



卫生政策 与项目影响评估

——量化方法与应用

杜本峰 著

 中国人民大学出版社

卫生政策 与项目影响评估

——量化方法与应用

杜本峰 著

中国人民大学出版社

· 北京 ·

图书在版编目 (CIP) 数据

卫生政策与项目影响评估：量化方法与应用 / 杜本峰著. —北京：中国人民大学出版社，2017. 10

ISBN 978-7-300-22364-3

I. ①卫… II. ①杜… III. ①卫生工作-政策-评估-研究 IV. ①R-01

中国版本图书馆 CIP 数据核字 (2016) 第 016026 号

卫生政策与项目影响评估

——量化方法与应用

杜本峰 著

Weisheng Zhengce yu Xiangmu Yingxiang Pinggu

出版发行 中国人民大学出版社

社 址 北京中关村大街 31 号

电 话 010-62511242 (总编室)

010-82501766 (邮购部)

010-62515195 (发行公司)

网 址 <http://www.crup.com.cn>

<http://www.ttrnet.com>(人大教研网)

经 销 新华书店

印 刷 北京玺诚印务有限公司

规 格 165 mm×238 mm 16 开本

印 张 17 插页 2

字 数 253 000

邮政编码 100080

010-62511770 (质管部)

010-62514148 (门市部)

010-62515275 (盗版举报)

版 次 2017 年 10 月第 1 版

印 次 2017 年 10 月第 1 次印刷

定 价 49.80 元

版权所有 侵权必究 印装差错 负责调换

前 言

自古以来，许多组织都致力于了解、描述和改善我们的物质和社会环境质量，通过系统创造和运用知识，提高个体和集体福利。评估活动可以追溯到三个世纪以前，霍布斯（Thomas Hobbes）及其同时代追随者曾努力用数量方法评估社会环境，探寻死亡率、发病率和社会解体的原因。

在不同时期，为了获得某些期望的结果，政策制定者、资助机构、规划者、项目管理者、纳税人或项目对象，都需要区分值得实施的和不值得实施的社会项目、启动新项目和改善既有项目，从而达到特定的预期目的。项目实施是否实现预期的影响或承诺，显然需要评估。测度福祉的变化是否来自项目或政策干预的实际作用效果是本书的主体。

影响评估包括定性和定量方法，以及事前和事后方法。定性方法相较定量方法而言，在于寻求估计项目可能产生的潜在影响和机制。事前设计通过模拟或构建模型确定干预的可能利弊，在给定个体行为和市场假设条件下，试图预测将要实施的政策变化所带来的结果；事前方法常常构建结构模型以确定不同的政策和市场怎样与受益人水平行为联系，更好地理解项目产生影响的机制；事前分析有助于项目执行实施之前对项目的再精炼，以及预测在不同的经济环境条件下项目的潜在效果。相反，事后影响评估是基于项目干预后或项目执行前后的实际收集的数据，测量由于项目实施，受益人获得的实际形成的影响。本书以讨论影响评估的事后定量方法为主。

本书回顾了影响评估的定量方法和模型，详述了影响评估所涉及的其他领域，包括监测和评估（monitoring and evaluation, M&E）、执行评估（operational evaluation）以及定性和定量分析相结合的混合方法。广义上，因果的问题使得影响评估不同于 M&E 和其他的评价方法，在缺乏反事实结果

(即项目参与者没有参与项目的结果)资料的情况下,主要通过将不同的方法应用到调查数据中,以便为项目参与者构造出可比较的对象。影响评估的主要问题是一种归因问题——从其他因素和潜在的选择偏差中分离出项目的效应。

虽然影响评估与其他方法的评估,如 M&E 可以有所区别,但影响评估不可能也没必要与 M&E 割裂开。影响评估常常需要知道项目是怎样设计的,是怎样有计划地帮助目标群体,以及项目是如何被执行实施的,而这些信息常常只有通过执行评估(作为 M&E 的一部分)来获取。M&E 对于了解计划的目标、干预发生的方式以及目标群体效果测量潜在指标和评估数据是必需的。

本书的结构如下:第 1 章讨论了有关评估的基本问题;第 2 章涉及卫生项目评估的特性;第 3 章讨论了影响评估与其相关的概念,主要关注评估者所关心的量化事后影响评估的一些基本问题。

有两种项目设计方案,一是实验方法(或随机化),另一种是非实验方法。第 4 章关注影响评估的实验设计,讨论其优点和不足。第 5 章到第 8 章讨论各种非实验方法:第 5 章考察了匹配方法,包括倾向值匹配技术;第 6 章涉及面板数据中的倍差方法,放松了在选择偏差潜在来源方面的一些假设;第 7 章回顾了工具变量方法,该方法进一步放松了自我选择假设;第 8 章考察了断点回归和管道方法(pipeline methods),利用断点附近样本的系统性变化来研究处置与其他变量之间的因果关系。

本书也涵盖了其他方法,以阐明不同参与者受益于项目的机制。第 9 章讨论了如何测度项目影响的分布,包括分位数回归方法;第 10 章讨论了项目评估的联立方程方法,包括评估项目的直接和间接效果的经济模型,讨论了实验和非实验方法的优点和缺点,也强调了影响评估在政策制定方面的有用性。

基于影响评估的文献资料,本书扩展讨论了不同的实验和非实验量化模型,包括较新的事前和事后方法以及它们的结合。本书强调的是评估各种旨在改善人类环境的项目,提供的方法或技术基本上适用于各种有组织的社会活动所提出的绩效问题。

本书是完整阐述政策分析与评估理论和方法学的一次尝试。作者收集了

国内外有关评估的理论和方法以及实际的案例，同时包括所授硕士生课程的课堂讨论与报告的部分内容。有的是直接引自国内外的相应原文（包括图表及例题）以反映真实情况。凡涉及此类问题的地方，都在参考文献中给出了相应的文献书目或论文题目。期待读者能够从本书的阅读中，理解政策分析和评估的理论，掌握评估方法，并在实践中运用这些理论和方法，不断提高我国政策分析与评估的技术水平。本成果受到中国人民大学“985 工程”的支持，也是国家自然科学基金项目（71373272）和教育部人文社科重点研究基地重大项目（15JJD840004）的阶段性成果。

对于对怎样利用统计软件使用这些模型感兴趣的研究者来说，本书也提供了用 Stata 进行数据分析的基本命令语句，这些内容将帮助研究者构建和解决在他们的评估计划中可能遇到的问题。

本书的读者对象包括学生、实际工作者、项目负责人、政策制定者，以及任何对意在改善社会状况的项目成败测量感兴趣的其他成员。

由于作者理论水平和实践经验有限，书中错误及不成熟之处在所难免，不当之处望不吝赐教。

第1章 项目、监测与评估 1/

- 1.1 项目、政策与评估概念 2/
- 1.2 监测和评估框架 7/
- 1.3 监测、结果和影响评估 24/

第2章 卫生项目评估与指标 33/

- 2.1 卫生项目与类型 34/
- 2.2 卫生项目评估与特性 35/
- 2.3 卫生项目评估指标 38/

第3章 影响评估的基本问题、理论与评估设计 53/

- 3.1 反事实问题与影响评估的基本理论 54/
- 3.2 影响评估中的研究设计与方法论 62/

第4章 随机化 75/

- 4.1 随机化设计 76/
- 4.2 随机实验逻辑与处理效应测量 82/
- 4.3 随机实验数据的统计分析 85/
- 4.4 实践中影响评估的随机化问题 93/
- 4.5 Stata 程序概述 101/

第5章 倾向值匹配 105/

- 5.1 控制协变量策略与数据平衡方法 106/
- 5.2 倾向值匹配理论 113/
- 5.3 倾向值匹配实施、估计与分析 118/
- 5.4 倾向值匹配法的优点与局限 133/
- 5.5 Stata 程序概述与运用 136/

第6章 倍差 145/

- 6.1 倍差：理论与模型 146/
- 6.2 倍差法的优缺点 157/
- 6.3 倍差法的拓展 159/
- 6.4 Stata 程序概述与运用 165/

第7章 工具变量估计 177/

- 7.1 工具变量与因果关系 178/
- 7.2 工具变量回归——两阶段最小二乘法 180/
- 7.3 工具变量有效性检验 183/
- 7.4 工具变量与异质性潜在结果 187/
- 7.5 工具变量的来源 193/
- 7.6 Stata 程序概述与运用 196/

第8章 断点回归设计 203/

- 8.1 断点回归的基本理论 204/
- 8.2 断点回归设计与估计 207/
- 8.3 RD 方法的优点与不足 213/
- 8.4 管道比较 214/
- 8.5 Stata 程序概述与运用 216/

第9章 测量项目影响分布 219/

- 9.1 观察项目影响分布的必要性 220/
- 9.2 项目影响分布研究：线性回归框架 221/
- 9.3 分位数回归 223/
- 9.4 数据收集问题 231/
- 9.5 Stata 程序概述与运用 232/

第10章 项目影响事前评估与预测 237/

- 10.1 结构与简约式方法 238/

10.2 政策效应模型构建与应用 241/

结 语 247/

参考文献 249/

项目、监测与评估

1.1 项目、政策与评估概念

1.1.1 项目/政策

国外不少学者对项目的定义进行了阐释。Stephen M. Shortell (1978) 等认为：项目就是有组织、有资源投入来解决一个或多个问题或实现一个或多个目标所确定的方案、计划、程序等。Carol H. Weiss (1999) 认为，项目就是意在改善个人、群体、组织或社区福祉的高度组织干预。Perter H. Rossi (1989) 认为，社会项目或项目干预是指设计出来用以解决社会问题或改善社会状况的有组织、有计划且经常是正在进行中的方案。

国内，吴之明等 (2000) 认为，项目是在特定条件下，具有特定目标的、一次性的任务。刘荔娟等 (2003) 认为，项目是指那些在特定环境内作为管理对象，按限定时间、预算和质量标准完成的一次性任务。戚安邦 (2006) 认为，项目是一个组织为实现既定的目标，在一定的实际情况、人员和资金约束下，所开展的具有一定独特性和不确定性的一次性工作。

从国内外学者既有的认知和表述中，可以发现项目有如下特点：(1) 目标明确性。任何项目都有一定的目标，包括成果性目标和约束性目标。成果性目标是指项目的功能要求，约束性目标是指限制性要求。如将我国的婴儿死亡率五年内降至指定水平，五年是约束性目标，指定水平则是成果性目标。(2) 一次性。这是项目与其他重复操作、日常运营的最大区别，项目的其他属性也是从这一最主要的特征衍生出来的。(3) 独特性。每个项目的过程具有自身的独特性。(4) 环境制约性。项目的实施离不开一定的自然和社会环境，它们都有可能对项目形成影响和制约。(5) 风险性。项目实施的预设条件往往会随着时间和环境变化，给项目进展带来不确定性，给资源投入带来极大风险。因此，必须对项目进行精心的设计和控制，以尽可能取得预期的

收效，这也是需要项目评估存在的一个重要依据。

实际上，项目是一个比较宽泛的概念，按照项目规模大小可划分为政策、规划或计划、课题等（Weiss，1999）。一项新的政策，一般需要由若干个项目的实施来实现，而每个项目又由若干个课题组成，如，要对现行的医疗保健制度进行改革，推行新的医疗保健制度的政策，就要对医疗保健的筹资、付费形式、人员培训、医疗保健服务的管理、医疗费用的控制、医疗保健质量的监督等进行评价，以上每一项就是一个内容广泛的项目，而每一个项目又可以分为不同的课题进行研究（吴耀春，2009）。项目效果的信息可以来自监测或评价系统，监测和评价是有效项目管理的基础。

1.1.2 监测

项目监测是在项目的实施过程中，系统和规律地记录项目执行情况 and 项目结果的主要信息，并建立一定的指标体系进行测定，以显示项目是否按照预期的计划运行的行为过程（见图 1—1）。

监测是通过收集、分析和利用信息来跟踪项目的进展以达到其目标和指导管理决策的系统过程。监测通常关注过程，如活动在何时何地发生、谁送达它们、送达给多少个体。监测是项目开始后进行并持续于整个项目实施期间。监测有时被称为过程、性能或形成性评估（Frankel and Gage，2007）。

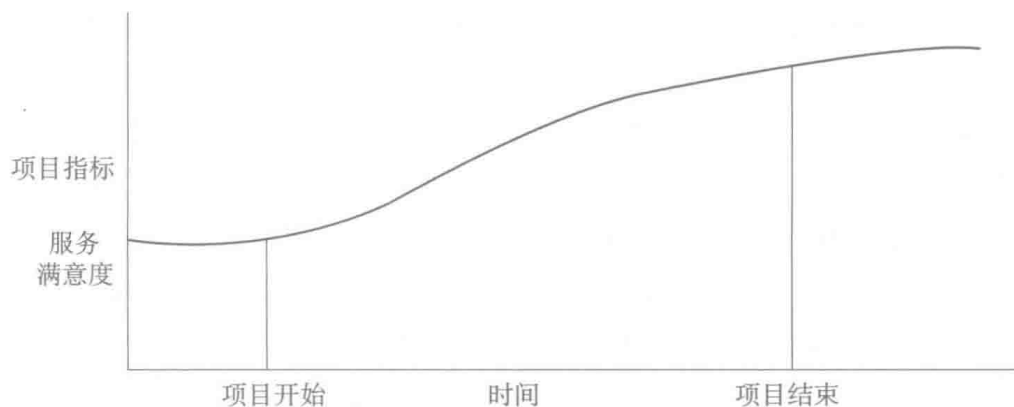


图 1—1 项目监测图解

为项目设立目的、指标和具体目标是监测系统的核心。结果信息和数据

可以用来评估项目干预的绩效。通过比较带有特定目标的结果，监测可以帮助提高政策的设计和执行效率，同样也能促进政策制定者和利益相关者增强责任感和交流。相反，评估则是对项目所获得的结果的一个系统的客观的评价。换句话说，评估旨在证明目标的变化仅仅来自具体政策的执行。

1.1.3 评估

评估活动可以追溯到科学的初创期。三个世纪以前，霍布斯及其同时代追随者曾努力用数量方法评估社会环境，探寻死亡率、发病率和社会解体的原因。要给评估下个定义也不是一件容易的事情，这里介绍几种不同的定义：

美国评估协会认为，评估是由不同利益和兴趣的人员所组成的专业性工作，主要是对项目、产品、人员、政策、绩效、标书、技术、研究、理论以及评估本身的评价。

Veney 和 Kaluzny 认为，评估是运用不同的方法对信息进行收集和分析，以决定项目行动的相关性（relevance）、适当性（adequacy）、进展（progress）、效率（efficiency）、效果（effectiveness）、影响（impact）和可持续性（sustainability）。实际上，该定义给出了项目评估的主要内容，相关性是指项目或政策是否必要，开展是否有意义，决定了项目存在的基本合理性（主要问题如，项目旨在解决的问题是什么？这个问题是否值得关注？问题和项目的定义是否恰当？项目是否针对所有问题？等等）。适当性是指从法律、政治、技术、社会、资源、组织等角度，分析项目或政策具体实施的可能性，以及根据先验判断，一个项目在多大程度上能解决针对的问题，如一个项目应该达到多大的规模、在多大范围内实施才能产生预期的效果等。进展是指对项目活动的追踪，主要问题是：项目是否在正确的时间、正确的地点，以正确的数量得到了所需的人力资源、设备物资和经费，以满足项目实施的需要？项目是否如期产出？产出的数量和质量如何？效果评估涉及的主要问题是：项目是否达到预期的目标？项目的提供者、受益者是否对项目的效果满意？项目的存在是否有益？影响评估涉及的主要问题是：观察效应是否是该项目产生的？如果没有该项目，观察的效应会发生吗？效率评估涉及的问题是：项目的效益是否超出成本？该项目的单位产出效益是否高于其他项目？

项目评估就是使用社会研究方法对社会干预项目的效果进行系统调查。具体来说,就是评估者使用社会研究方法对社会项目的所有重要方面进行调查、评审,并促使其改进,内容包括项目所针对社会问题的诊断、项目的概念化和设计、项目的实施和管理、项目的结果和效率(Rossi, 1989)。

项目评估是通过科学方法的使用来判断一个特定项目的价值,向决策者和处于一定层级的政策制定者提供信息,促使其对项目进行改造、推广,或者缩小或放弃项目以使资源得到优化配置。项目评估经常被用来作为政策研究的范例和政策分析过程的投入(Shortell et al., 1978)。

项目评估是一种特殊的政策分析,项目评估想了解的是,某些经济或社会项目对个人、企业或城市的影响(Jeffrey M. Wooldridge, 2007)。

评估是通过比照一套明确或含蓄的标准,对项目或政策的实施和结果进行评价,以此作为促进项目或政策改善的方式(Weiss, 1999)。

评估是对活动、项目、计划、策略、政策、主题、部门、操作地区或机构绩效的系统性评价。评估侧重于预期和取得的成就,检查结果链(输入、活动、输出结果和影响)、过程、情境因素和因果关系,以便评价所达到的成就。评估的目的是确定相关性、影响、有效性、效率和干预的可持续性以及干预对所得成果的贡献(Frankel and Gage, 2007)。

图1—2给出了有无项目干预的影响结果变化示意图。

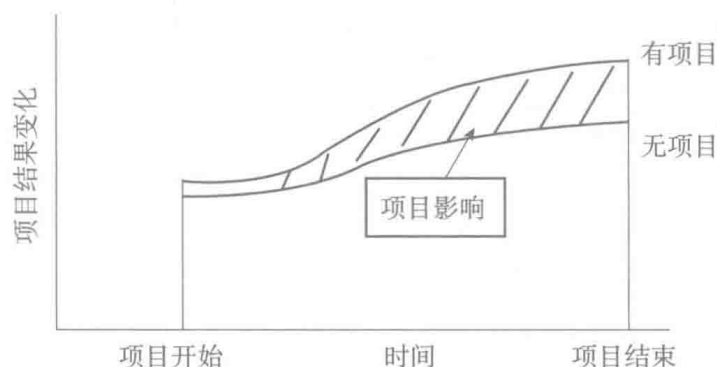


图 1—2 项目影响图解

由此可以看出,评估的主要目的是判定项目干预措施实现既定目标的相关度、效率、效果、影响和可持续性等方面的信息,明确项目干预措施和项目结果的因果关系、项目实现目标的效果和效率怎样、干预是否产生了未曾

预料到的结果，以及干预措施是否以最有效和可持续的方式实现某种目的。

我们认为，对于评估者而言，项目评估就是依据一套标准对项目的运作过程及其结果进行系统的评价。人们做出评估所依据的标准是建立在一定价值观基础之上的，即评估是指按照一定的价值准则，对评估对象及其后果进行的价值判断。正如美国学者 Gronland 给出的一个极其简洁的评估定义：

评估 = 量（或质）的描述 + 价值判断

所谓的“量（或质）的描述”，就是事实的记述；所谓“价值判断”，就是在事实判断的基础上，根据人的需要和愿望对客观事物做出的判断。

因此，评估的概念包含两个基本要素：事实挖掘和价值判断。评估并非纯粹主观或客观的活动，而是运用系统化的研究方法围绕评估对象（政策、项目、课题等）进行的信息整合与价值判断，评估活动是主观见之于客观的活动，是主观性和客观性的高度统一。

在评估研究文献中，人们会看到 evaluation 和 assessment 两个词频繁出现，且常常交替使用。根据《朗文现代英语字典》（*Longman Dictionary of Contemporary English*），evaluation 一词是指判断某事物的用处和价值的行动（the act of considering something to decide how useful or valuable it is）；assessment 是指对某人或情形做出判断的过程，对某事物的成本或价值的估算（a process in which you make a judgment about a person or situation, a evaluation about the cost or value of something）。在国际评估研究领域，“评估”主要指项目评估，而且这里的“项目”多是非营利性的社会公共项目，涉及卫生、教育、福利、生态、能源等多个方面。

在项目层面，实现监测和评估的目的是系统地跟踪执行情况和输出结果以及测量项目的有效性。它有助于准确地确定项目是否偏离目标以及何时可能需要调整，监测和评估形成了修改干预方式和评价活动质量的基础。

监测和评估可以用来说明项目产生的可衡量的预期成果。至关重要的是可以帮助管理者、规划者、实施者、政策制定者和捐助者获得他们所需要的信息，以便对项目操作做出明智的决定。

监测和评估有助于确定资源是否得到最有价值和高效的利用。给出关于项目在多大程度上可以判断“成功”的客观结论是至关重要的。监测和评估共同提供的必要数据可用来指导战略规划、项目和计划的设计与实施以及用

更好的方法进行资源的分配和再配置。

就具体干预来说,通过监测可以了解如下内容:项目拟包括哪些活动,为什么?为谁提供哪些服务,什么时候提供?频率、持续时间、服务内容如何?服务可及吗?质量是否达到要求以及服务是否送达目标群?有不可预见的结果吗?活动是否导致预期结果?干预或假设是否需要修改?

评估可以帮助回答如下的一些重要问题:干预是可行和可接受的吗?项目干预有影响吗?影响了谁?怎样影响?观察到结果吗?结果是否可信?测量或观察到的变化在多大程度上是因为干预?能否负担起以及是否划算?能否复制到其他情景?哪里可以复制以及哪里不可以复制?结果有推广价值吗?也就是说,项目干预可适应、复制或建立到更大的区域范围(更多的人口或不同地区)吗?如果是,怎样扩展?哪些方面可以扩展?

项目评估面临的挑战主要表现为:缺乏可比的指标和工具,因此很难跨地区进行比较;许多研究测量的是过程和结果而不是影响;也有很多衡量的变化是在个体层面而不是在社区一级;不同类型的干预方式(政策和法律改革,健康增进、保障和供养服务,社区动员,提高认识活动)和不同的环境需要不同的评估工具和方法;很难确定特定的策略对观察结果的贡献或影响,尤其是对于复杂的、多部门实施的综合的干预措施而言;很难定义特定的干预措施成功意味着什么;缺乏严格的统计方法(rigorous statistical methods are frequently not used);监测和评估计划往往缺乏清晰、适当的概念框架;数据解释通常需要专家解读;监测和评估往往没有分配足够的资源,这可能会花费整个预算的10%~40%,取决于项目的目的和目标、干预和活动的范围和类型。同时,在某些情况下,评估者取得资源支持并不是一件十分容易的事情。

1.2 监测和评估框架

不论在何种情况下进行影响评估,都需要一些使其有意义的先决条件,影响评估建立在早期其他形式的评估基础之上。评估不合理的、不可测量的

项目产出或者是尚未充分实施的项目，是一种浪费时间、精力和资源的做法。只有当项目干预已经介入了足够长的时间、消除了基本实施问题时，才应该考虑项目干预的影响。

1.2.1 项目周期、实施与评估

监测和评估是项目整体不可分割的一部分，并应该从整个周期考虑（包括评估、战略规划、设计、监控和评价，见图 1—3）。监测和评估不应被认为是项目之外的事，而应被视为项目的一部分，应在干预发生之前考虑，项目的不同阶段需要不同类型的评估。

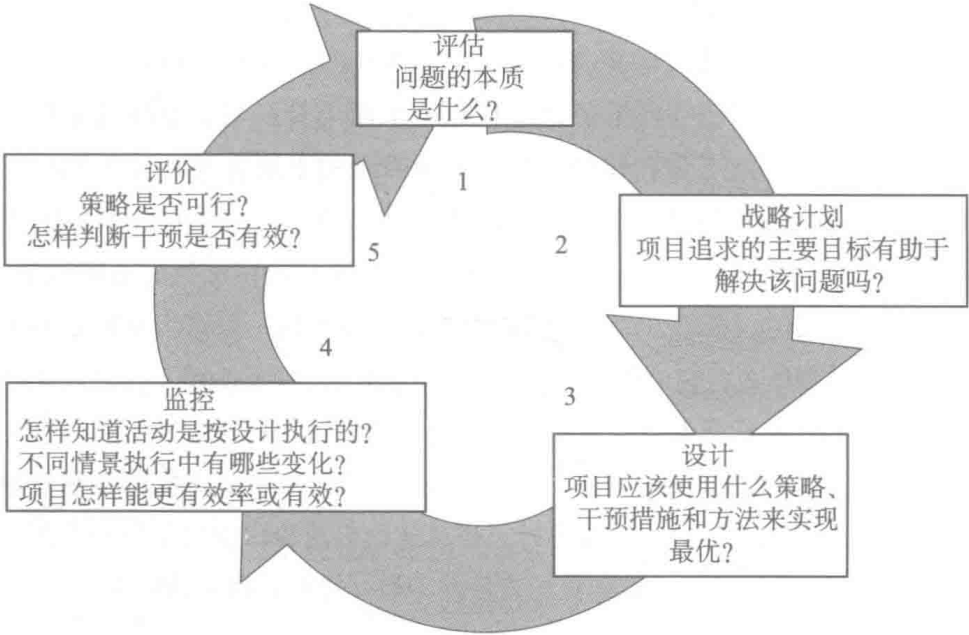


图 1—3 项目周期中的监测和评估

资料来源：Caro, Deborah, Manual for Integrating Gender into Reproductive Health and HIV Programs: From Commitment to Action (Second Edition), Population Reference Bureau on behalf of the Interagency Gender Working Group, funded by the U. S. Agency for International Development, Washington, D. C., 2009.

伴随着项目的计划 and 设计阶段的评估，被称为需求评估（need assessment）、情景分析（situation analysis）、项目构建（conceptualization）或项目论证（program initiation）等，处于这个阶段的评估是明确系统当前的主要