（比如，测量工具的类型，关系是否存在共同方法偏差等）。而这些可能的调节变量又通常分为两类：一类是类别变量（比如在美国进行的研究和在中国进行的研究，或者个体层次和团队层次的研究），而一类是连续变量（比如女性在各个样本中所占的比例，或者有效样本的回收率）。

针对类别变量，通常研究者们会采用分组比较分析（subgroup analysis）。也就是说对元分析中包含的所有样本根据某一类别型的调节变量进行分类。比如，Combs 等人（2006）对高绩效工作系统和公司绩效的关系研究中，就将行业类型作为一个类别型的调节变量，从而将收集到的实证研究分为来自制造业的研究和来自服务业的研究。分组之后，通过我们上面介绍的元分析基本方法，对各个组内的研究分别计算加权平均效应值，然后对效应值的差异进行比较。Chiaburu, Lorinkova 和 Van Dyne（2013）提供了一个对效应值均值差异进行比较的  $z$  统计量的公式（25），研究者们可以通过  $z$  的显著性水平来判定调节变量的作用是否显著。Hunter 和 Schimidt（2004, p. 423-424）指出，分组分析比其他检验调节变量的方法在统计上有更大的功效（power）去发现显著的调节作用的结果因此在组织和管理科学的元分析中得到广泛应用。但是分组分析每次只能针对一个调节变量进行，为了确保某一个调节变量的作用不被其他调节变量所影响，有时候研究者需要不断地对组别进行细分。如果用高绩效工作系统地例子来说，效应值可能同时受到行业类型和公司大小地影响，为了确保二者的调节作用不会彼此影响，研究者可在制造业和服务也内部又细分中小企业和大企业。

$$z = \frac{(\bar{\rho}_1 - \bar{\rho}_2)}{\sqrt{(SE_1^2 + SE_2^2)}} \\ SE = \frac{\bar{\rho}_*}{\bar{r}} * \frac{SD_r}{\sqrt{k}} \quad (25)$$

另外一种对调节变量进行检验的方法是加权回归分析 (weighted regression analysis)。从原理上来说, 研究者们把来自每个样本的效应值或者校正后的效应值当作因变量, 把潜在的调节变量当作自变量, 同时根据样本大小来对每一个研究在回归分析中占的权重进行赋值。如果调节变量在回归方程中的系数是显著的, 那么就说明调节变量的调节作用得到了验证。加权回归分析的好处在于, 可以直接对连续的调节变量进行分析, 并且能同时考虑多个调节变量的共同作用。然而, Hunter 和 Schmidt (2004) 指出, 元分析中包含的研究的数量通常比较小, 因此可用于加权回归分析的样本量会比较少, 有可能造成结果不准确或者很难发现显著的调节作用。

Steel 和 Kammeyer-Mueller (2002) 对两种方法进行了比较, 发现加权回归的方法更不容易受到多重共线性 (multicollinearity) 和异方差性 (heteroscedasticity) 的影响, 因此得出的结果比分组分析的结果更准确。在这里我们的建议是, 如果有足够多的样本的话, 首选加权回归模型的方法, 同时对于发现显著结果的调节变量再进一步采取分组分析的方法。这种模式类似于一般统计分析中先检验交互作用是否显著, 然后再比较在调节变量的高低取值时的斜率大小。二者结合能更具体的反映调节变量是否显著和作用方式。

## 六、元分析与结构方程模型的结合

Viswesvaran 和 Ones (1995) 提出了一种将元分析的结果和结构方程模型 (structural equation modeling) 或者路径分析 (path analysis) 相结合的方法。简单来说, 就是把元分析的结果用于估计结构方程模型。学者们 (Becker, 2009; Cheung & Chan, 2005; Viswesvaran and Ones, 1995 等) 通常把这种方法叫做元分析结构方程模型 (Meta-analytic structural equation modeling, 或者 MASEM), 为了方便期间, 我们用 MASEM 来指代元分析结构方程模型。MASEM 分为两个基本的阶段 (Cheung & Chan, 2005)。第一阶段, 研究者需要通过元分析来构建变量之间的相关矩阵 (correlation matrix) 或者协方差矩阵 (covariance matrix)。第二阶段, 用相关矩阵和协方差矩阵来估计结果方程模型。举例来说, Harrison 等人 (2006) 研究了工作满意度和组织承诺和五种工作行为之间的关系。为了检验这一关系, 他们首先用元分析的基本方法构建了一个  $7 \times 7$  的相关矩阵, 然后用相关矩阵中的系数来估计结构方程模型。根据 Aguinis 等人 (2011) 的统计, 有 52.8% 的元分析都采用了这种用相关矩阵来估计结构方程模型的做法, 而且这种方式有越来越流行的趋势。

然而, 有学者指出 (比如 Cheung & Chan, 2005), 这种用相关关系矩阵估计结构方程模型的做法有两个主要的局限性。第一, 普通的结构方程模型都是基于同一个样本得出的矩阵来进行分析的, 因此矩阵中各个关系所基于的样本大小都是相等的。然而, 在 MASEM 中矩阵中的元分析效应值都是基于不同的研究得出的, 因此每个研究的样本量并不相同。尽管 Viswesvaran 和 Ones (1995) 建议可以使用调和平均数 (harmonic mean) 来作为结构方程模型的样本大小, 但这种做法有可能会带来不准确的估计结果。第二, 通常的结构方程模型都是基于协方差矩阵而不是相关矩阵来进行的。使用相关矩阵来替换协

方差矩阵也会造成估计的不准确。因此，研究者在使用 MASEM 时要注意这两方面的局限性。Cheung 在后续的研究中提出了一个二阶段元分析结构方程模型的方法，由于篇幅的局限我们不做详细介绍，并推荐感兴趣的研究者参阅 Cheung (2015) 出版的 *Meta-Analysis: A Structural Equation Modeling Approach* 一书。

另外一种将元分析与结构方程模型进行结合的方式是通过回归方程的形式来比较不同自变量对于结果变量的相对大小。举例来说，Chiaburu 等人 (2011) 对大五人格 (big five personality traits) 和组织公民行为之间的关系进行了元分析。他们用大五人格和三种组织公民行为之间的元分析相关系数作为回归方程的数据发现开放性和随和性跟组织公民行为的关系比它们与任务绩效的关系更强。Jiang 等人 (Jiang et al., 2012) 也用了类似的方式发现以提升能力为导向的人力资源管理措施对人力资本的影响要比以提升动机和工作机会为导向的人力资源管理措施要大。通常这样的分析需要比较回归系数在不同回归方程之间的差距是否显著，同时可以计算每个自变量在解释因变量时候的相对权重 (relative weight, Johnson, 2000)。

## 七、元分析中的其他技术细节

前面的章节介绍了元分析中常见的一些分析方法，在这一部分我们讨论一下我们在文章评审和发表过程中遇到的一些需要注意的细节问题。

### 1. 合并同一篇文章中关于同一关系的不同效应值

在编码录入的过程中我们有时候会遇到一篇文章中包含了对同一个关系的不同效应值。比如说，一篇文章中可能汇报了外向型人格和两种组织公民行为之间的关系（比如，帮助个人和帮助组织），从本质上来说，这两个相关关系都反映了外向型人格和组织公民行为之间的关系，所以在进行元分析的之前，我们需要把这两个反映同一种关系的效应值进行合并。合并的主要理由在于，由于这两个效应值来自于同一篇文章，它们之间的关系不是相对独立的，因此如果把它们单独包含进来的话，就违背了样本独立性的原则。因此，Hunter 和 Schmidt (2004, p. 435-439) 对这一问题进行了讨论，并给出了具体公式将多个关于同一个关系的效应值进行整合，在最终的元分析中，针对某一个具体关系，一个实证研究只提供一个合并以后的效应值。如公式 (26) 所示， $r_{xy}$  代表合并后的两个变量之间的关系， $\sum r_{x_i} r_{y_j}$  代表所有关于这两个变量的效应值之和， $n$  和  $m$  分别代表有几个反映同一个自变量  $X$  和因变量  $Y$  的变量数， $\bar{r}_{x_i x_j}$  代表所有自变量  $X$  之间的相关系数均值， $\bar{r}_{y_i y_j}$  代表所有因变量  $Y$  之间的相关系数均值。

$$r_{xy} = \frac{\sum r_{x_i} r_{y_j}}{\sqrt{n+n(n-1)\bar{r}_{x_i x_j}} \sqrt{m+m(m-1)\bar{r}_{y_i y_j}}} \quad (26)$$

### 2. 将意思相近的变量进行归类

近年来，有很多元分析尝试将意思相近的变量归入一个大类，进而分析两个意思更为广泛的变量之间的关系。举例来说，Harrison 等人（2006）曾经建议，将工作满意度和组织承诺合并为工作态度，并将五种不同的工作行为合并为工作行为之后，对两者之间的关系进行分析的预测效度更高。Jiang 等人（2012）也将不同类型的人力资源管理措施笼统的分为三大类以提升能力为导向，以提升动机为导向，和以提升表现机会为导向的人力资源管理措施。在战略管理方面，Karna, Richter, 和 Riesenkaempff（2016）曾经把几十种不同的变量归类在一起统称为组织拥有的能力，并以此来研究组织能力和财务绩效之间的关系。针对这样的做法，不同的学者持不同的观点。比如 Eysenck（1984）将这种做法称作把苹果和橙子放在一起数数，得出的结论是没有意义的。但是有的学者认为，在一定程度上合并是有意义的。比方说，Rosenthal 和 DiMatteo（2001）认为，把苹果和橙子合并在一起来看是有意义的，尤其是当人们想研究的是水果这一大的种类。我们认为这两种观点都有一定的道理，关键在于变量合并的方式是否有理论以及是否与研究问题相符。比如说，如果有学者想研究工作态度和离职意向之间的关系，那么就很有必要将工作满意度，组织承诺，组织认同感，组织支持感和离职意向区别对待，看离职意向是如何受到其他工作态度的影响的。但是如果有的学者想看，员工对组织的依附感（attachment）和工作绩效之间的总体关系，那么我们就可以把离职意向反向记分后跟组织承诺合并为依附感，然后看它们跟工作绩效之间的总体关系。除此之外，我们还建议在将意思相近的变量合并之前，可以单独计算它们与某一个结果变量之间的加权平均效应值，比较效应值之间是否有显著的差异，作为一个辅助的证据来支持变量的合并。

### 3. 发表偏差

前文提到了在文献搜索的时候要尽量包含工作论文或者未发表的研究以避免发表偏差对研究结果的影响。Hunter 和 Schmidt（2004）中有单独一章对如何检验和校正发表偏差对结果的影响。对发表偏差进行检验的方法常用的有两种。第一种是 Rosenthal（1979）提出的文件柜分析（file drawer analysis），简单来说就是估计一下还需要存在多少未发表的研究才能将现有的研究结果从显著变得不显著了。这里对未发表的文章的估计被叫做 fail-safe N。第二种是由 Light 和 Pillemer（1984）提出的漏斗图法（funnel plot）。简单来说，漏斗图使用效应值和样本量作为坐标系将各个研究绘制在坐标系里的散点图。其理论依据是对于样本量越大的研究来说，它的效应值的估计也就越准确，反之其误差也越大。这样漏斗图里样本量大的研究集中在图的上方，平均效应值周围；样本量小的研究则散落在图的底部，离平均效应量值较远。通过观察图形的形状，我们可以来确定发表偏差是否存在。如果图形呈现一个倒着的漏斗（或者正着的火山）的形状，则预示这发表偏差可能不存在；但是如果图形有缺角，则表明发表误差可能存在。基于这一图形检测方法，Duval 和 Tweedie（2000）提出了迭代的方法（Trim and Fill）对不对称的漏斗图中的研究进行删减使其变成对称，并对校正后的样本重新计算。在利用元分析对效应值的调节变量进行分析的时候，研究者还可以将研究的发表状态作为一个类别变量，看其是否对效应值有影响，并以此来判断是否存在潜在的发表偏差。

#### 4. 样本大小

元分析中涉及到两个样本大小。第一个是元分析所收集到的研究的数量（通常用  $k$  来表示），第二是所有研究所包含的样本总量（通常用  $N$  来表示）。对于一篇成功的元分析来说，我们通常会希望两个数量都比较大，这样说明所选定的研究题目已经积累了足够多的研究，值得我们对其来进行一项元分析的研究。然而，从元分析方法本身而言，并没有一个绝对的判断标准来指明究竟需要收集多少篇关于某一选题的文章才能开展一项元分析。从论文发表的经验来看，我们见过在顶级管理期刊发表的元分析只包含了不到 50 个独立样本（比如，Wang, Waldman, & Zhang, 2014），我们也见过包含了超过 400 多个独立样本（比如，Judge, Rodell, Klinger, Simon, & Crawford, 2013）。因此，我们觉得只要研究者能够找到足够大的  $k$  对他们的研究问题提供稳定的估计，那么进行这些元分析就是有意义的。另外，样本大小对于元分析中不同分析所起到的作用也是不同的。比如对于检验调节作用而言，总体研究数量  $k$  在很大程度上决定了分组检验（subgroup analysis）的结果是否稳定，以及加权回归（weighted regression）中调节变量是否显著。对于这样的分析来说，我们会建议  $k$  越大越好。但是对于元分析结构方程而言，我们需要构建一个包含所有变量在内的元分析相关矩阵，对于相关矩阵中的每一个元分析相关系数来说，我们一般会建议至少有 3 个来自不同研究样本的效应值来确保对于该相关系数的估计是相对稳定的。

#### 5. 多种语言情境下的元分析

随着学者间的合作不断的深入，现在出现越来越多来自不同国家的研究人员共同开展一项元分析的机会。随之而来的一个问题就是，如果研究团队中的成员熟悉多种语言，是否一定要把元分析的对象限定为英文发表的文章呢？在这里，我们假定中国的研究者想把元分析的结果发表在英文期刊上。我们觉得至少在以下两种情况下，研究者是可以考虑把多种语言发表的文章整合在一起的。首先，元分析的目的是想对来自不同国家或文化的研究结果进行比较，这样尽可能地囊括来自不同国家的研究结果对实现元分析的目的是有帮助的。举例来说，Rabl 等人（2014）研究了高绩效工作系统的作用是如何受到国家文化的调节的。除了包含英文发表的文章以外，他们还包含了用巴基斯坦语和德语发表的文章，因为他们的研究团队中刚好有熟悉这两种语言的成员。此外，Oh 等人（2014）对个人-环境的匹配与工作态度和行为的关系进行了跨文化的元分析。他们在文中也包含了韩语发表的文章。对于这两个例子来说，包含其他语言的文章对于实现他们的研究目的本身是非常有必要的。其次，针对某一研究关系的英文文章数量不够，需要其他语言的文章进行补充。Liu 等人（2016）对个人创造力的中介机制进行了元分析。他们的模型中需要估计亲社会动机和个人创造力之间的相关关系。但是针对这一关系本身的英文文章很少，研究者为了让估计结果更加稳定，也考虑加入了中文和韩语发表的文章。除了上述提到的两种情况以外，我们建议在进行元分析的时候要谨慎考虑是否要加入中文发表的文章。一个可能的问题是，如果包含了中文的文章，那么不了解中文的研究者可能没有办法去重复或

者检验元分析的结果。但是不管选择什么方式处理，我们都建议在文章的方法部分将文搜索的过程和结果详细得汇报出来，让文章的评审和主编来做出他们的判断。

在元分析中还有很多需要处理的技术细节，鉴于篇幅的关系，我们无法一一展开来介绍，但我们鼓励对这一方法感兴趣的同行能够参阅相关的材料来解决他们在实际操作中遇到的其他问题。

## 八、常见的元分析工具的介绍

在本章的最后一个部分，我们介绍一下元分析常用的分析工具和软件。

### 1. 付费软件

Windows-Based Meta-Analysis Software Package 是 Frank Schmidt 和他的学生 Huy Le 编写的一个统计程序，现在是 2.0 的版本。这个软件包含了所有 Hunter 和 Schmidt (2004) 中涉及到的统计分析。Roth (2008) 曾对这一软件进行过评估，对其结果的准确性和分析的全面性进行了推荐。非常难得的是，这一软件从 2008 年之后售价一直保持在个人版本 150 美元和学院版本 350 美元。有感兴趣的同行可以联系 Frank Schmidt (frank-schmidt@uiowa.edu) 和 Huy Le (huy.le@utsa.edu) 进行购买。

Comprehensive Meta-Analysis (CMA) 是基于 Hedges 等人的方法开发的元分析软件。因此，如果用 CMA 得出的结果跟 Schmidt 的软件得出的结果是有一些差异的。其中的 Meta-regression 的模块对于检验元分析中的调节变量非常方便，值得推荐。相比 Schmidt 的软件来说，CMA 采用公司模式的运营，软件的售卖分不同的对象和版本，最低配置的学生版本售价为一年 125 美金，而最高配置的学术版本售价为一年 495 美金。高配置的版本允许研究者在分析中包含更多的变量同时进行更复杂的分析。有感兴趣的同行可以参阅其官方网站 <https://www.meta-analysis.com/index.php?cart=B338898796>。

### 2. 免费软件

Stata 和 R 是两种常用的可以进行元分析的软件。Stata 软件本身并不免费，但考虑到很多学校都会订购这一软件，我们将它归为免费软件。有些对统计方法感兴趣的研究人员将自己编写的 Stata 代码与同行分享，其中就包括针对元分析的代码。Jonathan Sterne 编写的 Meta-Analysis in Stata: An updated collection from the Stata journal 一书中对这些代码进行了详细介绍。但需要指出的是，这些代码的编写者并不是管理学或者心理学的背景，有很多是医学的背景。因此他们的分析方法并不是基于 Hunter 和 Schmidt 或者 Hedges 等人的方法，在组织管理科学的研究中应用较少。有感兴趣的同行可以参阅 <http://www.stata-press.com/data/mais.html>。

R 是今年来比较流行的开放共享型的统计分析软件。类似于 Stata，R 里的统计包都是由来自不同领域的研究者自己编写并与其他人共享的。其中有多位研究者都发表了各自

的软件包，比如 Dewey (2017) , Viechtbauer (2016) 和 Schwarzer (2017) 。我们需要首先下载 R 软件的应用程序，在下载各自的元分析软件包，根据其中的语句要求来进行分析。Cheung (2017) 也编写了 MASEM 的 R 软件包，感兴趣的同行可以参阅 <https://cran.r-project.org/web/packages/metaSEM/vignettes/metaSEM.pdf>。

### 3. 学者网站

除了软件以外，还有一些元分析的专家自己根据前文提到的公式编写了可以英语 SPSS, SAS, Stata 等软件里的编码。比如说 Practical Meta-Analysis 的作者之一 David Wilson 就在其个人网站上 (<http://mason.gmu.edu/~dwilsonb/ma.html>) 给出了用来计算元分析的代码。Andy Field 也在其网站上给出了包括估计加权平均效应值，检验调节变量，以及校正发表偏差在内的 SPSS 代码和 R 代码

([https://www.discoveringstatistics.com/repository/fieldgillett/how\\_to\\_do\\_a\\_meta\\_analysis.html](https://www.discoveringstatistics.com/repository/fieldgillett/how_to_do_a_meta_analysis.html))。此外，如果有熟悉 Excel 的同行，也可以根据我们上面提供的公式在 Excel 里自己编写语句，并将分析结果与上述软件的结果进行比对，以此作为自己进行元分析的工具。

## 九、结语

在本章中，我们对元分析的基本用途，原理，方法，和分析软件进行了介绍。随着实证研究在组织和管理科学中的快速发展，元分析技术也将得到越来越广泛的应用。我们希望本章的介绍能够帮助对这一方法感兴趣的同行更快地理解和掌握元分析的基本方法，并将这种方法应用到自己的研究中去。元分析研究法本身是一个系统的研究方法，介于篇幅的限制，我们对这一方法的介绍不能做到面面俱到。在此，我们鼓励大家在自己的研究中不断摸索更多的使用元分析的经验并与同行分享，共同推进元分析研究法在中国管理科学中的应用。

## 参考文献

- Aguinis, H., Pierce, C. A., Bosco, F. A., Dalton, D. R., & Dalton, C. M. (2011). Debunking myths and urban legends about meta-analysis. *Organizational Research Methods, 14*(2), 306-331.
- Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods, 15*(4), 624-662.
- Becker, B. J. (2009). Model-based meta-analysis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 377-395). New York: Russell Sage Foundation.
- Cheung, M. W. L. (2015). metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology, 5*, 1521.
- Cheung, M. W. L., & Chan, W. (2005). Meta-analytic structural equation modeling: a two-stage approach. *Psychological methods, 10*(1), 40-64.
- Chiaburu, D. S., Lorinkova, N. M., & Van Dyne, L. (2013). Employees' social context and change-oriented citizenship: A meta-analysis of leader, coworker, and organizational influences. *Group & Organization Management, 38*(3), 291-333.
- Chiaburu, D. S., Oh, I. S., Berry, C. M., Li, N., & Gardner, R. G. (2011). The five-factor model of personality traits and organizational citizenship behaviors: a meta-analysis. *Journal of Applied Psychology, 96*(6), 1140-1166.
- Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: a meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85*(5), 678-707.
- Combs, J., Liu, Y., Hall, A., & Ketchen, D. (2006). How much do high-performance work practices matter? A meta-analysis of their effects on organizational performance. *Personnel psychology, 59*(3), 501-528.
- Dewey, M. (2017) <https://cran.r-project.org/web/views/MetaAnalysis.html>
- Dickersin, K., Min, Y. I., & Meinert, C. L. (1992). Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *Jama, 267*(3), 374-378.
- Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-463.
- Eysenck, H. J. (1984). Meta-analysis: An abuse of research integration. *The journal of special education, 18*(1), 41-59.
- Field, A. P. (2001). Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed-and random-effects methods. *Psychological methods, 6*(2), 161-180.
- Field, A. P. (2005). Is the meta-analysis of correlation coefficients accurate when population correlations vary? *Psychological methods, 10*(4):444-467.
- Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63*(3), 665-694.

- Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3-32.
- Hafdahl, A. R., & Williams, M. A. (2009). Meta-analysis of correlations revisited: attempted replication and extension of Field's (2001) simulation studies. *Psychological Methods*, 14(1), 24-42.
- Hancock, J. I., Allen, D. G., Bosco, F. A., McDaniel, K. R., & Pierce, C. A. (2013). Meta-analytic review of employee turnover as a predictor of firm performance. *Journal of Management*, 39(3), 573-603.
- Harrison, D. A., Newman, D. A., & Roth, P. L. (2006). How important are job attitudes? Meta-analytic comparisons of integrative behavioral outcomes and time sequences. *Academy of Management journal*, 49(2), 305-325.
- Heavey, A. L., Holwerda, J. A., & Hausknecht, J. P. (2013). Causes and consequences of collective turnover: a meta-analytic review. *Journal of Applied Psychology*, 98(3), 412-453.
- Hedges, L. V. (1984). Estimation of effect size under non-random sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics*, 9, 61-85.
- Hedges, L. V. (1992). Meta-analysis. *Journal of Educational Statistics*, 17(4), 279-296.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3(4), 486-504.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park, CA: Sage.
- Jiang, K., Lepak, D. P., Hu, J., & Baer, J. C. (2012). How does human resource management influence organizational outcomes? A meta-analytic investigation of mediating mechanisms. *Academy of management Journal*, 55(6), 1264-1294.
- Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate behavioral research*, 35(1), 1-19.
- Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, 98(6), 875-925.
- Karna, A., Richter, A., & Riesenkauff, E. (2016). Revisiting the role of the environment in the capabilities-financial performance relationship: A meta-analysis. *Strategic Management Journal*, 37(6), 1154-1173.
- Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.
- Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

- Liu, D., Jiang, K., Shalley, C. E., Keem, S., & Zhou, J. (2016). Motivational mechanisms of employee creativity: A meta-analytic examination and theoretical extension of the creativity literature. *Organizational Behavior and Human Decision Processes*, 137, 236-263.
- Meyer, J. P., Stanley, D. J., Herscovitch, L., & Topolnytsky, L. (2002). Affective, continuance, and normative commitment to the organization: A meta-analysis of antecedents, correlates, and consequences. *Journal of vocational behavior*, 61(1), 20-52.
- Oh, I. S., Guay, R. P., Kim, K., Harold, C. M., Lee, J. H., Heo, C. G., & Shin, K. H. (2014). Fit happens globally: A meta-analytic comparison of the relationships of person–environment fit dimensions with work attitudes and performance across East Asia, Europe, and North America. *Personnel Psychology*, 67(1), 99-152.
- Park, T. Y., & Shaw, J. D. (2013). Turnover rates and organizational performance: a meta-analysis. *Journal of Applied Psychology*, 98(2), 268-309.
- Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, 2, 1243-1246.
- Pratt, J. G., Rhine, J. B., Smith, B. M., Stuart, C. E., & Greenwood, J. A. (1940). *Extra-sensory perception after sixty years: A critical appraisal of the research in extra-sensory perception*. New York: Henry Holt.
- Rabl, T., Jayasinghe, M., Gerhart, B., & Kühlmann, T. M. (2014). A meta-analysis of country differences in the high-performance work system–business performance relationship: The roles of national culture and managerial discretion. *Journal of Applied Psychology*, 99(6), 1011-1041.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin*, 86(3), 638-641.
- Rosenthal, R. (1991). Meta-analysis: a review. *Psychosomatic medicine*, 53(3), 247-271.
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual review of psychology*, 52(1), 59-82.
- Roth, P. L. (2008). Software review: Hunter-Schmidt meta-analysis programs  
1.1. *Organizational Research Methods*, 11(1), 192-196.
- Sagie, A., & Koslowsky, M. (1993). Detecting moderators with meta-analysis: an evaluation and comparison of techniques. *Personnel Psychology*, 46(3), 629-640.
- Schwarzer, G. (2017). <http://portal.uni-freiburg.de/imbi/lehre/lehrbuecher/meta-analysis-with-r/r-packages>.
- Steel, P. D., & Kammeyer-Mueller, J. D. (2002). Comparing meta-analytic moderator estimation techniques under realistic conditions. *Journal of Applied Psychology*, 87(1), 96-111.
- Viechtbauer, W. (2016). <http://www.metafor-project.org/doku.php/installation>. Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel psychology*, 48(4), 865-885.

Wang, D., Waldman, D. A., & Zhang, Z. (2014). A meta-analysis of shared leadership and team effectiveness. *Journal of Applied Psychology*, 99(2), 181-198.张翼, 樊耘, & 赵菁. (2009). 国外管理学研究中的元分析评介. *外国经济与管理*, 30(7), 1-8.

毛良斌, & 郑全全. (2005). 元分析的特点, 方法及其应用的现状分析. *應用心理學*, 11(4), 354-359.

王永贵, & 张言彩. (2012). 元分析方法在国内外经济管理研究中的应用比较. *经济管理*, 4, 182-190.

魏江, 赵立龙, & 冯军政. (2012). 管理学领域中元分析研究现状评述及实施过程. *浙江大学学报(人文社会科学版)*, 42(5), 144-156.