Ferramentas SUPPORT para a elaboração de políticas de saúde baseadas em evidências (STP)

18. Como monitoraro planejamento e avaliação de políticas?

Atle Fretheim[1*], *Andrew D Oxman*[2], *John N Lavis*[3] *and Simon Lewin*[4]

Fretheim A, Oxman AD, Lavis JN, Lewin S: SUPPORT Tools for evidence-informed health Policymaking (STP). **18. Planning monitoring and evaluation of policies.** Health Research Policy and Systems; 2009, 7(Suppl 1):S18 doi:10.1186/1478-4505-7-S1-S18.

http://www.health-policy-systems.com/content/pdf/1478-4505-7-S1-s18.pdf

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

* Autor responsável por comunicações (atle.fretheim@nokc.no)

Esta é a tradução de um artigo publicado no Health Research Policy and Systems, 2009; 7:Supplement 1 (www.health-policy-systems.com/supplements/7/S1).

Podem ser encontrados links das traduções desta série para o espanhol, português, francês e chinês no website do SUPPORT (www.support-collaboration.org). Opiniões sobre como melhorar as ferramentas nesta série são bem-vindas e devem ser encaminhadas para: STP@nokc.no.

A série de artigos foi preparada como parte do projeto SUPPORT, apoiado pelo 6º Programa-Quadro INCO da Comissão Europeia, contrato 031939. A Norad (Norwegian Agency for Development Cooperation), a AHPSR (Alliance for Health Policy and Systems Research) e o Milbank Memorial Fund organizaram um encontro de revisão por pares no qual se discutiu uma versão prévia da série. John Lavis recebeu salário como Canada Research Chair in Knowledge Transfer and Exchange (Catedrático de pesquisa no Canadá para a transferência e troca de conhecimento). A Norad, o satélite norueguês do grupo EPOC (Cochrane Effective Practice and Organisation of Care), o Norwegian Knowledge Centre for the Health Services, a AHPSR, a CHSRF (Canadian Health Services Research Foundation), a EVIPNet (Evidence-Informed Policy Network) e a Organização Pan-Americana da Saúde apoiaram a tradução e difusão dos artigos. Nenhum dos financiadores atuou na elaboração, revisão ou aprovação do conteúdo.

Este artigo foi traduzido para o português por Ocean Translations e contou com o apoio da Canadian Health Services Research Foundation (CHSRF) http://www.chsrf.ca/; Centro Rosarino de Estudios Perinatales (CREP) www.crep.org.ar; e Organização Pan-Americana da Saúde (OPAS) (www.paho.org/researchportal).

Resumo

Este artigo faz parte de uma série escrita para os responsáveis pelas decisões relacionadas a políticas e programas de saúde e para aqueles que dão apoio a esses tomadores de decisão.

O termo *monitoramento* é frequentemente utilizado para descrever o processo de coleta sistemática de dados para informar os formuladores de políticas, gerentes e outros interessados se uma nova política ou programa estão sendo implementados de acordo com suas expectativas Os indicadores são usados para o monitoramento com o objetivo de determinar, por exemplo, se os objetivos foram alcançados ou se os fundos alocados foram usados adequadamente. Algumas vezes o termo *avaliação* é usado intercaladamente com o termo *monitoramento*, mas o primeiro geralmente sugere um foco mais forte na obtenção dos resultados. Quando o termo *avaliação do impacto* é usado, isto geralmente implica que há um esforço específico para tentar determinar se as mudanças observadas nos resultados podem ser atribuídas a uma determinada política ou programa. Neste artigo, sugerimos quatro perguntas que podem ser usadas para orientar o monitoramento e a avaliação de uma política ou opção de programa. São elas: 1. É necessário o monitoramento? 2. O que deve ser medido? 3. A avaliação do impacto deve ser feita? 4. Como a avaliação do impacto deve ser feita?

SOBRE O STP

Este artigo faz parte de uma série escrita para os responsáveis pelas decisões relacionadas a políticas e programas de saúde e para aqueles que dão apoio a esses tomadores de decisão. A série se destina a ajudar essas pessoas para assegurar que suas decisões sejam devidamente sustentadas pelas melhores evidências de pesquisa disponíveis. As ferramentas SUPPORT e como elas podem ser usadas estão descritas de maneira detalhada na Introdução desta série [1]. Um glossário para toda a série acompanha cada artigo (ver Arquivo adicional 1). Podem ser encontrados links das traduções desta série para o espanhol, português, francês e chinês no site do SUPPORT (www.support-collaboration.org). Opiniões sobre como melhorar as ferramentas desta série são bem-vindas e devem ser encaminhadas para: STP@nokc.no.

CENÁRIOS

Cenário 1: Você um funcionário público sênior responsável por vários programas de saúde. Você quer assegurar que tem informações necessárias para avaliar como os vários programas são desempenhados e o impacto que eles têm.

Cenário 2: Você trabalha no Ministério da Saúde e foi instruído a preparar um memorando de várias questões que deveriam ser levados em consideração quando o programa nacional de vacinação for avaliado.

Cenário 3: Você trabalha numa unidade de apoio do governo e usa a evidência na formulação de políticas. Você está preparando um monitoramento e avaliação para o programa nacional de controle da tuberculose.

EXPERIÊNCIA

Para os formuladores de política (Cenário 1), este artigo sugere várias perguntas que suas equipes podem perguntar no planejamento do monitoramento e da avaliação de uma nova política.

Para aqueles que apóiam os formuladores de decisão (Cenários 2 e 3), este artigo sugere várias perguntas que devem ser consideradas no planejamento de como monitorar a implementação das políticas e programas e a avaliação de seus impactos.

Os formuladores de políticas e outros interessados muitas vezes precisam saber se a nova política ou programa foi implementado de acordo com suas expectativas. O lançamento do programa progrediu conforme planejado? Os objetivos estão sendo obtidos? E os fundos provisionados estão sendo gastos apropriadamente?

Monitoramento é o termo normalmente usado para descrever o processo de coletar dados sistematicamente para fornecer respostas a estas perguntas [2]. O termo *monitoramento de desempenho* é frequentemente usado quando o principal foco de uma avaliação é comparar "quão bem um projeto, programa ou política foi implementado contra os resultados esperados" [2].

Indicadores são frequentemente usados como parte do processo de monitoramento. Um indicador foi definido como um "fator quantitativo, qualitativo ou variável que fornece um meio simples e confiável para medir realizações, refletir as mudanças ligadas á uma intervenção, ou ajudar avaliar o desempenho" [2]. Um indicador pode ser simplesmente a contagem de eventos, por exemplo, o número de vacinações realizadas dentro de um período definido, ou uma estrutura com base em várias fontes de dados, por exemplo, a proporção de todas as crianças que foram imunizadas antes do primeiro ano de vida.

O termo *avaliação* é utilizado indistintamente com o termo *monitoramento*, mas o primeiro normalmente sugere um foco mais forte sobre o alcance dos resultados. Estes termos não são usados consistentemente e podem significar coisas diferentes para pessoas diferentes. O termo *avaliação do impacto* é frequentemente usado quando um esforço/tentativa é feito para avaliar se as alterações observadas nos resultados (ou 'impactos') podem ser atribuídas a uma determinada política ou programa.

PERGUNTAS A CONSIDERAR

1. É necessário o monitoramento?
2. O que deve ser medido?
3. A avaliação do impacto deve ser feita?
4. Como a avaliação do impacto deve ser feita?

1. É necessário o monitoramento?

A importância de monitorar depende da necessidade percebida entre os interessados para saber mais sobre o que está acontecendo 'na base'.

Determinar se um sistema de monitoramento de uma política ou programa deve ser estabelecido pode depender de vários fatores, inclusive:

- Se um sistema de monitoramento já está no local incluindo os indicadores pretendidos, ou se um novo conjunto de indicadores foi solicitado
- O custo provável de estabelecer o sistema solicitado. Por exemplo, poderiam ser adicionados alguns itens novos no procedimento de coletar dados, ou seria necessário conduzir um levantamento domiciliar adicional de larga escala ou ainda desenvolver uma ferramenta completamente nova?

- Se as descobertas podem ser úteis. Que ações deveriam ser tomadas se o monitoramento revelasse que as coisas não estão saindo conforme planejado?

O monitoramento não é vantajoso se os dados permanecem sem uso. Os dados são especialmente úteis se uma ação corretiva for tomada quando uma lacuna for identificada entre os resultados esperados e os reais. Estas descobertas podem resultar em expectativas a ser reconsideradas. Isto pode tomar formas de avaliações, por exemplo, se os planos iniciais foram muito ambiciosos, ou se a nova política deixou de ser tão eficaz quanto era esperado.

Ver Tabela 1 para dois exemplos ilustrativos de sistemas de monitoramento que foram colocados no local dentro de um sistema de saúde [3,4].

2. O que deve ser medido?

Os indicadores que tem como foco várias partes de uma 'rede de resultados' (por exemplo, entradas, atividades, saídas, resultados ou impactos – ver Figura1) são normalmente usados para monitorar a implementação de uma opção de programa ou política. Em algumas circunstâncias isso pode ser suficiente para monitorar entradas (por exemplo, a provisão de recursos como pessoal e equipamento). Em outra isso pode ser importante para monitorar atividades do programa ou resultados (tais como o número de crianças completamente imunizadas).

Vários fatores precisam ser considerados quando selecionamos qual(is) indicador(es) usar [5,6]:

- *Validade*: até onde um indicador mede com precisão o que pretende medir
- *Aceitabilidade*: até que ponto o indicador é aceitável para os que estão sendo avaliados e daqueles submetidos à avaliação
- *Viabilidade*: até que ponto estão disponíveis para coleta dados válidos, consistentes e confiáveis
- *Confiabilidade*: até que ponto há erro de medição mínima ou as descobertas são reprodutíveis, eles deverão ser coletados novamente por outra organização?
- *Sensibilidade à mudança*: até onde um indicador tem a capacidade de detectar mudanças na unidade de medida
- *Validade previsível*: até onde o indicador tem a capacidade de prever resultados relevantes com precisão

Os custos com relação aos dados coletados, a capacidade de análise e ao feedback dos dados aos administradores e fornecedores podem também limitar a escolha dos indicadores. Em locais onde os recursos analíticos são escassos, pode ser preferível selecionar um indicador simples mesmo que se esse não tenha a melhor validade previsível, em vez de um indicador que exija manipulação estatística.

Uma troca muitas vezes é evidente entre querer usar indicadores desejados e ótimos e ter que usar aqueles indicadores baseados em dados existentes. Há boas razões para não selecionar mais indicadores do que aqueles absolutamente essenciais. Essas razões incluem: a necessidade de limitar a carga da coleção de dados dentro de um sistema de saúde, evitar a coleção de dados que não são utilizados, e focar na coleção de dados de alta qualidade, mesmo que isto signifique coletar menos dados completos [7].

As informações coletadas diariamente no sistema de saúde podem fornecer dados valiosos que podem ser usados como fonte de dados para o monitoramento. Os dados podem também ser especificamente coletados para o monitoramento, por exemplo, através de pesquisas e entrevistas. Considerações devem ser dadas no nível de motivação entre aqueles estimados para coletar os dados. Em muitos lugares, a saúde pessoal precisará integrar coleção de dados de uma programação ativa diária. Portanto, se as informações coletadas foram poucas ou sem nenhum valor local, as motivações para o comprometimento destas tarefas podem ser baixas. Do mesmo modo, se incentivos ou penalidades forem associadas com as conclusões do processo de monitoramento (por exemplo, onde o pagamento dos fornecedores está ligado aos indicadores de desempenho), devem ser considerado o risco de manipulação de dados ou sistema de jogo.

3. A avaliação do impacto deve ser feita?

Uma das limitações de monitoramento de atividades, conforme descrito acima é o fato que tais atividades não indicam necessariamente se uma política ou programa apresentou impacto nos indicadores que foram medidos. Isto ocorre porque os indicadores usados para monitoramento quase sempre serão influenciados por outros fatores que não aqueles relacionados com intervenções específicas. Isso faz com que seja extremamente difícil determinar quais fatores causaram as mudanças observadas. Se o monitoramento revelar que o desempenho está melhorando, isto não significa necessariamente que a intervenção é (somente) o fator casual. É possível que os indicadores tivessem melhorado mesmo na falta de uma intervenção (ver Figura 2).

O estabelecimento de uma relação causal entre um programa ou política e as mudanças nos resultados é a essência do que é a avaliação do impacto. O que teria acontecido se com aqueles que receberam uma intervenção eles não tivessem de fato a recebido, é *a* pergunta principal da avaliação do impacto, de acordo com o Banco Mundial [8].

Pode haver fortes razões para esperar resultados positivos com base na documentação sólida de, por exemplo, avaliações prévias. Entretanto, muitas vezes falta esta evidência. Ou a evidência disponível poder não ser aplicada ao cenário atual. Assim, há um risco real que o novo programa possa se ineficaz, ou até mesmo pior, causar mais mal que bem. Essa questão é importante aos formuladores de políticas para esclarecer na implementação de novos programas. Isso também é importante devido ao benefício

que este conhecimento poderia trazer para o futuro da política de saúde tanto no cenário do programa quanto em outras jurisdições.

Conduzir avaliações de impacto poder ser custoso. Se estes estudos representam boa relação de qualidade e preço, estes podem ser constatados através da comparação das consequências do comprometimento de uma avaliação com as consequências do não comprometimento de outra. Por exemplo, um programa pode ser interrompido ou modificado se os resultados fossem negativos? Se a resposta for 'não' o valor de comprometimento de uma avaliação do impacto é claramente limitado.

Geralmente é mais provável representar a qualidade do investimento de uma avaliação de impacto quando os resultados podem ser obtidos enquanto a intervenção estiver sendo implementada. Nestas circunstâncias há uma oportunidade de melhorar ou interromper o lançamento com base nos resultados de uma avaliação de impacto conduzido nos estágios iniciais da implementação. Isto forneceria qualidade do investimento em duas instâncias: primeiro, quando um estudo piloto não for possível e, segundo quando for possível e prático modificar ou interromper o lançamento (se necessário) com base nos resultados.

O sistema de seguro de saúde do governo mexicano, Seguro Popular, é um exemplo de uma avaliação de impacto inserida na implantação de um programa [9-11]. Implementada em 2001, a estrutura foi estabelecida para aumentar a cobertura do seguro de saúde para quase 50 milhões de mexicanos que não estavam cobertos por nenhum outro programa existente. Aproveitando o calendário do uma implementação progressiva, o governo iniciou uma avaliação de comparação de resultados para as comunidades receberem o sistema para aqueles que ainda estavam esperando. Além de avaliar se a reforma obtinha os resultados pretendidos e não tinham efeitos adversos não intencionados, a avaliação também fornece aprendizado compartilhado.

Uma avaliação de impacto podem também ser útil após um programa ser totalmente implementado, por exemplo, quando há incertezas sobre continuidade de um programa. Por exemplo, o sistema de transferência de dinheiro condicional, Progresa (mais tarde conhecido como Oportunidades), que foi introduzido em meados-1990 fornecendo dinheiro "sob as condições de que as famílias preenchessem elementos essenciais de co-responsabilidade, tais como mandar as crianças para a escola em vez ao trabalho, fornecer a elas um suplemento nutricional especial formulado, e comparecer a clínica para receber pacotes específicos de intervenções para promoção da saúde e prevenção de doenças" [12]. Para finalidade de avaliação, 506 comunidades foram aleatoriamente designadas para ou entrar no programa imediatamente ou 2 anos mais tarde [13]. As descobertas desta avaliação do impacto informaram diretamente as decisões de política no México, convencendo o governo "não somente continuar com o programa, como também expandi-lo" [12].

4. Como a avaliação do impacto deve ser feita?

Atribuir uma mudança observada a um programa ou política exige uma comparação entre os indivíduos ou grupos expostos a esta mudança, e outros que não estão. São também importantes que os grupos comparados sejam os mais parecidos possíveis para eliminar outras influencias exceto a do próprio programa. Isto efetivamente pode ser feito por indivíduos colocados aleatoriamente ou grupo de pessoas (por exemplo, dentro de áreas geográficas) para receber ou não o programa, que é denominado *ensaios randomizados* Estas avaliações, geralmente, são conduzidas como projeto piloto antes de um programa ser introduzido nacionalmente. Mas elas podem também ser desenvolvidas em paralelo com implementação em escala, conforme ilustrado pelo exemplo do México dado acima.

Ensaios randomizados podem, entretanto, não ser sempre viáveis. As abordagens alternativas incluem a comparação de mudanças antes e depois de um programa de implementação, com mudanças observadas durante o mesmo período em áreas onde o programa não foi implementado (por exemplo, nas vizinhanças dos distritos e países). O que é denominado *avaliação controlada antes-depois*. Alternativamente, uma *série temporal interrompida* pode ser usada em dados que foram coletados em diferentes períodos, antes, durante e após o programa de implementação.

Na maioria das vezes não é recomendado simplesmente comparar o valor de um indicador antes e depois do programa implementado já que o risco de descobertas errôneas é alto – por exemplo, mudanças observadas.
HIV/AIDS – a incidência pode ser causada por fatores conhecidos e desconhecidos exceto aqueles relacionados no próprio programa (ver Figura 2) [14,15].

Uma perspectiva de vários projetos de avaliações é fornecida no Arquivo Adicional 2 no final deste artigo. As fraquezas e fortalezas de cada método descrito no Arquivo Adicional 2 estão esquematizadas no Arquivo Adicional 3.

As avaliações de impacto deveriam ser planejadas bem a frente da implementação do programa em conjunto com os interessados, inclusive os formuladores de política. Após um programa ter sido amplamente implementado é geralmente tarde demais para conduzir medidas de base ou estabelecer comparações de grupos apropriados. Por exemplo, usar distribuições ao acaso para decidir se as comunidades serão inclusas ou não no programa, não podem ser feitas após o programa ter sido nacionalmente implementado. As avaliações dos impactos que são incorporadas em um programa desde inicio são, portanto mais prováveis de produzir descobertas mais válidas do que aquelas conduzidas como reflexões posteriores. Além disso, se a avaliação do impacto for vista como parte integrante da implementação do programa, os formuladores de políticas e outros interessados podem ficar mais comprometidos em levar em conta as descobertas.

O número de indivíduos ou comunidades obrigatórios para uma avaliação de impacto também deve ser estimado em um estágio precoce. Isto garantirá que existe amostras de tamanhos suficiente grandes para significativas conclusões para ser projetadas das descobertas de avaliações.

Na saúde, como na maioria das outras áreas, os programas necessitam ser tanto eficazes *quanto* ter custo- benefício. Para avaliar os aspectos econômicos de um programa, o uso do recurso e os custos devem se estimado, preferivelmente com base nos dados coletados da implementação da vida real [16]. As decisões sobre quais dados econômicos coletar deveriam, portanto, também ser feitos em estágio precoce, antes de a avaliação iniciar.

As avaliações do impacto podem ser mais informativas se processo de avaliação for incluída. Um processo de avaliação pode examinar se o programa ou política foi entregue conforme planejado. Pode também investigar o processo de implementação e da mudança, pesquisar respostas para o programa, além de pesquisar razões para as descobertas da avaliação [17].

Ver Tabela 2 para exemplos de avaliações de impacto.

As restrições de orçamento, tempo ou dados podem atuar como desincentivos para garantir a implementação rigorosa. Estas restrições podem influenciar a confiabilidade das avaliações de impactos de várias formas:

- Através do comprometimento da completa validade dos resultados, por exemplo, devido ao planejamento ou acompanhamento insuficiente, ou através da escassez de dados básicos, confiança em fontes de dados inadequados, ou seleção do grupo de comparações inapropriada
- Através do uso inadequado das amostras, por exemplo, devido a seleção das amostras que são convenientes para experimentar, mas podem não ser representativas, em função das amostras serem pequenas demais, ou por falta de adequada atenção aos fatores contextuais

Estas restrições podem ser direcionadas no início do processo de planejamento antecipado ou nas formas de descobertas para reduzir os custos da coleção de dados. Entretanto é importante garantir, que nem as possíveis ameaças a validade dos resultados, nem as limitações de amostra, são tais que os resultados da avaliação serão incapazes de fornecer informações confiáveis. Antes de conduzir uma estimativa, uma avaliação deve em consequência ser feita como se uma estimativa apropriada fosse possível. Se não, uma avaliação precisa ser realizada como se o programa fosse implementado sem avaliação prévia, em face das incertezas sobre seus impactos potenciais [18].

As avaliações de impacto não são válidas se as descobertas não forem usadas. Os resultados podem ser usados para informar as decisões sobre se os programas

existentes devem continuar, mudar ou ser interrompidos. Claramente, outros interesses também precisam ser levados em consideração. Por exemplo, os formuladores de decisão podem eleger não enfatizar uma descoberta específica de determinadas avaliações quando tais descobertas conflitem com outros interesses que são tidos como mais importantes [19]. Contudo, é importante evitar a supressão de descobertas de avaliações de impactos, por exemplo, por razões políticas. Deixar de usar as descobertas na avaliação contradiz um dos principais objetivos de conduzi-las: aprender com a experiência e dividir o conhecimento que foi gerada. Usar partes independentes para conduzir avaliações de impacto pode diminuir o risco de ter as descobertas manipuladas ou restringidas do público.

CONCLUSÃO

Vários aspectos relacionados ao monitoramento e avaliação foram descritos neste artigo. Hoje em dia, muitos programas de monitoramento e aplicações de avaliação são geralmente feitas usando métodos que não produzem avaliações válidas de implementação de uma política ou programa ou estimativas válidas de efeitos. Algumas vezes estas avaliações não são feitas afinal. Tomando as questões descritas neste artigo em consideração, os formuladores de política e aqueles que os apóiam deveriam ser capazes de desenvolver planos que gerarão conhecimentos novos e úteis.

RECURSOS

Documentos úteis e leituras adicionais

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. 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.
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.
lnweb90.worldbank.org/oed/oeddoclib.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. www.who.int/alliance-hpsr/resources/Alliance%20%20HPSR%20-%20Briefing%20Note%202.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. 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

Links de sites

Independent Evaluation Group (IEG) at the World Bank: www.worldbank.org/ieg – IEG é uma unidade independente dentro do Banco Mundial. IEG avalia o que é ou não eficaz com relação às opções de políticas, como um plano mutuário pode funcionar e manter um projeto, e a duradoura contribuição do Banco a país em completo desenvolvimento

International Initiative for Impact Evaluation (3ie): www.3ieimpact.org – 3ie procura melhorar a vida das pessoas pobres em países de baixa – média renda fornecendo e resumindo a evidência relacionada a como as opções de política funcionam, quando e porque, e os custos envolvidos

Health Metrics Network: www.who.int/healthmetrics/en – The Health Metrics Network (HMN) tem como objetivo de estratégia aumentar a disponibilidade e o uso do tempo e informações precisas de saúde. Para alcançar isso, HMN identifica as estratégias para desenvolvimento e fortalecimento do HIS, apóia países em implementação reforma do HIS, e aumenta o conhecimento sobre os bens públicos globais através de pesquisa, inovações técnicas e compartilhamento de lições aprendidas

NorthStar: www.rebeqi.org/?pageID=34&ItemID=35 – NorthStar é uma ferramenta para planejar, conduzir e avaliar programa de melhoria de qualidade

CONFLITO DE INTERESSES

Os autores declaram que não têm interesses conflitantes.

CONTRIBUIÇÕES DOS AUTORES

AF preparou o primeiro esboço deste artigo. ADO, JNL e SL contribuíram para o projeto e na revisão.

RECONHECIMENTOS

Consulte a Introdução desta série para agradecimentos aos financiadores e contribuintes. Além disto, gostaríamos de agradecer a Ruth Levine pelos comentários úteis em uma versão prévia deste artigo.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Lavis JN, Oxman AD, Lewin S, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). Introduction.** *Health Res Policy Syst* 2009, **7 (Suppl 1:I1)**.

2. Development Assistance Committee Working Party on Aid Evaluation: *Glossary of Key Terms in Evaluation and Results Based Management*. Paris, OECD Publications. 2002.

3. Harries AD, Gomani P, Teck R, de Teck OA, Bakali E, Zachariah 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, Pilegaard H: **Data from a national lung cancer registry contributes to improve outcome and quality of surgery: Danish results.** *Eur J Cardiothorac Surg* 2009, **35:**348-52.

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". Copenhagen, World Health Organization, Europe. 2008.

6. Campbell SM, Braspenning J, Hutchinson A, Marshall M: **Research methods used in developing and applying quality indicators in primary care.** *Qual Saf Health Care* 2002, **11:**358-64.

7. MacKay K: *How to Build M&E Systems to Support Better Government*. Washington DC, The World Bank. 2007.

8. The World Bank: *Impact evaluation: Overview*. [http://go.worldbank.org/2DHMCRFFT2]. The World Bank. 2009.

9. Moynihan R, Oxman A, Lavis JN, Paulsen E: *Evidence-Informed Health Policy: Using Research to Make Health Systems Healthier*. Rapport nr. 1-2008. Oslo, Nasjonalt kunnskapssenter for helsetjenesten. 2008.

10. Frenk J, Gonzalez-Pier E, Gomez-Dantes O, Lezana MA, Knaul FM: **Comprehensive reform to improve health system performance in Mexico.** *Lancet* 2006, **368:**1524-34.

11. Gakidou E, Lozano R, Gonzalez-Pier E, Abbott-Klafter J, Barofsky JT, Bryson-Cahn C, et al: **Assessing the effect of the 2001-06 Mexican health reform: an interim report card.** *Lancet* 2006, **368:**1920-35.

12. Frenk J: **Bridging the divide: global lessons from evidence-based health policy in Mexico.** *Lancet* 2006, **368:**954-61.

13. Rivera JA, Sotres-Alvarez D, Habicht JP, Shamah T, Villalpando S: **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:**2563-70.

14. Savedoff WD, Levine R, Birdsall N: *When will we ever learn? Improving lives through impact evaluation*. Washington DC, Center for Global Development. 2006.

15. Shadish WR, Cook TD, Campbell DT: *Experimental and Quasi-Experimental Desings for Generalized Causal Inference*. Houghton Mifflin; 2002.

16. Oxman AD, Fretheim A, Lavis JN, Lewin S: **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, Lewin S: **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:**483-7.

20. Amuron B, Coutinho A, Grosskurth H, Nabiryo C, Birungi J, Namara G, 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:**21-7.

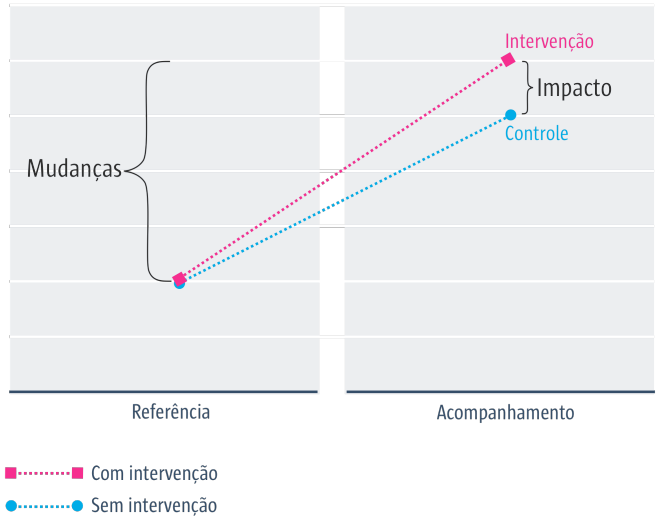
21. Jaffar S, Amuron B, Birungi J, Namara G, Nabiryo C, Coutinho A, et al: **Integrating research into routine service delivery in an antiretroviral treatment programme: lessons learnt from a cluster randomized trial comparing strategies of HIV care in Jinja, Uganda.** *Trop Med Int Health* 2008, **13:**795-800.

22. London School of Hygiene and Tropical Evidence: *Home-based HIV care just as effective as clinic-based care in Sub-saharan Africa*. [http://www.lshtm.ac.uk/news/2009/homeHIVcare.html]. London School of Hygiene and Tropical Evidence, University of London. 2009.

23. Fretheim A, Havelsrud K, MacLennan G, Kristoffersen DT, Oxman AD: **The effects of mandatory prescribing of thiazides for newly treated, uncomplicated hypertension: interrupted time-series analysis.** *PLoS Med* 2007, **4:**e232.

Figura 1. Resultados de modelo de cadeia (definições adaptadas do [2])

Figura 2. Comparar mudanças no desempenho em duas áreas: uma com intervenção e a outra sem*



* A Figura ilustra que atribuindo às mudanças de 'Base' para 'Acompanhamento' em respostas a intervenção é provável ser enganosa. Isto ocorre porque, nesta instancia, há também um aperfeiçoamento no 'Controle'.Mesmo com relação ao Controle, é incerto se a diferença entre a 'Intervenção' e 'Controle' (por exemplo o 'Impacto) pode, de fato, ser atribuído ao programa ou intervenção. Pode haver outras diferenças entre o cenário da 'Intervenção' e do 'Controle' que pode levar a diferenças nas observações do indicador medido.

Tabelas 1. Exemplos de sistemas de monitoramento no sistema de saúde

Incrementando a provisão da terapia anti-retroviral (TARV) em Malawi [3].

Quando as autoridades Malauianas da área de saúde decidiram colocar TARV disponível a uma grande proporção de HIV/AIDS – população positiva, um sistema foi colocado no local para monitorar a implementação desta nova política. Os princípios do sistema estão baseados na abordagem da OMS para o monitoramento de programas nacionais de tuberculose. A cada paciente que inicia a TARV é dado um cartão de identidade com um número exclusivo de identidade, e este número é mantido na clínica. As informações coletadas dos novos pacientes incluem nome, endereço, idade, altura, o nome de seus guardiões, e a razão do início do TARV. É solicitada a presença dos pacientes mensalmente para retirar suas medicações. Durante a visita, os pesos dos pacientes são registrados e eles são questionados sobre sua saúde geral, estado ambulatorial, trabalho e algum efeito colateral da droga. A contagem das pílulas também são realizadas e registradas como uma forma de garantir a adesão ao

tratamento. Além disso, os seguintes resultados mensalmente padronizados são registrados usando as seguintes categorias:

- *Vivo*: O paciente é vivo e retirou seu próprio medicamento – medicamento para 30 dias
- *Morto*: O paciente morreu durante a TARV
- *Ausência*: O paciente não foi visto por um período de 3 meses
- *Interrupção*: O paciente interrompeu o tratamento completamente ou devido aos efeitos colaterais ou por outras razões
- *Transferência*: O paciente foi transferido permanentemente para outro tratamento

Dados coletados podem ser analisados e usados numa variedade de formas como parte do sistema de implementação de monitoramento Malauiano da TARV Comparações dos resultados do tratamento dos pacientes que foram admitidos em períodos diferentes podem ser realizadas. Se, por exemplo, aumentar o índice de troca de regime da primeira para a segunda linha ou de mortalidade, a cause poderia ser um aumento da resistência à droga no regime de primeira linha. Se a taxa dos mortos ou ausentes diminuir, isto pode indicar que o gerenciamento do programa de tratamento da TARV está melhorando. Medidas são tomadas se os resultados são particularmente pobres em determinadas áreas geográficas ou clinicas.

Cirurgia do câncer de pulmão na Dinamarca [4]

As autoridades dinamarquesas emitiram diretrizes nacionais de clínica prática para gerenciar o câncer de pulmão sugeridos pelos maus resultados dos pacientes que se submeteram a cirurgia do câncer de pulmão. Para monitorar a implementação das diretrizes, foi estabelecido um registro de pacientes com câncer de pulmão no qual incluíam informações especificas sobre aqueles pacientes que se submetem a cirurgia. Os indicadores selecionados pelo Registro de Câncer de Pulmão Dinamarquês incluíram: o grau (ou 'estágio') do câncer no corpo, o procedimento cirúrgico utilizado, quaisquer complicações ocorridas, e o resultado de sobrevivência.

Os dados do Registro de Câncer de Pulmão Dinamarquês são usados, entre outras razões, para monitorar se as recomendações nacionais da cirurgia do câncer de pulmão estão sendo seguidas. As auditorias locais, regionais e nacionais são realizadas com a finalidade de identificar problemas ou barreiras que podem impedir a adesão das diretrizes nacionais. Com base nestas descobertas, especificas estratégias foram propostas para melhoria da qualidade.

Tabela 2. Exemplos de avaliações de impacto

Terapia domiciliar anti-retroviral (TARV) em Uganda [20-22]

A carência de equipe clinica e as dificuldades no acesso do serviço de saúde devido aos custos de transporte são os maiores obstáculos para incrementar o fornecimento da TARV em países em desenvolvimento. Uma solução proposta é o serviço de saúde domiciliar de HIV/AIDS, em que o fornecimento da droga, monitoramento do estado de saúde, e o apoio ao paciente é conduzido nas residências dos pacientes por equipe não-clinicamente qualificada. Porém, é totalmente incerto se esta estratégia é capaz de fornecer serviços de saúde de qualidade suficiente, inclusive períodos de consultas para assistência médica, ou se este sistema tem custo-benefício. Portanto, após a ampla implementação do programa domiciliar de serviços de saúde é importante que sejam avaliados o custo-benefício.

Para garantir a justa comparação entre TRAV sem recurso (domiciliar) e com recursos, os pesquisadores em Uganda conduziram um ensaio randomizado. A área do estudo foi dividido em 44 sub-áreas geográficas distintas. Em algumas destas áreas, o cuidado domiciliar foi implementado, enquanto em outras áreas continuaram a usar o sistema convencional de recursos. A seleção e a atribuição das áreas para receber ou não o sistema de serviço de saúde domiciliar, foi determinado aleatoriamente. Isto reduziu a probabilidade de diferenças importantes entre os grupos de comparações que poderiam influenciar o estudo se, por exemplo, os próprios distritos decidissem se implementaria o sistema domiciliar, ou se a decisão seria baseada na preparação existente para implementação dos serviços de saúde domiciliar. O sistema de atribuição aleatória usado foi também o modo mais justo de decidir onde iniciaria o sistema domiciliar já que cada distrito tinha chances iguais de serem escolhidos.

Os pesquisadores descobriram que o modelo de cuidado domiciliar com leigos treinados era tão efetivo quanto o cuidado hospitalar com enfermeiros e médicos.

Utilização obrigatória do tiazídico para hipertensão na Noruega [23]

Como medida de redução de custos, os formuladores de política da Noruega decidiram que o tiazídico seria prescrito como medicamento anti-hipertensivo ao invés de alternativas mais onerosas, nos casos em que as despesas com medicamentos eram reembolsadas. A política foi implementada nacionalmente uns meses após a decisão ser tomada. Porque os críticos continuaram argumentar que a nova política era improvável de conduzir aos resultados esperados, o Ministro da Saúde patrocinou um estudo para avaliar o impacto da política que tinha implementado.

A prescrição obrigatória de tiazídico para o tratamento de hipertensão foi implementada com urgência em toda Noruega de modo que foi impossível realizar uma avaliação rigorosa do impacto. Contudo, através do acesso eletrônico dos registros médicos de 61 clínicas em um estágio posterior, os pesquisadores extraíram dados de

prescrição variando de um ano antes para um ano depois de a nova política ser introduzida. Eles analisaram os dados usando série temporal interrompida. Taxas mensais de tiazídico prescritas e outros resultados de interesses foram analisados durante um tempo para verificar se alguma mudança significante poderia ser atribuída à implementação da política. As análises indicaram que houve um aumento acentuado no uso do tiazídico (de 10 a 25% sobre um período de transição pré-especificado de três meses), seguido da estabilização do uso.

Arquivo adicional 2. Estudos de avaliação (adaptados do Cochrane Handbook for Systematic Reviews of Interventions* (Cochrane Handbook para revisões sistemáticas de intervenções))

Ensaio randomizado controlado	• Um estudo experimental no qual indivíduos são aleatoriamente alocados para receber intervenções diferentes (por exemplo, mediante o lançamento de uma moeda ou uma lista de números aleatórios gerados por computador).
Ensaio coletivo randomizado	• Um estudo experimental no qual um grupo de pessoas (por exemplo, escolas ou hospitais) são aleatoriamente alocadas para receber distintas intervenções.
Ensaio controlado não randomizado	• Um estudo experimental no qual pessoas são alocadas a intervenções diferentes usando métodos que não são randomizados (por exemplo, pacientes admitidos durante a Semana 1 recebem a intervenção A, pacientes admitidos na Semana 2 recebem a intervenção B, e aqueles da Semana 3 recebem a intervenção A novamente, e assim por diante).
Estudo controlado antes e depois	• Um estudo no qual observações são feitas antes e depois da implementação de uma intervenção, tanto em um grupo que recebe a intervenção quanto em um grupo de controle que não a recebe. Normalmente, a coleta de dados deve ser feita simultaneamente nos dois grupos.
Estudo de séries temporais interrompido	• Um estudo usando observações em vários pontos temporais antes e depois de uma intervenção. As medidas são *interrompidas* pela intervenção. O estudo tenta detectar se uma intervenção teve um efeito significativo maior que qualquer tendência implícita ao longo do tempo.
Estudo	• Um estudo que compara um grupo de participantes que

historicamente controlado	recebem uma intervenção com um grupo similar do passado que não a recebeu.
Estudo de coorte	• Um estudo no qual um grupo definido de pessoas (o coorte) é seguido durante um tempo para examinar associações entre distintas intervenções recebidas e os resultados subsequentes. Um estudo de coorte *prospectivo* recruta participantes antes de qualquer intervenção e os acompanha no futuro. Um estudo de coorte *retrospectivo* identifica indivíduos de registros antigos, descrevendo as intervenções recebidas e os acompanha a partir destes registros.
Estudo de controle de caso	• Um estudo que compara pessoas com um resultado de interesse específico (*casos*) com pessoas da mesma população de origem, mas sem tal resultado (*controles*) para examinar a associação entre o resultado e a exposição prévia (por exemplo, recebendo uma intervenção). O estudo é particularmente útil quando o resultado é raro.
Estudo transversal	• Um estudo que coleta informações de intervenções passadas ou presentes e resultados atuais de saúde de um grupo de pessoas em um determinado momento. Este tipo de estudo examina associações entre os resultados e as exposições a intervenções.
Estudo qualitativo	• Um estudo realizado em um cenário natural que é normalmente pensado para interpretar ou compreender o fenômeno em termos dos significados que as pessoas atribuem a ele. Tipicamente neste estudo, coletam-se dados narrativos de indivíduos ou grupos de informantes ou de documentos. Estes são posteriormente interpretados pelo(s) pesquisador(es).

* Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions. Chichester: The Cochrane Collaboration and John Wiley & Sons Ltd.; 2008

Arquivo adicional 3. Pontos fortes e fracos selecionados de estudos de avaliação

	Pontos fortes	Pontos fracos
Ensaio randomizado controlado	• Amplamente considerado como o estudo mais sólido para estabelecer relações de causa e efeito, que é o principal foco do impacto da avaliação	• Pode ser demorado e representar desafios logísticos • Os resultados não são necessariamente transferíveis para cenários outros que o cenário de estudo
Ensaio coletivo randomizado	• Alguns pontos fortes em comparação a ensaios randomizados comuns. Além disso, o risco de "contaminação" é reduzido, por exemplo, que a intervenção A possa ser recebida por, ou afetar indivíduos alocados para receber somente a intervenção B. Por exemplo, se enfermeiros são alocados aleatoriamente para implementar uma nova rotina, outros enfermeiros podem ser influenciados por estas mudanças e passar a empreender as mesmas atividades. E, por isso, talvez seja melhor randomizar as alas, e todos os funcionários dentro dela, em vez de enfermeiros individuais	• Diferenças de base podem constituir um problema, já que o número de unidades (ou agrupamentos) randomizadas pode ser normalmente menor do que em ensaios nos quais indivíduos são randomizados. Isso pode ser demorado e constituir um desafio logístico, embora seja ainda mais difícil nos ensaios randomizados comuns

	Pontos fortes	Pontos fracos
Ensaio controlado não randomizado	• Pode ser mais fácil e prático de ser realizado do que uma ensaio randomizado controlado	• Quando a atribuição não é feita usando métodos randomizados, podem ocorrer tendências de seleção, pois pacientes e profissionais da área de saúde ajustarão seus comportamentos ao procedimento de alocação, se preferirem uma intervenção à outra, por exemplo
Estudo controlado antes e depois	• Pode ser a única opção prática, por exemplo, para intervenções de larga escala nas quais a randomização não é viável por razões práticas ou políticas	• Diferenças conhecidas ou desconhecidas entre os grupos que são comparados podem exercer mais influência nas descobertas do que no fato de que receberam diferentes intervenções. Consequentemente, tirar conclusões sobre relações de causa e efeito pode ser arriscado • Exige disponibilidade de dados de base
Estudo de séries temporais interrompido	• Pode ser viável e relativamente fácil de realizar se os dados necessários estiverem disponíveis. Não é necessário grupo de controle	• A dimensão do efeito é sempre difícil de avaliar neste tipo de análise, já que influências diferentes da intervenção sendo investigada podem impactar as mudanças observadas
Estudo historicamente controlado	• Pode ser feito de maneira rápida e fácil se os dados necessários estiverem disponíveis	• Diferenças conhecidas ou desconhecidas entre os grupos que são comparados podem exercer mais influência nas descobertas do que no fato de que receberam diferentes intervenções

	Pontos fortes	Pontos fracos
		Consequentemente, tirar conclusões sobre relações de causa e efeito é arriscado
Estudo de coorte	• Geralmente grandes estudos com um alto grau de validade externa (por exemplo, as descobertas podem ser generalizadas). São geralmente conduzidos por vários anos, o que possibilita detectar os efeitos a longo prazo de uma intervenção	• Estudos de coorte são tipicamente longos e onerosos, devido principalmente à necessidade de acompanhar uma grande quantidade (geralmente) de participantes • Diferenças conhecidas ou desconhecidas entre os grupos que são comparados podem exercer mais influência nas descobertas do que no fato de que foram expostos a diferentes intervenções. Consequentemente, tirar conclusões sobre relações de causa e efeito é arriscado
Estudo de controle de caso	• Realizado de maneira mais rápida e fácil do que o estudo de coorte	• A natureza retrospectiva destes estudos implica na coleta de informações sobre eventos que ocorreram anteriormente. Estes atrasos podem ser uma fonte de erro • Diferenças conhecidas ou desconhecidas entre os grupos que são comparados podem exercer mais influência nas descobertas do que no fato de que receberam diferentes intervenções. Consequentemente, tirar conclusões sobre relações de causa e efeito é arriscado

	Pontos fortes	Pontos fracos
Estudo transversal	• Não exige tempo de acompanhamento e, por isso, pode ser conduzido de maneira rápida e menos onerosa	• Diferenças conhecidas ou desconhecidas entre os grupos que são comparados podem exercer mais influência nas descobertas do que no fato de que receberam diferentes intervenções. Consequentemente, tirar conclusões sobre relações de causa e efeito é arriscado
Estudo qualitativo	• Permite a coleta de informações mais detalhadas do que outros estudos quantitativos. Permite saber se as intervenções e os programas estão funcionando ou não	• Não gera dados que podem ser usados para avaliar os efeitos de uma intervenção que estão além da percepção daqueles que são entrevistados ou investigados