心理学研究的元分析报告标准：现状与建议[1]

刘宇[1]，陈树铨[2]，樊富珉[1]，邸新[3]，范会勇[4]，封春亮[5]，甘怡群[6]，李会杰[7]，吕小康[8]，徐鹏飞[9]，袁博[10]，左西年[7]，胡传鹏[11,12*]

([1]清华大学心理学系，北京 100084)

([2]Department of Clinical and Counseling Psychology, Teachers College, Columbia University, New York City 10025, NY , USA)

([3] Department of Biomedical Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA)

([4]渤海大学教育与体育学院，锦州 121013)

([5]北京师范大学信息科学与技术学院，北京 100875)

([6]北京大学心理与认知科学学院，北京 100871)

([7] 中国科学院行为科学重点实验室，北京 100101)

([8]南开大学周恩来政府管理学院社会心理学系，天津 300350)

([9]深圳大学心理与社会学院，深圳 518060)

([10]宁波大学心理学系暨研究所，宁波 315211)

([11]Deutsches Resilienz Zentrum (DRZ), University Medical Center of the Johannes Gutenberg University, 55131 Mainz, Germany)

([12] Neuroimaging Center (NIC), Focus Program Translational Neuroscience (FTN), Johannes Gutenberg University, 55131 Mainz, Germany)

[1]*本文作者受到资助情况如下。徐鹏飞：深圳市孔雀计划（827-000235）。封春亮: 中国博士后创新人才支持计划 (BX201600019)与中国博士后科学基金面上资助(2017M610055)

通讯作者：胡传鹏: hcp4715@hotmail.com

摘要：元分析通过综合实证研究结果，积累证据、去伪存真，在研究与实践中有着非常重要的作用。元分析所用方法种类繁多，其选择对研究的质量影响较大，当前元分析报告迅猛增长使得其质量更应多加关注。为规范元分析流程、提升报告质量，生物医学（如由医疗卫生领域专家国际性小组制定的 PRISMA）及心理学（如美国心理学会推荐的 MARS）领域中均制定了元分析报告标准。然而，这类标准开放度不够而且没有中国数据支撑。结合心理学的特点以及最近开放科学的趋势，本文拟提出更具有开放性与透明性的元分析报告标准，共包括 7 部分 30 个条目。基于该标准的方法与结果报告部分的条目，本文对 2017 年国内心理学期刊上发表的元分析报告进行评估。结果发现：已有元分析报告存在方案预注册信息缺乏、研究方法与结果部分内容的混淆、数据来源与分析过程开放性不够等问题。本文提出的元分析报告标准，既可以帮助研究者实施元分析，又有助于提升国内元分析研究的质量。

关键词：开放, 可重复性, 元分析, 报告标准,

Reporting Standard for Meta-Analysis in Psychology: Status and Recommendations

LIU Yu[1]; CHEN Shuquan[2]; FAN Fumin[1]; DI Xin[3]; FAN Huiyong[4]; FENG Chunliang[5]; GAN Yiqun[6]; LI Huijie[7]; LV Xiaokang[8]; Xu Pengfei[9]; YUAN Bo[10]; ZUO Xi-Nian[7]; HU Chuan-Peng[11,12*]

(*[1]Department of Psychology, Tsinghua University, Beijing* 100084, *China*)

(*[2]Department of Clinical and Counseling Psychology, Teachers College, Columbia University, New York City* 10025, *NY*, *USA*)

(*[3]Department of Biomedical Engineering, New Jersey Institute of Technology, Newark* 07102, *NJ, USA*)

(*[4]College of Education Sport Science, Bohai University, Jinzhou* 121013, *China*)

(*[5]College of Information Science and Technology, Beijing Normal University, Beijing* 100875, *China*)

(*[6]School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing* 100871, *China*)

(*[7]CAS Key Laboratory of Behavioral Science, Beijing* 100101, *China*)

(*[8]Department of Social Psychology, Zhou Enlai School of Government, Nankai University, Tianjin* 300071, *China*)

(*[9]Institute of Affective and Social Neuroscience, Shenzhen University, Shenzhen* 518060, *China*)

(*[10]Department of Psychology, Ningbo University, Ningbo* 315211, *China*)

(*[11]Deutsches Resilienz Zentrum (DRZ), University Medical Center of the Johannes Gutenberg University,* 55131 *Mainz, Germany*)

([12]*Neuroimaging Center (NIC), Focus Program Translational Neuroscience (FTN), Johannes Gutenberg University,* 55131 *Mainz, Germany*)

Abstract： Meta-analysis is a cluster of statistical techniques to quantitatively synthesize empirical findings. By doing meta-analysis, the science community can accumulate existing evidence and distinguish true effects from false positives. Therefore, meta-analytic results were regarded as the high-level evidence in both basic research and clinical settings. Despite its importance, the qualities of published meta-analysis studies, however, is not always warranted. With the rapid increase in the number of published meta-analytic studies, the science community increasingly pays attention to the quality of meta-analytic studies, and several standards have been developed in different fields of science, for example, the Preferred Reporting Items for Systematic Reports and Meta-Analyses (PRISMA) were developed in medical science and the Meta-Analysis Reporting Standards (MARS) were developed in psychology. However, there is no standard for meta-analysis in Chinese, although many meta-analytic studies were published in Chinese every year. To fill this gap, we developed Preferred Reporting Items for Open and Reproducible Meta-analysis (PRIOR-MA). PRIOR-MA integrates items from PRISMA and MARS, as well as the Transparency and Openness Promotion (TOP) guidelines. Therefore, PRIOR-MA provides three advantages: (1) it is based on the most widely used reporting standard for meta-analysis - PRISMA, therefore have recommendation value; (2) it integrates items from MARS so that it has higher practical value than PRISMA for researchers in psychology and related field, and (3) it embraces the latest open science initiative and encourages opening data and analytic code. PRIOR-MA consists of seven sections, and 30 items in total. Based on PRIOR-MA, we evaluated meta-analytic studies published in Chinese psychology journals in 2017, and the results showed that many of these articles did not fully report important information that is necessary for evaluating their qualities, for example, there is lack of protocol and registration information, confusion between method and results, and unavailability of data and analytical code. These results suggested that Chinese researchers need a reporting standard like

PRIOR-MA. We recommend Chinese psychology journals adopt PRIOR-MA as part of their policy. We believe applying PRIOR-MA in the journal policy will benefit not only researchers who conduct meta-analysis but also the science community as a whole.

Keywords: open, reproducible, meta-analysis, reporting standard

1 引言

元分析（meta-analysis，又译为荟萃分析）是通过对统计指标进行搜集比较，将若干研究的结果进行整合分析的技术。而系统综述（systematic review，亦可译为系统评价）是指提供针对某一特定问题的文献评读与整合性总结，包含系统性搜索、选择与评价已有研究报告的方法(Higgins, 2008)。元分析与系统综述均对同一或者相似的研究问题，整合不同研究的数据结果，从而得到更加系统的结论。相较于系统综述，元分析使用定量分析能对上百个研究结论进行比较综合，这是叙述性系统综述难以完成的(Gurevitch, Koricheva, Nakagawa, &Stewart, 2018)。元分析/系统综述是研究者们迅速掌握关于某一研究问题当前进展的主要信息源。由于这种整合的效果，元分析/系统综述可以为相关政策制定及实践提供比单个研究更加有价值的参考。例如，牛津循证医学中心（Oxford Centre of Evidence-Based Medicine, CEBM, cebm.net）在证据分级中将元分析/系统综述推荐为 A 级证据（一级证据）（见图 1）(Gates, &March, 2016)。元分析/系统综述作为一种证据累积与综合的方法，在心理学与认知神经科学中同样起到非常重要的作用。

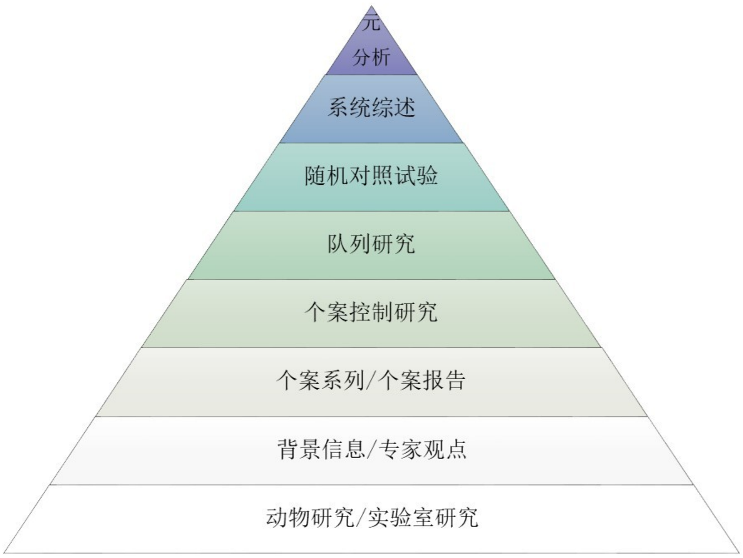

图 1 Oxford Centre of Evidence: Medicine Levels of Evidence 示意图

1.1 元分析/系统综述的报告标准及其重要性

由于元分析/系统综述在研究中的重要作用，同时，也因为元分析发表量的迅速增加，元分析/系统综述本身的严谨性和标准化也显得日益重要。一方面，元分析/系统综述的报告质量依赖于其在研究方法上的严谨性及其研究报告清晰程度。例如，Sacks 等(1987)对 86 篇使用英语发表的元分析从 6 个重要方面（包括 23 个条目）进行了详细评估，其结果表明，86 篇元分析中仅有 24 篇（28%）完整地报告了 6 个重要方面的内容。另一方面，元分析/系统综述的普及，带来了文献数量的迅速增长，但在方法上的差异，可能导致对同一研究主题的元分析/系统综述得到完全相反的结论(Schalken, & Rietbergen, 2017; Vrieze,

2018)，同时对后续的研究者产生干扰甚至误导。此外，在某些具体的研究领域，元分析的方法本身可能在不断发展变化(胡传鹏, 邸新, 李佳蔚, 隋洁, &彭凯平, 2015)。如果元分析报告缺乏清晰性与完整性，后续研究者可能难以评估该元分析的质量，也很难重复该分析。例如，神经成像领域，由于方法与软件的发展，分析软件本身的发展可能影响元分析结果的可重复性(Eickhoff, Laird, Fox, Lancaster, & Fox, 2016)。要解决这些问题，需要在元分析/系统综述中形成一个统一的、共同的标准，因此元分析/系统综述的报告标准（Reporting Standard，也可以称为报告规范）应运而生。

按照报告标准撰写的元分析/系统综述的一个明显好处是帮助审稿人和读者快速地甄别文章的质量。此外，报告标准还有三个重要的作用：第一，报告标准有助于进行元分析的研究者提前意识到在元分析过程中所涉及的各项元素，指导其实施元分析；第二，对后续研究者来说，当元分析的过程被清晰地报告时，该综合证据更具有可重复性，也更能够此基础上继续累积证据；第三，针对同一研究问题，使用相同的报告标准进行的元分析/系统综述之间可以进行更加准确地比较。

为了能够达到上述目的，一个好的报告标准应该包括三个重要的方面：推荐价值、指导作用与出版要求。也就是说一个好的报告标准，应能在方法上严谨而规范，促进报告的透明程度，避免研究中的偏倚（bias），具有被研究者推荐的价值；其有详细的说明与指导，能够对研究者实施该研究进行有效的帮助；同时将这些信息形成一系列具体的准则与标准，最后通过被期刊采纳，成为该类型研究发表基本要求。

1.2 当前主要的元分析/系统综述报告标准及现状

在心理学以及医学研究中，目前使用最为广泛的是 PRISMA 系列报告标准（Preferred Reporting Items for Systematic Reviews and Meta-Analyses），该标准的前身是由医疗卫生领域专家组成的国际性小组于 1999 年制定并发布的报告标准 QUOROM（Quality of Reporting of Meta-analyses statement）。2009 年，小组成员对 QUOROM 进行了修订和总结，加入了对系统综述的关注，并发布了 PRISMA 系列的第一篇报告标准 PRISMA 声明。该声明包括 7 个部分，27 个条目清单和一个四阶段的流程图。PRIMSA 满足一个好的报告标准的三个方面：首先，PRIMSA 标准得到 Centre for Reviews and Dissemination、Cochrane Collaboration、Council of Science Editors、National Evidence-based Healthcare Collaborating Agency (NECA)与 World Association of Medical Editors 5 个国际组织与几百种期刊支持；其次，PRISMA 能够对研究者提供切实有效的指导，其不仅涵盖了大部分随机对照试验，而且针对每一个条目，均有完备与清晰的阐述与解释性文件，并配有具体的例子(Liberati et al., 2009)；第三，PRISMA 是许多医学期刊元分析发表的必要条件，被写入期刊的投稿要求中。此外，从引用量上来看，PRISMA 是目前引用最多的标准：自 2009 年发表以来，截止到 2018 年 3 月 15 日搜索，PRISMA 的两篇介绍性文章(Liberati et al., 2009; Moher, Liberati, Tetzlaff, &Altman, 2009)在 Google 学术上的引用量分别为 30384 和 13083 次。

美国心理学会（American Psychological Association, APA）于 2008 年

首次推出了心理学领域的元分析报告标准（Meta-Analysis Reporting Standards, MARS）(A. P. A. Publications Communications Board Working Group on Journal Article Reporting, 2008)。最近，APA 对 MARS 进行了修订包括 7 个部分和 27 个条目 (Appelbaum et al., 2018)。但与 PRISMA 相比，MARS 作为一个报告标准仍然显得不太完整。虽然 MARS 得到了一些研究者的推荐，但是缺乏详细的说明，很难对心理学研究者起到切实的指导作用。同时，作为 APA 制定的报告标准，其从来未成为 APA 旗下杂志发表元分析研究的要求。自 2008 年推出以来，MARS 在 Google 学术上仅被引用 93 次。

当然，还有其他的元分析报告标准，例如 MOOSE（Meta-analysis of observational studies in epidemiology），其主要应用于观察性研究(Stroup et al., 2000)。不同的报告标准虽然内容有所差异，但其目的均是为了促进高质量的元分析/系统综述得以产生，保证循证科学证据的可靠性。随着研究方法、研究领域的不断深入与扩展，报告标准也处于不断更新与完善之中。

2 心理学研究的元分析报告标准的修订

虽然元分析的报告标准非常重要，但是国内心理学界极少有报告[2]。鉴于此，本文一方面介绍元分析的报告标准，以方便国内研究者参考；另一方面对方法的核心部分进行详细介绍，以期能够为研究者在进行元分析时提供实际帮助。这里在介绍元分析报告标准时遵循两个原则：第一，符合国际主流的元分析报告标准，即介绍的标准能够满足元分析报告标准的三个要求；第二，能够体现最近心理学研究中的新动向，即开放与透明的研究实践需求(Nosek et al., 2015; 朱滢, 2016)。

遵循第一个原则，本文介绍的标准以 PRISMA 为基础。如上所述，相较于其他标准，PRISMA 达到了好的报告标准的三个方面要求，同时具有很高的关注度与认可度。但由于 PRISMA 主要应用于医学研究，直接使用 PRISMA 对于心理学研究者来说并不满足“指导手册”的要求。因此，本文在 PRIMSA 的基础之上，整合最新版 MARS 中的相关条目。

遵循第二个原则，本文参考目前被众多杂志所采用的 TOP（Transparency and Openness Promotion, https://cos.io/our-services/top-guidelines/）标准(Nosek et al., 2015)，加入对透明与开放的要求。虽然 PRISMA 最近开始要求对元分析方案进行预注册(Moher et al., 2015)，但其并未要求对元分析中涉及到的数据与分析代码进行开放。因此，本元分析报告标准将 TOP 标准中相关的要求采纳进来。

通过本文专家组的讨论与整合，最终形成《开放与可重复元分析的报告标准》（Preferred Reporting Items for Open and Reproducible Meta-analysis, PRIOR-MA 见表 1)，并对核心部分进行介绍，以期为国内心理学研究人员撰写和报告元分析提供借鉴和参考。本标准共有 30 个条目，其中方法与结果部分的 21 个条目（条目 5 至条目 25）是其核心标准，对于元分析的方法部分具有指导意义，确保元分析中纳入的研究符合规范、偏倚最小并且采用有效的方

[2] 我们以几个元分析标准的论文为关键词（如 PRISMA），以“心理”为文献来源在 CNKI 中进行检索后发现检索结果为 0（最后检索日期：2018.05.14）。

法来综合信息。在这 21 个条目中，17 条为基础要求，4 条为拔高标准（表 1 中带*的条目），即鼓励研究者报告。考虑到方法部分与结果部分重合度较高，本文重点对标准的方法部分进行介绍。

表 1 PRIOR-MA 清单

章节/主题	#	清单条目
标题		
标题	1	陈述研究问题，并表明研究类型是系统综述、元分析或是两者兼具。
摘要		
结构化小结	2	提供结构化的小结，包括： 2.1 目标（研究的问题或假设）； 2.2 文献纳入标准（描述纳入研究的自变量、因变量与研究设计等特征）； 2.3 研究综合的方法； 2.4 结果（纳入研究与被试的数量及其特征，相关效应量与置信区间等）； 2.5 结论（描述文章中存在的优势与局限）。
引言		
问题提出	3	陈述该研究领域相关的问题及其关系，包括： 3.1 历史背景，已有的系统综述与元分析持有的观点； 3.2 该研究可能涉及的理论、政策以及实际应用中的议题； 3.3 该研究可能涉及的相关人群和背景； 3.4 说明理由：（a）研究设计选择的理由；（b）对文献结果进行选择与编码的理由；（c）对潜在调节或中介的结果进行选择与编码的理由； 3.5 对测量方法和其它变量在心理测量特征上的描述。

目的	4	陈述如何对假设进行检验，包括： 4.1 被试特征（包括动物）的问题，自变量（实验操作，处理或干预）相关的问题，排除可能产生混淆的变量，因变量（相关的结果或标准）及其研究设计的其它特征； 4.2 在元分析中使用的整合方法（例如，效应量测量，平均数分析的方法，同质性分析中使用的模型）。

方法

方案预注册	5	陈述是否提前对元分析研究方案进行预注册。如果预注册过，则需要提供相关信息（例如，公开预注册的网址）。
文献纳入标准	6	陈述文献纳入的标准： 6.1 自变量（例如，实验的操作，处理、干预或预测变量的类型）； 6.2 因变量（例如，对临床研究中存在的潜在积极与消极的效应进行综合后的结果）； 6.3 研究设计的标准（例如，样本抽样的方法或实验处理的方式）； 6.4 针对同一研究或同一样本存在多次报告或使用同一样本进行多次测量的情况，不仅需要陈述纳入标准的研究特征（例如，自变量、因变量和研究设计等），同时需要陈述报告特征（检索时间、语种、发表日期和发表状态等）； 6.5 研究纳入中存在的限制（例如，研究的年代、语言、地区或进行报告的方式）； 6.6 预注册时提交的文献纳入标准是否在研究开始后发生改变，或者研究过程中文献纳入标准是否发生改变，例如增加或减少纳入研究的变量，作者需要报告发生了什么改变，以及这些改变是什么时间发生，原因是什么； 6.7 如何处理文献中缺失信息，包括与纳入标准的信息相关信息的缺乏（例如，缺乏关于研究设计的信息）及与数据分析相关信息的缺失（例如，没有报告关于结果的量化数据）。

信息来源	7	描述搜索得到所有信息来源（例如，包括数据库中获取的文献信息，联系作者后获取的额外信息）和最后一次搜索日期。
搜索过程	8	至少完整呈现一个主要使用的数据库在线数据搜索策略，包括限定的范围，搜索时使用的关键词、搜索代码等，以确保搜索的可重复性。
研究筛选	9	陈述每一步研究筛选过程，包括： 9.1 每一步筛选文献的决策方式（如，通过阅读题目、摘要或全文）； 9.2 筛选过程中执行者的胜任力与人数。如果是一位执行者，如何评估提取数据过程的信度问题；如果有多位执行者，如何处理多人进行筛选可能存在的不一致问题。
数据收集过程	10	描述从报告中提取数据的方法，包括： 10.1 作为摘录数据依据的变量及其种类； 10.2 数据提取过程中执行者的胜任力与人数。如果是一位执行者，如何评估提取数据过程的信度问题；如果有多位执行者，如何处理多人提取可能存在的不一致问题。
数据条目	11	列出并定义纳入文献中所有与元分析相关信息（例如，文献的作者、发表年份、被试信息、研究设计或控制组信息等）。
单个研究偏倚*	12	描述如何对纳入研究的偏倚风险进行评估的方法（说明这种评估是在研究方法过程层面还是结果层面，或者两者都有），以及研究偏倚信息如何正在结果综合进行处理。
描述性统计指标	13	描述使用的效应量指标（比如，相关系数、Cohen’s d），包括用来计算单个研究效应量公式。如果同一研究存在多个效应量导致的非独立性问题（例如，同一研究中存在多个实验处理、多个控制组、多个相关系数或多个指标量等产生的非独立效应量），陈述处理的方法。

结果综合	14	描述进行跨研究效应量综合和对不同研究的效应进行异质性评估的方法（如固定效应模型或随机效应模型的稳健方差估计（robust variance estimation）），具体包括： 14.1 选取某一综合方法的原因； 14.2 对研究结果进行加权的方法； 14.3 估计单个研究与总体结果变异的方法（如，置信区间）； 14.4 描述所有的转换/校正（例如，小样本或非等组样本数据）、调整（例如，聚合，数据缺失，或建构不同水平间的关系）以及对这些进行的数据验证）； 14.5 如果进行了贝叶斯分析，报告先验分布的选择与模型拟合的评估。
跨研究偏倚	15	评估所有可能会影响累积证据的偏倚风险（例如，发表偏倚，选择性报告）。
额外分析	16	如果进行了额外分析，需要描述使用的方法（例如，亚群体分析，敏感性分析，元回归分析），指出哪些是预先计划的。
数据可获得性*	17	元分析中使用的数据（即从文献中提取的数据条目）是否公开。元分析的原始数据存放在一个公开并可长期存放的在线数据库中，方便其他研究者进行重复与在此研究基础上的累积。
分析过程可获得性*	18	元分析中使用的软件名称及版本号。元分析的代码是否公开，并报告在哪可以获得（例如：与数据存在同一在线数据库），方便其他研究者对分析过程的重复。

结果

研究筛选	19	给出筛选后得到的研究数量，给出每一步排除的原因，建议用流程图呈现。
研究特征	20	陈述纳入研究的特征，对每一个纳入的研究变量原则上包括： 20.1 自变量、因变量与主要调节变量的特征；

		20.2 重要的被试特征（年龄，性别，受教育程度等）； 20.3 重要的情境变量（例如，环境，日期）； 20.4 研究设计（例如，样本抽样的方法或实验处理的方式） 20.5 在什么地方可以获得该研究的全部数据特征（例如，与作者联系，在补充材料中或预注册网站中）。
单个研究偏倚*	21	尽可能提供每个研究存在偏倚风险的数据，以及质量评估的报告。（见#12）。
单个研究的结果	22	呈现每一个研究综合后的结果（包括阳性与阴性结果），推荐使用图表信息进行呈现（例如，森林图）。
结果综合	23	报告每个研究整合的结果（例如，效应量大小，置信区间与异质性程度）。
跨研究偏倚风险	24	描述跨研究偏倚风险，包括： 24.1 陈述是否是（a）未出版的研究与未报告的数据；（b）仅将报告的数据进行整合，如果是，陈述其合理性； 24.2 评估发表偏倚的影响大小（例如，数据检验模型于剪补法（trim-and-fill）等）； （3）是否存在选择性报告。
额外分析	25	如果做了额外分析，给出结果（例如，亚群体分析，敏感性分析，元回归分析[见#16]）。

讨论

证据总结	26	陈述主要发现，包括： 26.1 研究的主要结果，包括元分析的结果； 26.2 研究证据的总体质量； 26.3 结果存在的其它解释（例如，混淆变量（confounding），统计检验力）； 26.4 与先前研究进行整合的结果类似和不同之处。

局限性	27	研究中存在的局限（例如，不一致、不准确与偏倚风险）。
研究推广性与意义	28	对结论的推广性（外部效度）进行描述，并指出对未来研究的意义。
作者贡献与致谢		
作者贡献	29	详细说明各个作者在元分析之中的贡献。
致谢	30	关于对该提供帮助的个人、机构以及资助来源。

2.1 方法部分条目

以下为针对一些元分析方法部分具体条目的介绍。

条目 5. 方案预注册。研究者需要对整个研究进行详细计划，并确保整个计划在元分析实施之前就已经记录在案。报告标准要求作者报告研究是否存在方案如果存在，需要提供相应的预注册信息(Moher, Liberati, Tetzlaff, &Altman, 2009)。

方案预注册的主要目的在于避免研究者根据自己的需要选择性报告结果，避免摘取与使用数据过程中的随意决策对结果带来的偏倚。在医学元分析方面使用较为广泛的是 YORK 大学建立的 PROSPERO（International Prospective Register of Ongoing Systematic Reviews. www.crd.york.ac.uk/PROSPERO）系统综述国际预注册平台，它为研究者们发表方案和信息共享提供了极大的便利 (Booth et al., 2012)。随着方案预注册的便利性的提升，研究方案的透明性、准确性与完整性也变得极为重要，也出现了专门为预注册而定制的报告标准 PRISMA-P（Preferred Reporting Items for Systematic Reviews and Meta-Analyses protocols）(Moher et al., 2015)。PRISMA-P 由 17 个条目组成，用于指导作者如何对元分析/系统综述方案进行预注册。截止 2018 年 8 月，已有 42 本杂志可提供元分析预注册的报告，大部分杂志来自生物行为科学领域(Quintana, 2018)。在心理学研究中，目前研究者可以在 Open Science Frameworks (OSF , osf.io)上进行预注册。遗憾的是目前 OSF 尚未对元分析提供相应的模板。

条目 6. 文献纳入标准。元分析预先设定的纳入与排除标准是区别于叙述性综述的一个重要特点。纳入标准的清晰程度影响元分析的效度、可重复性以及最终的结论。元分析报告中对纳入标准的描述不充足会严重影响对稿件质量的评分(Fleming, Koletsi, Seehra, & Pandis, 2014)。PRISMA 基于临床干预研究的元分析，推荐作者按照 PICOS (participants, interventions, comparisons, outcomes, and study design chosen) (O'Connor, Green, & Higgins, 2008)的格式进行报告。在心理学研究的元分析中，作者需要在纳入标准中定义并至少报告如下关键内容：被试群体（年龄范围、社会经济地位的特征、精神健康状态

等）、自变量（研究者操纵的变量）、因变量（即主要的测量指标，比如正确率、反应时间、生理指标、激活脑区等）与研究设计（如抽样方式、实验设计、横向对照研究和干预训练研究等）。作者需要清楚地陈述纳入标准的以上特征与原因，以便读者能够对该报告的偏倚进行判断(Dwan, Gamble, Williamson, & Kirkham, 2013)。由于研究可能存在多次报告，一次报告中也可能存在多个研究因此文献纳入标准不仅需要陈述上述研究特征（例如，被试群体、自变量、因变量和研究设计等），还需要陈述报告特征（例如，时间、语种、发表日期和发表状态等）。

此外，元分析作者需要报告在纳入过程中可能存在的限制（例如，语言、区域或者报告类型）(Appelbaum et al., 2018)。

条目 7. 信息来源。 进行元分析时，为保证文献的全面性，通常需要通过搜索电子数据库。但任何一个电子数据库均无法涵盖所有的信息来源，因此作者还需从不同数据库与其它途径收集信息。来源细节的缺失容易导致其他人难以重复检索过程，因此作者需要对信息来源进行规范化报告。PRISMA 建议在数据库中获取信息至少需要报告数据库名称、平台或提供者，以及每次搜索开始与结束的日期 (Liberati et al., 2009)。报告搜索中最早与最晚的记录可以避免数据库更新后造成读者的错误解读。除了数据库的搜索之外，作者还应该采用其他方法对相关研究进行收集，比如私人通信等 (Liberati et al., 2009)。对信息来源的充分报告为读者提供了衡量文献覆盖程度的重要参考。

条目 8. 搜索过程。除了条目 7 中描述信息来源外，作者还需报告完整和全面的数据库搜索策略，以保证其他研究者能够重复该搜索过程。PRISMA 建议作者至少明确报告一个主要使用的数据库的搜索策略，包括布尔运算结构(boolean connectors)、关键词、自动扩展[3] (automatic explosion) (Appelbaum et al., 2018)。搜索实施中可能存在的限制同样需要报告，比如时间、资金支持与语言等原因导致某些研究的信息无法获取等。作者应该清楚、直接地描述这些限制，以便读者衡量报告的质量。

条目 9. 研究筛选。文献搜索完成后，作者需要从大量搜索到的信息中，根据研究文献纳入标准进行筛选。针对这一流程，PRISMA 与 MARS 均建议作者报告筛选过程中的每一步纳入与排除的决策依据（例如，通过阅读标题、摘要或全文，查看全文的频率如何，因为什么关键信息排除了某项研究）。PRISMA 标准的优势在于使用流程图（http://www.prisma-statement.org/）对文献筛选过程进行呈现，流程图清晰地将确认，筛选，合格性，纳入四个关键步骤进行呈现 (Moher, Liberati, Tetzlaff, & Altman, 2009) (具体流程图请见补充材料：https://osf.io/9vm34/)。通过筛选流程图在结果部分的呈现，读者对每一步纳入的文献数量，以及被排除的文献的原因能够一目了然。

此外，两位及以上研究者同时筛选能有效降低排除相关文献的概率(Edwards et al., 2002)，也有助于解决一些难以明确判断是否符合纳入标准的文献 (Cooper, & Ribble, 1989)。为提升报告的客观性以及在筛选过程中避免错误，作者应该报告每一阶段的执行者人数，以及他们的具体工作。如果是多

[3] 自动拓展指的是在数据库进行搜索时，为了保证文献的全面性，数据库自动对相关关键词进行拓展。例如在 Pubmed 中搜索 MRI 时，其会自动拓展为 Magnetic Resonance Imaging 等多个关键词进行搜索。

人共同完成，需要报告在筛选过程中出现的争议内容以及争议解决的方法。

条目 10. 数据收集过程。当作者完成对文献的筛选之后，接下来就需要从这些合格的研究中摘取与元分析研究问题相关的信息，以进行下一步的元分析或者讨论。同样，两位作者独立提取数据所犯的差错少于一位作者提取数据后由另外一位作者核实 (Buscemi, Hartling, Vandermeer, Tjosvold, & Klassen, 2006)。为了缩小误差与减少潜在的偏倚，建议由两位及以上的研究者独立进行数据收集并进行报告。

在数据收集的过程中，作者应描述为减少偏倚和错误的措施。例如，作者可能因为纳入研究的信息不清晰或者不完整而联系原始作者，那么作者应该描述如何联系原研究作者、询问了什么问题、获得了哪些必要的信息、哪些数据以及最终哪些信息被包括在元分析之中。

此外，一些数据存在多次出版的情况，比如在样本量、干预的对象或分析结果上存在微小变化而重复出版。这种情况很难被作者发现，且容易引起偏倚 (Von, Poglia, Walder, & Tramèr, 2004)。Liberati 等人（2009）建议作者将分散在多个报告的同一研究数据进行汇总，并对如何避免数据重复计算实施的措施进行报告（例如，将不同报告作者的姓名进行比较，对比干预措施之间的差异，查看样本量大小以及结果指标是否存在不同）。同时，在可能的情况下，可考虑与原作者联系进行确认。

条目 11. 数据条目。作者需要尽可能地列出纳入研究中收集到的所有变量数据（例如，作者、发表年份、被试信息、研究设计或控制组信息），并对变量进行清晰的定义。尽管有些数据不会在元分析中使用，但这些信息有助于读者了解作者纳入文献的所有信息。如果作者仅报告对他在元分析中使用的变量信息进行报告，而忽略那些他认为重要但又未能获取的变量信息，可能导致偏倚并误导读者 (Glasziou, Meats, Heneghan, & Shepperd, 2008)。

具体而言，作者在元分析实施之前应该对将要提取的变量信息在研究方案中进行陈述。清晰地说明这些信息可以有效避免提取的信息与研究预期不符的情况。此外，在预注册的研究方案中对变量/数据信息进行描述，有助于读者比较预注册方案与最后报告的研究在这些信息上的差异，从而对元分析的质量进行评估。如果作者在研究开始实施后，变量发生了变化，比如增加或剔除某些变量应该陈述这些变量变化的原因 (Liberati et al., 2009)。

条目 12. 单个研究偏倚。偏倚是研究结果或统计推断中的一种系统误差，或与真实值的偏倚 (Higgins, 2008)。如果纳入的研究在方法上存在高度偏倚，元分析的结论极有可能具有高度误导性。PRISMA 声明方法部分条目 12 与 15 均涉及偏倚，前者为单个研究的偏倚（包括该方法用于研究层面或结果层面），后者为研究间影响数据合并结果可能存在的偏倚（如发表偏倚）。

单个研究的偏倚评估主要指作者对纳入研究自身效度的衡量，Gates March (2016)认为作者对纳入研究进行评估时，需要关注三方面问题：（1）研究的结果在多大程度上是可靠的，或者这些研究结果值得信赖吗？（2）该研究包含什么类型的偏倚？（3）存在的偏倚如何对元分析的整体质量产生影响？在心理学研究中的偏倚可能出现在样本选取（主要是本科生样本 (Henrich, Heine, &

Norenzayan, 2010))、实施过程(是否有注意力检测(Oppenheimer, Meyvis, & Davidenko, 2009)、是否有操纵检验(manipulation check)、统计测量(统计方法是否得当，是否有进行多重比较校正)或写作报告过程之中。

值得注意的是，对单个研究的偏倚的评估，需要研究者对每个研究本身的质量进行评估。对于未经同行评议的文章来说，元分析作者有必要对其进行偏倚评估。但对于经过同行评议并在学术期刊上发表的文章，作者可以根据实际情况选择是否对每个研究进行偏倚评估。有时，元分析包含的文献数量较大时，对单个研究的偏倚评估可能变得过于耗时耗力。因此，我们将其作为拔高标准，即鼓励作者进行报告，但并非基本要求。

条目 13. 描述性统计指标。元分析需要报告使用的效应量指标是什么(例如，相关系数、Cohen’s *d*、风险比(risk ratio)或者神经成像中的激活坐标)。在行为研究中，也要说明效应量的转换/计算公式。与此类似，对神经成像的研究进行基于坐标的元分析时，需要报告不同坐标系统转换的标准。

条目 14. 结果综合。为了比较纳入研究的数据，除了报告效应量指标之外，还需要对结果进行综合。综合前需要报告同质性检验的结果，是否存在异质性，选择了何种模型(例如，固定效应模型或随机效应模型)进行综合，以及为什么选择该模型。

通常在元分析开始前需要报告预先设定好使用的描述性统计指标，并在研究方案中对结果综合部分进行报告(详见 PRISMA 声明中条目 5)，目的在于让读者可以理解结果的含义，并了解元分析实施中与最初计划发生了多大程度的偏离。

面对同一研究由于各种原因出现多个效应量的情况，经验不足的研究者常忽视元分析进行效应量综合时必须相互独立的原则，使用传统方法或将多个效应量进行简单平均后再处理均违背元分析效应量独立性的原则。针对该状况，作者需要报告处理非独立性问题采取的方法，例如多元元分析法(multivariate meta-analysis)、多元线性模型法和多元整合法等(叶子菁, & 石伟, 2010)。

条目 15. 跨研究偏倚。跨研究偏倚主要来自纳入研究的信息不完整，比如发表偏倚(Publication bias)。发表偏倚是指由研究结果自身特性及其指向影响的已发表或未发表的研究结论(Higgins, 2008)。通常，$p < 0.05$ 的结果比 $p > 0.05$ 阴性结果的研究更容易或更快地获得发表，这种“选择性”导致某些研究的发表受到压制，使得研究的发表过程不再是一个随机事件。除了未发表的研究导致信息不完整外，元分析作者纳入研究过程的信息缺失也可能导致结果偏倚，这些问题均属于发表偏倚。

Song, Eastwood, Gilbody Duley (1999)对医学领域元分析中存在的发表偏倚进行评估，认为约 50%的元分析中存在不同程度的发表偏倚(20%非常明显)，10%的元分析结果受到发表偏倚的影响。漏斗图通常是识别与检验发表偏倚的有效方法之一，它提供一个可视化的方法来确定偏倚的存在和来源，为后续元分析在方法上的选择提供一个当的理论依据(Sterne et al., 2011)。另外，Egger 线性回归检验、剪补法(trim-and-fill)与失安全系数(Rosenthal’s fail - safe number, N_{fs})也常用来识别与检验发表偏倚。值得注意的是，如果

文献中确实存在发表偏倚，采用统计方法进行校正并不能完全矫正偏倚。

如果在元分析实施前的研究方案中明确存在关于偏倚风险的评估，但却未真正地实施，作者需要对这一变化的原因做出说明，便于读者衡量与评估元分析的报告质量与效度(Liberati et al., 2009)。

条目 16. 额外分析。如果存在额外分析方法，比如敏感性分析或分组分析（subgroup analyses），在引言部分应该对该方法使用的合理性进行介绍，同时在方法部分对该方法的实施进行描述。

条目 17. 数据可获得性。数据是指作者在元分析中用于描述、阐释与统计分析的研究信息，包括来自条目 11 和 13 中收集到的信息。元分析作者需要报告在元分析中所用到的数据，是否能够获得，以及在什么地方获得。根据近年来国际心理学界对实验研究中透明与开放性的要求，作者需要公开其在研究中所用到的数据，如果不公开数据，需要对理由进行说明。

元分析一般是对已经发表的研究进行的分析，因此其数据一般不涉及伦理或者保密的问题，因此公开数据可能伦理上并无太大的阻碍。将元分析的数据公开，有利于研究者自己及其他研究者重复该元分析，同时有利于一段时间后继续在此基础上进行元分析，因此公开元分析数据对于科学知识的积累有极大的推动作用。但考虑到公开数据近年才逐渐受到研究者的关注，因时，本文暂时将其作为拔高标准，但学术期刊可以根据实际情况，决定是否要求作者将数据存放在可长期保存和容易获取的网络服务器中，并在文章中对数据的可获得性进行说明。

条目 18. 分析过程的可获得性。分析过程指的是对元分析中所有的步骤。包括从文献中提取数据的编码手册、将提取数据进行转化的过程（如 Excel 模板、R 代码等）、元分析的代码等。这些过程与代码对于重复元分析以及对该元分析的质量评估至关重要。同样，考虑到公开分析过程也是近年来出现的新趋势，因此也暂时将其作为拔高标准。学术期刊可以鼓励作者记录元分析的全过程，并将相应的脚本代码与数据一起存放于网络服务器中，并在文章中进行公开。

3. 国内元分析文献报告现状

为了解国内心理学元分析报告的现状，我们根据以上的新标准，对去年，即 2017 年[4]国内心理学期刊上发表的元分析进行评估。

3.1 研究方法

文献纳入标准 根据研究目的，本文纳入标准为：（1）元分析类研究，即有量化的综合结果的文献；（2）发表在中文主流心理学期刊上，排除非心理学期刊上的元分析文献；（3）选取完整已出版的文献，排除增刊、未公开发行与会议摘要文献；（4）发表于 2017 年。本纳入标准存在诸多限制：年份仅包括 2017 年，语言仅包括中文，文献来源仅包括期刊。但是本纳入标准主要与本报

[4] 仅选择 2017 年一年内心理学期刊上发表的元分析，主要原因是：（1）2017 年的元分析是较新的研究，能够代表国内元分析文章报告的最新水平；（2）工作量在可控范围内，由于本文并非要对国内所有的元分析的报告质量进行全面评估，而在于介绍元分析报告的标准，因此希望能够选择具代表性的元分析报告进行评估。

告的目的有关，见注释 4。

信息来源 所有文献来自中国知网数据库（cnki.net）和《心理学报》、《心理科学》、《心理科学进展》、《心理发展与教育》、《心理学探新》、《心理与行为研究》、《中国临床心理学杂志》七本期刊官方网站数据，文献覆盖时间为 2017 年 1 月 1 日至 2018 年 1 月 1 日，初次检索时间为 2018 年 4 月 12 日，最后一次检索时间为 2018 年 8 月 15 日。

搜索过程 以“系统综述”、“Meta 分析”、“元分析”、“Meta-analysis”、“荟萃分析”作为关键词进行检索[5]，共检索到 16 篇文献。整个检索过程由作者一独立完成。作者二同样以“系统综述”、“系统综述”、“Meta 分析”、“元分析”、“Meta-analysis”与“荟萃分析”作为关键词对七本期刊进行独立的手工检索，最终与作者一进行核对。文献搜索过程符合本研究的目的无纳入标准之外的限制。

研究筛选 通过文献题目与摘要，由作者一与作者二独立根据文献合格标准进行筛选。遇到分歧时与通讯作者进行讨论并处理分歧。

数据收集过程与编码 根据 PRIOR-MA，我们首先建立一份编码指南供评分参考（见补充材料，https://osf.io/wq26t/），同时也建立数据收集表格，表格中需要纳入文献中研究的基本特征，如题目、作者、发表时间与期刊来源等，以及需要进行编码并评估的内容。数据收集由作者一与作者二独立完成，并对收集内容一致性进行讨论与评估，无法统一时与通讯作者进行进一步讨论并做出决策。

编码指南由 PRIOR-MA 方法与结果部分条目 5-25 组成，每一条目需要被编码为“1”，“0”，“0.5”或“？”。“1”表示该条目得到完全报告。编码为“0”表示该条目完全未得到报告。“0.5”表示该条目未得到完全报告。“？”表示该条目在编码过程中很难判断，编码为不清楚的条目需要由作者一与作者二进行讨论并做决策，如无法决定时，继续与通讯作者进行讨论并最终决定。

编码过程最初由作者一按照编码指南对 16 篇元分析文献进行独立编码，初次编码后与作者二、通讯作者进行讨论，形成编码指南（见附录）。然后由作者一与作者二按照编码指南独立重新编码，两者编码完成后计算评价者一致性 Kappa 值(Mchugh, 2012)（计算过程见 https://osf.io/9vm34/）。Mchugh (2012)认为 Kappa 作为一致性指标，相较于百分比，更好地将偶然存在的一致性情况加以考虑，避免对一致性进行高估。根据先前的文献，我们对于 Kappa 值的判断标准为：0.40~0.59 之间为一致性好，在 0.60~0.74 之间为相当好，0.75 及以上为一致性非常好(Orwin, 1994)。

数据条目 文献信息摘取条目包括：（1）文献题目；（2）文献发表时间；（3）文献发表期刊；（4）文献作者；（5）文中涉及 PRIOR-MA 方法与结果部分的条目（条目 5-条目 25）。

[5] 具体的检索代码为：文献来源=('心理学报'或'心理科学'或'心理科学进展'或'心理发展与教育'或'心理学探新'或'心理与行为研究'或'中国临床心理学杂志') 和题名=('系统评价'或'系统综述'或'Meta 分析'或'元分析'或'Meta-analysis'或'荟萃分析')

描述性统计指标　按照 PRIOR-MA 标准，本文对 16 篇文献研究方法与结果部分，共 21 个条目进行评估。通过每篇文献总分评估其在 PRIOR-MA 标准中的完成情况。另外，为更好地了解目前心理学领域元分析报告质量，我们使用百分比来评估单个条目在 16 篇元分析中完成情况的百分比。

结果综合　本文仅进行系统综述描述性统计，无元分析过程，未进行结果综合。

单个研究偏倚与跨研究偏倚　本文主要目的是按照 PRIOR-MA 标准，了解国内元分析的完成状况，不以单个研究自身的研究目的为评估标准，主要存在的偏倚来自发表偏倚。

数据可获得性与分析过程可获得性 本研究的原始数据与分析过程均上传至 Open Science Frameworks 进行保存并公开，链接地址为：osf.io/dt84s。

3.2 结果

3.2.1 研究筛选与研究特征

共有 16 篇元分析纳入评估，经两位作者进行独立搜索与筛选，结果一致，无重复或需要排除的文献，16 篇文献全部纳入评估。筛选流程图见补充材料 3（https://osf.io/wq26t/）.

16 篇元分析均为中文，并发表于国内心理学杂志中，样本特征见补充材料（https://osf.io/9vm34/）。

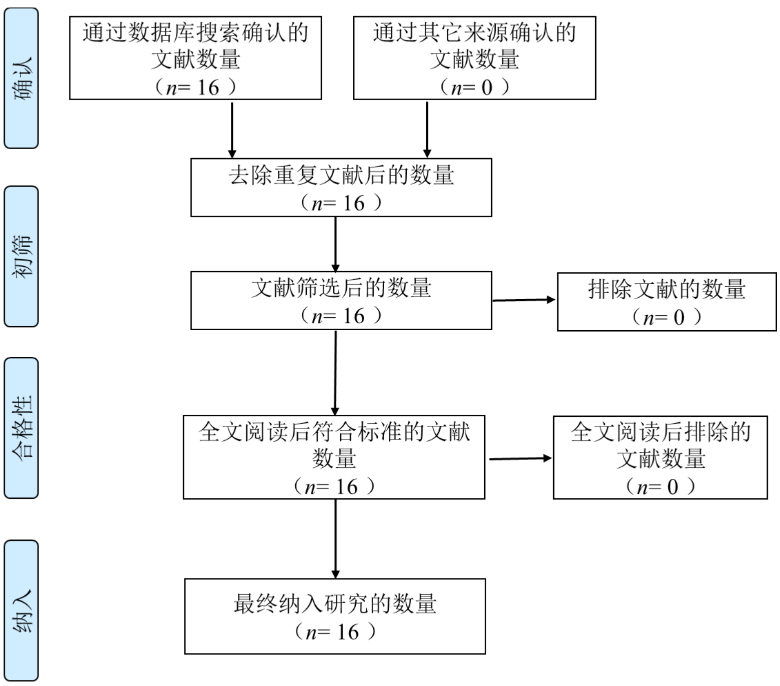

图 2　元分析文献筛选流程图

3.2.2 评分一致性

对作者一与作者二第二次独立编码的分数计算 Kappa 值，结果为 0.873，表明两名评分者的一致性达到了较高的水平。

3.2.3 总体报告质量

基于 PRIOR-MA 标准的元分析质量评估，两位编码者针对编码不同的条目进行再次讨论，最终统一编码得分。本文报告的结果为最终统一编码得分。

本次评估包括 21 个条目，总分为 21 分。16 篇元分析中，8 篇的研究方法为实验研究，其余 8 篇为问卷。所有元分析平均得分为 8.09 分，最高分为 12.5 分最低分为 5 分。每个条目完成情况见表二。纳入研究的元分析在条目 5、12、17、18 和 21 中平均分为 0 分，即所有元分析研究均未提及相关报告。条目 13 与条目 14 平均分分别为 0.84 与 0.78，表明国内研究在研究方法中的综合测量指标与结果综合部分完成情况较好。大部分条目平均分在 0.5 分左右，完成情况并不理想。

表 2 基于 PRIOR-MA 的 2017 年中国心理学期刊元分析报告质量（n=16）

条目	章节/主题	平均分	n（%）		
			0	0.5	1
方法					
5	方案预注册	0	16（100）	0（0）	0（0）
6	文献纳入标准	0.53	0（0）	15（93.8）	1（6.3）
7	信息来源	0.53	0（0）	15（93.8）	1（6.3）
8	搜索过程	0.53	0（0）	15（93.8）	1（6.3）
9	研究筛选	0.5	2（12.5）	12（75.0）	2（12.5）
10	数据收集过程	0.16	13（81.3）	1（6.3）	2（12.5）
11	数据条目	0.50	1（6.3）	14（87.5）	1（6.3）
12	单个研究偏倚	0	16（100）	0（0）	0（0）
13	描述性统计指标	0.84	2（12.5）	1（6.3）	13（81.3）

14	结果综合	0.78	3（18.8）	1（6.3）	12（75.0）
15	跨研究偏倚	0.59	6（37.5）	1（6.3）	9（56.3）
16	额外分析	0.59	6（37.5）	1（6.3）	9（56.3）
17	数据可获得性	0	0（0）	0（0）	16（100）
18	分析过程可获得性	0	0（0）	0（0）	16（100）
结果					
19	研究筛选	0.06	14（87.5）	2（12.5）	0（0）
20	研究特征	0.59	5（31.3）	3（18.8）	8（50.0）
21	单个研究的偏倚风险	0	16（100）	0（0）	0（0）
22	单个研究的结果	0.13	13（81.3）	3（12.5）	0（6.3）
23	结果综合	0.66	4（25.0）	3（18.8）	9（56.3）
24	跨研究的偏倚风险	0.56	7（43.8）	0（0）	9（56.3）
25	额外分析	0.53	7（43.8）	1（6.3）	8（50.0）

4. 讨论

使用元分析报告标准，有助于提升心理学领域的元分析质量，但目前国内对相关标准仍然缺乏了解。本文结合 PRISMA、MARS 与 TOP 标准，提出开放与透明的元分析报告标准 PRIOR-MA。如前所述，PRIOR-MA 是在 PRISMA 的基础上，结合心理学的特点而修订的，因此从某种程度上，其能够满足一个好的报告标准中的两点原则：推荐价值和指导作用。

以 PRIOR-MA 作为标准来分析了国内心理学期刊 2017 年发表的 16 篇元分析文献，结果表明，PRIOR-MA 在评估元分析报告标准时，有较好的评估者一致性。

同时，本次评估的结果也表明，根据 PRIOR-MA 能够探测出元分析文献在报告时不足和可以改进之处。具体如下。

第一，方案预注册信息报告缺乏。16 篇文献均未提及研究方案的预注册及其相关信息，表明国内研究者尚未意识到方案预注册的重要性。元分析方案可以保证元分析在开始之前是经过详细考虑与计划，将方案的详细计划公开记录在案，促进作者在研究实施中思想与行为的一致性，保证不同作者间责任明确，研究诚信以及公开透明。它也有助于降低作者由于从原始研究中有选择性地摘取数据与决策中的任意性所带来的偏倚(Moher et al., 2015)，也为比较预注册方案与报告之间的差异从而评估潜在问题提供了可能性。

第二，文献纳入标准、信息来源与搜索过程不够规范。虽然 16 篇元分析均对文献纳入标准、信息来源与搜索过程进行报告，但几乎均未达到 PRIOR-MA 标准，主要缺乏研究报告特征的陈述。在 PRIOR-MA 中，文献纳入标准条目指出，元分析需要报告纳入研究的研究特征与报告特征。研究特征在心理学中应包括纳入研究的被试群体，自变量与因变量等特征。而本文所评估的 16 篇元分析中很少提及纳入研究的报告特征，例如研究使用的语言与发表状态（例如，未公开出版或正在出版）。许多研究表明，使用不同语言的研究纳入元分析后会产生截然不同的结果(Moher et al., 2015)，即使作者纳入的研究语言仅为英语或中文，也有必要报告是否存在纳入其它语言的研究的可能性，而非默认为某一种语言。

在明确文献纳入标准后，作者需要根据标准收集相关信息。这一阶段涉及 PRIOR-MA 条目 7 信息来源与条目 8 搜索过程的报告。我们发现 16 篇文献均未报告最后一次数据库搜索更新文献的日期。报告搜索每个数据库开始与结束的时间是非常重要的，因为通过不同的时间标签搜索到不同的文献，会直接影响到元分析最终的结果(Shojania et al., 2007)。

在搜索过程报告部分，目前心理学领域的元分析主要对搜索使用的关键词与数据库进行报告，但许多文章并未报告搜索执行中可能存在的限制。事实情况是，几乎所有的搜索均存在限制，例如在时间、资金支持、因语言或搜索技巧而无法获取的信息(Liberati et al., 2009)，报告搜索中的限制一方面让读者能够衡量当前元分析结果的可靠性，另一方面也能让作者对可能存在的偏倚提高警惕。

第三，研究筛选和数据收集过程不够透明。本次评估结果表明，国内元分析在研究筛选上主要存在两个问题：一是筛选的过程不清晰，少数研究使用 PRIOR-MA 推荐的筛选阶段流程图来描述每一步纳入与排除研究的过程，更多研究仅直接呈现最终纳入研究的数量，读者并不清楚这些研究是如何进行筛选的；二是筛选多数研究未报告有几位筛选者，筛选者是谁以及如何处理筛选一致性问题，或者如何避免一位筛选者可能存在的错误。一位研究者进行筛选极有可能由于过于主观或者粗心而出现错误，同时也有研究表明，文献筛选过程也是一个极其困难的决策过程，一位研究者很难独立胜任(Cooper, & Ribble, 1989)。

在数据收集上，很少有研究明确对数据收集的过程（平均分为 0.16）进行报告。如上所述，数据收集过程的清晰报告可以有效减少错误与偏倚(Jones, Remmington, Williamson, Ashby, & Smyth, 2005)。该条目需要报告由哪位作

者进行数据摘取，如何解决数据收集不一致的问题(Liberati et al., 2009)。该条目也需要作者详细报告目标数据获取过程，比如通过联系原文作者获取文献中未知的信息（16 篇文献中仅有 1 篇提及），如何避免重复出版研究的多次计算等问题（16 篇文献中均未提及）。

第四，偏倚报告不够完整。在 PRIOR-MA 条目中，元分析存在的偏倚主要分为单个研究偏倚与跨研究偏倚。我们的评估发现，大部分元分析关注的焦点在于跨研究的偏倚。16 篇文献中有 9 篇完整报告发表偏倚。但是很少有研究涉及单个研究出现的偏倚，也就是说，作者并未报告纳入元分析的研究自身质量如何，衡量这些研究存在多大程度的偏倚。考虑到大部分的元分析是对已经发表的文献进行的分析，因此可以假定这些元分析中的原研究已经从某种程度上经过了同行评议的审核，无重大的偏差。但值得的注意的是，由于心理学近年来的可重复危机（胡传鹏 等，2016），研究者可能仍然需要从研究的内部效度、外部效度、构念效度与统计结论效度四个方面对原研究进行质量评估(Valentine, 2009)。

第五，研究方法与研究结果的表述存在混淆。我们发现，作者在写作时常将结果部分内容放入研究方法部分进行报告，最常出现的是将研究筛选结果（PRIOR-MA 中条目 19）、研究特征（PRIOR-MA 中条目 20）等内容放入研究方法部分。这样的报告方式可能让读者无法清晰地获取相应的信息，容易造成遗漏与误解。这也正说明通过清晰的报告标准，将有助于作者理清思路，方便写作。如果作者根据本文推荐的标准进行论文写作，则可能避免混淆研究筛选（条目 19）的结果与方法（条目 9）部分的内容。

第六，研究数据与分析过程可获得性较低。研究过程与结果的开放与透明程度在心理学的可重复危机后受到众多研究者的推荐，数据的可获得性也得到了许多学术期刊在政策上的支持（如 *Science*、*Nature* 和 *Psychological Science*）。此外，元分析研究数据与分析过程的可获得性也受到研究者的重视，例如 NeuroVault (Gorgolewski et al., 2016) 和 Archive of Neuroimaging Meta-analyses (ANIMA) (Reid et al., 2016) 鼓励研究者将元分析的结果上传。而 curatescience.org 则在线公开近年来许多存在重大争议的研究和重复实验的数据与元分析的结果。因此，我们根据在元分析报告标准中也加入了数据与分析过程可获得性的标准。但是考虑到该标准也是近年来才出现的标准，我们将这两个标准作为拔高标准。在我们考查的 16 篇元分析中均未提供数据或分析过程的可获得性也是意料之中。

总之，本次评估的结果表明，PRIOR-MA 对元分析进行评估时，评分者一致性较高，说明该标准使用起来相对容易，能够为广大的研究者所接受。同时，这个评估的结果也表明，即使是较新近发表的元分析报告（2017 年），仍然存在不足，而这些不足可能会影响元分析的质量。这意味着，如果国内心理学期刊采纳元分析报告标准，将有助于提升元分析报告的完整性和规范性，因此该标准具有推荐价值和指导作用。

5. 总结与建议

科学研究是不断积累的过程，元分析正体现了这种积累过程。在元分析中使用合理的报告标准，有助于研究者得到科学的结论，从而促进科学知识的积累。

然而在心理学领域，元分析方法的严谨性以及报告的标准化似乎仍然被忽视。即使是高影响力的心理学期刊如*Psychological Bulletin*，其发表的元分析文章中，统计错误也相当普遍(Cafri, Kromrey, & Brannick, 2010)，遵循元分析报告标准的文章比例也极少。因此，我们呼吁广大同行遵循严谨的元分析报告标准，以促进心理学知识的科学有效积累。

对于期刊而言，要提高元分析的质量，可以从审稿政策上加强对元分析文章报告的要求[6]。具体而言，期刊可以通过如下几步来提高其所发表的元分析报告的质量：第一、学术期刊可以需要根据实际情况，以本文提出的报告标准为基础，定制元分析的作者自检标准。在这个自检报告中，可将某些条目作为必要的要求，而对另一些条目是鼓励。例如，文献的纳入标准、结果综合的统计方法等应该严格要求作者进行清楚地报告，以保证元分析的质量；而对元分析中单个研究偏倚的评估、数据和分析过程可获得性，则可以作为鼓励性的拔高标准。第二、期刊在对元分析稿件的初步筛选时，可参照元分析报告标准，对关键条目进行核对，检查文章中是否包括了这些关键条目，以检查是该元分析是否满足元分析报告标准的基本要求，决定是否将该稿件送审。第三、期刊可建议审稿人参阅元分析报告标准，对文献进行审稿，确保各部分报告的合理性与准确性。或者可以参考*Psychological Science*的做法，期刊邀请一些专家组成方法审稿人小组，专门对研究方法部分进行审核。在这些方法审稿人中，包括元分析方法的专家，负责对元分析报告是否合理进行评估。

对于元分析作者而言，在从元分析开始计划到文章的写作全程中，可参考元分析报告标准或者其他类似的标准，有可能可以更全面地对元分析进行把握，避免遗漏。首先，可参考报告标准，将研究计划进行细化并进行预注册，尤其是对方法部分的标准进行具体的报告。其次，按照预注册计划进行元分析。在元分析在执行过程中，可能需要进行计划之外的数据处理和分析等，在这种情况下，作者可以根据具体状况对元分析的过程进行调整，而非刻板地按照不适用的原计划进行分析。最后，按照报告标准撰写元分析报告。如果在元分析的执行过程中与原计划有所偏离，作者需要描述这种偏离并给出解释。元分析报告标准最终目的在于帮助作者理清思路，减少偏倚，而非对研究过程进行限制。

作者贡献声明：

胡传鹏、刘宇提出本文的想法；刘宇、陈树铨、胡传鹏完成数据的收集与编码；刘宇、胡传鹏撰写本文初稿；所有作者均参与本文稿件的写作并同意本文的最终版本。

[6] 值得注意的是，本文重点关注了元分析的报告标准，关注尤其在实证研究上，可以适用于实验研究和相关研究，但是对于质性研究，可能不具有指导作用。

参考文献

A. P. A. Publications Communications Board Working Group on Journal Article Reporting, S. (2008). Reporting standards for research in psychology: why do we need them? What might they be? *American Psychologist, 63*(9), 839 - 851.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo - Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *Am Psychol, 73*(1), 3 - 25.

Booth, A., Clarke, M., Dooley, G., Ghersi, D., Moher, D., Petticrew, M., & Stewart, L. (2012). The nuts and bolts of PROSPERO: An international prospective register of systematic reviews. Systematic Reviews, 1, 2.

Buscemi, N., Hartling, L., Vandermeer, B., Tjosvold, L., & Klassen, T. P. (2006). Single data extraction generated more errors than double data extraction in systematic reviews. Journal of Clinical Epidemiology, 59(7), 697 - 703.

Cafri, G., Kromrey, J. D., & Brannick, M. T. (2010). A meta-meta-analysis: Empirical review of statistical power, type I error rates, effect sizes, and model selection of meta - analyses published in psychology. Multivariate Behavioral Research, 45(2), 239 - 270.

Cooper, H., & Ribble, R. G. (1989). Influences on the Outcome of Literature Searches for Integrative Research Reviews. Science Communication, 10(3), 179 - 201.

de Vrieze, J. (2018). The metawars. Science, 361(6408), 1184 - 1188.

Dwan, K., Gamble, C., Williamson, P. R., & Kirkham, J. J. (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias—an updated review. PLoS One, 8(7), 1 - 37.

Edwards, P., Clarke, M., DiGuiseppi, C., Pratap, S., Roberts, I., & Wentz, R. (2002). Identification of randomized controlled trials in systematic reviews: Accuracy and reliability of screening records. Statistics in Medicine, 21(11), 1635–1640.

Eickhoff, S. B., Laird, A. R., Fox, P. M., Lancaster, J. L., & Fox, P. T. (2017). Implementation errors in the GingerALE software: Description and recommendations. Human Brain Mapping, 38(1), 7 - 11.

Fleming, P. S., Koletsi, D., Seehra, J., & Pandis, N. (2014). Systematic reviews published in higher impact clinical journals were of higher quality. Journal of Clinical Epidemiology, 67(7), 754 - 759.

Gates, N. J., & March, E. G. (2016). A neuropsychologist's guide to undertaking a systematic review for publication: Making the most of PRISMA guidelines. Neuropsychology Review, 26(2), 109 - 120.

Glasziou, P., Meats, E., Heneghan, C., & Shepperd, S. (2008). What is missing from descriptions of treatment in trials and reviews? BMJ (British Medical Journal), 336(7659), 1472 - 1474.

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwartz, Y., Sochat, V. V., Ghosh, S. S., . . . Yarkoni, T. (2016). NeuroVault.org: A repository for sharing unthresholded statistical maps, parcellations, and atlases of the human brain. Neuroimage, 124(Part B), 1242 - 1244.

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta - analysis and the science of research synthesis. Nature, 555(7695), 175 - 182.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? Behavioral and Brain Sciences, 33(2 - 3), 61 - 83; discussion 83 - 135.

Higgins, J. P. T., & Green, S. (2008). Cochrane handbook for systematic reviews of interventions. NJ, USA: Wiley-Blackwell.

Jones, A. P., Remmington, T., Williamson, P. R., Ashby, D., & Smyth, R. L. (2005). High prevalence but low impact of data extraction and reporting errors were found in cochrane systematic reviews. Journal of Clinical Epidemiology, 58(7), 741 - 742.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P., C., Ioannidis, J. P. A., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta - analyses of studies that evaluate health care interventions: Explanation and elaboration. PLoS Medicine, 6(7), 1 - 28.

Mchugh, M. L. (2012). Interrater reliability: the kappa statistic. Biochemia Medica, 22(3), 276 - 282.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Annals of Internal Medicine, 151, 264 - 269.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . PRISMA - P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic Reviews, 4(2), 1.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., . . . Yarkoni, T. (2015). Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. Science, 348(6242), 1422 - 1425.

O'Connor, D., Green, S., & Higgins, J. P. (2008). Defining the Review Question and Developing Criteria for Including Studies: John Wiley & Sons, Ltd.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. Journal of Experimental Social Psychology, 45(4), 867 - 872.

Orwin, R. G., & Vevea, J. L. (2009). Evaluating coding decisions. In H. Cooper, L. V. Hedges, & J. C.

Quintana, D. (2018). Remedies for three common issues in meta-analysis. Retrieved August 27, 2018, from https://doi.org/10.31219/osf.io/5skuv

Valentine (Eds.), The handbook of research synthesis and meta-analysis (2 ed., pp. 177-203). New York, US: Russell Sage Foundation.

Reid, A. T., Bzdok, D., Genon, S., Langner, R., Müller, V.,I., Eickhoff, C. R., . . . Eickhoff, S. B. (2016). ANIMA: A data - sharing initiative for neuroimaging meta - analyses. NeuroImage, 124(Part B), 1245 - 1253.

Sacks, H. S., Berrier, J., Reitman, D., Ancona - Berk, V., & Chalmers, T. C. (1987). Meta -

analyses of randomized controlled trials. The New England Journal of Medicine, 316(8), 450 - 455.
Schalken, N., & Rietbergen, C. (2017). The reporting quality of systematic reviews and meta-analyses in industrial and organizational psychology: A systematic review. Frontiers in Psychology, 8, 12, 1 - 12.
Schulz, K., Altman, D., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. Journal of Pharmacology and Pharmacotherapeutics, 1(2), 100 - 107.
Shojania, K. G., Sampson, M., Ansari, M. T., Ji, J., Doucette, S., & Moher, D. (2007). How Quickly Do Systematic Reviews Go Out of Date? A Survival Analysis. Annals of Internal Medicine, 147(4), 224 - 233.
Song, F., Eastwood, A., Gilbody, S., & Duley, L. (1999). The role of electronic journals in reducing publication bias. Medical Informatics, 24(3), 223 - 229.
Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., . . . Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ: British Medical Journal (Online), 343(5).
Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., . . . Thacker, S. B. (2000). Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. JAMA, 283(15), 2008 - 2012.
Valentine, J. C. (Ed.). (2009). Judging the quality of primary research. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 129 - 146). New York, NY, US: Russell Sage Foundation.
von Elm, E., Poglia, G., Walder, B., & Tramèr, M.,R. (2004). Different patterns of duplicate publication: An analysis of articles used in systematic reviews. Jama, 291(8), 974 - 980.
胡传鹏, 邸新, 李佳蔚, 隋洁, & 彭凯平. (2015). 神经成像数据的元分析. 心理科学进展, 23(7), 1118 - 1129.
胡传鹏, 王非, 过继成思, 宋梦迪, 隋洁, & 彭凯平. (2016). 心理学研究中的可重复性问题:从危机到契机. 心理科学进展, 24(9), 1504-1518.
叶子菁, & 石伟. (2010). 多元整合元分析法评介. 心理科学进展, 18(3), 505 - 510.
朱滢. (2016). “开放科学数据共享软件共享”, 你准备好了吗. 心理科学进展, 24(6), 995 - 996.