

监督评价的基本原理以及适宜工具和方法的选择

石 光^{*1} 张 涛² 张保丽³

1. 卫生部政策法规司 北京 100044

2. 辽宁省鞍山市铁东区医院 辽宁鞍山 114001

3. 北京中医药大学附属东方医院 北京 100078

【摘 要】本文概括了用于卫生政策、规划和项目的监督评价的基本原理,系统分析了绩效指标法、逻辑框架法、基于理论的评价法、正式调查法、快速评估法、参与法、公共经费追踪调查法、成本效益和成本效果分析法、影响评价法等常用方法的目的、用途、优点和缺点、所需的成本、技术和时间等特点,作为选择适宜评价方法和手段的基本依据。结论是必须根据卫生政策、规划和项目的目标和内容,以及监督评价方法特点选择搭配,形成科学、系统、综合的监督评价体系。

【关键词】监督评价;卫生政策;卫生规划;卫生项目;工具和方法

中图分类号:R197.1 文献标识码:B 文章编号:1674-2982(2009)05-0057-06

The choice of an appropriate mix of tools and methods: a principal of monitoring and evaluation on health policies, programs and projects

SHI Guang¹, ZHANG Tao², ZHANG Bao-Li³

1. Department of Policy and Regulation, Ministry of Health, Beijing 100044, China

2. Tiedong District Hospital in Anshan, Liaoning Anshan 114001, China

3. Dong Fang Hospital, Beijing University of Traditional Chinese Medicine, Beijing 100078, China

【Abstract】This paper summarizes the basic principles and elements of monitoring and evaluation (M & E) on health policies, programs, and projects. Then, it systematically analyses the objectives and applicable implementation, the advantages and disadvantages, the cost required, skill and time needed in the following M & E methods and tools: performance indicators approach, the logical framework approach, theory – based evaluation, formal surveys, rapid appraisal methods, participatory methods, public expenditure tracking surveys, cost – benefit and cost – effectiveness analysis, and impact evaluation. The conclusion is that characteristics of the objectives and activities of health policies, health programs and health projects should be matched with the tools and methods of M & E in order to formulate a scientific, systematic and comprehensive system.

【Key words】Monitoring and Evaluation, Health policy, Health program, Health project, Tools and methods

《中共中央、国务院关于深化医药卫生体制改革的意见》和《国务院关于公布医药卫生体制改革近期重点实施方案(2009—2011年)》的发布,标志着深化医药卫生体制改革从方案制定阶段进入了全面实施阶段。医改方案确定了2020年建立起比较健全的基本医疗卫生制度、人人享有安全、有效、方便、价廉的基本医疗卫生服务的奋斗目标,部署了今后三年着力推动五项重点工作的目标和相关政策。要落

实好各项政策,实现医改的目标,不仅需要各级政府、各个相关部门、医疗机构和广大卫生人员不懈努力,共同奋斗,扎实推进,而且需要对改革的过程、进展、效果等进行科学有效的全程督导与评估。近年来,在经济社会发展领域,规划、项目和战略(program, project, strategy)的监督评估(monitoring and evaluation, M&E)理论和实践发展很快,用于监督评价的工具与方法更是纷繁复杂,因此,需要根据医疗

* 作者简介:石光,男(1963年-),博士,研究员,主要研究方向为卫生经济、卫生政策,Email:sg@nhei.cn。

卫生改革各个领域的具体目标、体系结构、工作过程和产出的特点,精心选择和设计,切实提高监督评价的效果。

1 监督评价的目的与要素

监督评价的目的:由于社会发展的规划、项目和战略各具特色,监督评价的目的也不尽相同。概括起来,监督评价的目的一般包括七个方面:发现问题、纠正偏差,总结教训、改善服务,调整规划、优化资源,传播经验、推进工作,宣传推广、争取支持,持续改进、实现目标。

监督评价的责任人:经济社会发展规划、项目和战略的发起者和所有者是监督评价的责任人。由于这些项目往往涉及教育、医疗等民生问题,涉及公众的利益,所以,项目必须向公众负责,监督评价的结果要向公众报告。具体到卫生项目或者改革来说,由政府相关部门向社区居民负责和报告。

监督评价的对象:包括整个规划、项目和战略的投入、程序、结构、产出、结果、影响等,但对于具体项目以及具体评估方法来说,则各有侧重,可以专注于绩效结果的评价,也可更加重视工作过程的监督,还可更加重视长远影响的评价。

监督评价的特性:在经济社会发展政策、战略、规划和项目确定之前,一般根据各国法律以及项目的规模、位置和内容等,要开展环境影响评估(Environmental impact assessment, EIA)^[1]、健康影响评估(Health impact assessment, HIA)^[2]和社会影响评估(Social impact assessment, SIA)^[3],这些评估是在项目立项前进行的预测性、综合性评估,是根据现有数据对环境、健康和社会的积极和消极影响的研究;与监督评价的时间节点不同,后者是在项目执行中或者执行了一定时期后进行的结果测量。二者的共同点是为后续决策和工作提供依据,均是循证决策的基础。

2 常见的监督评价方法

目前,常用的监督评价方法包括:绩效指标法(performance indicators)、逻辑框架法(The logical framework approach)、基于理论的评价法(Theory-based evaluation)、正式调查法(Formal surveys)、快速评估法(Rapid appraisal methods)、参与法(Participa-

tory methods)、公共经费追踪调查法(Public expenditure tracking surveys)、成本效益和成本效果分析法(Cost-benefit and cost-effectiveness analysis)和影响评价法(Impact evaluation)。^[4]这些监督评价方法并非适用于所有的规划、项目和战略,有些方法是互补的,有些是可以相互替代的;有些方法用途广泛,有些工具需要严格的应用条件,需要根据监督评价的目的,项目的利益相关方关注的重点,以及数据是否可以及时得到,监督的成本大小等挑选适合的方法和工具。为了便于根据监督评价的目的选择适宜的方法和工具,我们列举了各种方法的目的和用途,优点和缺点,所需的成本、技术和时间等特点。

2.1 绩效指标法

绩效指标是发展规划、项目和策略的投入(input)、过程(process)、产出(output)、结果(outcome)和影响(impact)的测量值。当具备良好的数据、分析和报告基础时,用绩效指标法,可以使管理人员追踪进度、弥补缺陷、改善服务。制定科学有效的绩效指标体系,不能没有各个利益相关者的参与,这样才能保证他们理解这些指标,并在管理决策中加以应用。

应用范围:制定绩效目标,并据此评估取得的进展;通过早期预警及时发现问题,采取补救措施;确定是否需要更加深入的评价(in-depth evaluation)或评估(review)。

优点:是测量目标实现进度的有效方法;可以为不同组织单位、社区或不同时期进行比较提供基础。

缺点:如果指标定得不科学,就不能很好地反映绩效;容易出现指标过多,或者指标没有数据来源支撑,由此导致指标体系成本过高、不实用,并很可能被束之高阁;往往要在理想指标和接受利用现有数据才能得到的指标之间做出艰难的取舍。

所需成本:可高可低,与选取的指标多少、评价的频率以及数据质量等相关,与指标体系的综合性有关。

所需技能:需要培训几天,培养确定实际指标、数据收集、分析和报告的技能,还要培训对信息系统的管理能力,才能执行绩效监督体系。

所需时间:从几天到几个月不等,取决于用于确定绩效指标的参与范围、规划的复杂性。实施绩效监督体系可能需要6~12个月。^[5]

2.2 逻辑框架法

逻辑框架法是一种辅助性监督评价方法。用于澄清发展规划、项目和政策的目标,有助于确定项目的逻辑关系(program logic),即预期的因果联系和后续的结果链:投入、过程、产出(包括覆盖率或者惠及的人群)、结果和影响等,从而确定各个阶段的绩效指标以及获得这些目标的内在风险。逻辑框架法还是吸收各个利益相关方澄清目标、设计活动的工具。在项目的实施阶段,也可作为评估进展、修正误差的有用工具。

应用范围:通过确定清晰具体的目标、使用绩效指标和评估风险等,改进规划和项目的设计;对复杂活动的设计进行总结;协助准备详细的实施计划;为活动的评估、监督和评价提供客观的基础。

优点:确保决策者提出最基本的问题,并分析假设和风险;吸引利益相关者参与计划和监督过程;在动态地利用该方法时,也是指导实施、监督和评价的有效管理工具。

缺点:如果僵化地使用,会扼杀创造性和创新性。如果执行时不能及时更新,就无法反映变化的情况;往往需要培训和追踪。

所需成本:较低或者中等,取决于用来支持该方法的参与范围和深度。

所需技能:对于辅导员来说,需要培训 3~5 天;还需要额外培训用于参与式规划和管理的辅导技能。

所需时间:从几天到几个月不等,取决于参与的范围和深度。^[4]

2.3 基于理论的评价法

基于理论的评价法是与逻辑评价法相似的一种监督评价方法,但它对规划和活动的工作关系,比如规划理论(program theory)或规划逻辑(program logic),有更深刻的理解。特别是这种评价方法,并不假设规划存在线性因果关系,因此分析更加深刻。例如,要实现降低孕产妇和婴儿死亡率的规划目标,需要多管齐下,从供需双方入手,克服贫困地区孕产妇不能得到安全产科服务的障碍,保障孕产妇和婴儿获得有质量的服务。不仅要加强对社区居民的生育知识教育,动员她们在医院分娩;还要进行文化的改造,改变一些地区孕产妇在家分娩的习惯;最根本的问题是解决贫困孕产妇住院分娩的资金,要由政府

和基本医疗保障制度承担;同时,还要培训合格的卫生人员,提高新法接生的能力;要加强服务体系建设,提高项目地区医疗机构的基本产科服务能力和处理难产的能力,以及处理新生儿窒息等抢救能力。基于理论的评价方法可勾画出这些因果关系及其影响的路线图,识别出保证成功的因素,确定应该监督和评价哪些环节和步骤。如果发现某个步骤成功的因素缺失了,可以有理由得出结论:这个项目可能无法实现其目标。

应用范围:为复杂活动的设计画出路线图;改进规划和管理。

优点:能够尽早反馈发挥作用的因素及其原因;只要问题一出现,就可采取纠正措施;有助于确定规划者不希望看到的副作用;帮助确定深入调查问题的优先重点,对此也许要用更加有力的数据收集方法,应用更加复杂的监督评价技术;为评估规划的可能影响提供依据。

缺点:如果活动很复杂,或者把所有因素和假设都放在一起考虑,很容易搞得太复杂;利益相关者也许不同意某些他们认为重要的决定因素,这很耗时间。

所需成本:中等水平,取决于分析的深度,特别是调查规划工作需要收集数据的深度。

所需技能:需要培训辅导员 3~5 天。

所需时间:差别很大,取决于分析的深度、规划或者活动的持续时间、监督评价的深度。^[5]

2.4 正式调查法

正式调查法是一组监督评价的方法,包括多个主题的家庭调查(Multi-Topic Household Survey,也叫作生活标准指标调查, Living Standards Measurement Survey, LSMS)、核心福利指标问卷(Core Welfare Indicators Questionnaires, CWIQ)、顾客满意度调查(Client Satisfaction Survey)和市民报告卡(Citizen Report Cards)等,用于收集标准化数据,调查对象是经过仔细抽样得到的人群或者家庭。这些调查往往收集能相互比较的数据,用于较大人群,尤其是特定人群的比较。

应用范围:为比较不同的策略、项目和规划的绩效提供基线数据;在某个时间节点上,不同人群之间的比较;比较实际情况与项目目标之间的差距;描述某个社区或者人群的情况;可作为正式评价某一规

划或项目的关键投入;评估贫困水平,为减贫策略提供依据。

优点:能够尽早反馈发挥作用的因素及其原因;只要问题一出现,就可采取纠正措施;有助于确定规划者不希望看到的副作用;帮助确定深入调查问题的优先重点,对此也许要用更加有力的数据收集方法,应用更加复杂的监督评价技术;为评估规划的可能影响提供依据。

缺点:除非有核心福利问卷,往往得不到长期结果;即使有计算机的帮助,数据处理和分析也是大型调查的瓶颈;生活标准指标调查和家庭调查成本太高,耗时太长;通过正式访谈,很难得到许多信息。

所需成本:核心福利问卷调查户均成本是 30 ~ 60 美元,生活标准指标调查的户均成本是 170 美元,如果没有成熟的国家抽样框,成本会更高。

所需技能:必须有扎实的技术、抽样、分析、问卷设计、数据分析和处理技能。

所需时间:取决于样本量,核心福利指标问卷需要 2 个月,生活标准指标调查需要 18 个月至 2 年。^[6]

2.5 快速评估法

快速评估法是一类快速而便宜的监督评价方法,包括关键知情人访谈(Key informant interview)、焦点小组讨论(Focus group discussion)、社区人群访谈(Community group interview)、直接观察法(Direct observation)和微型调查(Mini-survey),用于收集观点、受益人以及其他利益相关方的反馈意见等。

应用范围:为管理决策,尤其是在项目或规划层次的决策,提供快速信息;对于比较复杂的经济社会变革、高度互动的社会情况,或对人民的价值观、动机和反应等,提供定性理解;对更加正式的调查收集的定量数据,提供背景情况或做出解释。

优点:成本低;可迅速开展评价;为探索新思路提供灵活性。

缺点:调查发现常常与具体社区或者地区相关,因此,将发现一般化存在困难;比正式调查的针对性、可靠性和可信性差。

所需成本:较低或者中等,取决于采用方法的规模。

所需技能:需要非诱导性的访谈技术、小组辅导技能、现场观察、记录和基本统计技术。

所需时间:需要 4 ~ 6 个星期,取决于访谈人群的规模和位置以及观察地点的数量。^[5]

2.6 参与法

参与法是规划、项目和策略相关者积极参与决策过程的方法。参与法一般包括利益相关人分析(Stakeholder analysis)、参与式农村评估(Participatory rural appraisal)、受益人评估(Beneficiary assessment)和参与式监督评价(Participatory monitoring and evaluation)等,可使利益相关者在项目的监督评价以及提出建议的过程中有主人翁的感觉。

应用范围:了解当地的情况、人民的观点和优先重点,以便设计更加具有反应性和可持续的干预措施;在实施过程中,确认问题评价项目、规或政策;提供为穷人赋权的知识和技能。

优点:通过设计阶段吸纳关键角色的参与,检验相关问题;建立伙伴关系以及地方对项目的所有权;提高当地学习和管理的能力与技能;为管理决策提供及时而可靠的信息。

缺点:有时人们觉得参与法不够客观;如果让关键角色有效地参与,可能很花时间;某些利益相关者为了自己的利益,有独占或者误用该方法的风险。

所需成本:较低或者中等,差别很大,取决于应用的范围和深度,以及怎样珍视当地的资源。

所需技能:对于辅导员来说,至少要培训几天。

所需时间:差别巨大,取决于应用的范围和深度。^[4]

2.7 公共经费追踪调查法

公共经费追踪调查法是追踪公共经费流动,确定资源实际惠及目标人群程度的方法。公共经费追踪调查法一般要检查不同层级政府经费下拨的方式、质量和时间,特别要调查负责提供卫生和教育等社会服务的单位,常常作为更大的服务或机构调查的内容,关注服务的质量、机构的特点、管理和激励机制等。

应用范围:定量地诊断服务提供中的问题;提供关于公共经费拖延、流失和腐败的证据。

优点:当财务信息很少时,用于支持问责机制;通过找到公共服务资金流动中的官僚主义的瓶颈,改进管理。

缺点:政府部门可能不愿意开放自己的财务账

本;成本可能很高。

所需成本:在全国建立起公共经费调查能力之前,成本可能很高。例如,乌干达建立公共经费追踪调查法,在教育部门花了 6 万美元,在卫生部门花了 10 万美元。

所需技能:需要较强的抽样、问卷设计、数据处理和分析的技术与技能,很好地了解要评估的部门。

所需时间:需要 5~6 个月,调查本身就需要 1~2 个月。^[4]

2.8 成本效益和成本效果分析法

成本效益和成本效果分析法是评价规划、项目和策略的效果和影响以及活动的成本是否合理的工具。成本效益分析的成本和效果都用货币值表示,成本效果分析的成本用货币值表示,而结果用非货币量表示,比如发现报告艾滋病感染者数。

应用范围:提供更有效地配置资源的信息;确定能够提供最高投资效果的项目。

优点:是估计规划和项目效率的有效方法;可以使经济假设更加清晰,否则,项目设计时则较模糊或者高估结果;有效地说服决策者和投资人该项活动的收益很合理。

缺点:技术性很强,要求有足够的经济和人力资源;进行成本效益分析的数据有时没有,预测的结果有时高度依赖假设;对结果的解释必须小心谨慎,尤其在效益不能量化时。

所需成本:差别很大,取决于分析的范围和数据的可得性。

所需技能:两种分析方法往往需要很高的技术,需要经济分析的技能,要有相关经济和成本的数据。

所需时间:差别巨大,取决于应用的范围和数据的可得性。^[7]

2.9 影响评价法

影响评价法系统地确认规划、项目和策略对每个家庭、机构和环境带来的效果,无论是积极效果还是消极效果。常见的影响评价法包括随机对照试验前后评价(Randomized pre-test post-test evaluation)、项目前后以及与对照人群的类型试验评价(Quasi-experimental design with before and after comparisons of project and control population)、项目开展后与非严格对照组比较评价(Ex-post comparison of project and

non-equivalent control group)和项目后影响快速评价法(Rapid assessment ex-post impact evaluation)。影响评价法可帮助我们更好地了解某个项目救助穷人的程度以及对其福利的影响,可以根据大规模的正式调查,也可通过比较项目人群与对照人群在项目前后的差别,或者在项目干预过程中几个时点的差别;对于小规模快速评估和参与评价法来说,影响评价可综合小组访谈、关键知情人访谈、案例研究和现有二手数据。

应用范围:测量某个活动的结果和影响,并与其他外部因素的影响相区别;帮助澄清某项活动的成本是否合理;为项目、规划或政策的决策提供是否需要扩张、修改或取消的信息;为改进设计以及未来活动的管理提供教训;比较替代方案的效果;强化对结果的问责制。

优点:为不同年龄结构、地区和时间的人群,提供结果大小的估计值;对某些最核心的发展问题提供答案:带来了多大变化?结果到底如何?怎样才能做得更好?系统及严格的分析,可以使经理和决策者增加决策的信心。

缺点:有些方法价格昂贵、耗时过多,尽管进行了更迅速、更经济的尝试;当决策者很快需要信息时,会减少该方法的使用;保证清楚的陈述假设和评估成本效果的方法,以便使用者可以判断其有效性。

所需成本:世界银行一系列影响评价的成本在 20 万~90 万美元之间,取决于规划的大小、复杂性和数据收集情况;简单而快速的影响评价明显少于 10 万美元,有些评价个案成本仅为 1 万~2 万美元。

所需技能:需要很强的社会科学研究设计、管理、分析和报告的技术能力。理想情况是,评价团队要平衡定性研究和定量研究的技能。

所需时间:可能需要 2 年或更长时间,快速影响评价往往需要不到半年时间。^[7]

3 讨论

选择适宜的监督评价工具,不仅要了解各个工具和方法的特点与使用范围,还要从宏观经济社会发展和管理角度,综合考虑各种宏观和微观的局限因素,制定监督评价的方案,要为了保障政策、规划和项目目标的实现开展这项工作,而非局限于单纯的监督评价。

(1) 监督评价工作远远超出了方法和技术层面。也就是说, 仅仅选择好监督评价工具和方法, 并不能做好监督评价工作, 这是一项政治性、社会性和经济性很强的工作, 需要跳出技术能力和方法的局限, 要有政治支持、社会参与和专项经费, 才能搞好监督评价。比如, 经济社会规划和项目从设计阶段开始, 领导和政府部门就要考虑监督和评价的问题, 要建立科学、民主、有广泛群众参与的决策机制, 从评价指标的定义到监督评价工作的执行都需要群众参与, 监督评价结果还要向各个利益相关人报告, 以保证项目、规划和政策不偏离预先的设计目标, 真正做到惠民利民。没有良好的政治治理结构, 没有向群众负责的机制, 监督评价容易走过场, 成为一种政治姿态和做秀, 不仅无助于项目执行, 还会带来群众的反感, 挫伤群众参与的积极性, 导致项目失败。

(2) 需要在政策、规划和项目的整个运作周期中界定监督评价的切入点和节点。理论上, 监督评价可在计划工作中“计划、执行、评价和行动”的任何一个阶段切入, 但监督工作一般贯穿整个项目周期, 可根据项目的规模和持续时间, 确定每次监督的间隔; 而评价工作则在项目执行一定时期后, 在项目中期或者结束时进行。从监督评价的层次看, 可以是国家、省或地区级等战略或政策层次的监督评价, 也可以是项目层次的监督评价。因此, 选择监督评价需要考虑切入点和监督评价的主体。比如, 项目设计阶段更常用逻辑框架法和基于理论的评价法等, 项目执行后则应采取反映进展和结果的正式调查法等; 政策的监督评价一般采用较宏观的绩效指标法和快速评价等方法, 项目的评价则采用更加具体的成本效益、成本效果法和公共经费最终调查法等。

(3) 充分发挥各种监督评价方法和工具的作用, 必须满足一定的条件, 否则, 监督评价的效果就会大打折扣。但是, 现实工作中, 可能无法满足评价所需的理想条件, 因此, 需要向现实作出妥协, 根据项目

设计、数据基础、监督评价的能力和资源等现实条件, 找到最佳的监督评价组合, 设计有效可行的评价方案。比如, 随机对照影响评价法需要在两三个时点上收集实验组和对照组的数据, 进行严格的统计学分析, 控制其它影响因素。但在实际工作中, 因为经费、时间、方法和伦理的局限, 很少能满足这些设计要求而进行标准的评价, 多数影响评价使用成本更低或者不太严格的评价, 例如, 应用类试验比较评价或者不严格的试验后影响评价。

参 考 文 献

- [1] Munn, R E. Environmental impact assessment; principles and procedures[M]. Paris: ICSU-SCOPE, 1975.
- [2] Health Impact Assessment[EB/OL]. [2009-01-08]. <http://www.who.int/hia/en/>.
- [3] Burdge R J. A Conceptual Approach to Social Impact Assessment [M]. Middleton, Wisconsin: Social Ecology Press, 1994.
- [4] The World Bank. Monitoring and evaluation: some tools, methods and approaches[M]. Washington D. C: The World Bank, 2004.
- [5] THE World Bank. monitoring evaluation; some tools, methods and approaches. The International Bank for Reconstruction and Development[R]. 2004.
- [6] Monitoring & Evaluation News[EB/OL]. [2009-02-10]. <http://mande.co.uk/>.
- [7] Instituting Impact Evaluation Within the Framework of a Monitoring and Evaluation System[R]. Independent Evaluation Group. The World Bank. 2009
- [8] Insider Insights: Building a Results-Based Management and Evaluation System in Colombia[R]. Washington D C: The World Bank, 2008.

[收稿日期:2009-03-18 修回日期:2009-04-13]

(编辑 田晓晓)