

知证卫生决策工具之十八

——制定政策监测和评估计划*

SUPPORT Tools for Evidence-informed Policymaking in health 18: Planning monitoring and evaluation of policies*

Atle Fretheim¹, Andrew D Oxman², John N Lavis³, Simon Lewin⁴

1. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway; Section for International Health, Institute of General Practice and Community Medicine, Faculty of Medicine, University of Oslo, Norway;

2. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway;

3. Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, and Department of Political Science, McMaster University, 1200 Main St. West, HSC-2D3, Hamilton, ON, Canada, L8N 3Z5;

4. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway; Health Systems Research Unit, Medical Research Council of South Africa

摘要 “监测”一词常用于描述这样的过程,即系统地收集数据以告知决策者、管理者及其他利益相关者一项新政策或规划的实施是否符合其预期。出于监测的目的而使用指标作出相关判断,如是否达到预期目标,或是否恰当地使用所分配的资金。有时“评估”一词可与“监测”一词互换,但前者常更强调结果的获得。当使用“效果评估”一词时,常指特别尝试设法去确定观察到的结果变化是否可归因于特定的政策或规划。我们在本文中提出能用于指导政策或规划监测与评估的四个问题:① 是否有必要监测? ② 应测量什么? ③ 是否应开展效果评估? ④ 如何进行效果评估?

1 关于STP

本文是为卫生政策和卫生项目决策者及其支持者所写“知证卫生决策工具”系列文章之一。该系列文章旨在帮助他们确保其所做的决策基于当前可及的最佳研究证据。SUPPORT 工具及使用方法在本系列文章的导论中已有详述^[1]。每篇文后均附有本系列文章的所有术语(见附件1)。SUPPORT 网站上可找到本系列论文西班牙文、葡萄牙文、法文和中文翻译的链接(www.support-collaboration.org)。欢迎通过邮箱 STP@nokc.no 提出修改意见。

2 案例情景

情境 1 你是一名全权负责若干个医疗保健规划的高级公务员。你希望确保拥有必要的信息来评价各规划的实施情况及其影响。

情境 2 你在卫生部工作并奉命准备一个备忘录,关于评价国家免疫规划时应考虑的若干问题。

情境 3 你供职于一个支持政府知证决策的单位。你在准备国家结核控制规划的一项监测评估计划。

* 本文英文原文发表于 *Health Research Policy and Systems*, 2009, 7(Suppl 1):S18 doi:10.1186/1478-4505-7-S1-S18

作者简介: Atle Fretheim, Email: atle.fretheim@nokc.no

3 背景

对决策者(案例情境1)而言,当其下属在制定一项新政策的监测和评估计划时,本文提出了一些问题供决策者提问。

对那些决策支持者(案例情境2和3)而言,当其制定如何监测政策和规划实施计划及评估其影响时,本文提出需要思考的若干问题。

决策者及其他利益相关者常需知道一项新的政策或规划的实施是否与预期相符。首次展示的规划是否按计划进行? 是否实现目标及是否合理使用分配的资金?“监测”一词常用于描述系统收集数据以回答上述问题的过程^[2]。“绩效监测”一词常用于评价的重点是“与预期结果相比,一个项目、规划或政策实施情况如何”时^[2]。

“指标”常被用作监测过程的一部分。“指标”的定义为“定量或定性的因素或变量,可提供一个简单可靠的方法来测量业绩、反映与某干预有关的变化,或有助于评估绩效”^[2]。指标可以是一个简单的事件计数,如在一段时间内的疫苗接种数,或基于不同数据来源而构造的值,如所有儿童周岁前获得全面免疫接种的比例。

“评估”一词有时可与“监测”互换,但前者常更强调结果的实现。这些术语使用不一致且对不同的

人可能有不同含义。当试图评价结果中观察到的变化(或“影响”)是否可归因于特定政策或规划时,常用到“效果评估”这个词。

4 思考问题

- (1) 是否有必要监测?
- (2) 应测量什么?
- (3) 是否应开展效果评估?
- (4) 如何进行效果评估?

4.1 是否有必要监测?

监测的重要性取决于利益相关者欲更多地了解到正在发生什么而感觉到的需要。

决定是否应该建立一个系统来监测某政策或规划取决于若干因素,包括:

- 想要指标的监测系统是否已经存在,或是否需要一套新的指标;
- 建立所需监测系统的可能花费。如是否可以在已存在的数据采集程序中加入一些新的条目,或是否还必须进行大规模住户调查或是否有必要开发全新的工具?
- 结果是否可能有用? 若监测表明事情未按计划发展时应采取什么行动?

若其数据不被使用,则监测没有价值。若发现期望和实际结果有差距时就采取纠正行动,则监测数据就特别有用。上述结果可促成重新考虑我们的预期。如可采取评价的形式来评判当初的计划是否太不切实际,或是否一项新的政策已不如预期那样有效。

表 1 用两个实例来说明卫生系统内已存在的监测系统。

4.2 应测量什么?

关注“结果链”不同部分(如投入、活动、产出、结果和影响 - 见图 1)的指标通常用于监测某规划或政策的实施情况。在某些情况下,监测投入(如提供人力和设备资源)就已足够。其他情况下,重要的是监测规划实施活动或其结果(如儿童全面免疫接种人数)。

选择采用的指标时,需考虑若干因素:

- 有效性: 指标测量其欲测量事物的准确程度;
- 可接受性: 对被评估者和进行评估者而言,指标可被接受的程度;
- 可行性: 有效、可靠和一致的数据可被收集到的程度;
- 可靠性: 存在最小测量误差的程度,或由其他机构再次测量时结果的可重复程度;
- 变化敏感性: 指标能够探测测量单位变化的程度;
- 预测效度: 指标准确预测有关结果的程度。

收集数据的费用及分析数据并向管理者和提供者反馈数据的能力也会限制指标的选择。在分析资源缺乏的环境中,更倾向于选择一个简单指标而不是那种需进行统计处理的指标,即便前者预测效度不是最佳。

下面两者间的权衡通常显而易见:一方面想用理想和最佳指标,另一方面不得使用基于已有数

表 1 卫生服务系统内的监测体系举例

马拉维扩大规模提供抗逆转录病毒治疗(ART)^[3]

马拉维卫生当局决定让大部分 HIV 阳性人群获得 ART 时,使用了一个系统来监测这项新政策的实施情况。该系统的原则基于 WHO 监测国家结核病项目的方式。每个开始 ART 的患者都得到一个有唯一识别号的身份卡,并将该卡保存在诊所中。收集新病人信息包括姓名、住址、年龄、身高、监护人姓名、开始 ART 的原因。要求病人每月前来领取药物。来诊所时,记录他们的体重并询问其一般健康情况、可走动状况、工作情况及药物的所有副作用。作为确保药物依从性的一种方式,还计数并记录药丸数量。另外,使用以下类目来记录每月的标准化结果:

- 存活: 患者存活且获得其 30 天的药物供应情况
- 死亡: ART 治疗期间患者已死亡
- 失访: 有 3 个月根本未见到患者
- 停止: 因为副作用或其他原因患者已完全停止治疗
- 转出: 患者已永久地换为接受其他治疗

作为 ART 推广过程的马拉维监测系统的一部分搜集到的数据,可以多种方式对其进行分析和使用。可以对比不同时间征集患者的治疗结果。如从一线疗法转换到二线疗法的比例增加,或病死率也提高,则一线药物耐药增加可能是其原因。如病死率或失访率降低,则可能表明 ART 治疗方案的管理在改善。如某些地理范围或诊所的结果特别糟,则需要采取行动来解决这一问题。

丹麦的肺癌手术^[4]

接受肺癌手术患者的不良结果促使丹麦当局发布肺癌管理的国家临床实践指南。为了监测指南的实施,建立了肺癌患者登记系统,内容包括接受手术患者的具体信息。丹麦肺癌登记系统选择的指标包括机体癌组织严重度(或“分期”),采取的手术步骤,并发症发生情况及生存结果。

作为原因之一,丹麦肺癌登记系统所获数据被用于监测肺癌手术国家建议是否得到遵循。进行当地、地区和全国层面的审计以确定可能妨碍国家指南依从性的有关问题或障碍。基于这些发现,提出了特定策略来提高质量。

据的那些指标。除非绝对需要,不要选择多余的指标,有足够理由这样做。这些理由包括需要限制卫生系统内数据收集的负担,避免收集不需要的数据,及关注收集高质量数据,即使这样采集的数据总量更少^[7]。

卫生系统中常规收集的信息可提供有价值的数
据,能用作以监测为目的的数据源。也可为监测专
门收集数据,如通过调查或访谈。应考虑到数据收
集人员的动力这一问题。多数情况下,卫生工作人
员需将数据收集工作整合到繁忙的日常工作中。因
此,若所收集的信息对他们仅有很少或没有明显的
当地价值,他们承担这类任务的动力就会弱。类似
地,若激励或惩罚与监测结果相联系(如当提供者
的报酬与其绩效指标挂钩),则应考虑数据篡改或系
统博弈的风险。

4.3 是否应开展效果评估?

如上所述,监测活动的一个局限是这些行动未
必表明某政策或规划是否对所测量指标有影响。这
是因为用于监测的指标几乎总会受与特定干预无关
的各种因素影响,使得很难确定哪些因素引起了所
观察到的变化。若监测表明绩效在改善,并不一定
意味着干预是(唯一的)原因。可以想象,即便缺乏
干预,指标总是会改善(见图 2)。

建立规划 / 政策与结果变化之间的因果关系
是效果评估的核心。据世界银行观点^[8],如果那些
接受干预者实际上并未接受该干预,他们会发生什
么? 这是效果评估的核心问题。

基于可靠的资料,如来自以前的评价,可能很有
理由去期待积极的结果。但这样的证据常常缺乏,
或可得的证据不适用于目前的环境。因此,确实存
在这样的风险,即一个新的规划无效,甚至更糟,或
弊大于利。在实施新规划时,这是决策者需澄清的
重要问题。这些知识能使将来的决策获益,无论是
规划或其他管辖权,因此这一问题同样重要。

效果评估花费可能很高。这样的研究是否价有
所值? 可通过比较进行与不进行评价的结果来确
定。如当评价结果为阴性时,该规划是否可能停止
或修正? 如果回答“否”,则进行效果评估的价值显
然有限。

如干预执行过程中可得到评估结果,则效果评
估通常更具货币价值。这种情况下,有机会基于实
施早期开展的效果评估结果来改进或停止该干预。
这种货币价值体现在下面两种情况下: 第一,当不
可能进行试点研究时; 第二,当修改或停止基于评
价结果的早期干预(如需要)可能且可行时。

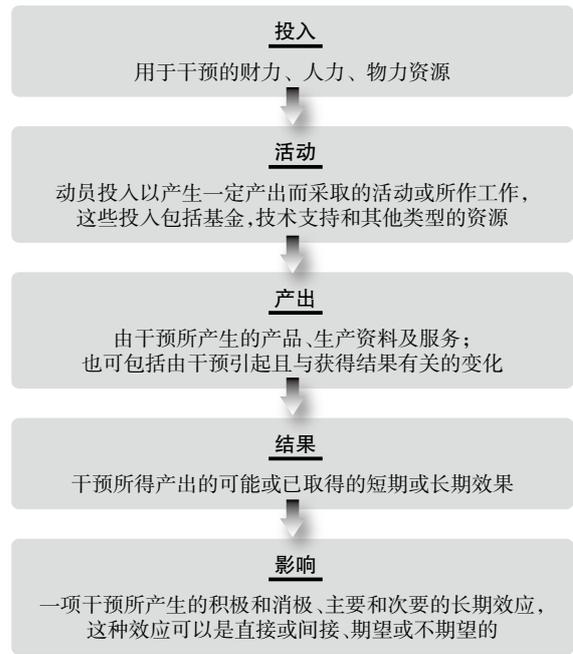


图 1 结果链-模型 (定义改编自文献[2])

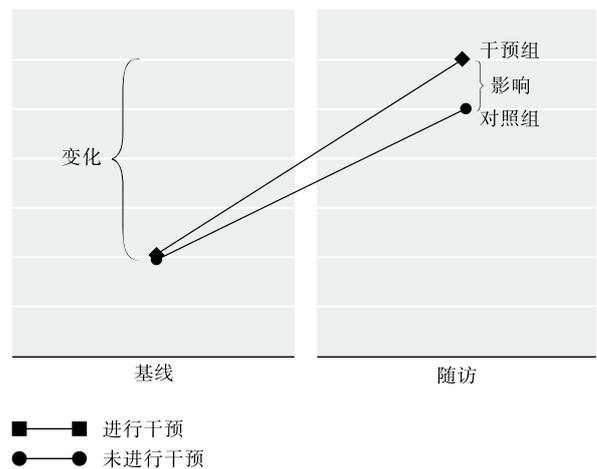


图 2 两个区域绩效变化的对比: 一个区域接受干预, 另一个区域未接受干预

* 此图展示了把从“基线”到“随访”的变化归因于干预,这可能是误导。因为,如此例所示,“对照组”也有改善。即便对对照组而言,干预组和对照组间的差异(即“影响”)是否可真正归因于规划或干预不确定。“干预”和“对照”组间的其他差异也可能导致测量指标上观察到的差异。

墨西哥政府的卫生保险计划“Segro Popular”
即是将效果评估内置于规划实施早期的一个例
子^[9-11]。该计划 2001 年实施,提出该计划旨在将卫
生保险扩大到涵盖近 5 千万尚未被现有计划覆盖
的墨西哥人。充分利用实施早期的进度表,墨西哥
政府通过比较接受与将接受该计划的社区的结果
来评价该计划。除评价改革是否取得了预期结果
且无不期望的不良影响外,该评价也提供了一种
分享式学习模式。

全面实施规划后进行效果评估也可能有用,
如不能确定是否应继续该规划时。如条件性现金

转移支付计划 PROGRESA (后来被称作 oportuni-dades), 该计划在 20 世纪 90 年代中期引进, “当家庭完成某些特定的共同责任, 如让孩子去学校而不是去工作, 给孩子提供特定配方的营养补充, 并到诊所接受指定的健康促进和疾病预防的干预包” 该计划提供现金^[12]。以评价为目的, 506 个社区被随机分配到立即进入该计划组和 2 年后进入该计划组^[13]。该效果评估结果直接为墨西哥的决策提供信息, 说服政府 “不仅要继续进行该规划, 且应推广之”^[12]。

4.4 如何进行效果评估?

要把观察到的变化归因于某规划或政策, 需比较暴露于该规划或政策的个人或群体与未暴露者之间的差异。为了排除规划之外的因素造成的影响, 比较组间应尽可能相似, 这一点也很重要。可通过在被称为 “随机试验” 的试验中将个体或群体 (如在地理区域内的) 随机分配到接受或不接受某规划来有效实现比较组间的相似性。通常在全国性引入该规划之前, 这种试验作为试点研究开展。但也可与规划的全面铺开平行进行, 如上述的墨西哥案例。

随机试验并非总是可行的。其他方法还有: 将规划实施前后的变化与在相同时间段内未实施该规划的地区 (如邻近地区或国家) 所观察到的变化进行比较, 称为 “前 - 后对照评价”。还可使用间断的时间序列研究, 从规划实施前、实施中和实施后的多个时间点收集数据。

通常不推荐简单比较规划实施前、后的某指标值, 因为得出误导结果的风险高 —— 观察到的变化, 如 HIV 发病率可能由与规划本身无关的诸多已

知和未知的因素所致 (见图 2)^[14,15]。

本文末的附件 2 提供了许多评估设计的概述。附件 2 中描述的每种方法的弱点和长处都在附件 3 中写明。

规划实施前应 与利益相关者一起, 包括决策者在内, 设计好效果评估。通常在规划已广泛展开后, 欲进行基线测量或建立适当的对比组为时已晚。如规划已在全国实施后, 就不能使用随机分配来决定某社区是否将进入项目组或非项目组。因此, 项目开始时就已内置于其中的效果评估比事后想起者更可能产生有效的结果。且如果效果评估被看做是规划实施的组成部分, 决策者及其他人可能更会致力于考虑研究结果。

早期也应估计效果评估所需的个人或社区的数量。这将确保足够少的样本量以从评估结果中得出有意义的结论。

像在其他领域中一样, 卫生服务规划既要有效, 又要划算。为了对项目进行经济学评估, 必须估计资源使用和成本, 最好是基于项目实际实施中收集到的数据^[16]。因此, 应该在评估开始前的早些时候决定收集哪些经济学数据。

如果纳入过程评估, 则效果评估可能是最能提供信息的。过程评估可以检查某规划或政策选择是否按所期望的那样被贯彻; 也可以调查实施过程及其变化; 探索对项目的反应; 并探讨得到相关评估结果的原因^[17]。

表 2 是效果评估举例。

预算、时间或数据限制可能成为保证项目严格实施的障碍。这些限制可能以多种方式影响效果评

表 2 效果评估举例

乌干达基于家庭的抗逆转录病毒治疗 (ART)^[20-22]

临床工作人员缺乏和交通费用问题造成的卫生服务利用困难是发展中国家推广 ART 的主要障碍。提出的一个解决办法即基于家庭的 HIV 护理, 这是由非临床专业的合格工作人员在病人家中提供给药、健康状况监测及患者支持的一种方式。但极不确定的是, 该策略能否提供足够质量的医疗卫生服务, 包括及时转诊以使病人接受所需医疗处理, 或该体系是否价有所值。因此, 广泛实施基于家庭的护理策略之前, 重要的是评估该计划的 (成本 -) 效果。

为了公正地比较基于家庭的 ART 和基于卫生机构的 ART, 乌干达研究人员进行了一项随机试验。研究区域被分为 44 个不同的地理亚区。其中一些地区进行家庭护理, 其他地区继续使用传统的基于卫生机构的这一系统。选择和分配哪些地区来接受或不接受基于家庭的护理系统随机确定。这样做减少了比较组间重大差异的可能性, 否则, 如存在下述情况, 这些差异可能影响研究: 参加地区自己决定是否实施基于家庭的护理, 或这种决定基于各地区现存的对实施家庭护理所做的准备情况。所用随机分配系统也是决定在哪里开始进行基于家庭的护理的最公正的方法, 因为每个地区入选机会均等。

研究者发现基于家庭的护理模式使用受过培训的非专业工作人员, 与使用护士和医生主导的基于临床的护理模式一样有效。

挪威强制性使用噻嗪类药物治疗高血压^[23]

作为成本控制措施, 挪威的决策者决定, 在药物费用可报销情况下, 噻嗪类药物应被处方作为抗高血压药, 而不是处方更昂贵的其他药。此决策做出后几个月在全国范围内实施。由于持续的批评一直认为新政策不可能取得预期结果, 卫生部发起一项研究来评价实施该政策的影响。

挪威全国范围内进行强制性处方噻嗪类药物治疗高血压的决策事出紧急, 不可能开展有计划、严格的效果评估。但通过获得后期 61 个诊所的电子病历, 研究者们提取了新政策引入前一年和后一年的处方数据。用有间断的时间序列研究法分析数据。对噻嗪类药物每月处方率及其他感兴趣的结果按时间进行分析, 看是否有任何有统计学意义的变化能归因于该政策的实施。分析表明: 噻嗪类药物的使用急剧增加 (在事先指定的 3 个月过渡期内从 10% 增长到 25%), 随后噻嗪类药物使用又趋于平稳。

估的可靠性:

- 通过放弃总体结果的有效性,如因计划或随访不足;或基线数据缺乏,对不充足数据源的依赖;或选择不恰当的比较组;
- 通过使用不充足的样本,如选择样本时进行方便抽样而样本缺乏代表性;由于样本含量太小;或由于对背景因素缺乏足够的关注。

上述限制可通过早期开始计划或想办法减少数据收集成本来解决。但重要的是确定既非对结果有效性的可能威胁,也非样本限制,使评价结果不能提供可靠信息。因此在做评估之前,应估计是否可能做一个充分的评估。如不可能,需要评估下述问题:某规划的潜在效果存在不确定性时,是否应该实施尚未开展事先评估的这一规划^[18]。

如果其结果不被使用,则效果评估不值得做。评估的结果若用于告知关于是否继续、改变或停止现有规划的决策,显然还应考虑其他利益。如当评估结果与其他被认为是更重要的利益相冲突时,决策者可以选择不关注某评估得到的特定结果^[19]。重要的是避免压制效果评估结果,如出于政治原因。不使用评估结果与进行评估的主要目的之一相矛盾,该目的即从实践中学习并分享已产生的知识。由独立机构进行效果评估可减少结果被篡改或向公众隐瞒的风险。

5 结论

本文描述了有关监测和评价的若干方面。目前,项目监测和评价所作诸多工作常采用的方法不能产生对政策或规划实施的有效评价,或不能有效评估效应。有时这样的评估并未开展。通过思考本文描述的问题,决策者及其支持者应能够制定出可以产生新的且直接有用知识的计划。

资源

有用的文件和补充阅读文献

- Segone M (ed). Bridging the gap: The role of monitoring and evaluation in evidence-based policy making. UNICEF, the World Bank and the International Development Evaluation Association. http://www.unicef.org/ceecis/evidence_based_policy_making.pdf
- MacKay K. How to Build M&E Systems to Support Better Government. 2007. Washington DC, The World Bank. http://www.worldbank.org/ieg/ecd/docs/How_to_build_ME_gov.pdf
- Monitoring and Evaluation (M&E): Some Tools,

Methods and Approaches. 2004. Washington DC. The World Bank. [http://lnweb90.worldbank.org/oed/oed-doelib.nsf/24cc3bb1f94ae11c85256808006a0046/a5efbb5d776b67d285256b1e0079c9a3/\\$FILE/MandE_tools_methods_approaches.pdf](http://lnweb90.worldbank.org/oed/oed-doelib.nsf/24cc3bb1f94ae11c85256808006a0046/a5efbb5d776b67d285256b1e0079c9a3/$FILE/MandE_tools_methods_approaches.pdf)

- Framework for Managing Programme Performance Information. 2007. National Treasury of South Africa. <http://www.treasury.gov.za/publications/guidelines/FMPI.pdf>
- Barber S. Health system strengthening interventions: Making the case for impact evaluation. 2007. Geneva, Alliance for Health Policy and Systems Research. <http://www.who.int/alliance-hpsr/resources/Alliance%20HPSR%20-%20Briefing%20Note%20.pdf>
- Savedoff WD, Levine R, Birdsall N. When will we ever learn? Improving lives through impact evaluation. Report of the Evaluation Gap Working Group. 2006. Washington DC, Center for Global Development. <http://www.cgdev.org/content/publications/detail/7973/>
- Grimshaw J, Campbell M, Eccles M and Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Family Practice* 2000; 17: S11-S18. http://fampra.oxfordjournals.org/cgi/reprint/17/suppl_1/S11

网址链接

- Independent Evaluation Group (IEG) at the World Bank: <http://www.worldbank.org/ieg> - IEG is an independent unit within the World Bank. IEG assesses what is effective or not effective with regard to policy options, how a borrower plans to run and maintain a project, and the lasting contribution of the Bank to a country's overall development.
- International Initiative for Impact Evaluation (3ie): <http://www.3ieimpact.org> - 3ie seeks to improve the lives of poor people in low- and middle-income countries by providing and summarising evidence related to what policy options work, as well as when and why, and the costs involved.
- Health Metrics Network: <http://www.who.int/healthmetrics/en> - The Health Metrics Network (HMN) has the strategic goal of increasing the availability and use of timely and accurate health information. To achieve this, HMN identifies strategies for Health Information System (HIS) development and strengthening, supports countries in implementing HIS reform, and increases

knowledge about global public goods through research, technical innovation, and sharing lessons learned.

- NorthStar: <http://www.rebeqi.org/?pageID=34&ItemID=35> – North Star is a tool for planning, conducting and evaluating quality improvement programmes.

附件

1. 术语: <http://www.biomedcentral.com/content/supplementary/1478-4505-7-S1-S18-S1.doc>
2. 术语: <http://www.biomedcentral.com/content/supplementary/1478-4505-7-S1-S18-S2.doc>
3. 术语: <http://www.biomedcentral.com/content/supplementary/1478-4505-7-S1-S18-S3.doc>

利益冲突、作者贡献及致谢

见原文。

参 考 文 献

- 1 Lavis JN, Oxman AD, Lewin S, *et al.* SUPPORT Tools for evidence-informed health Policymaking (STP). Introduction. *Health Res Policy Syst*, 2009, 7(Suppl 1): 11.
- 2 Development Assistance Committee Working Party on Aid Evaluation: Glossary of Key Terms in Evaluation and Results Based Management 2002 [<http://www.oecd.org/dataoecd/29/21/2754804.pdf>]. Paris, OECD Publications.
- 3 Harries AD, Gomani P, Teck R, *et al.* Monitoring the response to antiretroviral therapy in resource-poor settings: the Malawi model. *Trans R Soc Trop Med Hyg*, 2004, 98: 695-701.
- 4 Jakobsen E, Palshof T, Osterlind K, *et al.* Data from anational lung cancer registry contributes to improve outcome and quality of surgery: Danish results. *Eur J Cardiothorac Surg*, 2009, 35: 348-352.
- 5 Smith PC, Mossialos E, Papanicolas I. Performance measurement for health system improvement: experiences, challenges and prospects. Background Document for WHO European Ministerial Conference on Health Systems: "Health Systems, Health and Wealth" 2008 [http://www.euro.who.int/document/hsm/2_hsc08_ebd2.pdf]. Copenhagen, World Health Organization, Europe.
- 6 Campbell SM, Braspenning J, Hutchinson A, *et al.* Research methods used in developing and applying quality indicators in primary care. *Qual Saf Health Care*, 2002, 11: 358-364.
- 7 MacKay K. How to Build M&E Systems to Support Better Government 2007 [http://www.worldbank.org/ieg/ecd/docs/How_to_build_ME_gov.pdf]. Washington DC, The World Bank.
- 8 The World Bank: Impact evaluation: Overview 2009 [<http://go.worldbank.org/2DHMCRFFT2>]. The World Bank.
- 9 Moynihan R, Oxman A, Lavis JN, *et al.* Evidence-Informed Health Policy: Using Research to Make Health Systems Healthier. Rapport nr. 1-2008 2008 [<http://www.kunnskapsenteret.no/Publikasjoner/469.cms>]. Oslo, Nasjonalt kunnskapssenter for helsetjenesten.
- 10 Frenk J, Gonzalez-Pier E, Gomez-Dantes O, *et al.* Comprehensive reform to improve health system performance in Mexico. *Lancet*, 2006, 368: 1524-1534.
- 11 Gakidou E, Lozano R, Gonzalez-Pier E, *et al.* Assessing the effect of the 2001-06 Mexican health reform: an interim report card. *Lancet*, 2006, 368: 1920-1935.
- 12 Frenk J. Bridging the divide: global lessons from evidence-based health policy in Mexico. *Lancet*, 2006, 368: 954-961.
- 13 Rivera JA, Sotres-Alvarez D, Habicht JP, *et al.* Impact of the Mexican program for education, health, and nutrition (Progresa) on rates of growth and anemia in infants and young children: a randomized effectiveness study. *JAMA*, 2004, 291: 2563-2570.
- 14 Savedoff WD, Levine R, Birdsall N. When will we ever learn? Improving lives through impact evaluation 2006 [http://www.cgdev.org/files/7973_file_WillWeEverLearn.pdf]. Washington DC, Center for Global Development.
- 15 Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference Houghton Mifflin; 2002.
- 16 Oxman AD, Fretheim A, Lavis JN, *et al.* SUPPORT Tools for evidence-informed health Policymaking (STP). 12. Finding and using research evidence about resource use and costs. *Health Res Policy Syst*, 2009, 7(Suppl 1): S12.
- 17 Lewin S, Glenton C, Oxman AD. Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. *BMJ*, 2009, 339: b3496.
- 18 Oxman AD, Lavis JN, Fretheim A, *et al.* SUPPORT Tools for evidence-informed health Policymaking (STP). 17. Dealing with insufficient research evidence. *Health Res Policy Syst*, 2009, 7(Suppl 1): S17.
- 19 Scheel IB, Hagen KB, Oxman AD. The unbearable lightness of healthcare policy making: a description of a process aimed at giving it some weight. *J Epidemiol Community Health*, 2003, 57: 483-487.
- 20 Amuron B, Coutinho A, Grosskurth H, *et al.* A cluster-randomised trial to compare home-based with health facility-based antiretroviral treatment in Uganda: study design and baseline findings. *Open AIDS J*, 2007, 1: 21-27.
- 21 Jaffar S, Amuron B, Birungi J, *et al.* Integrating research into routine service delivery in an antiretroviral treatment programme: lessons learnt from a cluster randomized trial comparing strategies of HIV care in Jinja, Uganda. *Trop Med Int Health*, 2008, 13: 795-800.
- 22 London School of Hygiene and Tropical Evidence: Home-based HIV care just as effective as clinic-based care in Sub-saharan Africa 2009 [<http://www.lshtm.ac.uk/news/2009/homeHIVcare.html>]. London School of Hygiene and Tropical Evidence, University of London.
- 23 Fretheim A, Havelsrud K, MacLennan G, *et al.* The effects of mandatory prescribing of thiazides for newly treated, uncomplicated hypertension: interrupted time-series analysis. *PLoS Med*, 2007, 4: e232.

李幼平 总审校
杨晓妍 李鸿浩 译
本文编辑: 蔡羽嘉