心理学探新 2024,Vol.44,No.4,374-384
PSYCHOLOGICAL EXPLORATION

中英心理学期刊数据删除及删除标准的元研究

谢宜骏 杨忠静 吴 燕*
(杭州师范大学经亨颐教育学院心理学系,杭州 311121)

摘 要:数据删除在心理学研究中存在较大的操作空间,研究者借此获取显著统计结果的操作极大地威胁了科研结果的真实性和可重复性。研究以2000、2010、2020三个年度发表在《心理学报》和*Psychological Science*期刊上的实证研究为分析对象,结合量化和质性分析方法,揭示中英心理学期刊数据删除现状。结果显示,中英期刊存在数据删除的研究各占比为48.83%和35.56%,平均被试删除比例分别为16.31%和14.48%,在删除数据后未按相关标准进行报告的比例分别为11.48%和5.46%;在被试数据删除和观测值数据删除中,报告率最高的删除标准分别为基于极端值的删除(57.87%)和基于任务的删除(30.6%);研究报告的次级删除标准体现了一定的随意性。这些结果表明了建立更为严谨的标准化数据删除报告规范的必要性。

关键词:数据删除;元研究;有问题的研究操作;数据删除标准
中图分类号:B841.2 **文献标志码**:A **文章编号**:1003-5184(2024)04-0374-11

1 前言

近年来,文献信任危机从自然科学蔓延至社会科学(臧雷振,潘晨雨,2020),科研数据的真实性和可重复性受到质疑。学术操作不当的问题也引发了广泛讨论。*Nature*杂志2006年的一项研究表明,相较于严重的学术不端行为如伪造、捏造和抄袭,一些不明显的不当行为,如无依据地删除实验数据,更容易发生且对科学完整性的威胁更大(Martinson et al.,2006)。这类行为被称为有问题的研究操作(Questionable research practices,QRP)。心理学领域对问题性操作展开的研究调查中多次提及数据删除问题。研究表明,自我报告"知晓数据对结果的影响后再决定删除数据"的美国心理学家占比为38.2%(John et al.,2012);有同样行为的意大利科学家占比为39.7%(Agnoli et al.,2017);英国心理学家的自我报告比例为19.8%(Rabelo et al.,2020)。此外,并非所有的数据删除行为都会在文中进行报告。向Psych Disclosure.org提交研究信息的161名心理学研究者中,11.2%的人表示没有完整报告所有删除的观测值(LeBel et al.,2013)。由此可见,不当的数据删除在科研实践中并不少见。

心理学研究中,数据删除通常指向两个层面。一是对被试的删除,即部分被试未纳入后续数据分析。二是对观测值的删除,即将某一被试样本的部分观测数据进行删除。合理的数据删除对保障研究效度是有必要的,但基于p值操纵(p-hacking)动机的事后数据删除可能使假阳性的概率升高(Head et al.,2015)。在规范数据删除方面,CONSORT工作小组发行的《随机平行对照试验报告规范》(*Consolidated Standards of Reporting Trials*,CONSORT)和美国心理学会发行的《美国心理学会发表指南》(*Publication manual of the American Psychological Association*)都要求作者在文章中报告被试删除标准(American Psychological Association,2020;Moher et al.,2010)。虽然删除标准不能直接揭示数据删除过程中是否存在问题性操作,但可以反映数据删除过程的透明性和合理性等信息,便于后续进行复制研究。

心理学科研实践中,不当的数据删除频繁发生,但少有研究聚焦于这个问题。基于此,研究者采用了元研究的方法来探究国内外心理学期刊在数据删除上的科研实践。元研究综合运用多种定性和定量分析手段,对研究本身进行反思和探索,以帮助研究者实施、评估研究(Ioannidis et al.,2015)。元研究的主要步骤包括确定研究方案、选择文献来源并制定严格的文献纳入和排除标准、根据标准进行多轮次文献筛查、文献内容编码、对纳入研究进行质量评估以及最后的数据分析和总结(Heijden,2021)。已有的元研究探究了p值的误报是否具有普遍性(Lyu et al.,2020)、研究是否遵循了严格的实验设计规则

* **通信作者**:吴燕,E-mail:lewuyan@126.com。

(如是否使用双盲法、随机化法等)(Verhagen et al.,2022)、不显著结果的普遍性及其意义(王珺等,2021)等。

参照 Verhagen 等人(2022)比较2000年和2018年六大物理治疗期刊上物理治疗临床试验统计中显著性报告规范(是否报告置信区间、p 值、效应量等)的元研究,研究选取国内外有代表性的心理学综合期刊《心理学报》与 *Psychological Science*,分析其2000、2010与2020年发表的实证研究论文中,数据删除的比例和删除标准的报告情况,通过描述性统计得到平均删除比例(包括被试删除比例和观测值删除比例),并对删除标准进行质性分析,以对比国内外期刊的报告差异,从而探究当前心理学研究中关于数据删除可能存在的问题性操作,揭示其发生率和常见删除标准,以期为建立更合理的发表标准作出贡献。

2　方法

2.1　设计

研究聚焦于分析数据删除比例、删除报告情况以及删除标准。为探索国内外心理学研究实践的差异和发展变化,选取《心理学报》以及 *Psychological Science* 期刊于2000、2010以及2020年度发表的文章,在完成文章搜索、筛选工作后,以研究为单位(即同一文章的子研究相互独立)进行内容编码和质量评估,在此基础上对研究质量评分、数据删除占比、平均删除比例等指标进行量化分析,对具体的删除标准使用 NVivo 软件进行质性分析。

2.2　文献检索

在《心理学报》的期刊官网上,以及在知网上以《心理学报》为文献来源,搜索发表于2000、2010、2020年的文章,去除增刊、重复文章;在 Web of Science 和 Sage 两个数据库中,确定文献来源为 *Psychological Science*,同时选择发表时间为2000、2010和2020,删除重复文章。

2.3　文献筛选

两名独立的评估者首先根据标题和摘要筛选每篇文章,然后全文筛选。如果需要,由第三个评估人员解决冲突。文献纳入标准包括:(1)量化的实证研究;(2)招募被试(包含动物被试)并获得了实验数据。根据标准(1),排除社论材料、综述(包括元分析)、个案分析、访谈研究、会议宣传、理论方法介绍、心理学史等相关的研究或文章;根据标准(2),排除使用数据库数据或其他二手数据的文章。

2.4　数据提取

2.4.1　数据编码

编码过程分为3步,即初步编码、编码校对和分类编码及校对(图1)。从每个纳入的研究中提取以下信息:研究方法类型、样本容量、被试种类、被试的人口统计学信息、是否删除实验数据、删除数据后是否有报告、删除数据的标准。两名评估人员独立编码,冲突由第三名评估人员解决。

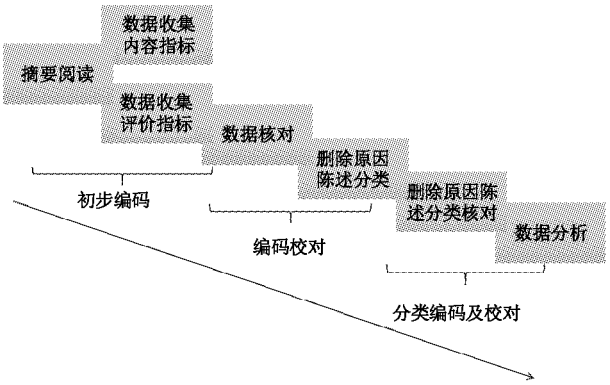

图1　文献编码和数据提取流程

2.4.2　研究质量评估

使用 CONSORT 量表2010版本的修改版对纳入研究进行方法学质量评估。CONSORT 工作小组于1996年公布初始版 CONSORT 量表(Begg,1996),随后在2001、2007和2010年分别进行了修订,由最先的几个维度逐渐发展到现在的9个维度、25个题项(www. consort - statement. org)。CONSORT 量表是临床医学领域对研究进行批判性评价和质量评估的常用量表(Moher et al.,2010),并非所有题项都适用于心理学研究,因此删除了不适用题项。修改后形成13个条目的心理学研究质量评估量表(见附录)。评估者对每项研究在每个条目上进行0~1打分(满足标准记为1,否则记为0),每个研究的分值范围为[0,13],分值与研究质量正相关。质量评估过程由两名评估者独立完成,评分中的冲突由第三个评估者解决。

2.5　数据分析

(1)研究质量评分分析

使用 F 检验,分析不同期刊、不同年份的研究质量评分是否存在显著差异。

(2)数据删除占比分析

通过研究给出的 F 值或者 t 值自由度计算实际纳入数据分析的被试数量,与研究中报告的被试数量进行对比,判断是否存在被试数据删除情况。通

过描述性统计，计算在所有研究中，有数据删除情况的研究占比，以及不同期刊、不同年份、不同类型研究中，有数据删除情况的研究占比。

(3)数据删除比例分析

数据删除比例是指删除的被试样本占总样本量的比值，或者删除的观测数据量占总的观测数据量的比值。研究将计算所有纳入研究的平均删除比例，并对不同期刊、不同年份、不同类型的平均删除比例进行差异分析。

(4)删除标准分析

将每项研究数据删除标准的相关陈述从编码文件中提取出来，按不同期刊分类形成单独文档导入 NVivo 12。根据程序化扎根理论(Strauss，1987)，逐条分析每项数据删除陈述并简化编码，得到若干自由节点，然后对自由节点进行整合，形成一系列子节点，即为第一轮开放编码。对子节点间的类属关系进行分析、归纳和总结，形成各类范畴(父节点)，即为第二轮主轴编码。因数据删除的分析角度(被试删除、观测值删除)是预先设定好的，故将各类范畴归入相应的分析角度之中，从而形成节点编码体系，即为第三轮核心编码。

NVivo 中的参考点数是指文本中的内容被编码为该节点的次数，本研究中该指标指某一删除标准出现的次数。对不同期刊的参考点数进行频数分析并使用卡方检验差异，可得到二者在具体删除标准上的异同。

3 结果

3.1 文章筛选结果

《心理学报》检索到 347 篇文章，经过筛选，共有 259 篇文章(428 个研究)纳入分析。*Psychological Science* 检索到 542 篇文章，经过筛选，共有 429 篇文章(824 个研究)纳入分析(图 2)。两个期刊的研究均以行为实验与问卷调查为主(表 1)。

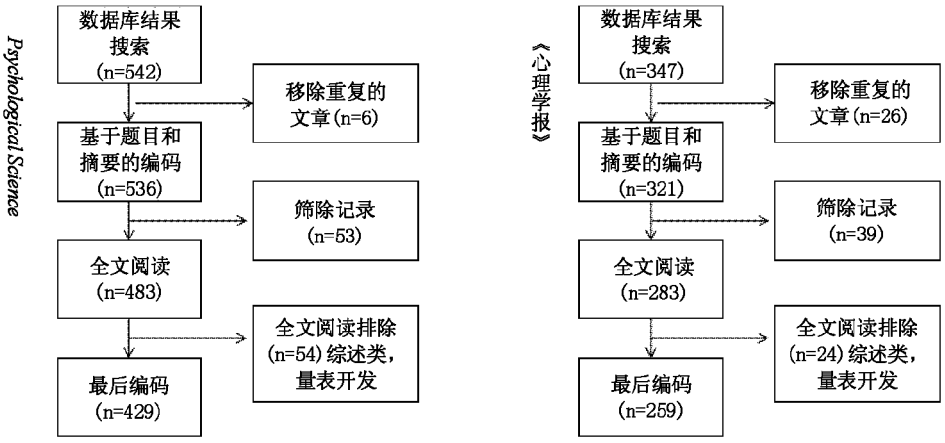

图 2 文献筛选流程

表 1 《心理学报》和 *Psychological Science* 三年中纳入研究的特征

期刊	研究类型	2000 年	2010 年	2020 年
《心理学报》	问卷(相关研究)	14	16	26
	行为实验	77	110	122
	脑成像研究	–	21	40
	其他	–	2	–
Psychological Science	问卷(相关研究)	8	48	33
	行为实验	104	422	178
	脑成像研究	8	18	2
	其他	1	–	2

注：其他指的是家庭录像观察等研究。– 指无相关类目。

3.2 CONSORT 质量评估

根据修改后的 CONSORT 量表对纳入研究进行质量评估。采用 2×3 被试间方差分析，检验不同期刊和发表年份对质量评分的影响。结果显示，期刊类型的主效应显著 $F_{(1,688)} = 52.56$，$p < 0.001$，$\eta_p^2 = 0.072$，《心理学报》的评分($M \pm SD = 9.56 \pm 1.05$)显著高于 *Psychological Science*($M \pm SD = 8.90 \pm 1.15$)；发表年份的主效应显著，$F_{(2,688)} = 71.63$，$p <$

0.001，$\eta_p^2 = 0.174$。多重比较显示，质量评分随年份升高，$M_{(2020)} = 9.76 > M_{(2010)} = 9.11 > M_{(2000)} = 8.39$。期刊类型和发表年份的交互作用不显著，$F_{(1,688)} = 0.59$，$p = 0.554$，$\eta_p^2 = 0.002$。

3.3　数据删除情况

3.3.1　数据删除在文献中的占比

在 1252 项研究中，有 502 项研究存在数据删除（包括被试删除、观测值删除以及两者都有的情况，总占比为 40.10%）。其中，《心理学报》存在数据删除的研究占比为 48.83%，*Psychological Science* 存在数据删除的研究占比为 35.56%。

在 502 项存在数据删除的研究中，462 项研究对删除情况进行了报告，报告率为 92.03%，其中《心理学报》的删除报告率为 88.52%，2000 年、2010 年、2020 年分别有 9 项、7 项、8 项研究有数据删除现象但未进行删除报告或报告情况不明（研究者仅就数据处理给出报告，如删除数据的标准，并未提及删除的比例或者数量）；*Psychological Science* 的删除报告率为 94.54%，在 2000 年、2010 年、2020 年分别有 12 项、3 项、16 项研究有数据删除情况但未报告删除或报告情况不明。

3.3.2　数据删除比例

《心理学报》和 *Psychological Science* 的平均被试删除比例分别为 16.31% 和 14.48%，二者无统计差

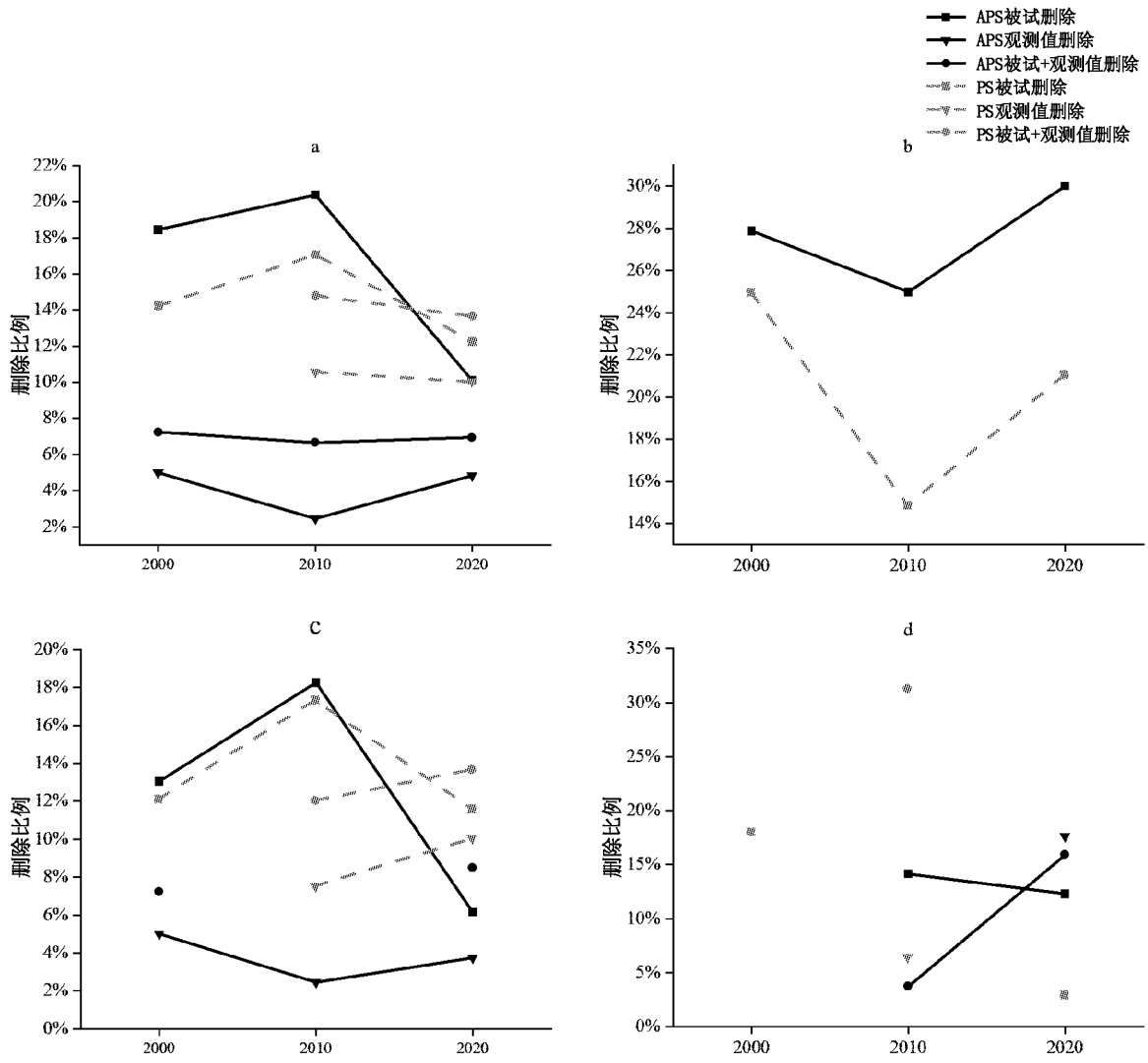

图 3　两种期刊在不同年份以及不同研究类型上的数据删除比例

a：两种期刊在不同年份的总数据删除比例；b：问卷调查研究；c：行为实验研究；d：脑成像研究。

注：存在某一研究类型在特定年份没有纳入的研究导致统计值空缺的情况。

APS =《心理学报》（*Acta Psychological Sinica*），PS = *Psychological Science*

异($t=0.51$,$p=0.51$,$d=0.011$)。而对于平均观测值删除比例、平均被试与观测值共同删除比例,*Psychological Science* 约为《心理学报》的两倍。时间上,2000 年至 2020 年,平均观测值删除比例、平均被试与观测值共同删除比例呈平稳趋势,而被试删除比例从 2010 年至 2020 年有大幅下降,《心理学报》较为明显。在研究类型方面,行为实验的数据删除随时间变化的趋势与以上所述基本保持一致。问卷调查类研究中,《心理学报》和 *Psychological Science* 都只出现了被试数据删除,《心理学报》的平均被试删除比例随时间变化不大,稳定在 30% 左右;*Psychological Science* 问卷调查类研究的平均被试删除比例从 2000 至 2010 年有约 10% 左右的下降,2020 年重新回归至 20% 左右。对于脑成像研究,平均被试删除比例在《心理学报》和 *Psychological Science* 中均呈下降趋势。

3.4 删除标准的质性分析

3.4.1 删除标准分类

根据数据删除在两个层面的含义,NVivo 操作下的研究者报告的删除标准编码结构如表 2 所示。

表 2 删除标准编码结构

核心编码	主轴编码	开放编码	参考点
被试删除	基于被试特征及表现的删除	被试特征不符合研究要求	49
		被试未按研究要求操作	24
		被试反应不符预期	48
		被试反应缺失或固定	33
		被试猜到实验目的、对实验表示怀疑	19
	基于任务结果的删除	未通过注意检测	40
		未通过操作检验	18
		未达到正确率要求	69
		未达到反应时要求	33
	基于任务完成度的删除	前后矛盾	8
		异常值	11
		未完成任务	64
		被试流失	17
	失误及其他因素	操作失误	62
		头动、眼动、伪迹	28
		语意模糊难以分类	62
观测值删除	基于异常值(极端值)的删除	平均值 ±3 个标准差	39
		平均值 ±2.5 个标准差	10
		平均值 ±2 个标准差	7
		反应时长低于或高于要求	41
		其他极端值	6
	失误及其他因素	没有反应或反应不正确	43
		头动、眼动、伪迹	17
		语意模糊难以分类	15

3.4.2 期刊的删除标准对比

在被试数据删除层面,基于任务结果的删除标准报告占比最高(30.6%),其次为基于被试特征及表现的删除标准报告(29.56%)(图 4)。具体的删除标准报告占比如图 5 所示,对不同期刊的删除标准占比进行卡方检验,结果如表 3 所示。

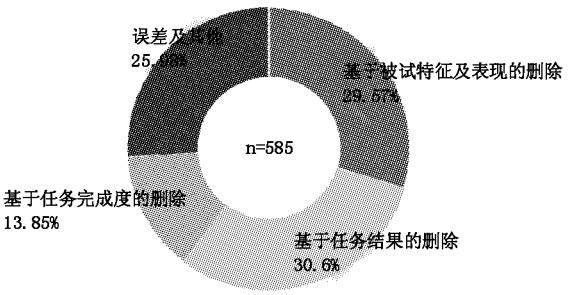

图 4 被试删除报告中各删除标准的占比

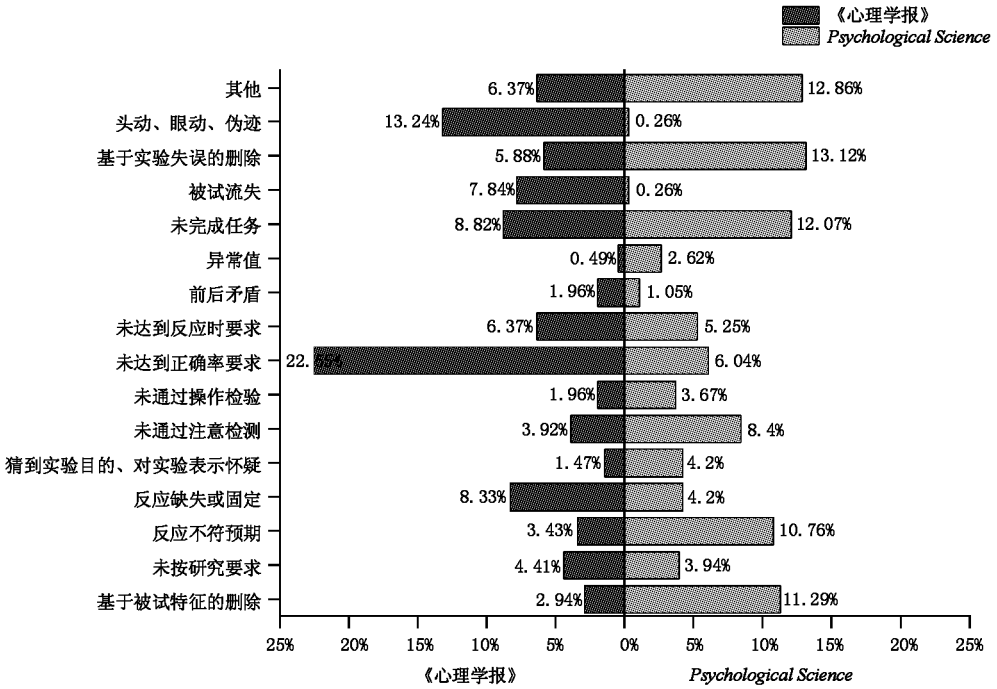

图 5　《心理学报》和 *Psychological Science* 中被试删除标准的报告

表 3　不同期刊被试删除报告中各删除标准的卡方检验

	删除标准	APS(n=204) 参考点数	APS(n=204) 报告率(%)	PS(n=381) 参考点数	PS(n=381) 报告率(%)	χ^2值	p值	Φ值
基于被试特征及表现的删除	被试特征不符合研究要求	6	2.94%	43	11.29%	12.06	0.001***	−0.144
	被试未按研究要求操作	9	4.41%	15	3.94%	0.08	0.783	0.011
	被试反应不符预期	7	3.43%	41	10.76%	9.48	0.002**	−0.127
	被试反应缺失或固定	17	8.33%	16	4.20%	4.27	0.039*	0.085
	被试猜测实验目的	3	1.47%	16	4.20%	3.15	0.076	−0.073
	未通过注意检测	8	3.92%	32	8.40%	4.18	0.041*	−0.085
	未通过操作检验	4	1.96%	14	3.67%	1.30	0.253	−0.047
基于任务结果的删除	未达到正确率要求	46	22.55%	23	6.04%	168.42	< 0.001***	0.244
	未达到反应时要求	13	6.37%	20	5.25%	0.32	0.575	0.023
	前后矛盾	4	1.96%	4	1.05%	0.28	0.596	0.037
	异常值	1	0.49%	10	2.62%	2.23	0.136	−0.075
基于任务完成度的删除	未完成任务	18	8.82%	46	12.07%	1.44	0.230	−0.050
	被试流失	16	7.84%	1	0.26%	27.06	< 0.001***	0.215
误差及其他因素	操作失误	12	5.88%	50	13.12%	7.35	0.007**	−0.112
	头动、眼动、伪迹	27	13.24%	1	0.26%	49.06	<0.001***	0.290
	语意模糊难以分类	13	6.37%	49	12.86%	5.90	0.015*	−0.100

注：APS =《心理学报》(*Acta Psychological Sinica*)，PS = *Psychological Science*

*p < 0.05，**p < 0.01，***p < 0.001

在观测值数据删除层面，异常值是最普遍的删除原因，占比为57.87%，主要报告为观测值在平均值±3个标准差之外、反应时过长或过短（图6）。具体的删除标准报告占比如图7所示，对不同期刊的删除标准占比进行卡方检验，结果如表4所示。

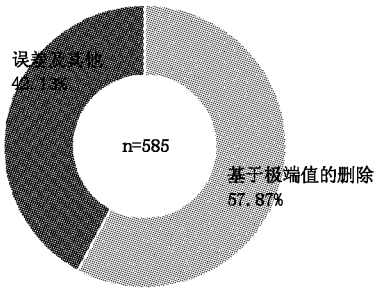

图6 不同类观测值删除标准在所有研究中占比

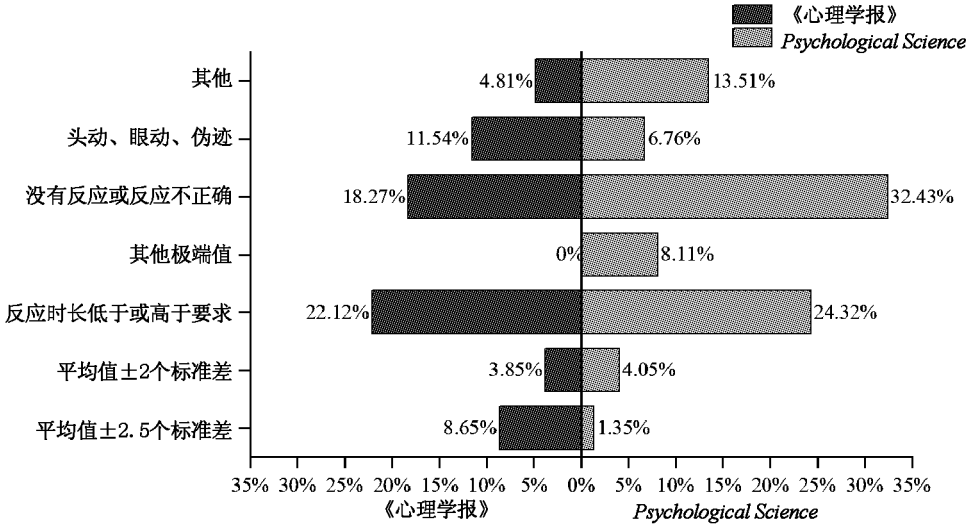

图7 《心理学报》和 *Psychological Science* 中观测值删除标准的报告占比

表4 不同期刊观测值删除报告中各删除标准的卡方检验

	删除标准	APS（$n=104$）		PS（$n=74$）		χ^2值	p值	Φ值
		参考点数	报告率（%）	参考点数	报告率（%）			
基于异常值（极端值）的删除	平均值±3个标准差	32	30.77%	7	9.46%	11.46	0.001***	0.254
	平均值±2.5个标准差	9	8.65%	1	1.35%	3.08	0.079	0.156
	平均值±2个标准差	4	3.85%	3	4.05%	Fisher	0.617	−0.005
	反应时长低于或高于要求	23	22.12%	18	24.32%	0.12	0.730	−0.026
	其他极端值	0	0.00%	6	8.11%	Fisher	0.005**	−0.221
误差及其他因素	没有反应或反应不正确	19	18.27%	24	32.43%	4.73	0.030*	−0.163
	头动、眼动、伪迹	12	11.54%	5	6.76%	1.14	0.285	0.080
	语意模糊难以分类	5	4.81%	10	13.51%	4.25	0.039*	−0.154

注：APS=《心理学报》（*Acta Psychological Sinica*），PS=*Psychological Science*

*$p<0.05$，**$p<0.01$，***$p<0.001$

4 讨论

4.1 数据删除占比

约有2/5的研究存在数据删除情况（包括被试删除、观测值删除以及两者同时删除），且《心理学报》和 *Psychological Science* 中各有部分研究删除了数据但没有报告删除的数值或者比例。可见，数据删除在心理学研究中较为常见，且确实存在未进行删除报告的研究实践。

4.2 数据删除比例

从2000到2020，两个期刊的平均观测值删除比例和平均被试与观测值共同删除比例呈稳定趋势，而被试删除比例在《心理学报》和 *Psychological Science* 中均有所下降。这可能由于学界对研究伦理关注增加和期刊对数据删除报告要求提高。研究人员在删除被试数据时变得更加谨慎，更倾向于删除异常或不可信的观测值。

不同研究类型中，行为实验的数据删除趋势与整体趋势基本一致，可能是因为行为实验的研究数量远超问卷调查的相关研究和脑成像研究。问卷调查的相关研究中，《心理学报》和 *Psychological Science* 都仅涉及被试数据删除。这可能是因为问卷调查很少单独分析特定题项的回答质量，研究者更倾向于直接删除回答不合格的被试数据，例如反应缺失、固定以及前后回答不一致的情况。两个期刊中，问卷调查研究的被试数据删除比例均高于各自期刊被试数据删除比例的平均值。这可能与研究时间跨度较长、被试流失较多有关。《心理学报》的平均被试删除比例稳定在30%左右，而 *Psychological Science* 的平均被试删除比例在不同年份均低于《心理学报》，且呈现先下降再上升的趋势。这可能与 *Psychological Science* 对数据删除报告的要求更加严格，研究者在被试删除方面更加谨慎有关，而2020年该期刊在平均被试删除数据上的上升可能是因为2020年有更多大样本研究（其中有9个研究初始被试量超过1000，5个研究初始被试量超过5000）。

《心理学报》和 *Psychological Science* 上脑成像研究的平均被试数据删除比例均呈下降趋势，总体低于平均观测值数据删除比例以及平均被试和观测值共同删除比例。这可能是由于脑成像研究中数据收集和分析的过程十分复杂，被试数量相对较小，研究者更加注重数据完整性。

总体来看，《心理学报》和 *Psychological Science* 中数据删除比例的差异受到伦理意识、期刊要求以及研究类型和方法特点的影响。

4.3　数据删除标准

被试数据删除标准的报告对提高研究的透明度和可重复性至关重要。在《心理学报》的所有删除标准报告中，未达到正确率要求而导致的被试删除远超其他各类别。不同类型的实验对正确率的要求不同，相应的删除标准可能存在差异。以反应时为主要分析指标的实验中，出于反应时－正确率权衡，往往会设定一个较高的正确率阈值，对平均正确率未达到阈值的被试数据进行删除；在简单任务中，正确率也可视为检验被试实验态度的指标。若正确率低于机会水平，被试很有可能未认真完成实验。如周爱保等人（2020）在比较精神分裂症患者和健康被试对自我与他人面孔识别能力差异的实验中，删除了正确率低于50%的患者和健康被试。但值得注意的是，很多研究在设定正确率要求时并未对其合理性进行解释，也没有参考标准，这就使得删除标准的设置具有随意性。如唐晓雨等人（2020）使用线索－靶子范式探究内源性空间线索有效性对视觉融合的影响，设定的被试删除标准是正确率低于90%；而任志洪等人（2020）使用敌意点探测范式的研究，设定的被试删除标准是正确率低于80%。明确报告删除标准和这些标准的设置依据可以减少使用模棱两可的删除标准，提升研究的可重复性。同时，删除标准的预注册可以一定程度上预防前文提到的“知晓对结果的影响后再决定删除数据”的科研实践。已有研究表明，预注册研究比传统研究具有更高的可复制性和质量（Chambers & Tzavella，2021）。

观测数据删除标准的分析可以帮助评估研究数据的可靠性和准确性。研究发现，在所有观测值数据删除标准的报告中，异常值是最主要的删除原因。同时，异常的任务结果也可能成为被试删除的原因。因此，正确处理异常值在心理学研究中至关重要。异常值可能涉及到研究者不关心的心理过程（如不加思考地猜测、注意力不集中等），选用合适的方法删除额外变量带来的异常值可以提高研究的信噪比（Berger & Kiefer，2021）。Simmons 等人（2011）指出，*Psychological Science* 期刊上的文章对异常反应时数据的处理存在不一致，说明异常值删除比较随意。研究发现，报告较多的异常值删除标准为观测值高于（或低于）阈值，以及超出Z分布临界值。对于前者，确实存在明确定义的阈值来区分有效反应和极端反应。例如对视觉刺激的反应时间（如Stroop任务），人们普遍认为反应少于200毫秒表示人为或软件错误（Ng & Chan，2012）。然而，大多数实验任务不存在这样的阈值，因而将异常值定义为与其余数据“不一致”或“相差太远”的数据点（Barnett & Lewis，1994）。心理学研究中常用的指标是Z分数（Andre，2022），但不少学者对此提出质疑。例如 Bakker 和 Wicherts（2014）的研究表明，根据Z分数大于某个临界值（这个临界值的常见值是2和3）来移除异常值，会使I类错误率上升。他们发现在移除阈值为 $Z = 2$ 的异常值后，I类错误率超过20%；也有学者怀疑在同一观察水平下通过Z分布识别和移除异常值意味着将此观察条件区别于其他条件，有违零假设检验的逻辑（Andre，2022）。

心理科研工作者有必要了解更多方法来识别异常值。结合王宏志等人（2019）、Massara 等人

(2023)的工作,研究者整理了当前科研工作中常用的异常值检测方法(表 5)。于此同时,并不是所有检测出来的异常值都需要进行删除,一刀切的删除处理可能会忽略有价值的异常值信息。相比总体,异常值数据较少,可能会让研究者低估其与理论的关联性(Gibbert et al. ,2021),忽视其违背预期可能是因为存在其它的解释,从而未能进一步探索新理论的构建。事实上,异常值的出现可能是新理论、新发现的契机,在很多学科中,异常值分析已经成为一种广泛使用的理论构建策略。

表 5 异常值检测方法

类型	定义	适用数据	优点	缺点	具体方法
基于统计的方法	依据统计模型和分布来识别异常值	假设特定分布的数据	简单易理解;适用于小规模数据集	对数据分布有假设;不适用于复杂或多变量数据	Z 分数、t 检验
基于距离的方法	计算数据点间的距离,距离远的点视为异常	数值型数据	直观,容易理解	高维数据效果不佳;计算量大	k 最近邻法
基于密度的方法	依据数据点周围的局部密度来确定异常值	数值型和混合型数据	适用于识别局部异常值	参数选择复杂;对不同密度区域敏感	局部异常因子(LOF)
基于聚类的方法	使用聚类算法分析数据,不属于任何聚类的点被视为异常	有自然群集倾向的数据	发现群集中的异常值	依赖于聚类算法的选择和参数设定	DBSCAN、K - means
基于模型的方法	使用机器学习模型来识别异常值,如隔离森林	大规模数据	对高维数据有效;计算效率高	随机性较强,结果可能不稳定	隔离森林、多模型集合方法

4.4 研究的不足与展望

研究采用元研究方法,选取国内外心理学的代表期刊《心理学报》和 *Psychological Science* 发表于 2000、2010、2020 年度的文章,分析其数据删除的报告情况。研究虽然揭示了数据删除的实践状况,但未能对所有的数据删除现象进行深入探讨,存在一些局限。

第一,纳入的期刊较少,仅分析了两个综合类期刊《心理学报》和 *Psychological Science*,其代表性有限。未来研究可聚焦不同的心理学分支,扩展期刊类型,以探讨不同分支的心理学研究在数据删除上的现状及差异。第二,在时间跨度上仅选取了 2000、2010、2020 三个年度进行分析,时间跨越性有限,可能不足以反映发展趋势。第三,研究使用的统计方法较为单一,仅使用了 t 检验,F 检验以及卡方检验来探究差异,后续研究可采用多元方法(例如贝叶斯分析)来验证结果。

基于对数据删除的量化和质性分析,参考系统性综述和元分析的 PRISMA 报告规范(Page et al. ,2021)以及《美国心理学会发表指南》(American Psychological Association,2020),研究提出关于数据删除报告的建议方案(表 6)。

表 6 数据删除报告方案

类型	报告内容
样本选择	阐明通过何种方式(如 G - power)确定计划的样本量大小
	描述被试纳入和排除标准
	表明何时、何地、由谁、通过何种方式收集数据
	描述数据收集的终止规则
数据删除	事先设定每条删除标准
	为每条删除标准提供合理性解释,如果可以给出例子
	根据删除标准,报告所有被删除的数据(如被删除的数据不存在对应的标准,需说明相应的删除理由)
	报告删除数据的数量以及占总体的比例
	通过统计手段,检验数据删除前后的研究结果是否存在差异

4.5 研究结论

数据删除在心理学研究中较为常见,《心理学报》和 *Psychological Science* 中有数据删除情况的研究占比较高,且存在少数研究删除数据后未按相关标准在文中报告;被试删除比例呈逐年下降趋势,而观测值删除以及被试和观测值共同删除比例随时间的变化较小;大多数研究者在设置删除标准时并未对其合理性进行解释,删除标准的设置体现了一定的随意性,因此有必要推动建立更加严谨的标准化报告规范。

参考文献

任志洪,赵子仪,余香莲,赵春晓,张琳,林羽中,张微.(2020).睾酮素与反社会倾向未成年犯的攻击行为:敌意注意偏向的中介和皮质醇的调节作用.*心理学报*,*52*(11),1288－1300.

唐晓雨,吴英楠,彭姓,王爱君,李奇.(2020).内源性空间线索有效性对视听觉整合的影响.*心理学报*,*52*(7),835－846.

王珺,宋琼雅,许岳培,贾彬彬,陆春雷,陈曦,等.(2021).解读不显著结果:基于500个实证研究的量化分析.*心理科学进展*,*29*(3),381－393.

臧雷振,潘晨雨.(2020).社会科学研究透明度:内涵、价值及其实现路径.*国外理论动态*,(5),82－92.

周爱保,谢珮,潘超超,田喆,谢君伟,刘炯.(2020).寻找丢失的自我:精神分裂症患者的自我面孔识别.*心理学报*,*52*(2),184－196.

Agnoli, F., Wicherts, J. M., Veldkamp, C. L. S., Albiero, P., & Cubelli, R. (2017). Questionable research practices among Italian research psychologists. *PLOS ONE*, *12*(3), e0172792.

American Psychological Association (Ed.). (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association.

Andre, Q. (2022). Outlier exclusion procedures must be blind to the researcher's hypothesis. *Journal of Experimental Psychology: General*, *151*(1), 213－223.

Bakker, M., & Wicherts, J. M. (2014). Outlier removal, sum scores, and the inflation of the Type I error rate in independent samples t tests: The power of alternatives and recommendations. *Psychological Methods*, *19*(3), 409－427.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd ed.). Wiley.

Begg, C. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *The Journal of the American Medical Association*, *276*(8), 637－639.

Berger, A., & Kiefer, M. (2021). Comparison of different response time outlier exclusion methods: A simulation study. *Frontiers in Psychology*, *12*, 675558.

Chambers, C. D., & Tzavella, L. (2021). The past, present and future of registered reports. *Nature Human Behaviour*, *6*(1), 29－42.

Gibbert, M., Nair, L. B., Weiss, M., & Hoegl, M. (2021). Using outliers for theory building. *Organizational Research Methods*, *24*(1), 172－181.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p－hacking in science. *PLOS Biology*, *13*(3), e1002106.

Heijden, J. (2021). Why meta－research matters to regulation and governance scholarship: An illustrative evidence synthesis of responsive regulation research. *Regulation & Governance*, *15*(S1), S123－S142.

Ioannidis, J. P. A., Fanelli, D., Dunne, D. D., & Goodman, S. N. (2015). Meta－research: Evaluation and improvement of research methods and practices. *PLOS Biology*, *13*(10), e1002264.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524－532.

LeBel, E. P., Borsboom, D., Giner－Sorolla, R., Hasselman, F., Peters, K. R., Ratliff, K. A., & Smith, C. T. (2013). PsychDisclosure. org: Grassroots support for reforming reporting standards in psychology. *Perspectives on Psychological Science*, *8*(4), 424－432.

Lyu, X.－K., Xu, Y., Zhao, X.－F., Zuo, X.－N., & Hu, C.－P. (2020). Beyond psychology: Prevalence of *p* value and confidence interval misinterpretation across different fields. *Journal of Pacific Rim Psychology*, *14*, e6.

Martinson, B. C., Anderson, M. S., & de Vries, R. (2005). Scientists behaving badly. *Nature*, *435*(7043), 737－738.

Massara, P., Asrar, A., Bourdon, C., Ngari, M., Keown－Stoneman, C. D. G., Maguire, J. L., … Comelli, E. M. (2023). New approaches and technical considerations in detecting outlier measurements and trajectories in longitudinal children growth data. *BMC Medical Research Methodology*, *23*(1), 232.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gotzsche, P. C., Devereaux, P. J., … Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal*, *340*, c869.

Ng, A. W. Y., & Chan, A. H. S. (2012, March). Finger response times to visual, auditory and tactile modality stimuli. In *International MultiConference of Engineers and Computer Scientists* (Vol. 2, pp. 1449－1454). IMECS.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* (*Clinical Research Ed.*), *372*, n71.

Rabelo, A. L. A., Farias, J. E. M., Sarmet, M. M., Joaquim, T. C. R., Hoersting, R. C., Victorino, L., … Pilati, R. (2020). Questionable research practices among Brazilian psychological researchers: Results from a replication study and an international comparison. *International Journal of Psychology*, *55*(4), 674－683.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False－positive psychology: Undisclosed flexibility in data collec-

tion and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359 – 1366.

Strauss, A. L. (1987). *Qualitative analysis for social scientists* (1st ed.). Cambridge University Press.

Verhagen, A., Stubbs, P. W., Mehta, P., Kennedy, D., Nasser, A. M., Quel de Oliveira, C., … McCambridge, A. B. (2022). Comparison between 2000 and 2018 on the reporting of statistical significance and clinical relevance in physiotherapy clinical trials in six major physiotherapy journals: A meta – research design. *BMJ Open*, *12*(1), e054875.

Wang, H., Bah, M. J., & Hammad, M. (2019). Progress in outlier detection techniques: A survey. *IEEE Access*, *7*, 107964 – 108000.

A Meta – research on Data Exclusion and Exclusion Criteria in Domestic and International Psychology Journals

Xie Yijun Yang Zhongjing Wu Yan

(Department of Psychology, Jinghengyi College of Education, Hangzhou Normal University, Hangzhou 311121)

Abstract: Data exclusion practices in psychological research provide researchers with considerable flexibility, which poses a significant threat to the validity and reproducibility of scientific findings. We conducted a quantitative and qualitative analysis based on 688 research articles published in the years 2000, 2010, and 2020 in the domestic journal *Acta Psychological Sinica* (APS) and the international journal *Psychological Science* (PS) for a better understanding of the research practices of data exclusion. The results showed that the proportion of studies employing data exclusion in APS and PS was 48.83% and 35.56% respectively. The average rates of participants exclusion were 16.31% and 14.48% for APS and PS, with 11.48% and 5.46% of studies failing to report exclusion criteria. The most frequently reported exclusion criteria were based on task (30.6%) and based on extreme values (57.87%). Flexibility existed in the setting of secondary exclusion criteria. These findings highlight the necessity of establishing more rigorous standardized reporting guidelines.

Key words: data exclusion; meta – research; questionable research practices; data exclusion criteria