

调查研究设计概述

崔 壮¹, 胡良平^{2,3*}

(1. 天津医科大学公共卫生学院卫生统计学教研室, 天津 300070;

2. 军事医学科学院生物医学统计学咨询中心, 北京 100850;

3. 世界中医药学会联合会临床科研统计学专业委员会, 北京 100029

* 通信作者: 胡良平, E-mail: lphu812@sina.com)

【摘要】 本文目的是扼要介绍调查研究包含的主要内容和调查设计的关键技术。不仅对调查研究中的调查对象、调查方法、抽样方法、调查前准备工作、调查的实施及调查过程中的注意事项进行了宏观概述, 还对调查设计方案、实施调查过程中的关键技术做了简要介绍, 这些内容可为科研工作者开展调查研究提供参考和借鉴, 以期提高调查研究工作的速度、质量和效益。

【关键词】 调查研究; 调查设计; 调查问卷; 随机抽样; 回收率; 应答率

中图分类号: R195.1

文献标识码: A

doi:10.11886/j.issn.1007-3256.2017.05.001

Overview for the design of survey study

Cui Zhuang¹, Hu Liangping^{2,3*}

(1. Department of Health Statistics, School of Public Health, Tianjin Medical University, Tianjin 300070, China;

2. Consulting Center of Biomedical Statistics, Academy of Military Medical Sciences, Beijing 100850, China;

3. Specialty Committee of Clinical Scientific Research Statistics of World Federation of Chinese Medicine Societies, Beijing 100029, China

* Corresponding author; Hu Liangping, E-mail: lphu812@sina.com)

【Abstract】 This paper aims at a brief introduction to the main contents in survey study and to the key techniques when designing a survey. In this paper, a macro overview was presented on the research objects, methods of investigation, sampling methods, pre-survey preparation, survey implementation and attentions during the survey. The key techniques in the survey process and design scheme of the survey were also briefly addressed. This paper could provide references for research workers in order to enhance the project schedule, the quality and the efficacy in survey research.

【Keywords】 Survey study; Survey design; Questionnaire; Randomized sampling; Recovery rate; Response rate

1 调查研究的概述

1.1 何为调查研究

调查研究 (survey research) 又称观察性研究 (Observational study), 是指在未施加任何干预措施的情况下, 客观地观察和记录调查对象的现状及其相关特征, 客观反映或呈现调查对象的实际情况, 如了解疾病的分布、流行特征和发生、发展规律等。

在调查研究中, 构成研究总体的最基本单位称为个体或观察单位。将这些个体的情况汇总, 可以描述由这些个体组成的群体 (样本), 还可外推至这一群体所代表的更大的群体 (总体)。研究者收集资料时直接询问的对象叫做调查对象, 观察单位的属性或特征称为研究内容或主题。比如, 在失独家

庭养老问题的研究中, 调查单位是失独家庭, 研究内容是养老问题, 调查对象是户主。

1.2 调查对象的种类

根据调查对象的属性或特征, 调查对象大致可分为以下八种类型^[1]:

(1) 个人: 调查研究中十分注重研究分析个人的社会角色和社会地位, 并基于个人特征来描述、解释和说明由个人行为所构成的更大的社会群体, 如调查对象为大学生^[2]、农民工等。

(2) 初级社会群体: 这类群体都是通过长期的、直接接触的和相互作用而形成的, 一般不是因长期的目标而组合, 如家庭^[3]、邻居、村落、非正式组织等。

(3) 社会组织: 这类群体是人们为了实现特定的目标, 通过直接或间接的联系, 按照一定的规则而结成的社会共同体, 如医院^[4]、工厂、公司等。

(4) 社会阶层: 社会成员按照一定等级标准划分为彼此地位相互区别的社会集团。同一社会集团成员之间的态度、行为模式和价值观等方面具有相似性。各阶层由于利益、欲望、态度、价值观念的差异, 对社会进程的作用和影响也不尽相同。对社会阶层进行调查研究具有重要的现实意义。例如管理者阶层、经理人员阶层、私营企业主阶层、专业技术人员阶层等。

(5) 民族: 调查研究把民族作为特定的对象, 探寻其文化、生产和生活方式等特征, 如语言、文字、风俗习惯、心理素质和行为方式等特点。

(6) 社区: 是指居住在一定地域内的人们所组成的社会共同体, 可分为城市社区、农村社区、小城镇社区、城乡联合体等^[5]。由社区研究一般可以进一步拓展到对整个社会的研究, 以促进全社会的协调发展。

(7) 社会行为: 即人类在社会中的行为, 包括各种类型的社会活动、社会关系、社会制度等。在对社会行为进行调查研究时, 具体对象不是作为主体的人, 而是侧重描述各类行为本身的特征。

(8) 社会产品: 指物化的人类行为的产物。它们既可以作为独立的个体调查单位, 也可以作为群体调查研究单位, 如学科、书籍、歌曲等^[6]。

1.3 调查方法的种类

研究者应考虑研究目的、目标人群的可及性、经费预算等方面, 选择不同的调查方法。调查方法可粗分为普查和抽样调查。

1.3.1 普查

对于小规模总体的调查, 可以采用普查的方式, 即对总体中的每一位个体开展调查。从理论上说, 普查不存在抽样误差, 但对质量控制提出了更高的要求, 在花费更多的时间、人力和物力的同时, 可能会出现漏查、重复调查等非抽样误差。

1.3.2 抽样调查

一般来说, 抽样调查在研究中应用更为普遍, 从总体中抽取有代表性的一部分观察单位组成样本, 依据样本的调查结果对目标总体做出估计和推断。与普查相比, 抽样调查省时省力, 能迅速获取所需的信息, 其调查量更小, 调查工作更容易深入细致地

开展。

1.4 抽样方法的种类

在实际工作中, 抽样调查包括概率抽样和非概率抽样两种方法^[7]。

1.4.1 概率抽样

在总体中, 每个观察单位都有被抽中的可能, 每一个个体被抽中的概率都是已知或可计算的。概率抽样具有统计的理论依据, 可计算抽样误差, 客观评价调查结果的精度, 在抽样设计时可对误差进行控制。概率抽样包括以下几种:

(1) 简单随机抽样: 将抽样框(即包含全部抽样单位的目录性清单)中的全部观察单位编号, 按等概率原则从总体中抽取部分个体, 所有观察单位有同等概率被抽中这是最基本的抽样方法, 也是其它抽样方法的基础。

(2) 系统抽样: 将总体中的 N 个观察单位编号排序后, 确定样本量 n , 令 k 为大于 N/n 的第一个整数, 然后在 1 到 k 之间随机抽选整数 R , 以 R 作为第一号的观察单位, 再按编号顺序等间隔 (k) 抽取观察单位, 即编号为 $R, R+k, R+2k, \dots, R+(n-1)k$ 的个体入选。

(3) 整群抽样: 将总体分为若干个“群”, 以“群”为抽样单位形成抽样框, 从中随机抽取若干个“群”, 将抽到群内的所有观察单位构成调查样本。群可以是自然的区划, 也可以是人为的区划, 如行政村、家庭等。整群抽样的优点是节省经费, 便于组织实施和控制调查质量。但与简单随机抽样相比, 同一群中的个体比由整个总体中随机选择的个体更趋于同质, 调查同一群里的个体, 在一定程度上造成信息重复, 导致对总体估计的精度下降。

(4) 分层抽样: 将总体中的所有观察单位按照某种特征或标志(如地域、职业、性别等)划分为若干个“层”, 采用简单随机抽样等方法从每个层中抽取子样本, 最后将子样本汇总起来形成样本。

值得注意的是, 当层间变异大、层内变异小时, 分层才是最有效的。由于每层的方差通常小于整个总体的方差, 分层抽样可以提高估计的精度。但在提高精度的同时, 也增加了调查的复杂程度和额外的管理成本, 在抽样前需要完整了解调查对象资料, 如可分多少层、每层有多少观察单位等。

(5)多阶段抽样:对于大规模的调查研究而言,很难通过一次抽样产生完整的样本,可以将抽样过程分成若干阶段,各阶段可采用相同或不同的抽样方法。

1.4.2 非概率抽样

样本的抽选不是按随机原则进行,不能根据样本计算抽样误差,也不能从概率意义上控制误差以及评价估计的准确性,但对于特定研究、特殊人群具有一定的应用价值,如吸毒者、同性恋人群等。常用的非概率抽样方法包括偶遇抽样(便利抽样)、判断抽样、配额抽样、雪球抽样、应答推动抽样等^[8-11]。

1.5 抽样调查前的准备工作

调查前的准备工作可以理解成为实现调查目标而进行的路线选择和工具准备,也称为调查设计阶段。路线选择是指根据调查目的进行的调查设计工作,包括从研究思路、策略到方式、方法和具体技术的各个方面。工具准备则主要是指调查所依赖的测量工具或信息收集工具的准备。当然,同时还应考虑调查信息的来源、调查对象的选取工作。准备阶段主要包括三个方面的工作:调查工作的总体规划、抽样设计、收集资料的工具与方法^[12]。

1.5.1 调查总体规划

调查的总体规划应以书面的形式,形成一份调查计划书,内容主要包括调查目的、调查内容和范围、调查经费预算和进度安排的说明。这种计划书有两种主要功能:一是作为研究者开展调查研究的行动指南和整个调查研究过程的备忘录;二是向他人说明调查研究目标、内容和研究的方法及可行性。为了能更有效地完成下一步的调查执行工作,最好能在调查准备阶段形成一份比较详细的调查工作流程表,对调查访问过程中所需进行的工作进行分类,尽量完整地列出每一类中的工作项目,并预设各工作项目的完成期限。

1.5.2 抽样设计

样本的代表性决定了调查结果外推至目标总体的可靠性和准确性。为了保证抽取到一份“优良样本”,需要精心的抽样设计。研究者需要认真了解调查的目标总体的现状和结构,了解收集调查总体万方数据

相关信息,了解接触调查对象时可能存在的困难和局限,同时结合现实条件(包括经费、时间、人力、关系等),选择和确定抽取样本的方案。

(1)选择合适的抽样方法。研究者要根据调查研究的目标要求,结合自身具备的现实条件,设计恰当的抽样方法和程序,获得有最大代表性的样本。具体抽样方法的选择一定要符合调查目标的内在要求。若一项调查研究是以了解现状、描述特征为主要目标,样本的全面性、覆盖的范围等因素往往是需要着重考虑的,多阶段抽样等方法的运用会相对多一些,样本含量也要尽可能大一些;若以分析关系、解释原因为主要目标,样本的针对性和典型性往往是最重要的,分层抽样、整群抽样的运用往往会多一些,样本含量也可以相对小一些。

(2)确定样本含量。影响样本含量的因素通常包括以下几个方面:一是用样本统计量推论总体参数值时的精确程度,即容许误差的大小。在其他条件不变的情况下,精确程度越高(容许误差越小),所要求的样本含量越大;二是推论的把握性程度(置信度)。若需进行差异性分析,还需考虑检验效能。在其他条件不变的情况下,所要求的这种把握度/检验效能越高,则所需要的样本含量也越大;三是目标总体的特征,如总体的标准差、总体率或总体相关系数等。总体标准差越大,总体率越接近 0.5,总体相关系数越小,所需样本含量越大;四是研究者拥有的人财物和时间,这些应是决定样本含量的客观条件。

1.5.3 设计收集资料的工具与方法

调查研究中收集资料的一个重要前提是可以通通过特殊的工具以及应用这种工具的特定方法来实现对调查对象的特征、行为、态度、倾向等进行系统的、客观的、准确的测量。因此,研究目标的确定、测量指标的选择、调查问卷的设计、资料收集方法的制定等,就构成了调查设计工作的另一项重要内容。

(1)确定基本变量并使之具有可操作性。在调查设计阶段,研究者一方面要依据调查研究的目标确立最基本的研究内容,包括研究者最关注的因变量,也包括作为解释变量的各种自变量;另一方面,研究者既要构建起一套需要进行测量的指标(operational terms)以满足调查研究需要,又要将这些指标细化为

一系列更为具体的测量指标 (measurable indicators), 以便在调查中收集相应的资料。前一方面的工作与调查的总目标密切关联, 后一方面的工作则是为问卷设计提供指引。

(2) 设计调查问卷。问卷是调查研究中用来收集数据的主要工具。问卷设计是调查设计中一项既相对独立、同时又与调查目标特别是内容具体化密切相关的一项工作。问卷设计的实质是以书面的形式将研究者希望了解的信息、希望收集的资料和希望询问的问题编制成询问表, 以便系统地、客观地、定量地收集被调查者的情况, 为探讨和回答研究问题收集实验材料和数据。研究者应该依据研究概念的框架和评价指标, 并遵循和运用问卷设计的基本原则和技巧, 将有关概念和测量指标转化为适合调查对象回答的各种具体问题。

(3) 设计资料的收集方式。调查研究中的资料收集方式主要有直接观察法和采访法两种, 通常是将二者结合使用。直接观察法是由调查人员到现场对调查对象进行直接观察、检查和测量以获取相关资料。采访法是通过被调查者的回答来收集资料, 多采用问卷调查的方式实现, 包括自填问卷法 (self-administered questionnaires) 和结构访问法 (structured interview)。

自填问卷法是指调查员将问卷发送或邮寄给被调查者, 然后由被调查者自行阅读和填答, 再由调查员收回资料。这种方法是调查研究中最常用的资料收集方法。

结构访问法是指调查员依据事先设计好的调查问卷, 采用口头询问和交谈的方式, 向被调查者了解有关情况、收集有关资料的方法。研究者在调查设计阶段, 应综合考虑调查总体和调查对象的性质, 调查样本的规模大小、调查内容的敏感程度、调查的目标和重点、调查课题完成的时间要求、调查者拥有的人力物力是否充足等, 选择合适的资料收集方法, 并做好相应的计划和安排, 以达到最好的调查效果。

与资料收集方法的选择有关的另一个问题是对调查过程以及调查的组织实施安排进行考虑, 包括调查人员的组织、培训、管理、经费的使用、时间的安排等。同时, 研究者在设计阶段要对所选择的资料收集方法在实施过程中可能遇到的各种困难、所面临的各种障碍做好充分准备, 并提前设计应对措施和处理方法。

1.6 抽样调查的实施

调查实施阶段也称作资料收集阶段或调查方案的执行阶段, 就是具体贯彻调查设计中所确定的思路和策略, 按照调查设计中所确定的方式、方法和技术进行资料的收集工作。在这个阶段中, 调查者往往要深入地接触被调查者, 调查工作中所投入的人力也最多。因此, 需要很好地组织和管理。另外需要注意的是, 由于社会现象的复杂性, 或者由于现实条件的变化, 事先所考虑的调查设计若在某些方面与现实之间出现偏差, 需要根据实际情况进行修正或弥补, 发挥研究者的灵活性和主动性。

1.6.1 调查的组织与实施

(1) 调查员的招募。调查员往往是调查研究中资料收集工作的主要承担者。可以通过登报、张贴海报或在互联网上发布公告的方式公开招募调查员。招募海报或公告的内容应包括: 调查项目的名称及内容简介、调查地点和预定的调查时间、调查员的应征条件、报名时间和方式、面试时间和地点等。对调查员的录取, 最好采取面试的方式。通过面试可以了解应征者的背景资料, 并对他们进行初步评估, 以判断其语言表达能力和调查意愿。在调查员的招募环节, 尤其要依据所研究的主题、社区的性质、被调查者的特点等因素来挑选调查员。比如从研究主题来考虑, 调查有关政治、经济问题时, 应选择男性调查员为主; 而在调查婚姻、家庭问题时, 则选择女性调查员更合适。又如从社区的角度来考虑, 所选择的调查员最好是当地的、同民族、同宗教信仰的人, 这样的调查员熟悉被访地区的风俗习惯、文化传统、语言特点等, 往往能够很顺利地开展访问调查工作。再如, 从被调查者的特点来考虑, 当被调查者为青年时, 应尽量选择青年调查员; 而当被调查对象年龄较大、资历较深、影响力较大时, 则应选择年龄较大的调查员。也就是说, 所选的调查员在年龄、职业、社会地位等背景条件上与被调查者越接近越好。

(2) 调查员的培训。首先, 研究人员要向调查员介绍该项调查的内容、意义、方法及其与调查项目有关的其他情况, 以便调查员对该项工作有一个整体性的了解。同时, 还要就调查访问的步骤、要求、时间安排、工作量、报酬等具体问题进行了说明。其

次,介绍和传授一些基本的调查访问技术,比如如何敲门、如何自我介绍、如何取得被调查者的信任、如何尽快与被调查者建立良好的合作关系、如何客观地提出问题、如何记录回答等。同时,要组织调查员集中学习调查员须知、调查问卷、调查员手册等材料,特别是要逐字逐句、逐条逐项地清楚调查问卷的内容、填写方法和注意事项等。最后要进行模拟调查或访问实习。最好是在一个小范围内,让每个调查员都按正式调查的要求和步骤,从头到尾实际操作一遍。然后认真总结模拟调查或访问实习中存在的问题,并通过讨论或讲解,解决这些问题。

(3)联系被调查者。调查的实施要求调查员与样本中的每一位被调查者都建立暂时的联系。如何顺利地使调查员为调查对象所接受,是每一项调查研究都必须面临和解决的问题。国外的做法是:一方面取得某种正式的或官方的允许;另一方面事先给调查样本中的每一位对象寄一封短信,介绍即将进行的调查项目,从而为调查员实际登门接触被调查者打下基础。结合我国情况,可以考虑以下途径:①通过正式机构;②通过当地有关部门;③通过私人关系;④直接与被调查者联系。

(4)调查过程的管理与质量控制。在调查实施阶段,一项重要的工作是加强对调查过程的管理和质量监控。这一工作的重点主要涉及以下几个方面:①合理组建调查队伍;②建立监督和管理办法及规定;③实地抽样的管理和监控;④实地访问的管理和监控;⑤问卷回收和实地审核的管理和监控。

1.6.2 保证调查的回收率

调查的回收率(response rates)也称为调查的应答率和回答率,是调查研究过程中研究者成功完成调查询问的个案数占计划完成的样本总个案数的百分比。调查研究中,研究者虽然可以通过科学的抽样设计,达到从总体中抽取具有代表性样本的目的,然而,这种抽样所得到的样本代表性并不能最终反映调查结果的代表性。在进行调查的过程中,会有许多因素导致抽样样本中部分个案缺失或失效,使得最终完成调查样本只是抽样样本的一部分。回收率与调查内容、数据收集方法和受访者相关特征有关。因此,最好在设计时就努力设法降低无应答率,在预调查、调查员培训、调查问卷设计、资料收集等环节做好准备。其次,也可通过回访和二次抽样的

方式提高回收率,并利用该子样本对其他无应答者进行推断。

1.7 实施抽样调查过程中的注意事项

1.7.1 样本代表性的问题

一个完美的样本应该是目标总体的“缩微版本”,能反映总体的所有特征。虽然现实研究的复杂性使得实际几乎不可能获得完美样本,但一个好的样本应该尽可能地再现总体中研究者感兴趣的特征,这也是通过样本客观、准确地推论和估计总体的根本保证。在调查研究中,围绕着感兴趣的目标总体,通常先设定抽样框,再从中进行抽样研究,这时的样本更多地代表了抽样总体。在理想的调查中,目标总体和抽样总体应该是一致的。因此,在总体估计和统计推断时,应充分考虑抽样框的界定、抽样方法、选择性偏差、应答率等因素,客观评价样本的代表性。

1.7.2 抽样方法的选择问题

抽样不仅仅是用部分覆盖代替总体覆盖,也是建立在概率论基础之上的对有用统计信息的可靠性进行控制和测量的科学。一个精心设计的有代表性的抽样研究,要优于对整个总体进行不可靠的或有偏倚的测量。研究者应充分考虑研究目标、受访者、样本含量、实际的人力财力及时间进度等方面,选择简单随机抽样,也可以进行包括分层抽样、整群抽样、不等概率抽样在内的复杂多阶段设计的抽样方法,必要时也可选择非概率抽样的方法。

1.7.3 警惕和评估偏倚的影响

除了不可避免的抽样误差,要警惕选择性偏倚、信息偏倚等对研究结果的影响。研究者应充分考虑和评估调查研究中偏倚的影响和效应,谨慎推论。

如果目标总体中具有某些特征的部分未包括在抽样总体中,就会出现选择性偏倚,它可能与研究者、调查员、调查对象有关。研究者的目标总体设定错误,或者未能把目标总体设定在抽样框中,或者设计调查时遗漏了感兴趣变量的某些特征,都会使目标总体的估计出现偏差;对于调查员,若调查过程中“有倾向性”地选择便利样本,或在原指定的总体成员的调查不易实施时,代之以总体中其他方便得到的成员,也会导致选择性偏倚;若研究样本是由志愿

者组成,属于自发性回应样本,这些特别乐意或自愿接受调查或测量的受访者的倾向、态度、意愿或经历等特征很可能与目标人群不同,也会产生选择性偏倚。

当调查收集的信息在某个或某些方面有偏离真值的趋势时,就会出现信息偏倚。问卷或量表等测量工具的某些问题的设置、措辞、排序都有可能影响受访者的理解、情绪,导致无法获取正确回答;受访者也可能因为忘记、曲解问题、故意隐瞒或编造等造成信息偏倚;调查员的工作能力和工作态度也会对信息收集产生影响。

1.7.4 应答率问题

应答率是影响调查样本代表性的重要因素。调查内容、调查时间、调查员、数据收集方法、问卷设计、应答者负担、激励和强制手段等因素都会导致抽样样本中部分个案的缺失或失效。评价应答率对研究结果的影响,主要取决于无应答的性质。若无应答的个案属于完全随机缺失,基本可以忽略无应答;若无应答者与应答者存在某种差别,则仅对应答者给出的问卷进行分析,所产生的结果偏差可能会影响调查结果的价值。一般来说,应在设计时考虑回访、二次抽样等方法以尽可能多地获取无应答者的信息,或者在分析时采用无应答加权调整、无应答的参数模型或缺失值插补等方法进行补救^[13]。

1.7.5 复杂调查中目标总体估计和统计推断问题

在很多调查研究中,抽样过程中除采用一阶段单纯随机抽样外,还会采取分层抽样、整群抽样、不等概率抽样或多阶段抽样等其他抽样方法或其组合的抽样方案,这种抽样叫做复杂抽样,这种调查属于复杂调查。复杂抽样中每个阶段的抽样方法不一定相同,其抽样误差的计算随着抽样阶段及抽样方法的增多变得极为复杂。然而,研究者在统计分析时,常忽略之前采取的抽样设计方法,将资料均视为来自单纯随机抽样设计下获得的资料来处理,会错误地估计标准误,最终可能得到错误的推断。目前,在 SAS 9.0 或更高的版本中可以通过 SURVEYMEANS、SURVEYFREQ、SURVEYREG、SURVEYLOGISTIC 等过程进行此类资料的分析。这部分内容在本文中不做详述。

1.7.6 调查实施和管理过程中的问题

调查实施和进程管理与调查质量控制相辅相成,调查的实施和管理事实上是调查负责人(督导员)与调查员之间的互动过程,调查负责人要合理地安排调查员的工作量,控制调查进度。

调查过程中要注意公共关系和宣传。根据调查主题、经费预算和目标总体的不同,大型调查应该采用不同的方法进行调查宣传活动。规模较小的调查可以通过致被调查者信、一般的机构介绍手册、电话等形式告知与被调查者积极合作的重要性,有助于被调查者配合调查。

对被调查者私人信息和隐私保密是调查研究的基本原则,事先告知保密原则有助于被调查者如实回答问题。同时,调查数据在传输和使用过程中的安全也必须注意,避免数据泄露。

2 调查设计方案内容概述

2.1 确定调查目的和调查方法

调查设计方案应依据调查研究目的进行撰写,调查目的也是选取调查指标和问卷设计的根本依据。研究者还应考虑采用何种调查方法获得目标对象,同时考虑通过何种调查工具获取每位调查对象的相关资料。

2.2 确定调查对象和抽样方法

调查对象的选定应在明确调查目的和调查方法的基础上,确定调查总体的同质范围,确定调查中最基本的观察单位,还需进一步细分对象人群,以期掌握更全面的资料。

2.3 确定调查指标和样本含量

围绕调查目的,确定调查项目,将其进一步细化为调查指标。调查指标应精练、完备、可获得,定义要准确,描述要通俗易懂,必要时可附加说明,便于统一标准。

2.4 确定编制调查表或调查问卷的原则

在调查研究的资料收集,为保证所获得的资料质量,问卷设计至关重要,需考虑目的性、适用性和可行性的原则。设计完成后需要进行预调查,对问卷的信度、效度和可接受性进行评估。

2.5 确定调查实施细则与要领

调查设计时,应当明确调查的时间和进度,明确调查经费和调查工具,明确调查机构和团队,对于实施过程中可能出现的问题准备相应的应急预案。在调查的组织实施过程中,调查核心小组、调查员和资金按时足额到位,动员受访者积极参与,保证现场资料的质量和调查进度。在质量控制中要抓紧原始记录的检查、补查和修正,及时总结经验教训,同时做好项目过程管理、评价和校正。

2.6 确定调查资料整理与分析的要领

现场收集的原始资料只有经过整理和分析、去粗取精,才能揭示事物内在的规律。制定资料的整理和分析计划包括:①调查问卷的回收和核查;②数据编码和录入;③设计数据整理表和数据分组;④数据分类汇总;⑤制定数据分析计划,说明研究指标的内涵、统计描述、统计分析和统计推断方法、需要做哪些探索性分析、如何控制混杂因素等。

2.7 呈现必要的附件

包括①调查表或调查问卷;②随机抽样方法的具体实现方法;③估算样本含量的依据和方法;④调查资料统计分析计划;⑤实施抽样调查的标准操作规程;⑥实施全程质量控制的策略等。

3 实施调查过程中关键技术的概述

3.1 调查问卷的设计与考评(思想和路径概述)

(1)围绕调查目的和研究假设,明确总体的核心概念(key concept)框架;(2)通过文献学习、专家小组讨论等方式,建立调查项目框;(3)采用专家咨询评分、小组讨论或将预调查的资料经因子分析等方法筛选调查项目;(4)将调查项目转化为问题,结合研究的理论框架,确定问题类型、答案形式、排列顺序等,形成问卷初稿;(5)对与调查对象特征相似人群开展小范围的预调查,修订问卷;(6)从信度、效度和可接受性等方面对问卷进行考评。

问卷设计时还应注意^[14]:①问题设计契合研究主题;②广泛阅读文献,避免遗漏重要指标或影响因素;③封闭式问题设计时,确保答案的穷尽性和互斥性;④问卷结构简单明了,问题排列层次清楚;⑤调查问题的表达清晰,适用于目标总体中的所有研究
万方数据

个体;⑥问题和答案设计时,应避免双重问题、含糊不清、抽象提问、诱导性问题;⑦设计敏感问题时要慎重。

3.2 抽样方法的具体实现(思想和路径概述)

在 SAS 软件中,SURVEYSELECT 过程提供了各种各样的概率随机抽样方法,既可以进行简单随机抽样,也可以进行包括分层抽样、整群抽样、不等概率抽样在内的复杂多阶段设计的抽样。用 SURVEYSELECT 过程选择样品并构成样本,在所输入的 SAS 数据集中需要包括抽样框。抽样单位是个体观测者或观测组(群组)。用户可以指定抽样方法、所需样本大小或抽样比例和其他一些选择参数。用 SURVEYSELECT 过程抽取样本后,产生包含抽样单位、抽样概率以及抽样权重的输出数据集。当进行多阶段抽样时,对每一个抽样阶段都需要调用这个过程来设定抽样框和选择参数。在等概率抽样中,抽样框或同一层内的每一抽样单位都有相同的概率被抽取。在等比例(PPS)抽样中,每一抽样单位被抽取的概率和它所在层的大小成正比^[15]。

3.3 样本含量的具体估算(思想和路径概述)

在调查设计时,必须确定完成调查所需的最基本观察单位的数量,在既定的科研经费下保证精确度和可靠性的同时,可以合理利用资源,保证统计推断的最大效果。不同的设计方案、研究目的、研究类型、资料类型、研究精度要求等对应的影响样本含量因素也不尽相同,样本含量的估计公式和方法也不尽相同,计算有简有繁。SAS 从 9.1 版开始增添了 POWER 和 GLMPower 过程,用以估算样本含量。还有一些专门用于计算样本含量的软件,如 PASS (Power Analysis and Sample Size)、Power and Precision、G * Power 等,大大减轻了计算量,提高了科研效率。

参考文献

- [1] 郝大海. 社会调查研究方法[M]. 北京:中国人民大学出版社, 2015: 10-14.
- [2] 黄晓光,何继继,田春辉. 大学生艾滋病防治知识态度行为调查[J]. 中国公共卫生, 2015, 31(2): 249-251.
- [3] 彭希哲,胡湛. 当代中国家庭变迁与家庭政策重构[J]. 中国社会科学, 2015(12): 113-132.
- [4] 张静,水黎明,江志琴,等. 宁波云医院的 SWOT 分析和对策

- 研究[J]. 中华全科医学, 2017, 15(2): 338-342.
- [5] 丁志伟,张改素,王发曾. 中原地区多尺度城乡收入的时空分异[J]. 地理研究, 2015, 34(1): 131-148.
- [6] 孟娟. 走向人文科学心理学[D]. 长沙: 湖南师范大学, 2009.
- [7] 杨士保. 医学科学研究与设计[M]. 2 版. 北京: 人民卫生出版社, 2013: 19-71.
- [8] Igberase O, Okogbenin E. Beliefs about the cause of schizophrenia among caregivers in midwestern Nigeria[J]. Ment Illn, 2017, 10, 9(1): 6983.
- [9] Lui PSC, Dunne MP, Baker P, et al. Sexual difficulties faced by men in the Solomon Islands: a mixed-methods study[J/OL]. Sex Health, <https://www.ncbi.nlm.nih.gov/pubmed/?term=Sexual+difficulties+faced+by+men+in+the+Solomon+Islands%3A+a+mixed-methods+study>, 2017-05-18.
- [10] Zhang Y, Chen P, Lu R, et al. Prevalence of HIV among men who have sex with men in Chongqing, China, 2006-2009: cross-sectional biological and behavioural surveys[J]. Sex Transm Infect, 2012, 88(6): 444-450.
- [11] Fedina L, Williamson C, Perdue. Risk factors for domestic child sex trafficking in the United States[J/OL]. J Interpers Violence, <https://www.ncbi.nlm.nih.gov/pubmed/27470203>, 2016-07-27.
- [12] 万崇华, 许传志. 调查研究方法与分析[M]. 北京: 中国统计出版社, 2016: 19-29.
- [13] Sharon L. Sampling: Design and Analysis[M]. Boston: Thomson Brooks Cole, 2009: 291-355.
- [14] Thwaites Bee D, Murdoch-Eaton D. Questionnaire design: the good, the bad and the pitfalls[J]. Arch Dis Child Educ Pract Ed, 2016, 101(4): 210-212.
- [15] 胡完, 胡良平. 基于 SAS 软件实现随机抽样及应用[J]. 四川精神卫生, 2016, 29(5): 391-395.

(收稿日期: 2017-08-17)

(本文编辑: 陈霞)



科研方法专题策划人——胡良平教授简介

胡良平, 男, 1955 年 8 月出生, 教授, 博士生导师, 曾任军事医学科学院研究生部医学统计学教研室主任和生物医学统计学咨询中心主任、国际一般系统论研究会中国分会概率统计系统专业理事会常务理事和北京大学

口腔医学院客座教授; 现任世界中医药学会联合会临床科研统计学专业委员会会长、中国生物医学统计学学会副会长, 《中华医学杂志》等 10 余种杂志编委和国家食品药品监督管理局评审专家。主编统计学专著 45 部, 参编统计学专著 10 部; 发表第一作者学术论文 220 余篇, 发表合作论文

130 余篇, 获军队科技成果和省部级科技成果多项; 参加并完成三项国家标准的撰写工作; 参加三项国家科技重大专项课题研究工作。在从事统计学工作的 30 年中, 为几千名研究生、医学科研人员、临床医生和杂志编辑讲授生物医学统计学, 在全国各地作统计学学术报告 100 余场, 举办数十期全国统计学培训班, 培养多名统计学专业硕士和博士研究生。近几年来, 参加国家级新药和医疗器械项目评审数十项、参加 100 多项全军重大重点课题的统计学检查工作。归纳并提炼出有利于透过现象看本质的“八性”和“八思维”的统计学思想, 独创了逆向统计学教学法和三型理论。擅长于科研课题的研究设计、复杂科研资料的统计分析与 SAS 实现、各种层次的统计学教学培训和咨询工作。