

# 法学研究中裁判文书运用的 方法论检讨

——以刑事法文献为例的内容分析

何 挺\*

---

**内容提要：**裁判文书上网为研究者提供了观察我国司法实践的便捷途径。通过对重要期刊发表的刑事法文献的内容分析发现，虽然裁判文书上网对法学研究有助推作用，但其尚不能完全满足学术研究探知经验的需求，相关研究在方法运用方面也存在短板。基于裁判文书的研究，在裁判文书获取、抽样、研究设计、数据分析与运用、研究发现与结论适用等方面，存在较多不规范之处，并且缺乏针对裁判文书这一特殊数据来源的特殊方法论关注。考虑到我国法律数据所受的现实环境制约，裁判文书仍将在较长时间内是实证研究的主要数据来源之一。因此，应当在遵循实证研究基本规范的基础上，正确看待裁判文书所能提供的信息，采取多种方式应对裁判文书总体不清、代表性存疑等问题，并迈向真正具备大数据思维的实证研究。

**关键词：**裁判文书 实证研究 内容分析

---

## 一、问题的提出

2013年11月，最高人民法院开通中国裁判文书网（以下称“裁判文书网”），建立全国统一的裁判文书公开平台；自2014年1月1日起，各级法院的生效裁判文书陆续上网公布。截至2020年12月，裁判文书网公开的裁判文书已超亿篇，访问总量超524亿人次，是世界上最大的裁判文书数据资源库。<sup>〔1〕</sup>裁判文书上网无疑提供了一个观察和研究中国司法实践的绝佳窗口。一方面，裁判文书蕴含的细节信息，提供了观察司法微观实践的途径。另一方面，上网裁判文书的巨大体量为宏观概览整体状况和大样本研究提供了基本条件；裁判文书蕴含的结构化或半结构化数据所具有的价值，远超法检工作报告或者法律年鉴的寥寥数项统计数据，为

---

\* 北京师范大学刑事法律科学研究院教授。

感谢李梦竹、杨涛和马栎在编码和数据整理方面的工作。

〔1〕 参见《中国裁判文书网：司法公开的“亿”道亮丽风景》，<https://www.chinacourt.org/article/detail/2021/02/id/5821367.shtml>，2022年3月25日最后访问。

所谓法律大数据提供了基本数据来源。更重要的是,运用上网裁判文书开展研究,是一个无需费力进入司法田野即可在案头探知经验的便捷途径。对于长期苦于我国逼仄司法数据环境,或者无充足资源开展具有一定覆盖面实证研究的研究者而言,裁判文书上网无疑是经验数据来源上的重大突破和值得深度挖掘的宝藏。因此,裁判文书上网应当被视作法学研究领域的一个大事件与重大“自变量”,其必然会对我国法学研究的样态、方法和路径等产生重大影响。

裁判文书这一新的数据来源的自身特征,也对如何运用这类数据提出了新的要求。首先引起广泛关注的是上网裁判文书是否具有代表性。通过比较公开裁判文书的数量与法院结案数可以发现,虽然裁判文书上网率逐年提升,但仍然存在“选择性公开”和“部分上网”的问题,与“应上尽上”的要求仍有差距,而且不同省份、不同案由之间上网率存在较大差别。<sup>[2]</sup>进言之,如果上网裁判文书不是应当公开的裁判文书的全部甚至大部,裁判文书选择性上网的实际操作隐性规则也不明晰,<sup>[3]</sup>那么,以上网裁判文书的全部或者部分为样本的研究能够得出可靠且普适的结论吗?有学者就认为,鉴于上述裁判文书上网的“选择性偏误”,对于根据裁判文书得出的研究结论,尤其是那些依赖统计结果得出的结论,需要考虑如何才能不受到研究时未公布的裁判文书的影响。<sup>[4]</sup>此外,裁判文书上传时间的滞后性,季末、年底“突击上传”等情况,也会影响到在不同时间选取的裁判文书的代表性。

其次是裁判文书包含的具体信息能否满足实证研究的需要。具体包括三个方面:一是裁判文书包含的具体信息不充分。裁判文书不可能精确记录司法实践中影响法院审判案件的所有法律或非法律因素及其在司法实践中真实的作用效果。<sup>[5]</sup>比如,当研究者试图揭示法官作出裁判的主观动机、思维推理时,鉴于当前刑事裁判文书说理不足、不反映推理过程的情况,适当的实证材料或许就不是裁判文书。<sup>[6]</sup>如果裁判文书未记载或未全面准确记载研究者希望发现的经验事实,基于裁判文书的实证研究还能揭示真实经验吗?二是裁判文书上网时可能会删除一些信息从而导致数据偏差。比如,部分法院在公开裁判文书时会对当事人身份信息以外的文书内容进行删减,这使得某些依靠裁判文书对特定问题作出的分析,存在不同程度的数据偏差。<sup>[7]</sup>三是如何将裁判文书记载的信息进行科学的操作化。裁判文书记录信息须遵循裁判文书的制作规则,不会因研究者的需求而改变。因此,裁判文书记录的信息能否代表真实经验是存在疑问的,并且可能与研究者试图研究的理论、概念存在一定距离。如何遵循实证研究的程序理性要求,通过抽象层到经验层的多维、多次下降操作,在理论与裁判文书记载的信息之间建立连接,是运用裁判文书开展研究必须要考虑的问题。<sup>[8]</sup>

此外,研究者还关注裁判文书与法律大数据研究的问题。一个较为一致的认识是,虽然裁判文书上网为中国的法律大数据提供了最重要的数据来源,但目前运用裁判文书的一部分研

[2] 参见马超等:《大数据分析:中国司法裁判文书上网公开报告》,《中国法律评论》2016年第4期;叶燕杰:《司法政策执行视阈下刑事裁判文书“部分上网”问题》,《山东大学学报(哲学社会科学版)》2019年第2期;杨金晶等:《裁判文书上网公开的中国实践——进展、问题与完善》,《中国法律评论》2019年第6期。

[3] 有学者指出,“裁判文书上网的真实标准,至今还是一个谜”,这导致使用海量裁判文书未必能得到“科学”信息。参见程金华:《迈向科学的法律实证研究》,《清华法学》2018年第4期,第161页。

[4] 参见唐应茂:《司法公开及其决定因素:基于中国裁判文书网的数据分析》,《清华法学》2018年第4期,第36页。

[5] 参见左卫民、王婵媛:《基于裁判文书网的大数据法律研究:反思与前瞻》,《华东政法大学学报》2020年第2期,第72页。

[6] 参见李强:《面向刑法教义学的实证研究》,《政治与法律》2021年第12期,第130页。

[7] 参见左卫民:《迈向大数据法律研究》,《法学研究》2018年第4期,第142页。

[8] 参见白建军:《法学研究中的实证发现——以刑事实证研究为例》,《政治与法律》2019年第11期,第13页。

究, 充其量只是“大样本”或者“大量数据”研究, 与具有深层指标机器学习、数据偏差技术调校等特征的大数据研究, 尚有不小差距。因此, 除了今后向真正的法律大数据研究发展之外, 现阶段还须注重传统“小数据”实证研究方法并遵循相应的规范要求。<sup>[9]</sup>

综上所述, 虽然裁判文书上网给法学研究带来了重大利好, 但研究者仍然要遵循社会科学研究的基本规范, 根据裁判文书及其上网方面的特殊性, 发展相应的特殊方法论。基于此, 本文拟以刑事法领域发表的、运用裁判文书开展研究的文献为对象, 尝试探讨下列问题: (1) 运用裁判文书开展研究的基本状况如何; 裁判文书上网对我国法学研究产生了什么样的影响。(2) 运用裁判文书开展研究的具体方法是否科学, 是否遵循了实证研究的一般规范要求。(3) 运用裁判文书开展的研究是否注意到了裁判文书及其上网方面的特殊性, 以及是否因此采用乃至发展了特殊的方法论。

## 二、研究设计与方法

### (一) 基本方法与文献综述

以文献为对象的研究, 通常有文献计量和内容分析两种方法。文献计量法是文献计量学采用的主要研究方法, 以文献的各种外部特征为对象, 采用数学与统计学方法描述、评价和预测科学技术现状与发展趋势, 是一种定量分析方法。<sup>[10]</sup> 文献计量法关注的文献外部特征, 通常包括文献数量、作者、关键词和引证等。内容分析则是对各种信息传播形式所显示的内容进行客观、系统和定量的描述与分析。研究者通过对文献显性内容特征的系统分析, 得到与之相关的潜在内容特征的推论。内容分析方法假定, 在传播材料中发现的行为模式、价值观念和态度, 反映出并影响着创造和接受这些材料的人们行为、态度。<sup>[11]</sup> 与文献计量是一种纯粹的定量研究不同, 内容分析被认为是一种定性与定量相结合的研究方法。内容分析以定性的问题假设为出发点, 通过对文献内容的定量分析, 找出能反映文献内容的一些本质且易于计量的特征, 将以语言表示的文献转换为以统计数字描述的资料, 并从统计数据中得出定性的最终结论, 以达到对文献内容更深刻、更精确的认识。<sup>[12]</sup> 本文旨在对运用裁判文书开展研究的文献进行系统分析, 因此主要运用内容分析法, 在涉及一些文献外部特征时则运用文献计量法。

在法学领域, 采用文献计量或内容分析方法开展文献研究已有不少尝试。一方面是利用文献引证情况开展的研究。其中既有对高被引作者和文献的研究,<sup>[13]</sup> 也有利用某类文献的引证数据研究某一具体问题,<sup>[14]</sup> 还有利用引证数据对有关学科或者学术期刊进行研究。<sup>[15]</sup> 在这类

[9] 参见左卫民:《一场新的范式革命?——解读中国法律实证研究》,《清华法学》2017年第3期;程金华:《未来还未来:反思中国法律大数据的基础建设》,《中国法律评论》2018年第2期;前引〔5〕,左卫民等文。

[10] 参见朱亮、孟宪学:《文献计量法与内容分析法比较研究》,《图书馆工作与研究》2013年第6期,第64页。

[11] 参见风笑天:《社会研究方法》,中国人民大学出版社2013年版,第208页。

[12] 参见刘伟:《内容分析法在公共管理学研究中的应用》,《中国行政管理》2014年第6期,第93页以下。

[13] 例如苏力:《从法学著述引证看中国法学——中国法学研究现状考察之二》,《中国法学》2003年第2期;郭旨龙:《谁是中国法学高影响论文作者——中国知网最新H指数的探索研究》,载苏力主编:《法律和社会科学》第14卷第1辑,法律出版社2015年版,第219页以下。

[14] 例如成凡:《从竞争看引证——对当代中国法学论文引证外部学科知识的调查分析》,《中国社会科学》2005年第2期;侯猛:《中国法学的实力格局——以青年学者的引证情况为分析文本》,《中国法律评论》2017年第5期。

[15] 例如刘馨宇、孙妹:《高影响因子法学期刊特征分析——基于“北大法宝—法学期刊库”引证研究》,《政治与法律》2019年第2期。

研究中,文献计量法占主导地位,即主要关注引证量、引证率以及相应指数,也有少数文献综合内容分析对文献引证的各方面特征通过编码进行定性与定量相结合的研究。<sup>[16]</sup>另一方面,通过分析文献的各方面特征,呈现法学研究的现状与问题。其中有的针对法学研究整体,<sup>[17]</sup>更多的则将研究范围限制在某个二级、三级学科乃至某一具体领域。<sup>[18]</sup>在这类研究中,除了文献计量,研究者还会运用其对某一领域的专业知识,在对文献的某一方面特征作出内容判断的基础上进行定量分析,例如文献所属学科与具体领域、所采用的研究方法等。有的研究还运用大数据技术,将法学文献的部分内容作为“语料”,研究法学论文“为谁而写”的问题。<sup>[19]</sup>

与本文研究主题更为接近的是,一些研究者对法学领域的实证研究文献进行了定量研究。赵骏、程金华主要采用文献计量方法分析了中国法学期刊刊发的法律实证研究文章的数量以及在不同学科的分布状况等,归纳总结了法律实证研究的实践现状及不足。<sup>[20]</sup>屈茂辉采用知识图谱的方法和citespace软件,对以裁判文书为数据来源的法学实证研究文献的产出数量和时间、研究主体、研究领域与关键词等进行了定量分析和可视化呈现。<sup>[21]</sup>

整体而言,现有研究主要限于对文献外部特征的分析,对一定数量的文献进行系统的、定性定量相结合的内容分析的研究并不多见。同时,尚未发现对包括运用裁判文书开展研究在内的实证研究文献的具体内容与细节进行内容分析的研究。

## (二) 样本选择

本文样本来自2014年至2019年发表于CSSCI法学来源期刊和《中国社会科学》杂志的2967篇刑事法文献,从中筛选出运用裁判文书开展研究的文献452篇。<sup>[22]</sup>刑事法文献与运用裁判文书开展研究的文献,均通过人工审读方式筛选。所谓“运用裁判文书开展研究”,是指通过裁判文书获得相应信息或数据,进而展开论述与研究。单纯提及某一案件的判决结果,或者引用指导性案例等以其他方式公布的案件的情况,不属此类。对于这一样本,需要作如下说明。

第一,以裁判文书网于2014年1月1日上线这一时间节点为样本选取的起点。虽然在2014年之前已经存在少量运用裁判文书的研究,但考虑到研究的工作量和重点关注的时间段,未将其纳入样本范围。

第二,本研究需要对文献呈现的研究方法和过程作细致评价,故对文献的规范性有较高要求。发表于《中国社会科学》杂志和CSSCI法学来源期刊的文献,其规范性和专业水平基本能够代表中国法学研究的最高水准。<sup>[23]</sup>

第三,研究对象限于刑事法文献,主要是基于对所需专业知识的考虑。进行内容分析,需要对文献内容作系统评价,很多编码需要研究者基于专业知识进行判断。囿于研究者自身的研

[16] 例如左卫民:《从引证看中国刑事诉讼法学研究》,《法学研究》2013年第5期。

[17] 例如赵磊:《当代中国法文学学术生态分布——基于〈中国社会科学〉(1980—2015)法学论文的考察》,《清华法学》2017年第1期。

[18] 例如林少伟、林斯韦:《中国商法学实证研究测评——基于2000年至2015年的法学核心期刊论文》,《法学》2018年第1期;许多奇、唐士亚:《中国财税法学研究路径审视与展望——基于2009—2017年CLSCI论文的统计分析》,《法学》2018年第5期。

[19] 参见周翔、刘东亮:《法学研究目标受众选择的大数据分析》,《法学研究》2020年第1期。

[20] 参见赵骏:《中国法律实证研究的回归与超越》,《政法论坛》2013年第2期;程金华:《当代中国的法律实证研究》,《中国法学》2015年第6期。

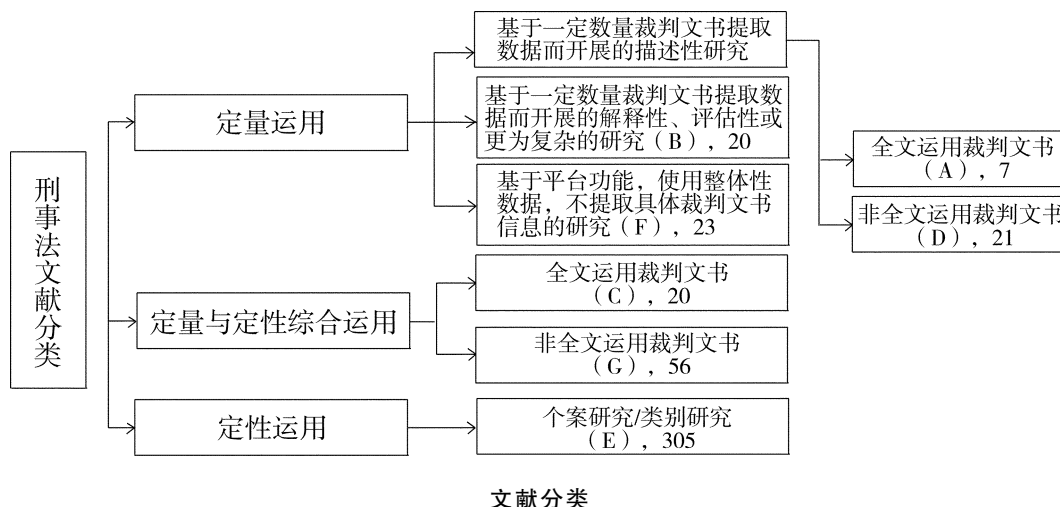
[21] 参见屈茂辉:《基于裁判文书的法学实证研究之审视》,《现代法学》2020年第3期。

[22] 刑事法文献的筛选采用相对宽泛的标准,与刑事法相关的交叉研究亦被纳入。

[23] CSSCI来源期刊目录定期更新,本文以某一年度某一期刊是否被纳入CSSCI法学来源期刊目录为标准。

能能力和专业限制，同时考虑到编码的工作量与质量，将样本限定在刑事法领域。

第四，将文献划分为不同种类是本文研究的基础。研究者使用裁判文书的方式极为多样，既可能是简单引用某一裁判文书作为司法实践例证，也可能是基于较大样本作定量研究。本文首先根据定性和定量这一基本分类，以文献运用裁判文书的方式为标准，区分定量运用、定量与定性综合运用、定性运用三大类。在定量运用中再区分三类：一是基于一定数量裁判文书提取数据而开展的描述性研究，并进一步区分为全文运用裁判文书和非全文运用裁判文书两类；<sup>[24]</sup>二是基于一定数量裁判文书提取数据而开展的解释性、评估性或更为复杂的研究，<sup>[25]</sup>这类研究均为全文运用裁判文书；三是未对一定数量裁判文书提取数据，而是基于数据库或网络平台提供的关键词搜索和分类统计功能，运用整体性的数量、比率等数据而开展的研究，这类研究都属于非全文运用裁判文书。在定量与定性综合运用中，区分全文运用裁判文书和非全文运用裁判文书两类。在定性运用中，区分个案研究和类别研究两类。上述分类及各类别文献的数量见下图。在所有类别中，A、B、C三类文献运用裁判文书最为充分、系统，对裁判文书数据及其研究价值有着更深程度的挖掘。



### (三) 编码与数据工作

内容分析在很大程度上依赖于编码。编码是将原始材料转换成标准化形式材料，以便于分析处理。编码可以分为对显性内容的编码（显性编码）和对隐性内容的编码（隐性编码），前者是指对文献可见的、表面内容的编码，后者是指对文献隐含意义的编码。隐性编码有助于开发文献的深层含义，但其代价是牺牲了可靠性和具体性，尤其当编码员多于一人时，不同的人可能采用不同的定义或标准。<sup>[26]</sup>

[24] 全文运用裁判文书是指整篇文献的逻辑架构和观点论证，均建立在裁判文书数据的基础之上；反之则指整篇文献中只有一部分内容是基于裁判文书展开的，其他部分为理论分析、教义学分析或者基于其他来源数据的分析。

[25] 一般而言，按照研究目的，可将研究分为探索性、描述性、解释性和评估性四类。探索性研究是指研究者在开展系统广泛的研究之前所作的初步探索，通常用于明确问题和形成初步观点。具体到本文语境，描述性研究是指运用裁判文书数据，描述呈现刑事司法某一方面状况的研究；解释性研究通常是指在描述性研究的基础上，运用裁判文书数据解释刑事司法中不同因素之间关系的研究；评估性研究是指运用裁判文书数据，评估某项法律、政策、改革或机制的实际效果是否达到预期的研究。

[26] 参见 [美] 艾尔·巴比：《社会研究方法》，邱泽奇译，华夏出版社 2009 年版，第 323 页。

本研究以单篇文献为分析单元,制定编码表及相应的含义说明,<sup>[27]</sup>并对不同类型文献进行不同层次的编码:(1)对所有文献均进行发表时间、刊物、作者、研究领域及文献类型等基本信息编码;(2)除定性运用的E类文献外,对其余147篇涉及定量运用裁判文书的文献进行裁判文书获取方面的编码,包括来源、检索方式等;(3)除使用平台功能而不提取具体裁判文书数据的F类文献外,继续对A、B、C、D、G五类共124篇文献进行样本形成方面的编码,包括样本数量、抽样方法等;(4)排除非全文运用裁判文书的D、G类文献后,继续对A、B、C三类47篇文献进行研究设计、数据运用分析、研究发现与结论三个方面的编码。之所以采用分层次编码,主要是考虑到文献类型的多样性以保证编码的有效性和针对性。

考虑到本研究中需要进行隐性编码的内容较多,为提高编码者间信度,对编码员人数作了限制,由三名刑事法专业的二年级硕士研究生担任编码员。三位编码员全程参与试编码及修改讨论过程,充分了解编码的具体要求。同时,为提高研究的整体信度和效度,对于编码内容最全且隐性编码内容最多的A、B、C三类文献,均由两名编码员同时独立编码并相互检验。两名编码员的编码结果不同时,由笔者最终确定编码结果。在所有文献编码结束后,再输入数据库进行统计分析。

### 三、数据分析与发现

#### (一) 运用裁判文书开展研究的整体情况

1. 整体数量与比例。如表1所示,自2014年以来,运用裁判文书开展研究的刑事法文献数量及其占全部刑事法文献的比例,始终呈现增长趋势;2019年占比最高,达到26.6%。这充分体现了裁判文书上网对刑事法研究的助推作用。虽然文献数量与占比呈不断增长的趋势,但更为系统运用裁判文书的A、B、C三类文献的数量与占比,却没有呈现出同步增长的趋势,整体上始终在低位徘徊,六年来只有47篇文献,仅占约10%。这一状况说明,研究者对裁判文书的运用整体上仍处较浅层次,裁判文书上网带来的研究价值受制于研究者研究设计与数据运用等方面的能力而未得到充分挖掘,运用的广度在提升,深度却未随之同步推进。

表1 2014—2019年运用裁判文书开展研究的刑事法文献的数量与占比

年份	运用裁判文书的文献	刑事法文献	运用裁判文书的文献在刑事法文献中所占的比例	A、B、C三类文献	A、B、C三类文献在运用裁判文书的文献中所占的比例
2014	33	456	7.2%	3	9.1%
2015	32	471	6.8%	3	9.4%
2016	57	461	12.4%	11	19.3%
2017	94	544	17.3%	9	9.6%
2018	94	501	18.8%	13	13.8%
2019	142	534	26.6%	8	5.6%
合计	452	2967	15.2%	47	10.4%

[27] 编码表由笔者和三名编码员共同进行三次试编码后确定,试编码针对23篇A、B、C类文献进行并作修改。

2. 具体研究领域分布。本文对文献的研究领域作了两个层次的编码。首先, 编码为刑事实体法、刑事程序法、犯罪学、刑罚执行、刑事法领域内交叉研究、刑事法与其他领域交叉研究六大类。之后, 在其中数量较多的实体法和程序法领域进一步作了二级编码, 前者分为定罪、量刑以及定罪量刑综合三类, 后者分为程序、证据以及程序证据综合三类。如表 2 所示, 有如下发现: (1) 从数量上说, 实体法是运用裁判文书开展研究最为集中的领域, 其中定罪研究文献几乎占全部刑事法文献的一半。这说明裁判文书提供的信息可能更便于开展实体法、尤其是定罪方面的研究。(2) 与实体法相比, 程序法领域运用裁判文书的研究不仅数量较少, 且运用程度也相对较浅, 只有 3 篇 A 类和 6 篇 C 类文献, 没有 B 类文献。这可能说明, 裁判文书记载的程序法信息不足, 无法满足研究者对司法过程性内容的需求, 研究者也尚未找到运用裁判文书开展程序法研究的更好方法, 且更多从其他渠道取得数据。(3) 文献主要集中于实体法和程序法两个部门法各自领域内。这说明研究者在面对信息内容本身较为综合的裁判文书时, 仍然倾向于从某一部门法的角度切入和提取信息, 刑事法学科内及学科外交叉严重不足。(4) 从研究深度上说, A、B、C 三类文献在量刑领域所占比例近 50%, 并显著高于其他领域。其中属于更为复杂的解释性或评估性研究的 20 篇 B 类文献中, 有 15 篇集中于量刑领域。这充分说明裁判文书提供的量刑信息相对较为充分, 适于开展较为复杂的统计分析。

表 2 运用裁判文书开展研究的刑事法文献的研究领域分布〔28〕

刑事实体法 (310, 31)	定罪 (216, 7)	刑事程序法 (102, 9)	程序 (44, 5)
	量刑 (37, 18)		证据 (41, 3)
	综合 (57, 6)		综合 (17, 1)
犯罪学 (16, 3)	刑罚执行 (2, 1)		
刑事法与其他领域交叉研究 (13, 0)	刑事法领域内交叉研究 (9, 3)		

## (二) 裁判文书的获取

1. 裁判文书的来源。裁判文书网上线前, 研究者主要通过其他数据库获取裁判文书, 或者自行搜集。裁判文书网上线后, 其成为裁判文书的主要来源, 但并未占据垄断地位, 裁判文书来源仍具有多样性。在全部 147 篇涉及定量运用裁判文书的文献中, 采用单一裁判文书来源的共 116 篇, 其中 68 篇直接从裁判文书网搜集, 43 篇通过其他网络平台搜集, 5 篇从自有数据库搜集或自行收集; 26 篇采用多样化的裁判文书来源, 其中 15 篇通过裁判文书网与其他方式共同搜集裁判文书; 5 篇文献未说明裁判文书来源。综上, 通过裁判文书网获取全部或部分裁判文书的文献, 占比为 56.5%。

2. 裁判文书的检索方式与检索范围。裁判文书网与其他网络平台通常会提供不同的检索方式, 可区分为简单检索和高级检索。前者是指对裁判文书全文进行简单的关键词检索; 后者是指运用平台提供的可以选择更多具体指标的检索方式, 例如进行法院级别、案由、文书的不同部分等限定后再作关键词检索, 有的平台还提供“包含”“排除”等逻辑选项更为复杂的检索方式。由于某一关键词在裁判文书中可能具有不同含义, 或者某一法律术语在裁判文书中可能以不同面目出现, 相对而言, 高级检索在精确性、帮助确定所需裁判文书方面更具优势。编

〔28〕 本表中括号内逗号前后分别为文献总数与 A、B、C 三类文献的数量。

码发现, 排除5篇自行收集裁判文书的文献后, 在142篇文献中, 采用高级检索方式的有24篇, 仅占16.9%; 简单检索的有60篇, 占42.3%。值得关注的是, 有58篇未说明检索方式, 占40.8%, 这与定量研究中应清楚交代样本来源的规范要求有较大差距。

研究者在检索裁判文书时会基于研究目的和研究能力等方面的考虑, 对检索的范围进行一定限制, 其中最主要的是地域和时间限制。如表3所示, 按照裁判文书所属地域进行检索的文献共16篇, 其中选择多个地域的有11篇, 占68.8%。研究者作出地域限制的理由主要包括: (1) 所选地域在某些方面具有代表性, 常见的是以地理位置或社会经济发展程度为标准选择多个代表性地域; (2) 研究便利需要, 如便于结合其他数据来源进行对照; (3) 在某一方面具有特殊性, 如法治程度较高或者某类犯罪发生较多; (4) 通常选择案件量相对较大的地域。按照裁判文书作出时间进行检索的文献共80篇, 其中以年为单位的55篇, 占68.8%。至于如何选择检索的时间, 研究者通常将与研究相关的某一法律、政策、司法解释或其他规范性文件的实施日作为所选时间段的起点, 根据所需的样本数量并考虑裁判文书的上网数量选择一定的时间段。值得注意的是, 在限制地域和时间范围进行检索的文献中, 有相当多未说明作出限制的原因。

表3 A、B、C、D、F、G类文献检索裁判文书的范围限制

按照地域进行检索 (16)	省级地域(8): 一省(2)/多省(6)		未说明选择地域的原因(8)/说明选择地域的原因(8)
	市级地域(4): 一市(3)/多市(1)		
	基层地域(3): 一个(0)/多个(3)		
	省市基层混合(1)		
按照时间进行检索 (80)	最短: 4个月	最长: 20年	未说明选择时间段的原因(51)/说明选择时间段的原因(29)
	以年为单位进行限制(55): 1年(16)/2年(3)/3年(6)/4年(5)/5年(6)/6-10年(12)/10年以上(7)		

考虑到包括裁判文书网在内的网络平台收录的裁判文书处于不稳定和不断变化的状况, 并且存在“突击上传”等情况, 研究者进行检索的时间无疑也会影响到研究对象的选择。在排除了5篇自行收集裁判文书的文献后, 在142篇定量运用裁判文书并在网络平台进行检索的文献中, 只有40篇说明了检索时间, 仅占28.2%。

### (三) 研究样本的形成

与裁判文书上网的具体方式有关, 在运用裁判文书进行的研究中, 通常存在三个层次的“裁判文书群”: 一是司法实践中某一类裁判文书的总体, 这一总体通常不可能为研究者所掌握; 二是在裁判文书网等网络平台上可以检索到的某一类裁判文书的研究总体; 三是对研究总体作进一步抽取而获得的研究样本。排除并不搜集裁判文书的F类文献后, 本部分对124篇涉及定量研究的文献, 在样本形成方面进行编码。

1. 研究者如何处理检索所得裁判文书与研究样本之间的关系。理论上, 研究可分为全样本研究和抽样研究, 前者不经抽样将所有对象纳入研究范围, 后者则按照一定抽样方案抽取一定的对象组成研究样本。对于研究者而言, 司法实践中客观存在的符合研究要求的裁判文书这一总体无法掌握, 其所能掌握的只是在裁判文书网或其他平台上检索得到的研究总体。因此,



对于未对检索所得裁判文书再作抽样而直接将其作为研究对象的研究,本文编码为全样本研究;〔29〕对于对检索所得裁判文书再作抽样从而确定研究对象的研究,则编码为抽样研究。编码发现,有92篇文献属于全样本研究,26篇属于抽样研究,有6篇无法判断是否对检索所得裁判文书进行了抽样。

2. 研究者如何处理检索所得中不符合研究要求的裁判文书。鉴于裁判文书的复杂多样性,无论在哪个平台、采用哪种检索方式,都无法确保绝对排除不符合研究要求的裁判文书,因此,研究者要进行相应的剔除工作,这种剔除有的采用人工方式,有的采用数据清洗等更为智能的方式。数据显示,绝大多数文献未说明剔除不符合研究要求的裁判文书的过程,只有48篇作了说明,占38.7%,其中全样本研究38篇、抽样研究10篇。在未说明剔除情况的文献中,实际上还包括研究者未进行大规模检索、而是主观选择部分裁判文书进行定量研究的情况。说明了剔除情况的10篇抽样研究文献,均属于先对检索所得进行抽样、再从抽样所得中剔除不符合研究要求的裁判文书,进而形成研究样本;不存在先剔除不符合研究要求的裁判文书,再进行抽样进而形成研究样本的文献。根据一般的抽样规则,抽样应当针对符合研究要求的对象,先剔除因不符合研究要求而不应纳入的对象之后再行抽样,其结果更符合概率论的要求。从研究可行性的角度看,这10篇文献经检索所得的样本总体数量相对较大,有的以万计,但究竟是先剔除还是先抽样,如何在可行性与科学性之间寻求平衡,还需要结合裁判文书的特殊性作进一步考量。〔30〕

3. 研究者如何看待样本规模与总体或研究总体的关系。这旨在观察研究者是否关注自己的研究对整体司法实践状况的代表程度。样本规模又称样本容量,是指样本包含的研究对象的数量。具体到本文的语境,即一项研究对多少篇裁判文书或裁判文书记载的案件、人或者罪刑关系进行了研究。基于法律适用的普遍性,研究者必然希望其研究结论能外推适用于整体,而研究结论的适用范围实际上取决于研究样本能在多大程度上代表总体或研究总体,因此,样本规模与总体或研究总体的关系就是需要关注的重点。编码发现,在124篇文献中,有9篇说明了样本规模占总体裁判文书的比例,2篇虽未说明具体比例,但说明了由于裁判文书未全部上网无法明确总体的数量等原因;对于检索所得的研究总体,研究者一般都能明确其具体数量,有89篇说明了样本规模占研究总体的比例,但仍有35篇未作说明。

4. 样本的数量。样本的数量是定量研究的基础性数据,也是研究者必须明确交代的内容。在124篇文献中,有113篇说明了样本数量,但仍有11篇未明确说明,占8.9%。在说明了样本数量的文献中,样本数量最小值是10,最大值超过300万,中位数为280;样本数量为500及以下的,占63.7%。在A、B、C三类47篇文献中,所有文献都说明了样本数量,样本数量最小值是14,中位数为750;样本数量为500及以下的,占38.3%。该三类文献的样本数量明显高于整体。〔31〕有19篇文献基于具体研究需要形成了多于一个的样本库,以进行

〔29〕当然,此处的“全样本研究”并不属于社会科学研究中将所有对象均纳入研究范围、研究结论具有普适性的全样本研究,还包括不少未明确说明检索方式、仅获得少量裁判文书并直接将其作为研究样本的文献,其科学性存疑。

〔30〕总结而言,检索、剔除、抽样进而形成研究样本的工作,在研究文献的操作中呈现为以下四种情况:检索后直接形成样本;检索后经过剔除形成样本;检索后经过抽样形成样本;检索后经过抽样再经过剔除形成样本。当然,还有很多文献缺乏对上述三项工作的明确说明,从而无法判断其样本形成情况究属何种。

〔31〕需要说明的是,此处样本数量的所指可能不同,有的文献指裁判文书的数量,有的指从裁判文书中提取的人(被告人、被害人)的数量,有的则指从裁判文书中提取的罪刑关系的数量。

比较或者对更为特定的样本进行分析。

表4 A、B、C、D、G类文献的样本数量分布

样本数量≤50:19	100≥样本数量>50:18	500≥样本数量>100:35
1000≥样本数量>500:11	2000≥样本数量>1000:14	10000≥样本数量>2000:9
样本数量≥10000:7		

5. 抽样研究中如何进行抽样、如何呈现抽样过程以及如何评价抽样方法。在抽样方法和规则中, 概率抽样和非概率抽样是两种基本分类。概率抽样也称随机抽样, 其按照概率原理进行, 能够保障在一个确定的总体内每个单位被抽取的机会是一样的。概率抽样的核心是“随机”, 根据概率抽样获得的样本一般被认为能够代表总体。非概率抽样也称非随机抽样, 是指不按照概率原理, 而是根据研究者的主观判断或条件便利等因素选择样本, 非概率抽样中每个单位被抽取的机会不是相同的。<sup>[32]</sup> 对于抽样研究, 研究者必须清楚呈现抽样过程以明确说明样本的质量和代表性。编码发现, 26篇抽样研究文献呈现出几个方面的情况(详见表5): (1) 在是否说明抽样方法方面, 只有3篇详细说明了抽样方法, 即读者能够明白整个抽取过程的每一步骤; 有8篇简单说明了抽样方法, 即对抽样过程有一些介绍, 但未具体到每一步骤; 有15篇未说明抽样方法和过程, 而是直接表述抽取了若干份裁判文书, 但该15篇中有8篇直接简单声称“随机”抽取。(2) 在如何评价自己的抽样方法方面, 有16篇表述为“随机”, 有10篇未表述为“随机”。表述为“随机”的文献中, 只有6篇分别使用了分层随机抽样、等距随机抽样和简单随机抽样等具体抽样方法的术语, 有10篇未提及采用的具体抽样方法。(3) 在实际抽样操作方面, 表述为“随机”的文献中, 只有7篇能判断为基本符合概率抽样的要求, 占43.8%; 无法判断是否随机的, 有9篇。有意思的是, 在未表述为“随机”的文献中, 也有2篇可判断为符合随机要求。

表5 抽样研究文献的抽样方法及其呈现情况

	抽样方法的说明程度	使用的抽样方法术语	实际操作是否随机
表述为“随机”的文献(16)	详细说明(2) 简单说明(6) 未说明(8)	无术语(10)/分层随机抽样(3)/简单随机抽样(1)/等距随机抽样(1)/系统抽样(1)	随机(7) 无法判断(9)
未表述为“随机”的文献(10)	详细说明(1) 简单说明(2) 未说明(7)	无术语(7)/分层抽样(1)/等距抽样(1)/方便抽样(1)	随机(2) 非随机(3) 无法判断(5)

(四) 研究设计

研究设计主要包括文献中呈现的研究类别、文献基础、假设与变量设置等内容。非全文运用裁判文书的文献很多缺乏整体的研究设计, 本部分编码仅针对全文运用裁判文书的A、B、C三类47篇文献展开。

1. 研究设计的基本类别。编码发现, 在47篇文献中, 描述性研究占半数以上(53.2%),

[32] 参见仇立平:《社会研究方法》, 重庆大学出版社2015年版, 第133页, 第168页。

共 25 篇；解释性研究共 16 篇，其中 12 篇的研究主题是量刑，量刑研究在目前的解释性研究中占据绝对主导地位；有 3 篇文献评估了某一政策或措施引入的效果，可以认为属于评估性研究；另有 3 篇文献兼具两类研究目的。<sup>[33]</sup> 值得注意的是，没有一篇文献自我评价只是对某一问题的初步探索性研究，即使其样本数量不足 100、甚至不足 30。

2. 研究的文献基础。研究设计需要建立在已有研究的基础之上，这一点通过两个方面进行评价。一方面是对研究设计的理论基础进行探讨的程度，即对相关研究现状、其他研究者的观点和既有争议进行探讨的程度，以及这些理论探讨与研究设计的关联程度。编码发现，有 14 篇文献的理论基础探讨较好，对已有研究进行了较为深入的分析，且理论基础探讨与研究设计关联度较强；24 篇的理论基础探讨一般，对已有研究的分析较为简单，或者理论基础探讨与研究设计关联度较弱；有 9 篇则未作理论基础探讨，或理论基础探讨与研究设计并无关联。<sup>[34]</sup> 另一方面是研究者是否借鉴、比较了其他类似或相关研究所使用的方法，是否在已有研究的基础上对方法进行了改进。编码发现，有 10 篇文献提及了其他类似或相关研究所使用的方法，但多数文献（37 篇，占 78.7%）未提及已有研究所使用的方法。

3. 研究是否明确提出了假设。假设是以已有事实材料和科学理论为依据，就未知事实或关系提出的一种推测性说明。<sup>[35]</sup> 假设是研究中需要运用数据进行检验的有关变量或者变量之间关系的预设，是研究的起点，对研究有指导作用。定量研究，尤其是解释性研究，都应当提出相应的假设。编码显示，在 47 篇文献中，只有 10 篇明确提出了假设，并且运用从裁判文书中提取的数据对假设进行了检验。尤其值得注意的是，在 19 篇解释性研究或者兼具解释性研究的文献中，只有 8 篇明确提出了假设。

4. 是否对变量设置作了具体说明。定量研究需要明确一定数量的变量及其不同属性，围绕变量收集、分析数据，进而描述变量或者解释变量之间的关系。变量的设定及其操作化，是研究设计的核心内容。对于运用裁判文书的研究而言，变量还必须与裁判文书记载的相应信息进行对接。因此，是否清晰、明确地说明了变量设置，也是评价研究设计的重要方面。编码发现，有 23 篇文献对变量设置作了详细说明，即专门列出了所涉及的变量并说明其内涵与属性；有 9 篇对变量设置作了简单说明，即未作专门集中说明，仅在行文时提及，但读者基本能够明白变量的指涉；有 15 篇未对变量进行具体说明。值得指出的是，19 篇解释性研究或者兼具解释性研究的文献都对变量设置进行了说明，其中 14 篇为详细说明，5 篇为简单说明。

5. 抽象概念与裁判文书所载信息是如何对接的。在运用裁判文书的研究中，除了将变量与裁判文书记载的信息对应外，如何把较为抽象的概念、理论与裁判文书记载的信息对接起来，也是研究设计及其操作化必须涵盖的内容。编码发现，在 47 篇文献中，有 13 篇涉及抽象概念，均对所涉及的抽象概念与裁判文书所载信息如何对接进行了论证说明。其中，9 篇文献的论证说明较为充分，涉及的抽象概念包括“减刑系数”“刑罚强度”“刑事审判共识度”等；4 篇的论证说明较为抽象，无法完全对接裁判文书记载的信息。

[33] 需要说明的是，解释性研究、评估性研究中通常会有描述性内容，描述性内容一般是服务于解释性或评估性研究的，此类文献会根据研究的整体目的编码为解释性研究或者评估性研究。只有描述性内容具有独立于解释性或评估性内容的自身价值时，才被编码为兼具两类研究目的。

[34] 需要说明的是，绝大多数文献都有对相应法律规范和司法实践背景的介绍，但仅有这些内容还不能视为对理论基础进行了探讨。

[35] 参见前引〔32〕，仇立平书，第 109 页。

### （五）数据运用与分析

对从裁判文书中提取的数据进行运用与分析，是得出研究结论的基础与依据。本部分编码针对全文运用裁判文书的 A、B、C 三类 47 篇文献展开。

1. 统计方法的层次。统计方法可以按照不同的复杂程度区分层次。一种分类是单变量统计分析与多变量统计分析，前者指对单一变量的描述，通常表现为频数、比率、均值、方差、标准差等；后者则是同时统计分析两个及以上变量，通常旨在解释、推断变量之间的关系，需要更为复杂的统计方法。法学研究者大多未接受过专业的统计训练，对于单变量统计分析通常能够理解并运用，对于多变量统计分析则存在理解与运用上的一定难度。编码发现，在 47 篇文献中，有 23 篇在单变量统计分析的基础上进一步运用了双变量甚至更多变量的统计分析。在使用的多变量统计中，有 20 篇使用了回归分析，包括 Logistic 回归和线性回归等，有 10 篇进行了 Pearson 相关系数分析，此外还涉及因子分析、T 检验、F 检验和卡方分析等。法学专业人员对多变量统计分析所涉及的专业统计术语和原理通常较为陌生，因此，在 23 篇运用多变量统计分析的文献中，有 16 篇尝试对所涉及的统计术语和原理作了一定说明。

2. 数据的展示方式。除了文字表述，图表是展示数据的重要方式。编码显示（详见表 6），在 47 篇文献中，既有 9 篇无图无表，也有 9 篇的图表总数超过 10，图表最多的文献的图表总数为 26。<sup>[36]</sup> 值得指出的是，除了普通的呈现数字的表格和常见的诸如饼状图、柱状图、折线图之外，运用较复杂统计方法并由专业统计工具生成的更复杂专业的图示极少出现，只有 4 篇文献在涉及回归分析和因子分析时使用了散点图和碎石图。

表 6 A、B、C 三类文献使用图表情况

使用表格情况	最大值 12/最小值 0（9 篇）/均值 4/中位数 3
使用图示情况	最大值 21/最小值 0（24 篇）/均值 1.7/中位数 0

3. 是否结合其他来源数据。鉴于单一数据来源可能存在缺陷，除了提取自裁判文书的数据外，研究者是否结合其他来源数据开展研究，也是全面评价研究科学性和质量的重要方面。编码发现，在 47 篇文献中，有 22 篇结合了其他来源数据，说明近一半研究者考虑通过多种来源的数据进行比较对照以提高研究结论的可靠性。在其他来源数据中，年鉴、各种类型的报告、官网发布等官方数据来源最多，共有 13 篇涉及。有 10 篇文献结合了研究者自行收集的其他数据，其中 7 篇结合访谈数据，各有 2 篇结合查阅案卷和问卷。另有 7 篇文献结合其他研究者收集的数据。

4. 大数据及相关方法的运用。虽然运用裁判文书开展的研究很多被称为法律大数据研究，但如前文所述，其中的绝大多数只能称为“大量数据”的研究。本文在编码过程中也关注研究文献在何种程度上开展了大数据研究，或者运用了大数据研究的相关技术。遗憾的是，目前只有 3 篇文献运用了自然语义识别技术（NLP）以及通过机器提取字段信息，与真正的大数据研究还相距甚远。

### （六）研究发现与结论

研究者如何得出研究结论并加以运用，在很大程度上反映了研究者如何看待自己的研究。

[36] 当然，图表数量上的差异，也可能受到不同期刊对使用图表的要求不同的影响。

本部分编码针对全文运用裁判文书的 A、B、C 三类 47 篇文献展开。

1. 如何看待研究结论的适用范围。任何一项非全样本的研究都有其适用范围，研究者应当理性看待研究结论的适用范围，并科学合理地将结论外推至适当范围。编码发现，有 21 篇文献论及研究结论的适用范围，占 44.7%；其中，有 20 篇认为研究结论可适用于超出样本的更大范围，只有 1 篇在反思研究的局限性时明确指出“研究结果能否推及到样本之外，还需作进一步的检验”。<sup>[37]</sup> 在 20 篇认为研究结论可适用于更大范围的文献中，有 19 篇分析了原因，主要是分析样本所具有的代表性，包括样本数量多少、抽样时选取的地域与时间段具有代表性、选取的案由具有代表性、属于随机抽样等；有的还专门针对裁判文书上网不全面的问题，分析指出这些问题不会对研究结论产生影响。<sup>[38]</sup>

2. 是否对比了自己的研究发现与其他实证研究发现。与类似或相关的研究发现进行对比，有助于佐证自身研究结论，或者经由可能存在的差异反思研究方法或提出更进一步的研究主题。编码发现，有 11 篇文献对比了自身研究发现与其他研究发现或结论，仅占 23.4%；其中，有 10 篇对比的是国内研究，有 1 篇同时对比了国内和国外研究。

3. 研究发现与对策建议的关系。在 47 篇文献中，有 42 篇明确提出了相应建议，占 89.4%。其中，28 篇提出了较为具体的建议，14 篇提出了较为宏观笼统的建议。进一步编码发现，在提出对策建议的文献中，有 3 篇文献的建议并非直接建立在研究发现的基础上，而是与研究发现有一定距离。

4. 对研究方法是否有所反思。受制于研究条件的限制，所有实证研究方法都存在某些方面的不足，研究结论的提出需要与对研究方法的反思相结合。编码发现，在 47 篇文献中，只有 10 篇对研究方法的局限性进行了反思，仅占 21.3%。研究者反思的内容主要包括：样本的代表性；裁判文书未记载的因素会影响研究的可靠性；裁判文书的文本挖掘可能存在误差；个别变量界定和赋值的准确性需要进一步探讨等。

#### 四、总结、反思与展望

前述发现不仅呈现了我国刑事法领域运用裁判文书开展研究的状况，也呈现了研究者对于裁判文书及其运用的认识与理解。这些发现为本文尝试探讨的三个问题提供了事实上的依据。

##### （一）裁判文书上网对我国法学研究的影响

如前述发现，运用裁判文书开展研究的文献数量及其占比逐年递增，无疑反映出裁判文书上网对我国法学研究具有助推作用。其提供的接近经验的便利机会，营造的基于裁判文书“经验”开展探讨的研究氛围，已经深刻影响了我国法学研究的整体样态乃至发展路径。在运用裁判文书的刑事法文献中，相对简单的定性运用占到近七成，其中包括不少选取个别裁判文书作为例证和论据的情形，非全文运用裁判文书的研究也占据了一定比例。这些在一定程度上说明，无论研究的具体路径与方法如何，在研究视野中纳入裁判文书并尝试运用已经是一种趋势。

然而，裁判文书上网对我国法学研究的影响，仍需理性看待。裁判文书包含的信息能否满

[37] 参见王剑波：《我国受贿罪量刑地区差异问题实证研究》，《中国法学》2016年第4期，第265页。

[38] 此处只是客观描述文献所述内容，不代表本文认可该文献所述理由。

足实证研究需要的问题,就在本文的研究中得到了部分反映:其一,编码发现,刑事实体法文献占绝对主导地位,数量是程序法文献的3倍以上。这说明较之实体法经验事实,裁判文书中天然缺乏过程性信息,无法满足探知程序法经验事实的需要。其二,在全部刑事实体法文献中,定罪研究远超量刑研究,但在A、B、C三类文献中量刑研究却占绝对主导地位。这说明裁判文书记载的信息中适宜开展较系统、复杂分析的数据并不多,且多集中在与量刑相关的问题上。其三,当研究涉及抽象的理论概念时,往往也无法在裁判文书记载的信息中得到完全对应。

总而言之,虽然裁判文书是探知经验的富矿,能够为法学研究注入新的动力,但其记载信息的广度和深度有限,公开过程和公开方式上的不足也导致其经验价值进一步消减。这些都决定了裁判文书上网并不能解决法学研究面临的所有经验和数据问题,因此,超越限度过度依赖裁判文书,反而会对法学的经验研究带来负面影响。

研究文献对裁判文书的运用程度,还折射出在我国法学研究中仍要进一步倡导对经验的认知和对接。无论持何种方法论立场,研究者可能都无法否认,法学研究需要在事实与规范之间形成良性互动。虽然法学研究对接经验的方式是多元的,运用裁判文书开展研究的状况也并非效率较高的测量对接经验程度的指标,但是,在长期逼仄的数据获取环境下,面对裁判文书随手可得的便利性,尤其考虑到裁判文书对法教义学研究同样具有非常重要的价值,运用裁判文书的状况还是能够作为一个重要参照,来评价法学研究认知和对接司法实践经验的程度。如前所述,在刑事法研究中,最多时也只有约四分之一的文献运用了裁判文书这一获取最为便利的经验材料。这其中固然有裁判文书记载的信息及其公开状况无法满足研究需要等原因,但确实也在一定程度上反映出仍要进一步倡导对经验的认知与对接。

运用裁判文书开展研究的现状,还充分暴露出我国法学领域在运用社会科学研究方法方面存在短板。比如,编码发现,A、B、C三类更为系统运用裁判文书数据的文献,其数量及占比始终在低位徘徊,这说明始终只有少数研究者具备充分挖掘裁判文书数据价值的的能力。更令人担忧的是,由于对研究方法的掌握先天不足,运用裁判文书的研究在推广普及实证研究的同时,还可能制造对实证研究的认识混乱。对裁判文书的不当甚至错误运用,还可能挤占其他更为耗时费力的实证研究的空间。比如,可能错误地认为基于几千份裁判文书开展的研究要必然优于对一个基层地区几十个案件的实地调查研究。再如,错误地认为在对大量裁判文书进行分析时选用的数学工具越复杂、越深奥,其结论就越科学。<sup>[39]</sup>

## (二) 运用裁判文书开展研究的科学规范性

裁判文书属于一种特殊类型的经验材料,无论是定性使用还是定量使用,都应当遵循社会科学实证研究的基本规范要求。裁判文书样本数量上的飞涨,可能从某种程度上造成了“数量为王”的假象,但实证研究的科学性仍然建立在是否遵从研究规范这一根基之上。“科学”的内容并不是研究的题材,而是研究的方法和规则;能把所有科学研究整合起来的是其采用的方法,而不是那些用于研究的材料;科学研究遵守一系列能够确保推论有效性的规则。<sup>[40]</sup>正如白建军所指出的,法学实证研究的程序理性,主要体现在呈现实证研究过程和结果的几个必备要素,即问题、已有学术成果中的竞争性理论、样本、核心概念的操作化、假设、因变量和

[39] 参见白建军:《论刑法教义学与实证研究》,《法学研究》2021年第3期,第120页。

[40] 参见[美]加里·金等:《社会科学研究中的研究设计》,格致出版社、上海人民出版社2014年版,第7页。

自变量、假设的检验逻辑、分析方法、结果与发现、讨论、新的理论与问题等等。这个模式为研究者提供了一条轨道，严格按这条轨道前行的，未必有出色的实证发现，但未能满足这些要素要求而偏离轨道或跳过某个环节的，就不大可能得到一个像样的实证发现。<sup>[41]</sup>然而，编码发现，运用裁判文书开展的研究，至少在以下几个方面距离基本的规范要求仍有较大差距。

第一，裁判文书的获取方式交代不清甚至存在错误。“提高数据质量的首要准则是记录和报告产生数据的过程。没有这些信息，就无法保证使用标准程序分析数据进而得到无偏的推论。只有知道产生数据的过程，我们才能获得有效的描述性或因果关系推论。”<sup>[42]</sup>对于来源多样、浩如烟海的裁判文书，按照什么样的标准选择具体对象及其理由，无疑要交代清楚。编码显示，在定量运用裁判文书的147篇文献中，对于最基本的获取裁判文书的来源，有5篇未作交代。在非自行收集裁判文书的142篇文献中，约40%完全未说明如何检索获得裁判文书。有50%的文献未说明对检索的地域范围进行限制的原因，有超过60%的文献未说明对检索的时间段进行限制的原因。鉴于裁判文书上网的现状，检索时间对可供检索的裁判文书范围有重大影响，但在前述142篇文献中，超过70%未说明在网络平台上的检索时间。在检索的科学性方面，大多数文献采用的是关键词简单检索，但是，因受到文字表述及其他多样因素的影响，关键词检索的命中率较低。<sup>[43]</sup>在刑事法文献中，也可以观察到因关键词简单检索导致样本代表性和结论存疑的情况。<sup>[44]</sup>

第二，对样本代表性和抽样方法的关注意识明显不足。上网裁判文书对于整体裁判文书的代表性一直存疑，其中可能存在系统偏误，从中抽取一部分作为样本又可能存在抽样误差。对于定量研究而言，此种情况下更需要关注样本的代表性。编码发现，在124篇定量研究文献中，对于样本数量，仍有约9%的文献未作明确说明；也存在不少未进行大规模检索或者未说明检索方式，主观选择部分乃至少量裁判文书开展定量研究的情况；样本数量虽然并非越多越具有代表性，但整体看，50%文献的样本数量在280以下，超过60%文献在500以下，样本数量整体偏低；对于检索所得裁判文书，超过60%文献未说明是否以及如何剔除其中不符合研究要求的裁判文书。关注样本的代表性，还必须关注样本与总体之间的关系。所谓抽样误差正是指样本值与总体值之间的偏差。诚然，运用裁判文书开展研究时面临的是一个总量不明确、边界不清晰的裁判文书总体，但是，已经公开并可在网络平台检索到的裁判文书研究总体的数量一般是可知的，有的情况下还有官方发布的数据等其他来源可为对照，即便如此，在124篇定量研究文献中，仍有近30%未说明样本与研究总体之间的关系。此外，抽样方法使用混乱，进一步严重影响了样本的代表性。在26篇抽样研究文献中，只有3篇详细说明了抽样方法。在将抽样表述为“随机”的16篇文献中，有8篇未说明抽样的具体方法，只有6篇说明了具体

[41] 参见前引〔8〕，白建军文，第12页。

[42] 前引〔40〕，金等书，第21页。

[43] 有学者指出，因为生态破坏类案件在裁判文书中被大量遮蔽，仅凭案由等关键词无法直接检索出该类案件。参见张忠民：《生态破坏的司法救济——基于5792份环境裁判文书样本的分析》，《法学》2016年第10期，第120页。

[44] 一项研究使用“非法证据”这一关键词在网络平台进行简单检索获得裁判文书，经过分析得出结论：司法实践中绝大多数非法证据排除申请都能启动对证据合法性的审查，其中约有一成左右的申请得到支持从而将非法证据予以排除（参见易延友：《非法证据排除规则的中国范式——基于1459个刑事案例的分析》，《中国社会科学》2016年第1期）。另一项时间接近但采用对多个不同层级法院进行实地调查方法的研究则发现，非法证据排除规则在实践中遇冷，不仅法院很少依职权启动非法证据排除程序，而且被告方提出启动程序申请的比例也较低，很少有证据被认定为非法并予以排除，即便被排除也难以对案件处理产生实质影响（参见左卫民：《“热”与“冷”：非法证据排除规则适用的实证研究》，《法商研究》2015年第3期）。

抽样术语。另有9篇无法判断样本抽取是否符合随机要求，有“随意抽样”之嫌。

第三，A、B、C三类47篇文献较为系统地运用了裁判文书，但其研究设计的多项要素缺乏或者呈现不足。首先，理论基础探讨不足。编码发现，超过70%的文献在理论基础探讨上需要加强。其次，对研究方法的比较参考严重不足，未从已有研究中汲取方法上的经验教训，近80%的文献未提及相关研究使用的方法。再次，缺乏明确的假设。尤其对解释性研究而言，缺乏明确的假设会使研究设计和数据运用缺乏起点和终点，也较难与抽象的理论相对接。只有约20%的文献明确提出了假设并进行了验证；在解释性研究中，明确提出假设的也仅占约40%。最后，对变量和概念的操作化呈现不足。对于裁判文书，研究者需要把理论上的概念、变量与裁判文书中记载的信息对接起来，进而收集裁判文书中的信息作为概念的具象表现和变量的属性。如果缺乏对变量和概念如何操作化的明确说明，研究的效度将高度存疑。编码发现，超过30%的文献没有对变量设置进行具体说明；一些文献对抽象概念的说明，无法完全对接裁判文书记载的信息。

第四，数据分析与运用在多个方面有待提升。首先，运用的统计分析方法整体上处于初级阶段，在A、B、C三类47篇文献中，超过50%开展的是单变量描述统计。其次，具体统计方法、工具的运用及对结果的解释是否合理，需要进一步观察。法学研究者缺乏统计学知识，这可能影响到统计分析的科学性，甚至在运用统计分析方法时出现“硬伤”。<sup>[45]</sup> 本文编码虽然无法对每一项研究使用统计方法是否合理进行个别化确认，但也能观察到与统计分析科学性相关的若干现象：只有极少数文献在进行回归等某类统计分析前，会论证说明样本数据符合开展这类统计的前提要求，例如是否正态分布；用于检验结论稳定性和可靠性的稳健性检验等分析方法，只在极个别文献中有所涉及。最后，数据的呈现方式主要是常规图表，对图表的使用也存在繁简无序的问题。事实上，并不是要把所有表格都放在研究报告的最终稿中，而是应该在保证向读者充分展示信息的前提下，选用尽量少的图表和表格。<sup>[46]</sup>

第五，对研究结论的认识存在偏颇。首先，研究者极少关注研究结论的适用范围以及研究结论能否外推适用。出于对象异质性的考虑，社会科学研究尤为关注基于一定范围样本开展的研究的结论能否外推到样本之外的群体。对此，除了确保样本选择的随机性等方法外，还形成了专门用于评价研究发现与结论能否外推至更大群体的效度指标——外在效度。在裁判文书本身以及研究样本的代表性存疑的情况下，如何客观评价研究结论就显得尤为重要。编码发现，在A、B、C三类47篇文献中，超过50%的文献没有论及研究结论的适用范围；在21篇论及结论适用范围的文献中，有20篇认为研究结论可适用于样本之外的更大范围，其中包括一些样本选取交代不清、无法判断是否随机抽样、甚至样本数量少于200的文献。其次，研究结论缺乏与其他研究的对话。研究结论的比较对照不仅能够互相验证，还可能催生出新的问题。编码发现，在A、B、C三类47篇文献中，只有约20%与其他研究的结论进行了对照。再次，

[45] 这一点在域外法律实证研究中同样存在。美国学者指出，美国法律实证研究中统计分析存在“硬伤”很普遍，包括错误界定P值和统计显著性，对回归分析的结果进行错误的解释（例如把相关关系解释为因果关系，把无效结果解释为没有因果关系），错误应用统计检验以及忽略产生有效分析结果所需要满足的前提条件等。See Kathryn Zeiler, *The Future of Empirical Legal Scholarship: Where Might We Go From Here?*, 66 (1) *Journal of Legal Education* 85 (2016).

[46] 参见[美]劳伦斯·纽曼：《理解社会研究——批判性思维的利器》，胡军生、王伟平译，人民邮电出版社2015年版，第351页。



编码发现,有少量文献提出的对策建议与其研究发现之间缺乏紧密关系。最后,对研究方法的反思和问题披露不足。任何研究都有局限性,研究者应该比任何人都熟知自身研究的缺点和错误。同时,研究者对同行负有的学术伦理责任也要求,研究者有责任使读者了解这些缺失。只要研究者把进行某项研究时经历过的所有困难与问题和盘托出,就可以对自己的同行、甚至对整个科学探索有所贡献。<sup>[47]</sup>在法律领域,相关研究有可能成为司法或立法决策的基础,这一披露义务就显得更为重要,更需要在一篇文章中提请受众注意样本选取、验证过程、结论的适用性等方面可能存在的局限性、弱点以及替代性解释和政策建议。<sup>[48]</sup>运用裁判文书的研究同样不能回避对研究方法问题的反思和披露,而且由于裁判文书这一数据来源存在先天不足,对问题的反思和披露就显得更为重要。但是,编码显示,在A、B、C三类47篇文献中,只有约20%对研究方法的局限性进行了反思。

鉴于目前裁判文书的运用在我国法律实证研究中的重要地位,上述五个方面的问题,其实也充分反映出我国法律实证研究整体上在遵循基本研究规范方面存在严重不足。这些问题需要在今后通过对方法和技术规范的倡导、相应学术伦理规则的建立来逐步解决。

### (三) 运用裁判文书的特殊方法论

概括而言,裁判文书作为实证研究数据来源时,主要有两个方面的特殊性:记载的信息较多但又有明显的局限性;数量规模巨大、易于获取但代表性存疑。除了需要遵循的基本规范要求外,如何更好地运用裁判文书,还需要发展出针对这些特殊性的方法论。

第一,正确全面看待裁判文书所能提供的信息,实现多方面数据的组合。裁判文书所能提供的信息受到多方面限制,有其天然缺陷,无法单独满足探知司法经验的所有需求,但是,裁判文书包含的信息又是包罗万象的,并有大规模数量作支撑。除了目前主要关注的罪、刑、证据等要素外,裁判文书实际上还记载了很多其他方面信息。单独来看,这些信息可能与法律研究的核心问题相距较远,但可以从学科交叉、视角交叉的不同角度,将这些信息与其他数据组合起来使用。因此,在明确裁判文书提供的信息中哪些有效可用这一前提的基础上,基于更为开放、多元和交互的视角,实现信息的多方面组合使用,能够充分挖掘裁判文书的数据价值:(1)不同学科的交叉使用。编码显示,刑事法领域内的交叉研究、刑事法与其他领域的交叉研究,数量都非常少。学科视野的扩大,可以更好地利用裁判文书中记载的有效信息。(2)与其他类别数据结合使用。不同类别的数据有各自的用途,裁判文书数据作为定量数据使用时,即使样本数量较大也只能反映数据间的相关性,要解释此种相关性,还要依靠其他素材,特别是离不开对一些典型个案的具体剖析。在克服脱离个案细节这一问题上,可以多采取定性方法与定量方法相结合。<sup>[49]</sup>(3)与其他来源的数据结合使用,包括官方数据、社会调查数据等。鉴于单一来源数据可能存在缺陷,研究者应当对基于某个单一数字系统得出的有关人类行为的任何说法都持谨慎态度,无论这个数字系统的规模有多大。<sup>[50]</sup>但编码显示,在A、B、C三类47篇文献中,只有不到50%的文献尝试结合了其他来源的数据。

第二,采取多种措施应对裁判文书总体不清和代表性存疑的问题。在裁判文书上网尚无法

[47] 参见前引[26],巴比书,第69页。

[48] 参见陈若英:《中国法律经济学的实证研究:路径与挑战》,载苏力主编:《法律和社会科学》第7卷第2辑,法律出版社2010年版,第15页。

[49] 参见周翔:《作为法学研究方法的大数据技术》,《法学家》2021年第6期,第68页。

[50] 参见[美]马修·萨尔加尼克:《计算社会学:数据时代的社会研究》,中信出版社2019年版,第49页。

做到全部公开和各地一致的情况下,裁判文书总体不清和代表性存疑的问题就是定量研究始终要考虑和应对的问题。除了对自己研究结论的适用范围保持谨慎外,可以考虑以下几个方面:(1)运用统计学工具和技术提升样本的代表性。一些研究已经开始运用事后分层方法、<sup>[51]</sup> bootstrapping 算法<sup>[52]</sup>等技术方法提升样本的代表性,这就需要法学研究者更好地与统计学等专业开展合作研究。(2)通过完善检索方式提升裁判文书检索的命中率。如前所述,关键词检索的命中率较低且会直接影响样本的代表性。相对而言,基于大数据技术的检索能较好地处理同一法律术语或同一情形的汉语表达方式多样的问题,目前一些网络平台已经开始提供类似的检索服务。例如,以“要素”代替“关键词”检索,系统对术语进行语义识别,匹配到特定的知识体系中,再通过知识图谱支撑的法律认知引擎技术,在更大范围内尽可能准确匹配检索需求。<sup>[53]</sup>(3)从某种角度看,选择较小范围内全部可获得的裁判文书作为研究样本,可能要比在较大范围内选择一部分裁判文书作为研究样本更为可靠。这是因为前者能够提高数据收集率,尽可能避免抽样误差的影响,也更易于与在小范围内通过其他来源收集的数据相结合。(4)与选择一定样本并将结论外推相比,不同样本之间或样本内的比较更值得关注。裁判文书总体不清和代表性存疑的问题,导致结论外推错误风险与法律适用普遍性之间的矛盾无法完全消除,但是,在裁判文书内部选择不同样本进行比较或在样本内部进行比较,则可以回避这一问题。例如,以不同地域或不同时间段为标准选择样本进行比较,或者采用倾向得分匹配(propensity score matching)方法选择除了某些关键自变量外其他方面可以进行对照比较的多个样本库进行比较,或者在样本内部进行部分个体与整体之间的比较。<sup>[54]</sup>

第三,运用裁判文书的研究应当向真正具有大数据思维的研究迈进。编码显示,现有的研究中,除了极个别文献运用大数据相关技术处理或识别裁判文书记载的信息外,更多还是运用裁判文书来描述一种司法实践状况或者验证研究者已有的基于理论或逻辑推演出来的预设,这些都与大数据思维距离甚远。虽然法律大数据不应停留在对裁判文书的使用这一初级阶段,但裁判文书的最大优势——海量与公开,确实可以为大数据思维的研究提供最基础的前提。未来可期的一种研究路径可能是,基于对海量裁判文书的数据挖掘,由数据自行呈现出研究者并不掌握的要素、变量等之间的关联,再经由相应的研究设计并结合其他来源的数据进行验证与解释,最终助推理论发展和立法司法完善。

## 结 语

本文通过对运用裁判文书开展研究的刑事法文献的内容分析,尝试对裁判文书这一特殊数

[51] 事后分层技术的主要思想是利用关于目标总体的辅助信息,帮助改善源于样本的评估结论。这一方法被广泛用于调整有覆盖面误差的概率样本,有助于修正不平衡样本。参见前引[50],萨尔加尼克书,第98页。

[52] bootstrapping 算法主要指利用有限的样本资料经由多次重复抽样,重新建立起足以代表母体样本分布的新样本。这种方法已经在运用裁判文书的研究中使用,但其科学性可能有待进一步探讨。See Y. Xia, T. Cai, and H. Zhong, *Effect of Judges' Gender on Rape Sentencing: A Data Mining Approach to Analyze Judgment Documents*, 19 (2) *China Review* 125-149 (2019).

[53] 参见华宇元典法律人工智能研究院编著:《让法律人读懂人工智能》,法律出版社2019年版,第185页。

[54] 例如,为了验证个体法官的量刑倾向是否偏离于法官集体的量刑倾向这一假设,研究者收集了某一地区五类刑事案件判决书,利用裁判文书记载的法官姓名这一数据,以审理的法官不同为标准,比较了法官个体的刑罚裁量与法官量刑集体经验。参见吴雨豪:《量刑自由裁量权的边界:集体经验、个体决策与偏差识别》,《法学研究》2021年第6期,第109页以下。

据来源被引入法学研究的现状及引发的问题，进行较为细致的呈现和分析。受制于内容分析方法本身的局限性和研究能力，本文可能存在如下不足：（1）内容分析法局限于考察已经呈现的传播内容，<sup>[55]</sup> 本文将已发表文献作为研究对象，无法考察期刊发表文献的过程，文献中最后呈现的有关裁判文书运用状况的描述同样会受到期刊审稿标准、篇幅要求和表述方式等方面的影响，这些因素都在本文研究范围和考察能力之外。（2）隐性编码不能完全排除主观性。虽然本文采用了提高编码者间信度的方法，但隐形编码，尤其是研究设计部分涉及程度性的编码，不能完全排除编码者的主观影响。而且，由于很多文献对裁判文书运用过程语焉不详，缺少编码的充足依据也可能导致偏差。（3）本文未对单篇文献运用裁判文书是否科学进行评价。一方面，这需要结合研究主题和相关专业知识，即使限于刑事法领域也超出了笔者的能力范围；但另一方面，这也并非一般内容分析研究关注的重点。

考虑到我国当前的法律数据环境，在探知司法实践经验方面，裁判文书仍将在较长时间内占据举足轻重的地位。除了呼吁提升裁判文书上网率、明确不上网的例外规则、充实裁判文书记载的内容，以提供更好的数据来源外，如何在遵循基础研究规范的基础上，针对裁判文书的特殊性，更加科学合理运用裁判文书，同样是法学研究者共同面临的课题。

---

**Abstract:** The online disclosure of judgment documents provides researchers with a convenient way to observe judicial practice. Research findings based on content analysis of criminal law literature published in journals show that online disclosure of judgment documents promotes the development of the jurisprudential study, but cannot fully meet the needs for the exploration of judicial experience, and they also expose the shortcomings of the jurisprudential study in China in the application of research methods. Researches using judgment documents have many deficiencies in document collection, sampling, research design, data analysis, and application of research findings and conclusions, and they fail to pay attention to the special methodology for using judgment documents as a special data source. In view of the actual legal data environment in China, judgment documents will remain the main data source of empirical study for a long period of time. On the basis of following basic research norms of empirical study, Chinese researchers should correctly understand data from judgment documents, adopt various methods to deal with such problems as the questionable representativeness of judgment documents, and move towards research truly with big data thinking.

**Key Words:** judgment document, empirical study, content analysis

---

---

[55] 参见前引 [26]，巴比书，第 328 页。