

# 刑事司法实证研究： 以数据及其运用为中心的探讨

何 挺<sup>\*</sup>

---

**内容提要** 实证研究数据的核心内涵在于研究者基于自身经验的直接发现,其外延应基于这一核心内涵延展,而不受形式、载体等外在表现的过多限制。用数据进行测量是实证研究的基本观念,测量的实施需要经由概念化、操作化。刑事司法领域的测量具有自身的特点,如何在纷繁复杂的刑事司法实践中开展科学有效地直接测量,是实证研究的核心环节和难点。依据不同的标准将刑事司法实证研究涉及的数据划分为不同的类别,能够提供很多观察不同种类数据特征、优劣与运用规则的视角。刑事司法实证研究中数据的运用应当遵循寻求较佳数据、量力而行、组合运用以及不迷信数据的原则。我国刑事司法实证研究的数据运用亦存在多方面的问题值得反思。现阶段大数据尚未对刑事司法实证研究产生质的影响,但却能提供多方面的有益启示与思考。

**关键词** 刑事司法 实证研究 数据 测量 大数据

---

近十年来,我国法学领域的实证研究取得了长足的发展,在这一过程中,实证研究的独特功能与价值得到了广泛的认同,实证方法的运用至少在名称层面已经蔚然成风,在刑事司法领域尤其如此。虽然实证研究的名称(有的称为“经验研究”<sup>①</sup>)、含义及具体包括的研究类别尚存不少争议,但就其区别于纯粹思辨研究或理论研究的显著特点,却已然达成共识:研究者基于不同渠道“发现”的资料或事实进行逻辑推演进而得出“认为”的观点与结论,而不是不经“发现”即通过理论上的推演直接“认为”。从某种意义上来说,经由“发现”并据以得出结论的资料或事实——通常称之为数据(data)——是实证研究区别于其他类型研究的本质特征。进一步来说,某项实证研究所依赖的数据决定了这项研究的具体方法,同时也决定了这项研究所能达到的深度与广度,数据及

---

<sup>\*</sup> 北京师范大学刑事法律科学研究院副教授、法学博士。本文系国家社会科学基金青年项目“刑事司法改革中的实验研究”(项目批准号:12CFX036)和北京高等学校青年英才计划“刑事司法实证研究方法”(项目批准号:YETP0239)的阶段性成果。

<sup>①</sup> 关于实证研究名称与含义的不同观点,参见刘建宏《“经验”方法与法学研究》,载《西南政法大学学报》2007年第3期。

其运用之于实证研究方法不啻具有生命线一般的意义。

## 一、实证研究中的数据

### (一) 核心内涵:研究者基于自身经验的直接发现

研究方法是指在研究中发现新现象、新事物,或提出新理论、新观点,揭示事物内在规律的工具、手段,是运用智慧进行科学思维的技巧。<sup>②</sup> 随着新类型研究方法的不断涌现,如何按照不同的标准对各种研究方法进行分类有不同的观点,而研究所依赖的“论据”不同则是分类的一个重要标准。区分思辨研究与实证研究的关键亦在于思辨研究主要依据研究者主观上的逻辑或价值推演,而实证研究则依赖于数据这一研究者所“发现”的经验事实。

数据是指“进行各种统计、计算、科学研究或技术设计等所依据的数值”。<sup>③</sup> 这一界定从日常汉语的一般性理解出发,并因此将数据的范围界定为可以量化的数值。但作为实证研究所依赖的数据,则不应采取如此狭义的界定。从语义分析的角度,与数据相对应的英文单词“data”则被解释为作为依据的事实与信息。<sup>④</sup> 在国外研究方法教科书中,还将数据分为“软数据(soft data)”与“硬数据(hard data)”。所谓软数据,指的是以感受、文字和照片等形式呈现出来的数据,而硬数据则是指以数字的形式呈现出来的数据。对于这两种数据,应当按照各自的特点在研究中发挥其作用。<sup>⑤</sup> 事实上,一项研究是否基于可以量化的数值并非判定其是否属于实证研究的关键,而是用于区分定量研究与定性研究的标准。在基于可以量化的数值所开展的定量研究中,如果使用的是研究者基于自身经验所发现的数值,则属于实证研究,相反,如果完全基于非自身经验的间接渠道所获的数值则不能称之为实证研究。实证研究强调的是获取数据的途径与研究者自身经验之间的密切程度,而非数据的形式。实证研究的独特价值,除了在“发现”的基础上进行逻辑推论,更重要的其实在于这种“发现”是研究者自身直接而非间接的发现。正如学者所指出的,经验研究是一种研究的进路,这种进路强调对现象、事实的清楚把握,对事件、行动发生的条件、逻辑进行深入考察,对法律、制度运行的实际有真切把握。<sup>⑥</sup>

之所以强调实证研究数据需由研究者直接获取,主要基于以下几方面:首先,研究者基于具体研究的需求设计相应的数据收集方法,只有研究者直接获取的数据才与研

② 参见叶继宁等编著《学术规范通论》,华东师范大学出版社2005年版,第84页。

③ 《现代汉语大词典》,上海辞书出版社2009年版,第2333页。

④ Data 被解释为“facts or information, especially when examined and used to find out things or to make decisions”。参见《牛津高阶英语词典》(第六版),商务印书馆·牛津大学出版社(中国)有限公司2004年版,第318页。

⑤ See Peter B. Kraska & W. Lawrence Neuman, *Criminal Justice and Criminology Research Methods*, at 148-149 (Pearson Allyn and Bacon, 2008)。

⑥ 参见王启梁、张剑源主编《法律的经验研究:方法与应用》,北京大学出版社2014年版,第3页。

研究者的需求最为“匹配”，而间接获得的数据则可能是其他研究者为了其他目的而收集的，不一定完全符合研究者的需求；其次，对于间接获得的数据，研究者通常不清楚其样本的选择、收集的具体方法和过程，对其是否经过筛选、加工、整理甚至篡改，研究者也往往不具有判断的能力；再次，任何数据都有其优势与缺陷，而在真实的刑事司法情境中收集任何种类的数据更可能因各种原因未达预期。对于研究者而言，只有对自身直接获取的数据才清楚其内在的缺陷与优势，才能更为精确地经由数据推导出适当的结论。因此，实证研究数据的核心内涵应当在于研究者基于自身经验的直接发现。在一项实证研究所运用的数据群中，必须包含源自研究者自身经验直接发现的数据，或者从更高的要求来说，这些直接发现的数据应在数据群中居于核心，否则就不能称之为实证研究。

当然，虽然强调实证研究数据的核心内涵是直接获取，但并不完全排斥间接获取数据的运用。相反，在一项实证研究开始阶段的探索性研究以及研究结论的比对与推广方面，间接获得的数据甚至是不可或缺的。

## （二）实证研究数据的外延

实证研究数据不仅不限于数值，而且应当围绕上述核心内涵对其外延进行充分地延展。美国学者曾对实证研究的数据外延进行了非常宽泛的界定。“正如自然科学家和社会科学家都认同的，实证研究的范围很广。‘实证’一词指的是基于观察或者经验而获得有关客观世界的证据。这种证据既可以是数量化的（定量的），也可以是非数量化的（定性的），数量化或非数量化的证据在‘实证’的程度上并没有高低之分。一项研究只有建基于对客观世界的观察才能称之为实证研究，易言之，在实证研究中，数据是一个用于表述客观事实的术语。……数据可以是精确的，也可以是模糊的；可以是相对确定的，也可以是非常不确定的；可以是直接观察所得的，也可以是经由收集工具获得的；可以是人类学的、解释性的、社会学的、经济学的、法学的、政治学的、生物学的、物理学的甚至是自然的。只要研究包括观察所得或所期望获得的数据，这就是一项实证研究。”<sup>⑦</sup>这一界定为实证研究可资运用的数据范围提供了很好的启发。某种意义上，研究者不应事先对一项实证研究可以运用的数据的形式、载体、来源等外在表现作过多地限制，而应该用最大的想象力来拓展其范围。研究者只应该按照两项标准来筛选可供运用的数据：一是某项研究的具体情境能够提供或者产生出哪些数据，或者说，在研究的具体情境中，基于具体的数据收集方法以及客观条件、研究伦理等的考量，哪些数据是可以获得的；二是某项具体研究需要哪些数据，或者说，某项研究所提出的问题或假设，需要哪些数据去回答或验证。如果符合上述两方面标准，任何形式的数据都具有成为实证研究数据的资格。在刑事司法实践中，数据的具体形式同样可以非常多样，除了常见的法律文书和案卷材料以外，包括办案系统内的记录、庭审录像、讨论记录、相关人员

<sup>⑦</sup> Lee Epstein & Gary King, *Empirical Research and the Goals of Legal Scholarship: the Rules of Inference*, 69(1) U. Chi. L. Rev. 1, 2-3(2002).

的评价和表达、发表的文献、甚至涉案人员的言谈举止抑或办案机关工作场所的陈设,在针对特定研究主题时,都可以成为数据的来源或载体。如果这些数据是经由研究者自身经验而直接获得的,则会使这项研究成为真正的实证研究。

在考虑实证数据的外延时,还需要发展的眼光:随着技术、法律或其他影响因素的变化,实证研究数据的外延也可能发生相应的变化。

## 二、实证研究的基本观念:用数据进行测量

实证研究数据的核心内涵和外延仅仅解决了数据在实证研究中的基本“准入”资格,至于在一项特定的研究中应该运用哪些种类的具体数据及如何运用,还需要在数据与拟回答的问题或拟验证的假设之间建立联接,这一联接的过程实际上体现了一种用数据进行测量(measure)的观念。

### (一) 刑事司法实证研究中的测量

科学研究中的测量,则是指对特定的研究对象进行有效地观测与量度,并在观测与量度的基础上描述研究对象。与日常生活中的测量通常具体、直接且易于理解相比,科学研究中的测量更为复杂。绝大多数情况下,科学研究就是要通过测量的方法,使原本不能直接观察到的事物得以清晰呈现;或者使原本抽象、难以理解的事物变得具体,可以为更多的人所理解;或者使原本不能直接比较的事物变得可以等量齐观。科学研究中的测量延展了人类观察所及的范围,并为通过归纳或演绎形成普适的理论提供了基础。

自然科学与社会科学研究中的测量也有着显著的差别。在自然科学研究中,虽然也会采用间接的方法测量一些“肉眼”无法观察到或无法直接测量的研究对象,但总体而言,其测量对象大多属于客观存在的事物,测量的过程相对稳定并可不断重复,而且通常能够建立起公认、通用的测量标准(例如公里、克、赫兹等)。与此相比,社会科学以社会现象为研究对象,其测量具有更大的弹性和不确定性。一方面,社会科学研究中,人是最基本的研究和测量对象,人所具有的主观意识、价值观念和个性特征,都会使测量处于不稳定的状态,人与人之间的异质性,也会使社会科学的测量无法通过完全的重复提升其可靠性。另一方面,社会科学研究涉及的概念与理论大多抽象且存在于观念之中,将其具体化的过程更为复杂且歧义丛生,很难达成如自然科学一般公认、通用的测量标准。例如,有关经济发展状况,就有各种不同的测量标准。此外,虽然“测量”一词天然地与数量相关联,但在社会科学的定性或质性研究中,同样存在测量。与定量研究在准备与计划阶段就将概念转换为特定的测量方案之后再行测量不同,定性研究中的测量往往与思考概念和收集数据的过程混合在一起,即在收集数据的过程中,才尝试

将数据与概念联系在一起。<sup>⑧</sup>

刑事司法实证研究中的测量同样具有很大的弹性和不确定性,并有自身的特点。一方面,刑事司法以特定的人——与犯罪有关的人和案件为基本的研究和测量对象,人所具有的主观特性和案件与案件之间的异质性,同样使测量处于不稳定的状态且不能完全重复。而且,由于与犯罪有关的人所处的特殊状态(例如被害人不可接触、被告人被羁押)、特定案件种类的限制、案件处理过程不可重现以及法律适用的平等性等因素的影响,可能使得最佳的测量对象与测量方法不可实现,因而不得不转向次级的测量对象与方法。除此以外,刑事司法领域还必须面对理论的归纳、法律的规定与实践的操作三者之间不可完全消弭甚至巨大的鸿沟。另一方面,刑事司法的许多概念与理论同样是以观念的形式抽象存在的,很多概念与理论的内涵尚未达成共识,将之与刑事司法实践中的具体表现联接的过程更是争议满途,因此可能形成这样的状况:不同的研究者采用不同的标准和方法对同一主题进行测量并得出不同甚至截然相反的结论。<sup>⑨</sup>例如,目前国内讨论颇为热烈的法治指数测量也面临着各种不同测量方法和标准之间的争议。在实证研究相对发达的犯罪学领域,缺乏统一测量标准的问题同样存在。正如有的学者在考察世界犯罪学实证研究后认为,验证犯罪学理论面临着缺乏标准化的测量量具的严重问题。<sup>⑩</sup>因此,如何科学地进行测量是刑事司法实证研究的核心环节和难点所在,同时亦是其魅力之源。如何在纷繁复杂的刑事司法实践中寻觅最为恰当且具有可操作性的测量标准,并采用相对较为科学的方法直接进行测量,是实证研究所独具的研究体验。基于社会科学研究很难就测量标准完全达成一致这一前提,刑事司法实证研究中对于相同或相关的研究对象存在不同的测量标准和方法亦属正常。在某种意义上,刑事司法实证研究的深化恰恰体现在测量标准与测量方法的演进上:测量标准通过争鸣与讨论逐步形成一定的共识,而测量方法则通过比较与实践进而更为科学。

## (二) 测量的基本过程

测量是一个将抽象的概念、理论与具体的实证数据相联接的过程,也是一个用实证数据来描述研究对象的过程。按照社会科学研究的一般方法,测量必须经历概念化与操作化两个步骤始能付诸实施。

### 1. 概念化

在实证研究中付诸测量的概念,并不等同于教科书上所下的定义,而是包括所有经过概念化后的抽象观念。例如刑事司法实证研究中可以测量“逮捕”、“审前羁押”等教科书中的概念,但可能更多的是测量类似于“审前羁押对案件处理的影响”、“审判公开的程度与效果”这种有别于普通概念,但却以抽象观念的形式存在于研究者思维之中的

<sup>⑧</sup> 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123-124页。

<sup>⑨</sup> 一个可以类比的例子是大学排名,不同的机构采用不同的测量标准和方法会得出不同的大学排名,有的还可能差异较大。

<sup>⑩</sup> 参见曹立群、周慷嫻《犯罪学理论与实证》,群众出版社2007年版,第36页。

内容。

所谓概念化,并不仅指对研究对象或研究主题给出一个定义,而是指对抽象的概念进行清晰的界定与详述的过程。概念化需要完成两方面的任务:一方面是表达、澄清和界定概念,将存在于观念之中的概念用文字予以明确的表述,并划清其与相近概念的界限,指出概念包括什么、排斥什么;另一方面则是对概念的内涵进行详述,细化出能够体现概念内涵的不同维度。例如,在英国学者开展的对公设辩护质量的评估研究中,将“公设辩护的质量”这一较为抽象的概念详述为案卷、交流、信息与事实收集、法律意见与协助、工作/协助、效率、职业伦理与整体表现等七个维度。<sup>⑪</sup>在我国学者开展的隔离式量刑程序试点研究中,在确定隔离式量刑程序概念的基础上,进一步将其效果或给刑事诉讼所带来的变化详述为透明度、满意度、所需时间和量刑结果四个维度。<sup>⑫</sup>概念化是整个测量的基础。“很多人将测量视为刑事司法研究中最重要同时也是最困难的阶段。之所以困难,在某种程度上是因为,在刑事司法领域中有如此多的基本概念不像我们希望的那样进行了明确的界定。如果我们不能进行实现概念化,我们会发现操作化和测量都会极富挑战。”<sup>⑬</sup>概念化的结果是使研究主题明确、清晰、丰满,并具有与经验事实及实证数据相联接的可能性。

## 2. 操作化

与概念化相比,操作化更为接近实际测量的过程。操作化将抽象的理论与具体的经验现象联接起来,使科学知识能够被经验事实所检验。操作化的结果通常是将抽象的概念转化为操作定义。所谓操作定义,就是建立一些具体的程序或指标来说明如何测量一个概念。<sup>⑭</sup>

操作化的核心内容是确定测量的标准与具体方法。测量标准也可称为测量指标,是指用具体的经验现象衡量是否符合抽象概念及其程度的标准。上文提及的研究案例同样在概念化的基础上确立了具体的测量标准。隔离式量刑程序研究中,在隔离式量刑程序所需时间这一维度上,用量刑过程持续时间、庭审时间、开庭次数、定罪量刑之间休庭时间、立案至开庭再至结案时间和羁押时间等六方面的具体标准来测量其与普通审判程序的差异。<sup>⑮</sup>通过将概念中不同维度的内涵进一步具体化为测量标准,实际上将经过概念化后仍属抽象的研究主题进一步“转化”或“翻译”为更为具象的表现,并指向那些在刑事司法过程中可以观察到并记录下来的经验事实。测量指标的确定实际上已经“圈定”了实证研究拟收集的数据类别与范围。

⑪ 参见 [英] 李·布里奇斯等《英国公设辩护质量评估:基于同行评议的方法》,载何挺等编译《外国刑事司法实证研究》,北京大学出版社2014年版,第241-242页。

⑫ 参见陈卫东、程雷《隔离式量刑程序实验研究报告——以芜湖模式为样本》,载《中国社会科学》2012年第9期。

⑬ Micheal G. Maxfield & Earl R. Babbie, *Research Methods for Criminal Justice and Criminology (international edition)*, at 130-131 (Wadworth Cengage Learning, 2011).

⑭ 参见袁方主编《社会研究方法教程》,北京大学出版社1997年版,第12页。

⑮ 参见前引⑫,陈卫东、程雷文。

测量的具体方法则指在刑事司法情境中将测量付诸实施的具体方法,其要解决的是测量哪些具体对象以及如何具体测量的问题,通常意义上所谈论的实证研究收集数据的具体方法均在其列:第一,选取测量的地点和时间段;第二,确定分析单元,以使所有的测量标准都能在分析单元中呈现,并符合刑事司法实践的具体状况;第三,选择具体的研究对象及无法开展全样本研究时采用不同的抽样技术抽取样本;第四,在有多种数据来源时选择更好、更具操作性的数据;第五,选择具体的数据收集途径;第六,设计收集数据的具体工具,包括问卷与量表、观察记录表格及访谈提纲等;第七,设计比较的具体方案;第八,在测量的过程中尽量符合伦理的要求等。经由对这些具体问题的回答,测量的过程实际上已经延伸至具体收集的数据,并在抽象的研究内容与数据之间建立了联接。

“一项具体社会研究的结果,同它所采用的操作化方式及其所产生的测量指标密切相关。”<sup>①⑥</sup>在刑事司法实证研究中,测量标准与测量方法实际上决定了研究成果的呈现形式和研究结论的依据,并进而决定了研究的结论,其重要性显而易见。基于社会科学研究测量标准难以完全达成一致这一前提,刑事司法实证研究当然也应当允许存在对同一研究主题采用不同的测量标准与测量方法,而且相较于其他社科领域,刑事司法领域的测量标准更可能因为不同法域的不同传统与法律规定、地点、实践等因素的影响而有较大的差别。例如,有关辩护律师作用的测量标准,国内的研究者采用会见、举证、质证的频度、内容、方式,辩护意见提出的频度和内容,庭审耗时,辩护意见采纳率,定罪率/无罪判决率与撤诉率和量刑的刑种与量刑幅度等为标准,而国外不同的研究者则分别采用了辩护律师提出的审判动议数量,诉讼周期,辩诉交易率、定罪率、撤案率和量刑严重性等标准。<sup>①⑦</sup>但这并不意味着测量标准与测量方法可以随意确定,而应当从研究的目的出发,参酌理论的预设、法律的规定、实践的操作与其他研究者的已有测量标准和方法。测量标准与测量方法的选择应当尽量提高测量的效度和信度。<sup>①⑧</sup>当然,在刑事司法实证研究中,测量标准与测量方法的选择还受制于司法实践的状况、研究的能力与经费预算等客观因素,在一些具体领域甚至受限较大。因此,在受限于各种客观条件的情况下,如何在客观条件允许的范围内选择效度和信度相对更高的测量标准与方法,是刑事司法实证研究所必须面对和解决的问题。

### 三、刑事司法实证研究的数据分类

数据分类,是针对刑事司法实证研究中可能运用的数据,根据数据的来源、表现形

<sup>①⑥</sup> 风笑天《社会研究方法》(第4版),中国人民大学出版社2013年版,第88页。

<sup>①⑦</sup> 参见马静华《辩护律师作用实证研究——以D县为个案》,载《南京大学法律评论》2012年春季卷。

<sup>①⑧</sup> 效度指的是测量的有效度,所要回答的问题是“我所测量的正是我想要测量的东西吗?”信度则指测量的可靠性,所要解决的问题是“当我采用这种测量方法进行测量时,是否每一次都能得出相同的结果?”

式、产生时间等方面的区别,考虑其可能对实证研究产生的影响,依据不同的标准对数据进行的逻辑上的分类。对刑事司法实证研究数据的分类借鉴了证据法学中证据分类的研究思路,目的在于从不同的角度揭示各类数据的特点、优劣与运用规则,拓展在实证研究中组合运用不同类型数据的视野,提升实证研究具体方法的科学性。

### (一) 定量数据与定性数据

定量数据与定性数据的分类源于定量研究与定性研究的分野。定量研究,也称为量化研究,简而言之就是用数值或数量来进行描述和分析的研究;定性研究,也称为质性研究,指用非量化的方法从事物的性质方面开展的研究。定量研究与定性研究的最重要区别就在于依据的数据不同。定量数据是定量研究所主要依据的数据,是基于一定规模或数量的研究对象而获得,以数字的形式呈现出来并可以进行统计分析的数据;定性数据则是定性研究所主要依据的数据,即以感受、文字和照片等形式呈现出来的对研究对象的性质、特征进行描述或概括的数据。两者的最大差别在于是否数值化及能否基于一定数量的样本进行量化分析。

在刑事司法领域,基于一定数量的案件或一定数量与犯罪有关的人所收集到的同一内容指向的数据大都属于定量数据,包括:与案件有关的人的特征,既包括年龄这样本身就数值表现出来的特征,也包括经过归类赋值后的特征,例如将犯罪嫌疑人的家庭经济状况归纳为“较好”、“一般”和“较差”三类并赋予不同的值;一定数量案件所产生的总数、频数、比率等数值,例如审前羁押率、无罪判决的数量等;犯罪客观方面所呈现出来的各种数值,如受贿犯罪的数额、毒品犯罪的毒品数量;案件办理的延续时间、判决和执行的刑期等。刑事司法领域(包括其他法学领域)中法律适用的普遍性和案件的规模化为获取定量数据开展定量研究提供了最基本的前提——具有一定数量规模的样本,因此与社会学、人类学领域相比,刑事司法领域中的定量数据可能更为常见。<sup>①</sup>

刑事司法领域中常见的定性数据包括:对当事人、办案人员等人员访谈所获得的资料;判决书等法律文书记载的文字性内容;法律法规文本;案件办理过程中产生的录音录像资料;个案资料等。定性数据还可以借助内容分析法中的编码技术转化为定量数据。例如,在一项有关社区矫正的实证研究中,在分析缓刑官与矫正对象谈话内容时,把每次谈话的录音分为若干个五分钟的片段,再由编码员按照不同的标准进行编码(计数与打分)。通过这样的编码方法,就把谈话这一内容纷杂、形式多样的定性数据转化为标准化的、可供比较的定量数据,从而可以对缓刑官与矫正对象的谈话内容做定量的分析。<sup>②</sup>

相对于定量数据基于大量的样本所具有的“规模性效应”,定性数据的功能与作用

<sup>①</sup> 根据统计学原理,定量数据还可以根据测量水平和可以进行的运算种类进一步划分为称名数据、顺序数据、等距数据和等比数据。其中称名数据、顺序数据只具有分类或排序的功能,不能进行四则运算,等距数据只能进行加减运算,等比数据则可以进行加减乘除四则运算。在刑事司法领域,有许多原始的定量数据都属于前三类,因而可以进行推断统计的种类相对较为受限。

<sup>②</sup> James Bonta, Guy Bourgon & Tanya Ruge et al., *An Experimental Demonstration of Training Probation Officers in Evidence-based Community Supervision*, 38(11) *Criminal Justice and Behavior*, 1127, 1133(2011)。



易被忽视。“有些人错误的认为,定量数据总是更有说服力,而质性数据则总是说服力不足。只有数据标准才能决定数据是否有力,而不是数据的定量或定性的形式。可靠、有力的数据与其定量或定性的形式无关,而是由研究者搜集数据时的小心谨慎和系统性程度决定的。”<sup>②</sup>定量数据与定性数据本身并无高低之分,而是各具优势与缺陷,选用哪一类型数据完全取决于研究的需要和现实可操作性。定量数据更为标准化、精确化,可以进行比较并借助统计分析得出直接的结论,但其数值化的过程可能会损失刑事司法有关背景、过程等的丰富含义;定性数据的代表性和普适性相对较差,不能直接进行比较,但往往更为形象具体,能够深入现象内部。一项理想的刑事司法实证研究应当综合运用定量数据与定性数据,集二者之所长。

## (二) 主观性数据与客观性数据

以数据来源的不同为标准,可以分为主观性数据与客观性数据。主观性数据,指的是源自调查对象主观认识与判断的数据;客观性数据,则是指源自客观发生的现象、行为、状况等的数据。划分主观性数据和客观性数据的标准并不是数据描述的内容,而是数据的来源。在刑事司法实证研究中,一般而言,通过问卷调查、访谈和心理测量量表等所获得的数据是主观性数据,而无论其描述的内容是什么,因为其所收集的是被调查者主观的认识与判断。而有关案件的处理情况和程序进展、证据的运用状况、办案机关所记录的系统数据或台账数据、通过调查与回访所了解到的与犯罪有关的人员的特征与生活状况(例如是否重新犯罪)等则属于客观性数据。这些客观性数据通常通过查阅案卷、调取办案机关记录的数据、观察、回访等方式收集。

之所以划分主观性数据与客观性数据,是因为这两类数据的特征决定其各擅胜场,主观性数据用于测量相关人员的主观认识,客观性数据则用于测量客观状况,不适当地使用甚至调换会明显降低测量的效度和研究的科学性。从某种意义上来说,一项实证研究的内容甚至名称,基本上已经决定了此项研究所主要依据的是主观性数据还是客观性数据。例如,“新刑事诉讼法实施状况研究”的目的是调查新刑事诉讼法的适用状况及存在的问题,毫无疑问应以客观性数据为主。“司法公信力研究”或“公众对于死刑存废的态度研究”等研究的目的在于了解公众或特定人员对于某一事项的主观态度与判断,毫无疑问应当以源自调查对象主观认识的主观性数据为主甚至以其为唯一的数据来源。<sup>②</sup>如果在有关某一法律制度实施状况的调查中,主要采用访谈办案人员的方法来收集数据,甚至完全依据主观性数据得出结论,其测量的效度和研究的科学性是相对较低的;反之,在主要应当依据主观性数据的研究中,不适当地加入甚至主要依据客观性数据亦同。当然,这是在理想状况下的最佳选择,刑事司法实践状况和现实可行性也会影响到数据种类的选择。

<sup>②</sup> 前引⑧, [美] 劳伦斯·纽曼书,第12页。

<sup>②</sup> 例如,研究者认为,“检察公信力测评”的评价方式具有主观性特点。参见谢鹏程、邓思清《检察公信力测评指标体系研究》,载《人民检察》2014年第23期。

还需要说明的是,主观性和客观性数据的分类有时也会存在模糊的中间地带,即某一数据虽然呈现的是客观的状况,但却是经过主观整理和判断的结果。例如,根据心理学的研究成果,将未成年人父母的教养方式划分为冷漠型、宽容型、溺爱型和专制型四类,而在研究父母教养方式与未成年人犯罪及转处措施的关系时,需要研究人员(或开展社会背景调查的人员)针对每个个案的客观情况,根据四种类型的特征进行主观判断,最终呈现的数据其实是一种主客观相结合的数据。这种情况在刑事司法实证研究中并不少见,尤其是需要将客观的状况(定性数据)判断、归于某一类别(定量数据)时,例如判断犯罪嫌疑人的家庭经济状况、社会危险性等等。对于此类主客观相结合数据的收集与运用,需要在具体操作技术上予以特别的关注,例如明确和细化判断标准以及增加作出判断的人数等。

此外,基于特定的研究需要,还可以主动将客观性数据转化为主观性数据。例如,在研究法律职业人员办理案件的质量时,考虑到法律职业人员的工作是以异质的个案为基础的,案件具体情况影响甚至决定了办案的具体过程和结果,无法用整齐划一的客观标准对其质量进行评价,因而需要由同行专家通过阅读办案材料或者直接观察,转化为等级、评分等专家的主观判断,从而将客观性数据转化为主观性数据,而这种处理同时也实现了定性数据向定量数据的转化。

### (三) 原生数据与新生数据

根据数据产生的时间与开展实证研究的先后关系以及数据的产生是否依附于实证研究为标准,可以将刑事司法实证研究中涉及的数据分为原生数据与新生数据。所谓原生数据,即原本就会产生并以原始状态存在的数据,包括在刑事司法实证研究开展之前即已存在的数据,以及在实证研究期间未受实证研究影响而自行产生的数据;所谓新生数据,指由于开展实证研究才新产生的数据。区分原生数据与新生数据的关键在于,数据的产生是否是实证研究所导致的,或者说数据是否是基于实证研究的需要而新产生的。在刑事司法领域中,常见的原生数据包括官方发布的数据、办案机关的系统或台账数据,以及案卷材料与法律文书中记录的数据等。更为常见的则是因为实证研究而新产生的新生数据,包括问卷调查、访谈所获的数据,观察所记录下的数据以及在实验研究中记录下来的描述实验过程并验证假设的大部分数据。

不同类型的研究对于原生数据与新生数据的倚重有所不同。有的研究主要甚至完全依赖新生数据,例如,上文提到的对公众开展的问卷调查研究,主要依赖于经由问卷调查所获得的新生数据——这些数据如果不开展这项问卷调查是不会产生的。有的研究则主要依赖原生数据,例如主要依赖官方发布的数据和办案机关系统或台账数据开展的对我国刑事诉讼运行状况的研究。在调查法律制度实施状况的实证研究中,也可以综合灵活运用原生与新生数据。例如,在“合适成年人”制度实证研究中,既使用了办案机关之前所记录下来的原生数据,也使用了通过问卷调查、访谈在研究过程中收集的

### 新生数据。<sup>②</sup>

划分原生数据与新生数据的目的主要在于凸显这两类数据与实证研究的匹配度上的差异。在实证研究收集所需数据之前,都要进行研究设计,而研究设计的核心就是基于测量的需要确定数据收集的具体方法。新生数据产生于数据收集具体方法确定之后,基于实证研究的需求并根据事先设计的规范严格的方法进行收集,因而与实证研究对数据的要求具有较高甚至完全一致的匹配性。一般而言,如何对新生数据进行分析运用在数据收集之前也已基本成型,除非数据收集过程中遇到了事先未能虑及的影响因素,按照设计好的数据收集方法所获的新生数据可以直接进入数据分析的阶段。原生数据产生于研究设计之前,或者虽产生于研究设计之后却并非基于研究的需求,因而与实证研究的数据要求匹配性较差,规范性也相对较低,收集之后需要进行系统的筛选、分类、整理甚至技术处理之后才能运用,还可能存在原生数据缺失的情况。例如,在利用案卷和办案机关的系统数据进行研究时,常常会遇到案卷不齐、系统中记录的数据不全的问题。而且在刑事司法领域,有些实证研究所需的数据可能只有在案件办理过程之中予以记录才会留存,一旦案件办理终结则无法回溯补充。基于此,主要依赖于原生数据的研究在进行研究设计时很大程度上还受到存在什么样的原生数据可供收集的掣肘,甚至有时研究的科学性也不得不因此妥协。

划分原生数据与新生数据给刑事司法实证研究带来的另一个重要启示是,在主要依赖于新生数据的研究中,应当在研究设计阶段根据研究的需求全面思考、反复推敲,务必使数据收集方案和收集工具涵盖所有拟收集的新生数据,否则,本与研究匹配度极高的新生数据可能会在研究过后,因为时过境迁“退化”为匹配度较低的原生数据。

#### (四) 关系主体数据与其他主体数据

根据数据来源主体与研究主题之间是否具有利害关系,可以分为关系主体数据和其他主体数据。关系主体数据,是指来源于与研究主题有利害关系的主体的数据。“利害关系”,指的是特定主体所具有的身份和其在刑事司法中所处的地位,会影响其对于特定研究主题的认识与看法,并进而影响其所提供的数据。在开展这些主题的研究时,来自这些关系主体的数据属于关系主体数据。其他主体数据则是指来源于与研究主题没有利害关系主体的数据。

之所以如此划分,是因为这两类数据具有两方面截然不同的特点。关系主体由于对研究主题所涉较深,其对研究主题的了解或思考也通常较为全面、深入,因此关系主体数据通常覆盖面广且内容丰富,而且有很多研究必须要有来自关系主体的数据才能体现“经验性”。同时,关系主体在刑事司法中所具有的特定身份也会影响数据的真实性、客观性和中立性。例如,当事人对于某项刑事司法制度的看法与感受,很大程度上会受其自身案件处理结果的影响,因此无论是问卷调查还是访谈,都应当通过具体的设问予以区分和细化,并与多种来源的数据进行甄别、核实。其他主体数据的特点则截然

<sup>②</sup> 参见何挺《“合适成年人”参与未成年人刑事诉讼实证研究》,载《中国法学》2012年第6期。

相反,由于其他主体对研究主题所涉不深,因此了解有限,数据的信息量较小,但同时其可靠性与中立性却相对较高。在一些研究中,其他主体数据甚至是不可或缺的核心数据。当然,是否“有利害关系”并非一个客观明确的界分标准,应当由研究者基于研究的具体情况判断,这一数据分类的目的在于提醒研究者对于来自不同主体的数据,需要根据各自的特点区别对待。

#### (五) 一手数据与二手数据

基于数据是否由研究者所直接收集,可以划分为一手数据与二手数据。一手数据,也称为原始数据,即研究者为了某项研究所直接收集的新的数据。二手数据,则是指研究者所运用的其他研究者为了其他目的而收集的数据。<sup>②④</sup> 一手数据与二手数据的划分是相对的,研究者在某项研究中所直接收集的一手数据在其他研究者的研究中则为二手数据。在刑事司法领域,常见的二手数据包括官方(全国或地方)公布的数据、非政府组织发布的报告中的数据、其他研究者收集的数据和其他相关调查报告中的数据等。

正如前文所指出的,实证研究强调由研究者直接获取一手数据,但并不排斥二手数据的运用。与二手数据所具有的“非直接性”这一“天生缺陷”伴生的则是其在收集、获取方面的便利性这一“天生优势”。与研究者走入真实的司法情境直接收集数据需要投入的经济、时间和人力成本相比,收集二手数据所需要的更多是查阅、检索和筛选资料,在降低数据收集成本和提升效率方面具有明显优势。“社会学家不必总是出去收集新资料。以前别人所作的研究在经过重新提炼、整理和分析之后,又变成了回答新问题的主要信息来源。”<sup>②⑤</sup>

二手数据在刑事司法实证研究的多个方面都有运用的价值。首先,在开展正式研究之前的探索性研究中可以通过二手数据了解研究现状、帮助形成研究假设和确定研究的具体问题;其次,可以运用二手数据为拟研究的具体问题描绘出整体背景与状况;再次,可以使用二手数据进行效果的比对,例如在某地开展司法改革试点的效果可以与未开展试点地区的二手数据进行比较;最后,可以使用二手数据来比较、印证、调整研究结论,并有助于将研究结论推广适用于更大的范围。虽然二手数据的收集比较方便,但二手数据是别人或其他研究机构为特定目的收集的,不一定符合研究者的需要,研究者或者要将就二手数据的限制,或者要综合多种数据来源进行分析。<sup>②⑥</sup>

与二手数据相关的是二次分析(secondary analysis)。二次分析也称为二手分析,是利用现有二手数据的一种非介入性研究。二次分析指的是对最初为其他研究目的所收集的数据进行再次分析,是在数据收集环节为研究者节省时间的极好方法,也是对现代社会堆积如山的数据进行有效利用的方法。<sup>②⑦</sup> 只有其他研究者收集的原始数据才能进行二次分析,即二次分析并不是对其他研究者经过统计分析、处理后的结论性数据的引

<sup>②④</sup> 参见麦高伟、崔永康主编《法律研究的方法》,中国法制出版社2009年版,第61页。

<sup>②⑤</sup> [英]戴维·波普诺《社会学》,李强等译,中国人民大学出版社2007年版,第59页。

<sup>②⑥</sup> 参见前引⑩,曹立群、周榛桐书,第24页。

<sup>②⑦</sup> Frank E. Hagan, *Research Methods in Criminal Justice and Criminology*, at 237 (Pearson Education, 2003).

用或再次利用,而是对其他研究者基于其他研究目的所收集到的未经统计分析的“原生态”数据,采用不同于数据收集者的分析方法所进行的新的分析,或者为了新的研究目的所进行的再次分析。

是否能够运用其他研究者的原始二手数据进行二次分析研究,取决于原始二手数据公开和可供获取的机制与整体的数据环境。美国学者认为,在犯罪学与刑事司法领域开展二次分析研究的机会几乎是无限的,因为已经收集了大量与犯罪有关的数据并建立了数据库。研究者对于他们感兴趣的题目,通过这些数据库能够获得数以千计的数据来源,而且由于这些数据是由国家提供的研究资助所收集的,因此这些数据来源常常是公开的或者是可以免费获取的。<sup>②⑧</sup>在我国刑事司法领域,无论是官方公布的数据还是国家资助的研究项目,对于收集到的原始数据都很少全面提供并公开,在我国刑事司法领域进行利用二手数据的研究尚待数据公开机制的完善和整体数据环境的建设。

#### (六) 真实数据与模拟数据

按照数据产生情境的不同,可以将数据分为真实数据与模拟数据。真实数据产生于真实的刑事司法情境和过程之中,模拟数据则源自为开展实证研究所人为模拟出来的情境。真实数据与实证研究所强调的“经验性”或“直接性”更为契合,但特定研究主题无法或不便在真实的刑事司法情境中开展研究和收集数据时,以模拟研究的方式收集模拟数据也是一种必要的替代。

常见的模拟研究可以分为三类:一是“真实主体”+“模拟事项”,即由真实的法官、检察官等刑事司法主体判断、处理模拟出来的刑事司法事项。例如,在对有证人出庭的案件进行真实审理的同时,由另外的真实的法官进行模拟的书面审理,以比较证人是否出庭对于定罪量刑的影响。<sup>②⑨</sup>二是“模拟主体”+“真实事项”,即由模拟出来的刑事司法主体对真实的刑事司法事项进行判断、处理。例如,在真实案件审理时,自始至终在旁听席上听审的影子陪审团对案件如何裁判进行评议,以研究陪审团的评议过程。<sup>③①</sup>三是“模拟主体”+“模拟事项”,即主体和事项都系模拟的研究,例如招募大学生组成陪审团对非真实的案件进行模拟审判研究。<sup>③②</sup>

与真实数据相比,除了数据收集方法本身是否科学外,模拟数据的质量以及模拟研究中测量的效度,在很大程度上取决于模拟的情境与其所代表的真实情境之间的差距,差距越小则质量和效度越高。上述三种模拟研究中,前两种由于具有一半的真实因素,其与真实情境之间的差距通常要小于第三种模拟研究。即使是同一种模拟研究,也应当在具体细节上尽量缩小其与真实情境之间的差距。

#### (七) 官方数据与研究者自行收集数据

<sup>②⑧</sup> Richard D. Hartley (ed.), *Snapshots of Research: Readings in Criminology and Criminal Justice*, at 313 (Sage Publication, 2011) .

<sup>②⑨</sup> 参见左卫民《证人出庭作证程序:实证研究与理论阐释》,载《中外法学》2005年第6期。

<sup>③①</sup> 参见 鞠 李哉协等《韩国影子陪审团评议过程实证研究》,何挺译,载《国家检察官学院学报》2016年第1期。

<sup>③②</sup> 参见 [日] 藤田政博《日本裁判员制度的实证研究》,肖萍译,载前引①①,何挺等书,第123-137页。

按照数据来源的不同,可以将数据分为官方数据和研究者自行收集的数据。如上所述,官方数据是一种二手数据,其优点在于覆盖的样本规模较大、内容全面、具有一定的权威性并且常常是公开的,方便研究者获取。但其缺陷也显而易见,尤其是官方数据通常会进行相应的统计分析与处理,研究者无法获知其统计口径与处理过程。在刑事司法领域,官方数据通常源自办案机关的记录,但对没有进入办案机关视野或者以不需要记录信息的方式处理的案件则无法覆盖,这也是运用官方数据所必须注意的问题。因此,“使用任何一种官方的统计资料时都要小心”。<sup>②</sup>与发达国家通常有较为系统、内容详实且定期发布的有关刑事司法的官方数据相比,我国刑事司法数据统计工作仍处于初级阶段,全国性的官方数据屈指可数,地方性的官方数据标准不一、质量参差不齐。因此,在我国官方刑事司法数据薄弱的背景下,主要甚至完全建立在分析官方数据基础上的研究无疑在科学性上是存有问题的。

#### 四、刑事司法实证研究数据运用的基本原则

##### (一) 寻求较佳数据

与思辨研究相比,实证研究是一种充满“遗憾”和“不完美”感受的研究过程,这种感受很大程度上源自于研究者在面对数据时常常感到“力不从心”:数据收集无法实现尽善尽美,最为理想、质量最佳、测量效度最高的数据常因各种原因难以企及。在刑事司法领域,研究对象的特殊性、司法过程相对的隐蔽性、案件的异质性以及政治的密切关系等因素会使获取数据的途径进一步受限。因此,运用数据不能以获得最佳数据为目标,而应当在虑及各种客观制约因素的前提下,寻求“较佳”数据。或者说,真正体现研究者对于实证研究方法掌握程度的,并非洞悉什么是最佳数据,反而是如何选择较佳数据。例如,最为科学的简单随机抽样在刑事司法领域通常不可实现或没有必要,如何在科学性低于简单随机抽样的系统抽样、分层抽样、整群抽样等一系列方法中,结合研究主题和刑事司法的具体情境,选择相对较佳的抽样方法,才是研究者应当重点关注的问题。以寻求较佳数据作为数据运用的一项基本原则,既可以避免脱离刑事司法实际“过度”追求完美无瑕的数据,也可以为研究者提供一项衡量实证研究科学性的务实标准。

##### (二) 量力而行

不仅最佳数据常常不可实现,即使在有可能实现的数据中,也存在着不同的种类和层次,如何从中作出选择,需要从研究者开展实证研究的能力来考量。研究者走入真实的司法情境后,其开展实证研究的能力更多地受到研究者思考能力以外的其他因素的影响。首先是研究经费,研究经费甚至直接决定了实证研究收集数据的范围与具体方式。其次是研究人手,人手充裕时可以开展多地点、大范围的研究,这将扩大数据的来

<sup>②</sup> Frank E. Hagan, *super note* 38, 119.

源并提升测量的信度,但人手不足时则只能开展单一地点和小范围的研究。再次是研究时间,研究的持续时间和最终期限会限制收集和分析数据的时间,并最终影响到数据收集的来源和具体方式。最后是研究者与实务机关的关系,刑事司法实证研究大多需要相关实务机关的配合,研究者与实务机关的关系会在很大程度上影响到前往哪些实务机关及以何种方式收集数据。因此,在思量上述因素对研究能力限制的基础上,量力而行地收集和运用数据,应当是刑事司法实证研究数据运用需遵循的另一项原则。

### (三) 组合运用数据

为了使不同种类数据的优劣能够互补并进而提升研究整体的科学性,组合运用各类数据是应当遵循的另一项基本原则。在这一方面,定量研究与定性研究的优劣之争提供了很好的启示。在社会科学领域,定量研究与定性研究及相应的定量数据与定性数据的优劣是一个争论不休的话题。但在争论的过程中,有的研究者却能抛开二元对立的观察视角,在肯定定量研究与定性研究各自优劣的基础上,以实用主义和相容理论为立场,发展出混合定量与定性方法的混合方法,并得到了广泛的认可。<sup>③</sup> 这种混合方法实际上将定量数据与定性数据融于一项研究中实现其组合运用。除了定量和定性数据可以组合使用外,其他不同来源、不同表现形式的数据也同样可以且应当组合运用。以刑事司法实证研究著称的美国维拉司法研究所在解释其研究方法论时也强调,他们倾向于从不止一个来源收集数据。通过从若干不同来源收集到的信息,研究人员可以从不止一个角度来理解问题,而且必须同时收集定量数据与定性数据。<sup>④</sup> 事实上,在社会科学研究方法中,这种组合运用多种数据和方法的策略还有一个专门的术语——三角测量(triangulation)。三角测量的核心观点在于,如果研究结论经由多种研究方法和数据得出,其可靠性更高。三角测量既包括数据种类的三角测量(data triangulation),也包括数据收集方法的三角测量(method triangulation),前者是指针对同一研究主题收集不同种类的数据,后者则是指在一项研究中采用多种方法收集数据,以抵消每种方法所自带的影响有效性的缺陷。<sup>⑤</sup>

因此,为了提高刑事司法实证研究的科学性,研究者应当以更宽广的视野和互补的思维看待可供收集与运用的数据,而不用执念于理论上某一类型的研究应当运用某一类型的数据。在大多数实证研究中,数据的组合运用应当成为一项原则。

### (四) 不迷信数据

数据是实证研究的立论之本,但是,如果在分析和运用数据的过程中完全被数据所

<sup>③</sup> 参见 [美] 阿巴斯·塔沙克里、查尔斯·特德莱《混合方法论: 定性方法与定量方法的结合》,唐海华译,重庆大学出版社 2010 年版,第 10-18 页。

<sup>④</sup> 参见 [美] 吉姆·帕森斯等《试点与改革: 完善司法制度的实证研究方法》,郭志媛译,北京大学出版社 2006 年版,第 20 页。

<sup>⑤</sup> 三角测量还包括调查者的三角测量(Investigator triangulation),即在一项研究中应当由一位以上的研究人员收集数据。See Eugene McLaughlin & John Muncie, *The Sage Dictionary of Criminology* (3<sup>rd</sup> edition), at 474-475 (Sage Publication, 2013)。

引导甚至误导,却可能陷入迷信数据的泥潭之中。与思辨上的错误相比,这种迷信数据的错误又是以“客观”的数据为表象和支撑的,因而更具欺骗性。除了研究者违反研究伦理故意篡改、编造数据而得出结论这种“有意为之”的错误之外,迷信数据的“扎实表象”还会使研究者犯下很多“无心之失”。首先,研究者所收集到的数据本身就可能是不可客观或者不全面的,尤其是经过统计处理后的二手数据。有些数据虽然符合客观实际,但却只展现了研究对象的某一侧面或某一部分,完全依据这种数据得出结论会产生“瞎子摸象”的效果。其次,数据本身是可靠的,但如果对这些数据的解读缺乏理论的指引或者原因的探究,也可能陷入迷信数据的窠臼之中。在实证研究中,数据本身只是论证的工具,数据本身不直接产出结论,即使是推断统计所得出的“P 值”是否具有显著性也不等同于研究结论本身,仍然需要研究者运用理论和逻辑思维进行解读和推导结论。例如,在体现办案状况的数据中,某一年度的数据相较于其他年度出现了大的跳跃,并不能简单地认为这一方面的办案状况发生了很大的改变,而必须经由其他途径去探究数据跳跃背后的原因。最后,数据的“效力”是有限的,一般而言,数据只对其产生的主体、情境和时间等具有“解释效力”,其“效力”的范围不能被随意扩展。例如,基于产生于某地的数据所推导出的结论并不能直接地认为适用于全国。随意扩展数据的效力范围实际上稀释甚至切断了数据与结论之间的逻辑关系。

因此,依靠数据但又不迷信数据,应该是对待数据的正确态度。关于如何避免迷信数据,以下几方面的工作似为重要:一是对于非研究者直接收集的数据保持必要的警惕;二是研究者应当直面自己收集的数据的缺陷,并在推导结论时考虑在内;三是多来源多类型数据之间的相互验证,即上文所述的组合运用数据;四是对数据的分析运用应与理论相结合,避免单纯基于数据得出结论;五是对于从数据推导出来的结论应从多个角度进行反思。

## 五、我国刑事司法实证研究数据运用的初步反思

客观而言,我国刑事司法实证研究面临着较为逼仄的数据环境,这使实证研究中的数据运用处于一种先天不足的状态。首先,官方数据并不能给实证研究提供充分的支撑。正如学者在考察我国犯罪统计数据之后认为,我国犯罪统计数据的真实性质疑(谎报数据)、可靠性折扣(统计标准多变)、系统性不足(数据过于单一宏观)和公开共享机制缺失等如同重重迷雾使观察者和研究者都无法看到我国犯罪形势清晰的“面孔”。<sup>③⑥</sup>官方数据的缺位使研究者更多依赖于自行收集的数据,这无疑增加了数据收集的难度和成本,更重要的是,研究者自行收集的数据往往限于个别区域,因为缺乏全国性的官方数据作为背景或进行比对,因而其研究结论的适用范围饱受质疑。其次,我国法律实证研究开展时日尚短,相应研究主题的数据积累刚刚起步,有关数据收集工具与运用方

<sup>③⑥</sup> 参见卢建平主编《中国犯罪治理研究报告》,清华大学出版社2015年版,第59-60页。



法的展示和讨论尚未成为学术争鸣的重要内容,原始数据的公开共享机制亦未建立,这些都使数据收集与运用在某种程度上处于一种自我说服和自我满足的状态,缺乏方法上的批判和砥砺前进的互动机制。最后,我国刑事司法的公开透明程度和司法环境对于研究者收集数据限制较多,研究者进入真实司法情境并获取数据的途径受到很多主客观因素的影响。我国目前的状况是,在宏观层面上,已经就实证研究有助于完善刑事司法的功能达成了共识,但在微观层面上,对于实证研究如何介入具体的刑事司法情境和实证研究所要求的数据及其科学性却缺乏必要的准备和理解。

数据环境对数据运用的巨大影响不容否定,在我国数据环境的很多方面短期内无法改变的情况下,更为重要的可能是如何在观念和技术层面提升数据收集与运用的科学性,营造关注数据收集与运用细节的学术氛围。以我国刑事司法实证研究数据运用的总体情况而言,下列问题值得反思与改进:

第一,对于实证研究数据核心内涵的理解不准确。考察国内研究者的相关研究成果,往往存在两种误区:一是从数据的形式进行界定,最常见的就是将定量数据等同于实证数据,并进一步将定量研究等同于实证研究;二是忽视研究者的自身经验对于实证研究的决定性作用,将完全或主要运用非直接发现数据的也称之为实证研究。

第二,测量的观念尚未确立,测量标准与测量方法未得到应有的关注。研究者开始在司法实践中测量不同的研究主题,但对同一或相关主题的测量并未形成可供争鸣与讨论的不同测量标准与方法。更为重要的是,大量的刑事司法实证研究只重视研究结论,对于研究过程、测量标准和方法却缺乏应有的关注。<sup>⑤</sup> 实证研究结论的科学性是建立在测量标准与方法科学性基础之上的,仅有结论而无测量过程并不足以让其他研究者确信其研究结论的科学性。因此,当前我国刑事司法实证研究所面临的迫切问题,并不在于测量标准的多元与不统一,而在于如何鼓励对相同或近似研究主题采用不同的测量标准与方法并进行争鸣与讨论,以及如何鼓励在研究成果中充分论证测量标准并说明测量方法的细节。

第三,数据视野相对狭窄,运用的数据种类较为单一,组合运用数据的原则尚未确立。考察研究文献可以发现,大量刑事司法实证研究运用的数据集中在源自典型案例或法律文书的案件情况与处理结果数据、源自个别地区办案机关所记录的原生数据、问卷调查所获得的主观性数据以及部分官方数据等。这些类型的数据不过是真实刑事司法情境所能产生的多样化数据中的一部分而已。更为重要的是,研究成果往往倾向于集中运用某一类数据得出结论,而非组合运用各种数据来弥补单一数据类型所天生具有的缺点。因此,如何结合研究主题和司法环境的客观条件拓展运用的数据种类,并通过不同种类数据的组合运用提升研究结论的科学性是我国刑事司法实证研究在技术层面迫切需要解决的问题。

第四,对某些种类数据的特点与运用规则的理解并不准确。例如,在一些单纯测量

<sup>⑤</sup> 一个例证是,法学期刊所发表的刑事司法实证研究文献大多对研究方法和测量过程从略。

主观态度和评价的研究中不适当地引入了客观性数据,而在一些应当依据客观性数据的情形却又主要基于主观性数据得出结论。正如学者针对某项研究所指出的,有关法官学历、业绩、知识结构、学术成果等内容的客观性数据采用问卷调查这种主观性数据的方式进行收集不必要且不合理,应当采用查阅档案等方式收集。<sup>③</sup>

第五,缺乏对自身收集和运用数据方法的反思。实证研究的“不完美”源自数据的“不完美”,每一项实证研究收集与运用数据的方法都或多或少因各种主客观原因存在缺陷。与域外实证研究文献通常会对数据收集与运用方法进行自我反思并说明其对研究结论的影响相比,我国刑事司法实证研究文献在这一方面却着墨甚少。事实上,只有建立在研究者对数据收集与运用方法的自我反思与学术共同体的共同反思基础之上,经由具体方法的科学化,我国刑事司法实证研究才能逐步走向成熟,并真正发挥其在推动理论发展和完善法律制度方面的独特功能。

## 六、大数据与刑事司法实证研究

大数据近年来已成为各领域的热点话题,法学领域亦不例外,但真正主要运用大数据开展的刑事司法领域的专业研究尚不多见,在国内甚至可能尚付阙如。一些实务部门发布的名为司法大数据的数据分析究其本质仍为通过传统方法收集和分析的数据,与大数据思维与分析方法相距甚远。

一般认为,大数据具有“4V”特征,而这些特征使利用大数据进行的社会科学研究有别于传统方法:(1) 巨量( Volume) 数据,而与巨量数据相关的是大数据分析通常都被视为是全样本研究,区别于传统方法中的抽样研究。同时,大数据分析的前提是将一切量化为可供运算的数据,因此可以视为一种“升级版”的定量研究;(2) 高速( Velocity), 高速不仅指对巨量数据的高速运算,更重要的是数据收集和运算均可实现实时,即数据在产生时即被收集并可以随时进行运算,这与传统方法通常情况下数据收集和分析的滞后性或非实时性具有明显区别;(3) 多样( Variety), 与传统社会科学定量研究通常收集结构性、形式统一且单一的数据相比,大数据分析处理的则是形式多样的非结构化数据,或者可以称之为“混杂数据”,而这种非结构化的混杂数据在传统社会科学研究中则可能被认为不符合数据的基本要求而无法进行量化分析;(4) 价值( Value), 既指单一数据的低值,也指巨量数据分析所具有的高值,即能挖掘出传统社会科学研究所无法发现的价值。此外,大数据分析不再关注传统研究所关注的因果关系,不再通过研究验证假设,不再追求精确性,只关注相关关系,能发现很多未被注意的相关性,并基于这些相关性进行预测,而预测正是大数据分析的核心。正因为大数据分析具有以上迥异于传统研究的特征——尤其是颠覆了传统社会科学通过抽样代表全体和探究因果关系这两方面基础,因而被认为对包括刑事司法和犯罪学在内的社会科学研究整体带来了极大的

<sup>③</sup> 参见徐昕《司法的实证研究:误区、方法与技术》,载《暨南学报》(哲学社会科学版)2009年第3期。

冲击。

具体到刑事司法领域,如果现在就得出结论,认为大数据将极大地改变刑事司法研究的范式与方法确实为时尚早。即使放眼国际,真正运用大数据在这一领域开展研究也并不多见。根据国外学者的考察,目前在犯罪学领域运用大数据的研究主要有两类:一是将大数据作为传统研究方法所运用的一种新的数据种类,尤其是社交媒体数据和手机通讯数据;二是将计算机建模或算法作为风险评估或犯罪预防的预测工具。但无论哪一类现阶段运用都很少。<sup>③⑨</sup>而且,上述两种运用大数据的方式仍然与传统的研究方法紧密结合在一起,或者说大数据并未完全替代传统研究方法所收集的数据,或者依然需要传统研究方法所需要的理论或假设指引。例如,上述第一种方式虽然将社交媒体中用户实时发布的信息或者手机通讯信息作为大数据纳入研究之中,但仍然需要将其与传统方式收集的人口统计数据 and 办案机关所记录的犯罪数据合并进行分析,除了分析的数据体量因大数据的引入而大幅增长外,其遵循的仍然是传统的研究路径。<sup>④①</sup>从某一角度来看,此类社交媒体上的数据甚至可以视为传统方法所收集的研究对象自行报告的数据。第二种预测犯罪或评估审前释放与判处非监禁刑风险的研究方式对于犯罪学与刑事司法研究者也并不陌生,在大数据的概念产生之前就已经有过犯罪热点等预测研究,大数据的引入不过是增加了预测的途径和提升了数据的体量,同样需要依据犯罪学或刑事司法的相关概念和理论去寻找预测的因素。<sup>④②</sup>而运用来自社交媒体等大数据同样会遇到传统刑事司法实证研究所遇到的样本偏差问题,即进行这样的大数据分析及其结论只适用于使用这些社交媒体的人,而对不使用社交媒体的人则不适用。

当然,刑事司法领域纯粹的大数据分析也并非完全不能实现,但可能主要集中于民众有关刑事司法的认识等相对公开、并能与互联网或计算机系统关联的事项。但是,如果缺乏来自刑事司法体系内“大数据”的支撑,大数据分析目前似乎是无法探究刑事司法与犯罪的核心问题的。或者从另一个角度来说,正是因为刑事司法与犯罪领域的相对专业性或相对隐蔽性,才阻碍了在商业或其他公开领域如鱼得水的大数据分析的“渗透”。进一步来说,或许大数据分析在刑事司法领域的扩展尚取决于三个前提条件的实现:一是刑事司法与犯罪领域所有过程与事件的量化与数据化;二是能够实现对这些数据化信息的实时收集与实时分析;三是建立公开和分享的机制,并使研究者有机会获得这些司法“大数据”。如果以这三个条件进行衡量,我国刚刚起步的司法信息系统及公

<sup>③⑨</sup> See Janet Chan & Lyria Bennett Moses, *Is Big Data challenging criminology?* 20(1) *Theoretical Criminology* 21, 25-27 (2016).

<sup>④①</sup> 例如在英国进行的一项利用在 twitter 上发布的位置信息,对城市中心犯罪热点地区会因流动人口而改变的研究。See Nick Malleon & Martin Andresen, *The impact of using social media data in crime rate calculations: shifting hot spots and changing spatial patterns*, 42(2) *Cartography and Geographic Information Science*, 112(2015).

<sup>④②</sup> 当然,也有论者认为可以完全抛开理论上有关因果关系的探讨,完全采用大数据思维的相关性来预测犯罪。例如,如果犯罪者所穿的鞋码大小经过大数据分析发现与重新犯罪有相关性,那么鞋码大小就是预测重新犯罪的因素,至于鞋码大小为什么能预测重新犯罪则根本不重要。See Richard A. Berk & Justin Bleich *Statistical Procedures for Forecasting Criminal Behavior: A Comparative Assessment*, 12(3) *Criminology & Public Policy*, 511, 517(2013).

开机制的建设距离大数据思维尚有很远的路要走。

虽然现阶段大数据尚不能对刑事司法实证研究产生质的影响,但仍然可以为我们现在所倡导的实证研究提供多方面的启示,并提出一些值得思考的问题。首先,大数据思维同样体现了经验事实在研究中的决定性地位,这一点与实证研究并无二致。“大数据的核心是尊重经验真实,敬畏经验真实,在乎经验的代表性。”<sup>⑫</sup>其次,大数据思维为我们提供了一种更为宽广甚至没有边际的数据视野,来看待刑事司法领域中存在并可资研究运用的数据。“将世界看作信息,看作可以理解的数据的海洋,为我们提供了一个从未有过的审视现实的视角。它是一种可以渗透到所有生活领域的世界观。”<sup>⑬</sup>如何借鉴大数据思维,树立数据无处不在的观念,将刑事司法过程及相关事件中的所有片段,视作并分解为承载着经验事实的并可供运用的数据,是拓展和深化刑事司法实证研究要重点解决的问题。再次,大数据思维还可能拓展实证研究中研究者直接获得数据的路径,实时生成并记录的大数据收集方式是否是一种研究者直接获得数据的路径值得思考。最后,大数据分析中的个人信息保护还与实证研究伦理所要求的研究对象“自愿参与”、“匿名保密”及“有限欺骗”的原则相关,在运用并非专门为某项研究收集的用户信息等大数据开展研究时,是否违背了上述原则及是否需要在大数据背景下重新界定这些伦理原则也有待进一步研究。

---

**Abstract:** The core intension of empirical research data lies in researchers' direct findings on the foundation of their own experience, while its extension should be unfolded based on this kernel rather than excessively restricted by outward expressions like form, carrier, etc. The basic conception of empirical research is data measurement, whose implementation should be conceptualized and operationalized. Measurement in the criminal justice field has its own characteristics. How to scientifically and effectively conduct direct measurement in complex criminal justice practice is the core and difficult part of empirical research. Data classification based on different standards in empirical research could provide diverse perspectives to rule application as well as characters observation and quality assessment of various kinds of data. Data utilization in empirical research of criminal justice shall follow the principles of "preferable data selection", "to act within capability", "combined application" and "reasonable dependency on data". There are still many problems of data application in China's criminal justice empirical research need to be reflected. Despite the fact that Big Data analysis has not posed qualitative effect on empirical research of criminal justice till now, it can provide various helpful enlightenments and reflections.

---

(责任编辑:李 游)

<sup>⑫</sup> 白建军《大数据对法学研究的些许影响》,载《中外法学》2015年第1期。

<sup>⑬</sup> [英]维克托·迈尔-舍恩伯格、肯尼思·库克耶《大数据时代:生活、工作与思维的大变革》,盛杨燕、周涛译,浙江人民出版社2013年版,第126页。