

S H E H U I K E X U E Y A N J I U F A N G F A J I N G D I A N Y I C O N G

SHIYONG SHUJU
ZAIFENXIFA

万卷方法

社会科学研究方法经典译丛

SHEHUI KEXUE YANJIU FANGFA JINGDIAN YICONG

实用数据再分析法

■ 马克·W. 利普西 戴维·B. 威尔逊 著

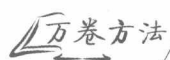
■ 刘 军 吴春莺 译



重庆大学出版社

<http://www.cqup.com.cn>

SHIYONG SHUJU
ZAIFENXIFA



社会科学研究方法经典译丛

SHEHUI KEXUE YANJIU FANGFA JINGDIAN YICONG

■主编 沈崇麟 夏传玲

实用数据再分析法

■马克·W. 利普西 戴维·B. 威尔逊 著

■刘 军 吴春莺 译

重庆大学出版社

Authorized translation from the English language edition, entitled PRACTICAL META-ANALYSIS, by Mark W. Lipsey, David B. Wilson, published by Sage Publications, Inc., Copyright© 2001 by Sage Publications, Inc.

All rights reserved, no part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher. CHINESE SIMPLIFIED language edition published by CHONGQING UNIVERSITY PRESS, Copyright © 2006 by Chongqing University Press.

实用数据再分析法。原书英文版由 Sage 出版公司出版。原书版权属 Sage 出版公司。

本书简体中文版专有出版权由 Sage 出版公司授予重庆大学出版社, 未经出版者书面许可, 不得以任何形式复制。

版贸渝核字(2006)第 15 号

图书在版编目(CIP)数据

实用数据再分析法/(美)利普西(Lipsey, M. W.),
(美)威尔逊(Wilson, D. B.)著;刘军,吴春莺译. —重庆:重
庆大学出版社, 2008. 1

(万卷方法. 社会科学研究方法经典译丛)

书名原文: practical meta-analysis

ISBN 978-7-5624-4296-7

I. 实… II. ①利…②威…③刘…④吴… III. 社会科学—数据
处理 IV. C37

中国版本图书馆 CIP 数据核字(2007)第 165277 号

实用数据再分析法

马克·W. 利普西 戴维·B. 威尔逊 著

刘 军 吴春莺 译

责任编辑:唐启秀 版式设计:雷少波

责任校对:夏 宇 责任印制:张 策

*

重庆大学出版社出版发行

出版人:张鸽盛

社址:重庆市沙坪坝正街 174 号重庆大学(A 区)内

邮编:400030

电话:(023) 65102378 65105781

传真:(023) 65103686 65105565

网址: <http://www.cqup.com.cn>

邮箱: fxk@cqup.com.cn (市场营销部)

全国新华书店经销

自贡新华印刷厂印刷

*

开本: 787 × 1092 1/16 印张: 14.5 字数: 276 千

2008 年 1 月第 1 版 2008 年 1 月第 1 次印刷

印数: 1—4 000

ISBN 978-7-5624-4296-7 定价: 37.00 元

本书如有印刷、装订等质量问题,本社负责调换

版权所有,请勿擅自翻印和用本书

制作各类出版物及配套用书,违者必究

马克·W. 利普西 (Mark W. Lipsey) 1972 年获得约翰霍普金斯大学 (Johns Hopkins University) 心理学博士学位, 现任范德比尔特 (Vanderbilt) 大学公共政策系教授。他的研究领域: 青少年犯罪的风险和干预问题以及项目评估研究中的方法论质量问题。其著作有《评估: 方法与技术》(第 6 版)^①等。利普西教授是《评估动态》(*New Directions for Evaluation*) 的前主编, 同时也是《美国评估杂志》(*American Journal of Evaluation*) 等多家著名刊物的编委会成员。获美国评估协会保罗·拉扎斯菲尔德 (Paul Lazarsfeld) 奖。

戴维·B. 威尔逊 (David B. Wilson) 1995 年获得克莱尔蒙研究大学 (Claremont Graduate University) 应用社会心理学博士学位。2005 年至今任乔治梅森大学 (George Mason University) 副教授, 致力于预防犯罪及犯罪者改造效果的再分析研究。1999 年获得美国评估协会颁发的马西亚·古藤塔格 (Marcia Guttentag) 奖。

刘军 1970 年生, 博士后, 教授, 博士生导师, 哈尔滨工程大学人文学院社会学系主任。2003 年获北京大学社会学博士学位, 北京大学 2006 年优秀博士学位论文二等奖获得者。主要著述:《社会网络分析导论》(2004)、《法村社会支持网络》(2006)、《社会网络分析法》(2007, 译著), 发表“一般化互惠”(《社会学研究》2007 年第 1 期) 等论文近 20 篇。研究领域: 社会网络分析、社会学方法论。完成国家社科基金项目等 3 项课题。目前主持教育部留学回国人员科研启动基金项目和黑龙江省留学回国人员择优资助项目。

吴春莺 1972 年生, 博士, 哈尔滨工程大学人文学院社会学系副教授。2006 年获哈尔滨工程大学经济管理学院管理学博士学位。主要研究领域: 经济社会学、福利经济学与福利制度。

① 本书中文版已由重庆大学出版社出版, 详见: 万卷方法——社会评估与测量方法译丛。

总序

社会研究方法的 现状及其发展趋势

近年来,社会调查技术和社会研究方法都有很大的发展。在调查技术方面,自20世纪70年代以来,社会变迁多次横断面的跟踪调查研究,几乎成为所有国家和地区了解社会结构转变和社会发展状况的基础性调查。这种调查不仅对社会学的研究有很大促进,对整个社会科学的研究都产生了重大影响,而且这些调查结果有的已成为政府有关部门决策的重要依据。国际上比较著名的此类调查有:美国芝加哥大学全国民意调查中心(National Opinion Research Center,简称 NORC)的“社会综合调查(General Social Survey,简称 GSS)”,英国埃塞克斯大学调查中心进行的“全国家庭生活和社会变迁调查”,法国经济和社会调查所进行的“全国经济社会调查”,日本社会学组织进行“全国社会分层与社会流动调查(简称 SSM)”。中国台湾“中央”研究院社会学研究所,也每两年进行一次“台湾社会变迁基本调查”。美国的“社会基础调查”,现在已成为年度性的调查项目,它是美国国家基金会目前资助的最大的社会科学研究项目。以上这些调查,除美国的调查外,一般均因经费原因采用纵向的间隔性重复调查法,即每隔一段时间,进行一次全国规模的抽样调查。每次调查除保留社会研究所需的基本项目外,每次都有不同的主题。在间隔若干时间后,再重复同一主题的调查,这样的研究设计,使社会变迁研究在可以涉及更为广泛的研究领域的同时,具有更好的积累性和可比性。多年来,这些基础性调查获得的资料,滋养着大批的社会科学研究者,有时一项调查就有很多名博士生用来写博士论文,以此所取得的研究成就,其可靠性受到社会科学界的广泛认同。例如1997年出版的,以台湾社会变迁基本调查数据为基础的研究报告集《90年代的台湾社会,社会变迁基本调查研究系列二》收论文16篇,内容涉及社会生活的各个方面,在台湾引起了极大的反响。

国内社会科学界在这方面也有了长足的发展。笔者所在的中国社会科学院社会学研究所的社会调查和方法研究室,组织或参与了多项与社会变迁有关的大规模抽样调查,取得了一定的研究成果,并积累了大量有关社会变迁的宝贵数据资料,其中主要有:

1. 城乡家庭变迁系列调查:该课题是由中国社会科学院社会学研究所牵头,联合北京大学和地方社科院的研究人员展开的一项类似多次横断面的城乡家庭变迁调查。这一调查始于1981年的“中国五城市婚姻家庭调查”,而后有1988年的“中国农村家庭调查”、1991年的“中国七城市家庭调查”、1998年的“中国城乡家庭变迁调查”。

2. 有关中国城乡社会变迁的系列调查:调查始于1991年的第二批国情调查,

然后有1992年的“中国城乡居民生活调查”、1993年的“第三批国情调查”、1995年的“第四批国情调查”和1997年的“中国沿海发达地区社会变迁调查”。上述调查虽然还不是严格意义上的多次横断面的纵贯研究,但研究者已在研究设计中尽量考虑到纵贯研究的基本原则,如调查队伍的稳定、指标的可比性和样本空间的延续性等。

3. 中国城乡社会变迁调查:这一调查开始于2000年,为中国社会科学院重大课题。目前已经完成第一期第一次调查和第二次调查,今后将把这一调查发展为连续的、定期进行的社会变迁调查。

在纵向调查技术取得长足进步的同时,上世纪末至今,电话调查技术也有很大发展。电话调查涉及的范围几乎与个别(面对面)访谈同样全面。电话调查中使用的一系列方法,是在20世纪70年代后期和面对面调查一起发展起来的。在20世纪80年代中,电话调查开始变得很普遍,且成为许多场合中各种调查方法的首选。正如某些学者所言,一种在公共和私营部门被人们用来帮助提高决策效率的收集信息的有效方法为人们所普遍认同时,这一现象本身就具有方法论上的意义。不仅如此,电话调查还有很大的实践意义,因为它为研究者提供了更多的控制调查质量的机会。这一机会包括抽样、被调查人的选择、问卷题项的提问、计算机辅助电话访谈(CATI)和数据录入。正因为如此,今天在各种社会调查中,如果没有发现其他重要的足以放弃使用电话调查的原因,电话调查由于其独特的对调查质量进行全面监控的优点,常常成为各种调查方式的首选。由笔者翻译,重庆大学出版社出版的《电话调查方法:抽样、选择和督导》一书,也于2005年面世。

无论是纵向调查抑或电话调查,实际上都是收集研究资料的方法,而应用社会科学的发展,不仅在于调查技术,即收集资料技术的发展,还在于研究方法和分析技术的发展。近年来,无论是定性研究方法,还是定量研究方法都有了长足的发展。

首先,计算机技术的发展可谓突飞猛进,它对当今社会生活的各个方面产生了巨大的影响,在悄悄地改变着社会科学的研究风格和研究方式的同时,也大大提升了社会科学学者的研究能力。这种影响表现在研究过程的各个阶段,从理论建构(概念映射)、问卷设计(专业的问卷设计软件)、调查实施(计算机辅助访谈、计算机辅助电话访问系统、网络在线调查系统)、数据录入(光学标记识别软件)到数据分析(包括文本、声音、图像资料的处理),甚至延伸到写作发表阶段。这样的过程发生在如社会学、经济学、政治学、心理学、教育学中,促进了学科之间的相互借鉴和交叉融合,至少在研究方法上呈现这种趋势。随着计算机计算能力的大幅度提高,20世纪80年代之后,统计学领域内发生了一场“革命”,主要表现在对定类和定序变量的建模能力的大幅度提高上,以及与分布无关的统计分析模型的发展之上,特别是基于“Resampling”(包括Bootstrap、Jackknife、Monte Carlo模拟等)的建模技术^①。同时,计算能力的提高还带动了基于神经网络、动态模拟、人工智能、生态进化等新兴的分析和预测模型的发展。这些进展都为定量社会科学研究提供了更

① 对于当前一些国内尚无公认译法的模型、软件等的名称,本“译丛”都只给出了英文的表述,以免造成称谓上的混乱。

多的可供选择的工具。

亚德瑞安·E·拉夫特里(Adrian E. Raftery)依据社会学家所处理的数据类型,将定量社会学在美国的发展划分为三个时代:第一代起始于20世纪40年代,交互表是其主要处理对象,研究重点是关联度和对数线性模型;第二代起始于20世纪60年代,主要处理单层次的调查数据,Lisrel类型的因果模型和事件史分析是其研究重点;第三代起始于20世纪80年代后期,开始处理诸如文本、空间、社会网络等非传统的数据类型,目前尚没有形成成熟的形态。拉夫特里的综述,虽然更强调定量社会学研究对统计学的贡献,但也大致勾勒出定量社会学在国外的发展脉络。

从分析模型的角度来看,定量分析在以下几个方向有了突破性发展:

1. 缺失值处理:由于社会生活的复杂性,社会调查数据常常出现缺失值,传统的处理方式是忽略这些缺失值,或者用均值替代。但现在则倾向于用多重插值法(multiple imputation)或者其他基于模型的方法进行处理。这些技术的发展,不仅会增强我们对数据的处理能力,而且将改变我们设计问卷的方式。基于这些技术,我们在不增加被访者负担的前提下,大大增加调查问卷的内容:每个被访者只回答问卷的一部分,然后通过对缺失值的处理,获得他们对未回答部分的估值。

2. 非线性关系:线性假定是经典定量分析的一个常见假定,但在实际研究当中,线性假定只能被看作是对社会现实的一个逼近和简化。面对具体的研究数据,如果没有理论上的明确指引(不幸的是,我们常常没有中程理论的指引),我们是无法在线性模型和非线性模型之间做出取舍的。但MARS模型的出现,让我们可以从经验数据当中获得最为拟合的变量之间的函数关系,而不必预先做出线性假定。这样,理论思考 and 数据分析就可以实现一个互动的循环过程,定量分析就不单单是对理论和假设的简单证伪过程,而是理论思维一个重要组成部分。

3. 测量层次:20世纪六七十年代的统计模型,大多要求数据的测量层次在定距以上,如因素分析,但社会学的调查数据却大多为定类或定序数据。对应分析、Loglinear、Logit、Logistic Regression、潜类分析、Ordinal Regression、Normal Ogive Regression等统计模型的出现,大大提高了定量社会学处理定类和定序数据的能力。

4. 测量模型:基于文化、社会、心理和认知等方面的考虑,在社会学界仍有人对问卷调查在中国的效度提出质疑。抛弃“本土化”的文化执著,我们更应当关注的是问卷调查的项目反应理论(item response theory),即被访者回答问卷题项时的过程模型。这方面的进展主要表现在两个方面:一是分解测量量表的成分,如Rasch model、IRT分析、Mokken分析等,二是将测量模型与因果模型或其他分析模型结合在一起,明确把测量误差引入到分析当中,充分评估它们对分析结果的影响,如结构方程模型。

5. 潜变量模型:与测量模型相关联的另外一个发展方向是潜变量模型,例如,潜变量分层分析(latent class analysis)、潜变量结构分析(latent structure analysis)、潜变量赋值分析(latent budget analysis)等。“潜变量”这一概念表明,我们可以通过测量“显变量”来测量无法直接观察的理论概念,如权力、声望、地位等。这样,理论和现实之间,通过“潜变量”到“显变量”的映射(测量过程),就有了联接的

桥梁。

6. 分析单元的层序性:在定量分析当中,我们常常强调要避免出现“生态谬误”,即分析单元的层次和结论或推论的层次不一致。与其相关的方法论争论是“宏观和微观”的问题。随着多层次模型的出现,我们可以同时考察多个层次上的问题,我们可以把个人放在其家庭背景中,再把家庭放在社区的背景下,考察个人层次的变量对社区变量的效应,或者社区层次的变量对个体行为的具体影响。在定量分析模型当中,“宏观和微观”的联接获得了建模技术上的支持。在这个领域当中,还有一个方向也值得关注:分析宏观层次的数据,对微观层次进行推论。

7. 社会网络模型:区分“关系数据”和“属性数据”,是把分析重点从个体/群体等社会单元转移到这些社会单元之间关系的第一步,社会网络模型是目前发展较快的一个定量分析领域,其理论根基是结构主义。社会网络分析目前仍然具有较浓厚的“形态学”特征(基于图论的缘故),但却为我们理解社会关系在社会空间上的形态奠定了基础,通过计算机模拟和研究社会网络的历期数据,研究社会结构的“发生学”性质模型也处在萌芽状态当中。

8. 系统动力学:如果说社会网络模型是在社会空间上拓展定量社会学的研究手段,那么,社会过程在时间上和物理空间上的属性,则是事件史模型、事件数模型、历期分析、Cox 回归、时间序列分析、Cohort 分析、状态空间模型等模型的研究对象。在这个领域,计量经济学为定量社会学研究提供了许多有益的范例。

9. 预测模型:上述模型仍然是在分析主义的范式下。有些社会学的应用研究,更强调模型的预测精度,而不是模型的认知价值,例如,社会趋势的预测。由于计算能力的提高,神经网络、基因算法、人工智能、模式识别等数据挖掘技术有了长足发展,已经出现了许多拟合经验数据的预测模型,比较成功的应用出现在计量经济学领域(如对股市的预测)。

10. 计算机模拟:对于社会学应用研究而言,研究的对象具有历史性、规模大、变迁的过程不仅漫长且表现某种渐进性,且因社会隔离/社会伦理原因无法接近或有实验禁忌等,无法直接进行观察和研究,这时计算机模拟就成为一个可供选择的替代方案。计算机模拟主要有两个类型,一是基于计算机网络的模拟:每台微机作为一个代理,整个网络作为“社会”,实时演化,如法国的 Swarm 计划;二是基于概念模型的系统,在计算机时间上,按照既定规则运行,较有名的研究是罗马俱乐部的《增长的极限》,常见的软件有 Simul, Arena 等。自然科学家对此方向似乎比社会学家更有兴趣。

定性研究方法一直是社会学研究领域比较传统的研究方法,在社会学研究的古典时期,它甚至是社会学家手中唯一的研究方法。但随着定量研究方法在社会学研究中的广泛应用,定性研究方法就似乎越来越不受到人们的重视。但需要澄清的事实是,在定量分析模型取得飞速发展的同时,在过去的二十多年里,定性研究方法也有了长足的进步。主要表现在以下六个方面:

1. 研究素材日益扩大:除了传统的参与观察、深度访谈、专题小组访谈之外,会话、交谈、电视、广播、文档、日记、叙事、自传 (autobiography) 等社会过程中自然产生的素材,甚至社会学理论本身(理论的形式化),也开始进入定性分析的视野当中。所有这些资料,不仅可以以文本的格式存储,而且,新型的多媒体介质,如图像、声

音和视频,作为原始的分析素材,也日益成为定性分析的新宠。

2. 分析方法更加多样:定性方法的种类在最近的二十多年中,更是有了一个质的飞跃。在比较传统的、源自语言学的方法,如内容分析、话语分析、修辞分析、语意分析、符号学、论据分析等方法之外,社会学家也创造出自己独特的定性分析方法,如施特劳斯(Strauss)等人的扎根理论、海斯(Heise)的事件结构分析、拉津(Ragin)的定性对比分析、Abbott 和 Hrycak 采用最优匹配技术的序列分析、亚贝儿(Abell)的形式叙事分析(formal narrative analysis)、鲍尔(Bauer)等人的语库建设、Attride-Stirling 等人的主题网络分析和神经网络技术应用的定性分析领域。所有这些方法的一个共同特征是,把定性研究向更加系统、更加精确、更加严格、更加形式化的方向推进。

3. 认识论基础更加多元化:现象学、释义学和本土方法论(ethnomethodology)的认识论,一直是定性分析的大本营,但近年来,实证主义也开始逐渐为定性分析所接纳,解释和阐释之间,由激烈的对立关系,逐渐演变为相互融合。

4. 研究过程更加客观规范:定性分析的一个主要问题在于阐释过程中不可避免的主观性。为了尽可能消除“解释者偏见”和主观选择性,定性分析开始遵循严格的程序模板或程序规则,并尝试引入定量分析中的“信度”、“效度”、“代表性”等概念,通过编码和对比,再加上传统的定性分析标准,如可解释性、透明性和一致性,使得定性研究的过程更加规范、阐释的结果更加客观,研究的结论更加可信。

5. 研究过程更加有效率:这主要应归功于大量计算机辅助定性数据分析(CAQDA)软件的涌现。从20世纪80年代以来,定性分析过程的数字化和计算机化,已经是一个不可逆转的大趋势。这种发展趋势与定性研究者的理论取向无关,不管他们的理论立场是实证主义、符号互动论,还是本土方法论,大多数定性研究者都在自己的研究当中,开始采用计算机来辅助定性资料的分析过程。据不完全统计,目前已经有二十多种定性分析的软件,分别隶属于德国、英国、法国、美国等国家。其中,有一些软件是国外研究机构的科研成果,可以免费使用,但比较成熟的定性辅助系统大多是商业软件。这些定性分析的辅助系统,不仅使得研究者从处理大量文字材料的繁复劳动中解放出来,而且能够让研究者共享他们各自分析的细节,从而改变了定性研究的流程和研究集体之间的合作方式。同时,由于采用数据库结构,定性资料的管理也更加方便,这就为组织大型定性研究项目(包括多个研究地点、多个研究对象、历时的定性研究)提供了新的可能性。越来越多的定性研究人员开始走出他们的摇椅,坐到计算机屏幕前,湮没在访谈资料和故纸堆中的定性社会学家的形象已经一去不复返了。

6. 定性研究和定量研究的结合更加紧密:在定量分析方法的教材中,定性研究常常被看作是定量研究的前期准备工作,但定性研究者却持完全相反的观点,他们一般认为定性方法是自成一体的,可以完成从形成概念到检验假设的全部研究过程。在实际的应用研究中,定性方法和定量方法常常是交织在一起的,例如,克劳(Currall)等人在研究组织环境重要的群体过程时,通过内容分析把5年的参与观察资料量化,然后用统计分析来检验理论假定。格雷(Gray)和邓斯坦(Densten)在研究企业的控制能力时,利用潜变量模型把定性方法和定量方法有机结合在一起。雅各布斯(Jacobs)等人在研究比利时的家庭形态对配偶的家庭劳动分工影响时,

首先用定量方法对纵向调查数据进行分析,从定量分析的结果中,又延伸出对核心概念的定性研究。这三个研究分别代表了定量和定性方法相互融合的三个方向:①克劳等人的研究代表着定性方法的实践者试图将定性数据尽可能量化的取向,近年来涌现出的处理调查数据中开放题器的编码问题的工具软件(如 Words at, Smarttext 等,注意:它们都是由著名的统计软件公司出品的处理定性资料的软件),处理定性资料的传统内容分析软件(如 Nvivo、MaxQDA、Kwalitan 等)也开始提供将定性资料转换到常用统计软件的数据接口,这些工具上的革新将加快这种趋势的发展。②格雷和邓斯坦的工作代表了“方法论多元论”的取向,即在应用研究过程中,通过核心概念的测量模型,把定性研究和定量研究结合在一起。③雅各布斯等人的工作则代表了一部分定量研究者对过度形式化的定量方法的不满,并试图通过定性方法加以弥补。在定量研究领域中,对“模型设定”问题的关注,是定量方法重新试图返回定性研究这种取向的另外一种表现。

与社会调查技术和社会研究方法突飞猛进的现实相比,我国学术界在这些方面的论著的出版似乎显得有些迟缓。虽然已经翻译了美国的一小部分经典定量分析教材,如布莱洛克(Blalock)和巴比(Babbie)的教材,也有自己编写的一些教材,如袁方等人的《社会研究原理和方法》、卢淑华的《社会统计学》等,此外,偏重软件操作的还有郭志刚的《社会统计分析方法——spss 软件应用》、郭志刚的《logistic 回归模型——方法与应用》、阮桂海的《spss for windows 高级应用教程》等。在《社会学研究》等专业杂志上,也常常有一些定量分析的应用研究,可是专门的方法和模型研究却没有,也没有专门的方法研究期刊。仅就定量研究方法的介绍而言,也存在一些缺陷,主要表现在:

1. 原理和操作脱节。
2. 过分依赖某些商业软件,不全面。
3. 与中国的实证研究相脱节。
4. 不能反映当前方法研究的最新进展。

与定量研究方法相比,由于各种原因,定性研究方法的引进和介绍都比较少。在福特基金会资助的方法高级研讨班上,曾讨论过一些定性研究方法。在定性方法研究方面也有少数专著,如袁方和王汉生 1997 年出版的教程,陈向明 2000 年出版的专著。但总体说来,我们对定性研究方法还停留在初步介绍的阶段,主要的介绍也局限在定性研究的研究设计和资料收集的阶段上,对定性分析方法的介绍则没有能够反映出当代定性方法的最新进展。特别是,在定性分析工具(定性分析软件)的引进和研究上,基本上还是一个空白。虽然不乏一些出色的定性研究报告,但从方法研究上讲,我们才刚刚起步。当然,我们同时还应当注意到,在历史学领域,我国对定性资料的鉴别、考据和分析,积累了大量的经验和知识,这也应当是定性方法研究的知识来源之一,应努力加以发扬光大。

令人欣慰的是,社会研究方法的引进和出版方面相对滞后的状况终于有所改观。重庆大学出版社的编辑,以独到的学术眼光,逆当前出版界唯利是图的不良选题风气,投入了大量的人力物力,组织出版“万卷方法”。自 2004 年至今,已引进社会科学研究方法方面的专著十余种,在我国社会科学界已经引起了一定的反响。

然而,更为可贵的是,重庆大学出版社并未以已经取得的成绩而自满,而是再接再厉,在原有“万卷方法”的基础上,进一步组织出版“万卷方法—社会科学研究方法经典译丛”。按我们的设想,“译丛”应该是一个开放的体系,旨在跟踪社会科学研究方法发展的前沿,引进和介绍这一方面的经典著作和最新成果。

“译丛”第一批有《实用数据再分析法》、《社会科学研究设计原理》、《社会科学研究测量原理》、《社会科学研究分析技术》、《问卷设计手册》、《回归分析法》、《数据再分析法》、《实用数据再分析法》、《社会网络分析法》、《广义潜变量模型》、《定性变量数据分析》和《复杂调查设计和分析方法》(书名也许有变化)等十余种,几乎囊括了研究设计、测量和分析方法的所有领域,涵盖从基础的回归分析到最前沿的潜变量分析和多水平模型等各种分析方法。无论是社会科学各专业的本科生、研究生,还是社会科学研究的学者都将从中有所收获。

“译丛”由中国社会科学院社会学所社会调查和方法研究室的多位研究人员担纲,主译者都是在社会研究方法各个领域中具有相当造诣的教师和研究人员。“译丛”的译者不仅仅把翻译看作是一个“翻译”,而且也把它看作是一次再学习和再创新。

我们期待“译丛”的出版能对社会研究方法的研究、应用和教学有所推动。

沈崇麟 夏传玲

2006年12月于中国社科院社会学所社会调查与方法研究室

Meta 分析:过程及意义

Meta 分析是一种数据再分析法。它是对关于同一个问题的多项独立的定量研究结果进行再分析,进而得出更具普适性的结论。Meta 分析已经有三十多年的历史,最初主要应用于医学领域,随后广泛应用于社会科学中。这里简要介绍 Meta 分析的过程、优缺点以及应用于社会学时遇到的困境。

近半个世纪以来,随着定量研究范式的扩张,关于同一个问题的定量研究结果越来越多。各种文献所采用的方法、样本之间可能存在很大差异,得到的结论也可能大相径庭。如何梳理这些文献?能否从中得到一个统一的结论?数据再分析法(Meta-analysis)是解决上述问题的一种有效的方法。

一、Meta 分析的含义

Meta 分析是对关于同一个问题的多项独立的定量研究进行再分析,进而得出一般性的结论。该方法的思想可追溯到 1904 年,卡尔·皮尔逊分别计算了 5 个独立样本中伤寒接种和死亡率之间的相关系数,并求出均值。20 世纪 40 年代到 70 年代,Meta 分析在缓慢发展。在国外,Meta 分析方法最早应用到医学有关的文章发表于 1955 年,作者综合了 15 份独立研究结果,对 1 000 多名不同疾病患者服用安慰剂的疗效进行分析,得出了安慰剂具有 35% 疗效的结论(转引自何杰、刘树贤、刘殿武,2000)。该方法随后在临床医学中得到了大量运用。该方法出现在教育学的时间是 1976 年,这一年美国教育学家格拉斯(Glass)首次将该方法命名为 Meta 分析,真正创立了 Meta 分析法。在国外,以“Meta 分析”为关键词在很多著名的全文数据库中检索,会得到几千篇文献,国外有关这方面的专著也不下十几部。在国内以“Meta 分析”为关键词在 CNKI 上检索,会得到近 250 篇论文(截至 2007 年 9 月 18 日),数量显然少于国外,并且绝大多数文章都发表在医学杂志上,少量发表在地质学、生态学、心理学、教育学等杂志上。以“Meta 分析”为关键词在《中国优秀博硕士学位论文全文数据库》上检索,得到近 20 篇硕博论文,全是医学方面的论文(2007 年 9 月 18 日)。

Meta 分析在诊断、治疗、危险度评价、干预措施、预防对策等方面起着独特的作用(赵宁、俞顺章,1993)。随着社会科学量化研究的加深,该方法已经从 20 世纪 70 年代开始逐渐渗透到生态学、心理学等社会科学领域(彭少麟、郑风英,1999;郑风英、彭少麟,2001;Lipsey & Wilson, 2001),在社会学领域中的应用却较少(Goldschmidt, 2006)。

Meta analysis 这个词在国内有多种译法,如荟萃分析、元分析、汇后分析、综合分析、后设分析等,台湾学者认为应该翻译为整合分析,并认为这是首选的译法。实际上,即使在英文文献中,类似的研究也有几种表达方式,如 overview, quantitative synthesis, research synthesis, meta analysis 等,但是 meta analysis 是得到公认的叫法。

考虑到 Meta 分析是对数据的再一次分析,因此本书把它翻译为数据再分析法。数据再分析最初的含义是从关于同一个问题的文献中搜集足够多的定量研究结果,经过统计分析后加以汇总。关于 Meta 分析的具体含义,格拉斯提出的定义是:对由多项研究结果构成的集合进行统计分析,目的是对已有的发现进行综合。萨克斯(Sacks)等提出的定义是:对以往的研究结果进行统计学的合并和严谨的综述方法。1991 年,弗莱斯和格罗斯(Fleiss and Gross)给出其公认的定义为:整合分析是一类统计方法,用来比较和综合针对同一学科问题所取得的研究结果,比较和综合的结论是否有意义,取决于这种研究是否满足特定的条件。

二、Meta 分析的过程

Meta 分析的步骤与一般的量化研究异同点并存,下面简要论述。

(一)确定课题

首先,研究者根据自己的兴趣确定 Meta 分析研究的课题,即拟解决那些存在分歧争议的问题。Meta 分析关注的是存在争议的领域。

(二)收集文献

明确单个研究的人选及排除标准,也要明确查阅文献的方法及要采取的统计分析方法等。在具体检索的时候,可借助计算机,也可进行手工检索,特别是年代久远的研究更可能需要手工检索。需要注意,检索的内容要广泛,包括过刊、现刊、综述性文献和未出版的文献如会议论文、私人交换资料、硕博论文等。在检索过程,不得遗漏重要文献,必要时还需要向专家咨询,请专家列出文献清单,特别是重要文献的清单。

(三)质量评定和筛选

应根据具体的目的和专业知 识等制定评价标准,如确定受试对象的标准、样本量、随机分组方法、盲法观察、变量之间的关系(单变量、双变量、还是多变量关系)、自由度等。剔除不满足标准者,以确保 Meta 分析的有效性。

(四)文献编码

收集完文献之后,要逐一检查,从如下方面进行编码(Lipsey & Wilson, 2001: 85)。

1. 有关文献的实质性方面。包括样本的来源;人口统计学特征(如社会经济地位、年龄、性别、教育、族群),个人特征(如认知能力,人格特质),诊断特性(如临床病人,少年犯);自变量(如干预或治疗),如理论取向、所描述的层次(如剂量、强度、持续时间等)、组织的特征(如年龄、规模、行政结构)、治疗实施的模式、干预人员的特征等。

2. 量化研究的方法和步骤。包括抽样步骤或方法(如随机概率抽样),调查设计(如邮寄法、电话法、访谈法、历时性研究、横剖性研究、预测性研究、回顾性研究等),统计功效,测量的性质,数据分析的形式,自变量(如干预或治疗),安排各种条件的方法,控制组的性质(如没有受到治疗,得到安慰剂,预备性治疗),设盲(blinding),实验者扮演的角色等。

3. 对文献的来源进行描述。出版的形式(杂志、书、博士论文、技术报告等),出版的年份,出版的语言,研究的发起方和/或资金来源,研究者的特征(性别、学术机构等)。

(五)资料的综合

由于研究的性质各不相同,各项研究的指标不尽相同,因此 Meta 分析要将其转化为统一的指标,即效应值(effect sizes),它是 Meta 分析的核心概念。Meta 分析收集的定量信息可以有很多类,对应于每一类也存在不同的效应值。给出各个效应值之后,应该分析其分布,计算其均值,计算置信区间,对同质性进行评价。各种效应值统计量及其标准误参见(Lipsey & Wilson, 2001:72)。

1. 平均效应值。对这些值进行综合加权,计算出合并后的平均统计量。平均效应值的计算是通过每个效应值(ES_i)根据其方差的倒数(w_i)加权进行的,即每个效应值乘以其各自的权重,求总和后再除以权重之和。

2. 计算置信区间。一个平均效应值的置信区间是以均值的标准误和 z 分布的一个临界值为基础的。均值的标准误等于方差权重倒数之和的平方根,表示为(Lipsey & Wilson, 2001:114):

$$SE_{\bar{ES}} = \sqrt{\frac{1}{\sum w_i}},$$

有了均值的标准误,就可以计算置信区间,其下限和上限分别为:

$$\overline{ES}_L = \overline{ES} - z_{(1-\alpha)} (SE_{\bar{ES}}) \text{ 和 } \overline{ES}_U = \overline{ES} + z_{(1-\alpha)} (SE_{\bar{ES}}),$$

3. 同质性检验(homogeneity test)。Meta 分析有一个前提条件,即多个独立研究之间应该相似。如果各个独立的研究之间具有同质性,便可以将多个统计量进行加权合并;若不一致,可以考虑剔除特大、特小的或是方向相反的统计量后再综合;如果经过这一步仍然达不到要求,则不能利用 Meta 分析方法了。常用的同质性检验方法有:

(1) 图示法:大多数常见的作图技术,如直方图、茎叶图、散点图、误差条形图和盒须图(box-and-whisker plots)等都可用来分析同质性。

直方图和相关的图形技术,如茎叶图,能够有效地传递一个效应值分布的集中趋势、变动量和正态性。在用再分析数据诊断一些问题,如极度偏态性和极端值的时候,这些图特别有用。在再分析中最常见的图是效应值相对于样本量的散点图,称之为关于预期的散点形状的漏斗图(funnel-plot)。漏斗图可用来探测由于拥有小对象样本的研究的代表性小而造成的潜在的偏差。

对于一个定类变量的不同层次来说,为了同时比较其分布的集中趋势和分散程度,一个常用的图形展示便是盒须图。盒须图可以针对两组或多组效应值,展示其中位数、第1个四分位点、第3个四分位点、极差和极端值。

作图技术并没有给出具体的同质性数值,因而还需要另外一种检验方法,即 Q 检验。

(2) Q 检验:以 Q 统计量为基础的卡方统计量服从自由度为 $k-1$ 的卡方分布,其中 k 是效应值的数量。

Q 检验的原假设为 $H_0: Y_1 = Y_2 = \cdots = Y_k$, 即全部效应值都来源于一个总体。公式为 $Q = \sum w_i (ES_i - \overline{ES})^2$, 其中 ES_i 是个体效应值, i 从 1 到 k (效应值的数量), $\overline{ES}$ 是 k 个效应值的加权平均效应值, w_i 是 ES_i 的个体权重。若 Q 大于自由度为 $k-1$ 的卡方分布的临界值,则表明同质性原假设被拒绝。因此,在统计上显著的 Q 意味着一个异质性分布 (Lipsey & Wilson, 2001:115-116)。对具有同质性的统计量进行加权合并,综合估计出平均统计量,对该统计量进行统计检验和推断,最后确定结论。如果统计量不具有同质性,可采用分组分析进一步分析。

(六) 灵敏度分析(sensitivity analysis)

除了综合若干研究结果的效应均值之外,Meta 分析还有一个目的,即分析研究的特征与效应值之间的关系。研究特征包括实质性特征和方法论特征。前者包括研究的人、时间、地点、研究本身的特征等;后者包括文献筛选、测量指标和权重的方式等。通过这种分析,可以揭示各项研究之间的差异的原因,评价 Meta 分析的质量和有效性。

灵敏度分析是评价 Meta 分析真实性的指标之一,灵敏度高说明结论正确性强,外在真实性好即应用范围广,有较高的参考价值。评价灵敏度可从以下几方面考虑:1. 原始文献的质量评价(quality assessment)。Meta 分析的灵敏度取决于各个原始文献的科学性和严谨性,如果原始资料较差,那么分析所得结论也不会是高质量的。原始文献的质量主要以研究设计是否严格符合所规定的要求进行衡量;2. 对各项研究的方差进行评价;3. 抽样偏差的大小。下面重点分析抽样偏差问题。

在一个给定的再分析中,抽样偏差是否大到影响结论的程度?在评估这一偏差的潜在大小的时候,一个简单的方法是比较再分析中未发表研究与已发表研究的平均效应值。例如,如果未发表的研究的平均效应值为 0.40,已发表的研究为 0.50,那么很可能由于省略未发表的研究而引起的抽样偏差不

会超过 0.10。这一方法要求必须有足够数量的已发表的和未发表的研究,以便获得每一子类研究的均值的可靠估计值。

我们还可以计算由罗森塔尔(Rosenthal)给出的失安全数 N (fail-safe N) 统计量(Lipsey & Wilson, 2001:165-166)。失安全数 N 估计报告了无效结果的未发表的研究的数量,用它把贯穿各项研究的累积效应减少到非显著性的程度,或者说需多少阴性研究结果可能使结论逆转。

(七)总结成文

结论中应详细陈述分析的目的,文献查找方法及取舍标准,所综合的单个研究的特征,所应用的统计分析方法,提供包含有各个研究统计结果的图表,提供灵敏度分析结果;结论可能遇到的偏倚及处理方法;讨论分析结果应用价值等。

三、Meta 分析的优缺点

(一)Meta 分析方法的优势

Meta 分析方法主要解决以下的问题(赵宁,俞顺章,1993;Lipsey & Wilson, 2001:5-10)。

1. 增加统计功效。由于再分析整合了多项研究结果,因而可以有效地降低甚至排除单一研究结果中存在的测量误差、抽样误差等,从而可以提高结论的论证强度。由于单个试验往往样本太小,难以明确肯定某种效应,如果要求从统计学上来肯定或排除这些效应,则需要较大的样本,而若采用 Meta 分析方法要比一项大规模的研究更为可行,而且把许多具有可比性的单个试验结果进行合并分析,可以改善对效应的估计值。不需要研究者进行实际调查就可以得到真正大样本的研究结果,有处理大量文献的能力,不受研究数量的限制。

Meta 分析可用于对样本量小的多项研究进行再分析,能够处理关于同一问题的多个文献的全面整合。单项定量研究结果是进行 Meta 分析的初始资料,因此,Meta 分析是更高层次的研究。特别是,当研究的项数超过某个临界值时,再分析法的系统编码程序,以及为了记录结论信息而构建的数据库可记录来自每一项研究的详细信息,并且能够包含大量的研究。通过这种系统的再分析过程,研究者就能评估作者的假定、程序、证据和结论,而不是无条件地相信结论的正确性。在医学领域里应用 Meta 分析方法,可以定量地综合分析多个研究结果,得出科学、合理、可信的结论,为疾病的预防、治疗、诊断等方面提供更全面更可靠的依据。在社会学领域,国外有学者利用 Meta 法对关于离婚问题和移民问题的定量研究进行了再分析(Goldschmidt, 2006)。

2. 评价各项研究结果之间的不一致性。对同一个研究问题,各个试验结果可能不一致,甚或存在分歧争议,利用 Meta 分析方法可以得到对该问题的全面认识,做出科学的结论。

3. 寻求新的假说。Meta 分析方法可以回答单个试验中尚未提及或不能回答的问题,尤其用于对随机对照试验设计所得的结果进行综合评价,可以提出一些尚未研究的新问题,发现一些隐藏在其他汇总性的研究中的效应或关系。再分析会比单个研究产生拥有更大统计功效的综合性的效应估计值。

(二) Meta 分析方法的缺陷及争议

1. 再分析法的一个缺点是它比较专业,需要相当数量的专业技术。特别是恰当的效应值的选择和计算及应用于其中的统计分析等都需要专业知识。

2. 苹果和橙子问题(apple and orange problem)。有些学者认为,Meta 方法将很多各自采用不同的技术、步骤、检验方法和样本的研究放在一起分析,就好像把苹果和桔子混合起来比较,这没有什么意义。有人认为,如果一系列研究在各个方面都相同,那就没有必要去比较,因为除了统计误差外,它们应该有相同的结果。如果一些统计量是对不可比的研究结果的汇总,那么它们是没有意义的。有学者把范围广泛的多种精神疗法研究包含在一起进行,因为他们对精神疗法的总有效性问题感兴趣。但是,批判者认为,利用平均效应值来汇报诸如行为疗法、心理动力疗法和格式塔疗法这样性质截然不同的研究结果会引起误导。

实际上,Meta 分析的基础是各研究在某些方面相同,在此基础上分析各研究具备的共性(水果的性质)和区别。在这种意义上说将苹果和桔子放在一起分析是必要而有意义的。对于在研究方法上存在较大差异的诸多单项研究来说,一个较好的 Meta 分析法应该将这种差异考虑在内,设置必要的调节变量。

另外,在表述一项再分析中的多项性质截然不同的研究结果的子类时,可以把它们分别取出,各个效应值的分布以及相关的统计量也可分别报告,从而允许在它们之间进行比较。另外,再分析技术的发展也使得对同质性进行统计检验成为可能,进而决定来自各类研究的一组效应值表现出来的变异量是否比仅仅来自抽样误差的期望变异量多。它也提供了一种经验检验,即检验这些研究是否表现出如此异类的结果,以至于似乎不能假定它们是可比较的。从另外一个角度讲,当代的再分析法越来越专注于效应值分布的方差,而不是这些分布的均值。也就是说,关注的主要问题常常与区分出各项研究结果之间的差异的根源有关,而不是把各个结果聚集在一起得出一个总的均值。这个专注点针对性质不同的研究结果的子群体进行了认真的处理,并且较少受到批判。