



# **Guia Metodológico de Avaliação e Definição de Indicadores Doenças Crônicas Não Transmissíveis e Rede Carmen**

**MINISTÉRIO DA SAÚDE**  
**Secretaria de Vigilância em Saúde**  
**Departamento de Análise de Situação de Saúde**

# **Guia Metodológico de Avaliação e Definição de Indicadores Doenças Crônicas Não Transmissíveis e Rede Carmen**

**BRASÍLIA / DF**

© 2007 Ministério da Saúde

Todos os direitos reservados. É permitida a reprodução parcial ou total desta obra, desde que citada a fonte e que não seja para venda ou qualquer fim comercial.

A responsabilidade pelos direitos autorais de textos e imagens desta obra é de responsabilidade da área técnica.

A coleção institucional do Ministério da Saúde pode ser acessada na íntegra na Biblioteca Virtual do Ministério da Saúde: <http://www.saude.gov.br/bvs>

## **Série A. Normas e Manuais Técnicos**

**Tiragem:** 1.ª edição – 2007 – 20.000 exemplares

### **Elaboração, distribuição e informações**

MINISTÉRIO DA SAÚDE

Secretaria de Vigilância em Saúde

Departamento de Análise de Situação em Saúde

### **Produção: Núcleo de Comunicação**

Esplanada dos Ministérios, bloco G,

Edifício Sede, 1.º andar, sala 134

CEP: 70058-900, Brasília – DF

E-mail: [svs@saude.gov.br](mailto:svs@saude.gov.br)

Home page: [www.saude.gov.br/svs](http://www.saude.gov.br/svs)

Impresso no Brasil / *Printed in Brazil*

## **Ficha Catalográfica**

---

Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise de Situação de Saúde.

Guia metodológico de avaliação e definição de indicadores : doenças crônicas não transmissíveis e Rede Carmem / Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Análise de Situação de Saúde . – Brasília : Ministério da Saúde, 2007.

233 p. : il. – (Série A. Normas e Manuais Técnicos)

ISBN 978-85-334-1316-0

1. Doença Crônica. 2. Vigilância em Saúde. I. Título. II. Série.

NLM WT 500

---

**Catálogo na fonte – Coordenação-Geral de Documentação e Informação – Editora MS – OS 2007/0065**

### **Títulos para indexação:**

Em inglês: Methodological Handbook on Evaluation and Definition of Indicators: non transmissible chronic diseases and Rede Carmem

Em espanhol: Guía Metodológico de Avaluación y Definición de Indicadores: enfermedades crónicas non contagiosas y Red Carmem

## Sumário

<b>Apresentação</b>	<b>7</b>
O desafio do Brasil na construção do Guia de Avaliação Carmen e Vigilância de Doenças Crônicas Não Transmissíveis <i>Gerson Oliveira Penna</i>	
<b>Introdução</b>	<b>11</b>
Guia Metodológico de Avaliação e Definição de Indicadores – Doenças Crônicas Não Transmissíveis e Rede Carmen <i>Deborah Carvalho Malta</i> <i>Otalíba Libânio de Moraes Neto</i>	11
Guia Metodológico de Avaliação e Definição de Indicadores – Doenças Crônicas Não Transmissíveis e Rede Carmen <i>Iná S. Santos</i>	17
A iniciativa Carmen	18
<b>PARTE I</b>	<b>21</b>
Revisão da Literatura <i>Iná S. Santos, Marcelo Capilheira e Carmen Moreira</i>	21
Metodologia da Revisão Bibliográfica	22
Programas de Prevenção de Doenças Crônicas Não Transmissíveis em Países Desenvolvidos	25
North Karelia Project	26
The Stanford Community Projects	31
The Minnesota Heart Health Program (MHHP)	32
The Pawtucket Heart Health Program (PHHP)	34
Countrywide Integrated Noncommunicable Diseases Intervention (Cindi)	36
Multifatorial Monitoring of Trends and Determinants in Cardiovascular Disease (Monica)	40
Programas de Prevenção de Doenças Crônicas Não Transmissíveis em Países em Desenvolvimento	42
Who's Interhealth Programme	43
Programa Mirame!	43
Tianjin Project	44
Programa de Prevenção de Infartos na Argentina (Propia)	44
Conjunto de Acciones para la Reducción Multifactorial de Enfermedades No transmissibles (Carmen) – OPAS	45
Financiamento dos Programas	50
Perspectivas	50
Intervenções Atuais	52
Conclusões	53

	<b>PARTE II</b>	<b>67</b>
	Indicadores de Avaliação <i>Iná S. Santos</i>	67
	Referencial Teórico	68
Exemplos de Indicadores utilizados em Avaliações de Programas / Intervenções Preventivas de DCNTs		71
	Indicadores de Oferta	72
	Indicadores de Utilização	73
	Indicadores de Cobertura	73
	Indicadores de Impacto	73
Relação entre os Indicadores do Marco Referencial de Habitch <i>et al</i> e os de Avedis Donabedian		76
	<b>PARTE III</b>	<b>81</b>
	Grau de Inferência da Avaliação <i>Iná S. Santos</i>	81
	Avaliações de Adequação	82
	Avaliações de Plausibilidade	84
	Avaliações Probabilísticas	88
	Decidindo Quanto a Que Grau de Inferência Usar	90
Exemplos de Estudos Utilizados para Avaliação de Acordo com o Grau de Inferência		93
	Avaliação da Adequação do Desempenho (Processo) do Programa / Intervenção	94
	Avaliação da Adequação do Impacto do Programa ou Intervenção	94
Avaliação de Plausibilidade do Desempenho (Processo) do Programa ou Intervenção		96
	Avaliação de Plausibilidade do Impacto do Programa ou Intervenção	97
	Com Grupo Controle Histórico	97
	Com Grupo Controle Interno	97
	Com Grupo Controle Externo	98
	Avaliação Probabilística do Desempenho (Processo) do Programa ou Intervenção	100
	Avaliação Probabilística do Impacto do Programa ou Intervenção	101
	<b>PARTE IV</b>	<b>103</b>
	Planejando a Avaliação de Programas / Intervenções <i>Iná S. Santos</i>	103
	Avaliações com Dados Primários	111
	Usando Estudos Transversais	112
	Usando Estudos Quase-experimentais	127
Usando Estudos Longitudinais com Análise Tipo Antes-e-depois, sem Grupo Controle		140
	Usando Estudos Tipo Casos e Controles	147
	Usando Ensaios Randomizados	154

Avaliações com Dados Secundários	166
Usando Sistemas de Vigilância de Fatores de Risco e Doenças Crônicas como Instrumento de Avaliação	167
Conclusões e Recomendações	171
<b>PARTE V</b>	<b>175</b>
Manual de treinamento	175
<i>Iná S. Santos, Deborah Carvalho Malta, Otaliba Libânio de Moraes Neto, Lenildo de Moura, Luciana Sardinha, Elisabeth Duarte e Denise Bontempo</i>	
PASSO 1: Definir que decisões serão tomadas em decorrência da avaliação e quem serão os tomadores de decisão	178
PASSO 2. Definir quem vai fazer a avaliação	180
PASSO 3. Definir quais indicadores serão medidos	181
PASSO 4. Com base nos dados anteriores, escrever o protocolo da avaliação	198
PASSO 5. Preparar os instrumentos de pesquisa, coleta e análise	202
PASSO 6. Escrever o relatório final	203
PASSO 7. Disseminar os resultados e as recomendações da avaliação	206
<b>Referências Bibliográficas</b>	<b>207</b>
<b>Anexo</b>	<b>221</b>
Glossário de Termos	221



# **Apresentação**

**O desafio do Brasil na construção do  
Guia de Avaliação Carmen e Vigilância  
de Doenças Crônicas Não Transmissíveis**

Foram muitos os desafios enfrentados na construção de um guia de avaliação da Rede Carmen e de intervenções de Vigilância de Doenças Crônicas Não Transmissíveis (DCNT). A proposta foi delegada pela OPAS ao Ministério da Saúde do Brasil, em 2003, durante o I Fórum Global Mundial, realizado no Rio de Janeiro, Brasil. Seguiram-se duas reuniões com a OPAS em 2004 e 2005, para definir os objetivos da construção do Guia, a expectativa dos países e o alcance da proposta. Em outubro de 2005, apresentamos aos países e à OPAS na Reunião Bianual da Rede Carmen, no Chile, uma proposta inicial do marco conceitual que norteou o trabalho, e, finalmente, a versão final contendo a contribuição brasileira objetivando a institucionalização da avaliação das DCNTs.

8

Nas discussões entre a OPAS e o Ministério da Saúde, várias questões foram abordadas, ilustrando a complexidade metodológica da tarefa, seja na definição do marco conceitual, seja pela diversidade das experiências dos países e a dificuldade em avançar em uma proposta que pudesse ser factível nas diferentes realidades das Américas. Alguns desafios e questionamentos nortearam o trabalho: Como fazer um guia de avaliação que se tornasse operacional ao conjunto dos países, em contextos e realidades de implantação de experiências tão distintas? O que seria mais adequado na condução de uma avaliação: operar com avaliadores externos ou internos? Como combinar metodologias quantitativas e qualitativas? Como avaliar experiências em níveis de implementação distintos: local, nacional e regional? Como tornar a avaliação um instrumento aplicável no contexto dos serviços? Como utilizar os dados secundários disponíveis ou os sistemas de informação existentes? Quando utilizar estudos realizados no âmbito dos serviços, como, por exemplo, as análises de séries históricas? Como atribuir as mudanças encontradas no desempenho dos programas e serviços? Como sensibilizar os governos e atrair financiamento para avançar nestes estudos? Como o contexto socioeconômico-político poderá influir nos resultados?

Algumas destas questões são abordadas neste Guia Metodológico de Avaliação e Definição de Indicadores de DCNT/Rede Carmen. O texto apresenta inicialmente a revisão bibliográfica de avaliação trazendo experiências de avaliação conduzidas nos diferentes países em relação às DCNTs. Apresenta ainda o referencial teórico adotado e a metodologia dos estudos epidemiológicos de avaliação segundo Habitch et al (1999). São apresentados exemplos de como se empregar indicadores de oferta, utilização, cobertura e impacto. O guia ainda discute o grau de inferência necessário para responder à pergunta avaliativa (adequação, plausibilidade e probabilística) que está diretamente relacionado ao estágio de desenvolvimento do

programa e seus recursos disponíveis. Apresenta-se ainda um guia prático de avaliação aplicável às situações de serviço.

A Secretaria de Vigilância em Saúde do Ministério da Saúde do Brasil envolveu a Universidade Federal de Pelotas – UFPEL, parceira na viabilização desta proposta, por meio da Prof<sup>a</sup>. Dr<sup>a</sup>. Iná Santos, como pesquisadora responsável pela realização dessa tarefa. Foram muitos meses de trabalho, que contaram com a participação da equipe da Coordenação Geral de Doenças e Agravos Não Transmissíveis do Ministério da Saúde, da OPAS – Brasil e da equipe da UFPEL.

Esperamos que o trabalho possa servir como indutor da avaliação de serviços, contribuindo na sua institucionalização no Brasil e nas Américas. Com esse intuito temos a grande satisfação de apresentar à OPAS e aos demais países das Américas o atual trabalho.

Gerson Oliveira Penna  
Secretário de Vigilância em Saúde  
Ministério da Saúde Brasil



# Introdução

## Guia Metodológico de Avaliação e Definição de Indicadores – Doenças Crônicas Não Transmissíveis e Rede Carmen

*Deborah Carvalho Malta<sup>1</sup>*  
*Otaliba Libânio de Moraes Neto<sup>2</sup>*

- 
- 1 Coordenadora Geral de Doenças e Agravos Não Transmissíveis, Secretaria de Vigilância em Saúde – Ministério da Saúde Brasil e docente da Universidade Federal de Minas Gerais – UFMG.
  - 2 Diretor do Departamento de Análise de Situação de Saúde, Secretaria de Vigilância em Saúde – Ministério da Saúde Brasil e docente da Universidade Federal de Goiás – UFG.

O desafio de avaliar as intervenções relacionadas às Doenças Crônicas Não Transmissíveis – DCNT já estava presente na criação da Rede Carmen, buscando “*desenvolver, implementar e avaliar: políticas voltadas à redução concomitante de um conjunto de fatores de riscos comuns às doenças crônicas não-transmissíveis; mobilização social e intervenções na comunidade; vigilância epidemiológica das condições de risco; e práticas de prevenção que possam contribuir na redução das desigualdades em saúde*” (OPAS, 2003). A Rede também preocupa-se, desde o início, em melhorar a articulação entre os países das Américas e, por conseguinte, o desenvolvimento de ações que promovam a melhoria do estado de saúde das populações, mediante a redução dos fatores de risco associados às DCNT.

Na avaliação de serviços e intervenções de saúde há um processo de expansão e diversificação conceitual e metodológica que se constitui no instrumento dos administradores para a tomada de decisão. A avaliação de uma intervenção deve permitir não somente o julgamento de sua eficácia, mas também dos fatores explicativos dos resultados obtidos, tendo em vista modificações posteriores (Hartz, 1997; Novaes, 2000).

- 12 Segundo Contandriopoulos *et al.* (1997, p. 40), “*a avaliação é uma atividade tão velha quanto o mundo, banal e inerente ao processo de aprendizado. Hoje também é um conceito em moda, com contornos vagos e que agrupa realidades múltiplas e diversas*”.

Logo após a Segunda Guerra Mundial, apareceu o conceito de avaliação dos programas públicos. No decorrer dos anos setenta a necessidade de avaliação das ações sanitárias se impôs, criando-se, em diversos países, organismos encarregados de avaliar as novas tecnologias. Essa proliferação é, certamente, o sinal de uma necessidade, mas também é sinal de complexidade da área (Contandriopoulos *et al.*, 1997; Denis; Champagne, 1997).

A avaliação é conceituada como “*Processo de determinação, sistemática e objetiva, da relevância, efetividade, eficiência e impacto de atividades fundamentadas em seus objetivos. É um processo organizacional para implementação de atividades e para colaborar no planejamento, programação e tomada de decisão*” (OMS, 2000).

As dificuldades do processo de avaliação são inúmeras, tanto conceituais quanto operacionais. Avaliar é atribuir valor, entretanto nem sempre estão estabelecidos os critérios sobre os quais se determina esse valor e tampouco os objetos que serão

valorizados. Destaca-se também que a avaliação não é imparcial, dotada de suposta neutralidade técnica, mas depende da visão de mundo daqueles que a realizam. A maneira como se estrutura o sistema de informação subordina-se a alguma concepção de modelo assistencial. A própria construção da análise destas informações igualmente é dotada de intencionalidade e direcionalidade. Da mesma forma não há um único caminho avaliativo, ou seja, não existe uma verdade única na esfera dos fenômenos sociais; às vezes, são selecionadas algumas dimensões para a análise em detrimento de outras. Os seus resultados são aproximações da realidade e podem ser operacionalizados por vários caminhos (Malta, 2001).

A avaliação é um dispositivo de produção de informação e, conseqüentemente, fonte de poder para os atores que a controlam, por isso o avaliador deve fazer o papel de agente facilitador e de pedagogo, na utilização dos resultados. A informação é uma ferramenta de negociação entre interesses múltiplos e não verdade absoluta (Malta, 2001).

Torna-se importante avaliar as intervenções, visando verificar seus acertos e limites. O resultado da avaliação é importante para os dirigentes, no sentido de validar a intervenção, provendo adequação de rumos, ampliação de seu escopo, ou mesmo descontinuidade no caso de resultados desfavoráveis.

A OMS (2000) justifica a importância da prática da avaliação por possibilitar as melhores escolhas por parte dos indivíduos e pela opção de políticas públicas mais efetivas para a coletividade. A avaliação é adequada *“para proteger as pessoas e ajudá-las a se protegerem. Os governos precisam ter capacidade para avaliar os riscos e escolher as intervenções com melhor relação custo efetividade, bem como financeiramente viáveis, para se evitar os riscos”* (OMS, 2000).

Autores como Habicht, Victora e Vaughan (1999) discutem a importância da avaliação no processo de influenciar decisões. *“O grau de complexidade e de precisão da avaliação tem relação intrínseca com o responsável pelo processo decisório e os tipos de decisões que serão tomadas em função dos resultados obtidos”*.

Victora *et al* (2004) questionam qual seria a melhor evidência no campo da saúde coletiva. Por evidência entende-se *“sólido conhecimento (ou o melhor disponível) sobre a frequência e distribuição das doenças”*, sobre seus determinantes e conseqüências, e sobre a segurança, eficácia e efetividade das intervenções e seus custos.

Habicht, Vaughan e Victora (1999) discutem os níveis de evidências necessários para se tomar decisão em Saúde Pública. Os autores explicam que os Ensaios Clínicos Randomizados (ECR) e as meta-análises são os estudos de escolha na clínica, mas nem sempre os ECR são factíveis em função de questões éticas e obstáculos legais. Os autores relativizam a premissa do padrão ouro dos ECR, mostrando que os mesmos são usualmente apropriados para resumir resultados de estudos clínicos, podendo ser pouco recomendável para a avaliação de programas complexos, pois o efeito de fatores contextuais locais é tão importante, que a premissa de homogeneidade de resultados entre diferentes estudos deixa de ser válida.

Ensaios clínicos randomizados podem ser desnecessários, não apropriados, impossíveis ou inadequados para avaliar a performance e impacto de intervenções em larga escala (saúde coletiva) (Victora *et al.* 2004; Black, 1996).

Santos e Victora (2005) debatem a necessidade de se avançar no desenho de estudos de avaliação adequados à realidade da Saúde Coletiva, ou Saúde Coletiva Baseada em Evidências, levando em consideração as redes de causalidade complexas nos programas aplicados à Saúde Pública e que respondam a questões éticas, quando se trata, por exemplo, de programas já implantados, com evidências de efetividade, mas que não tenham sido submetidos anteriormente ao rigor dos ECR.

Apoiado nesse referencial teórico é que se fundamenta esse guia, avançando na discussão da metodologia dos estudos epidemiológicos de avaliação, segundo Habicht *et al.*, (1999). São adotados indicadores de oferta, utilização, cobertura e impacto e discute-se ainda o grau de inferência necessário para responder à pergunta avaliativa (adequação, plausibilidade e probabilística).

O avanço da avaliação em Saúde Pública deve ser acompanhado do emprego de construções metodológicas diversas, visando apreender a realidade e as transformações oriundas de uma determinada forma de intervir em saúde. Nesse caso, torna-se necessário avançar tanto na utilização de métodos de coleta de dados quantitativos quanto qualitativos, possibilitando integrar enfoques metodológicos diversos. No caso desse Guia, optou-se pelo enfoque da metodologia quantitativa, em função do maior registro da literatura de experiências prévias em prevenção de DCNT e por se pretender focar o arsenal acumulado no campo da Epidemiologia. Entretanto este não é o único enfoque possível e devemos trabalhar no sentido de avançar em outros trabalhos com o método qualitativo. Importante lembrar, por exemplo, os avanços da Rede Carmen no Observatório de Políticas de Preven-

ção de DCNT, no estudo coordenado pela Agência de Saúde Pública do Canadá, em parceria com a OPAS, Brasil e Costa Rica, no qual a opção metodológica foi a da metodologia qualitativa, desenvolvendo estudos de caso nos três países (OPAS, 2002). Torna-se importante ressaltar que estes enfoques devem ser complementares e o ideal é trabalhar integrando metodologias diversas. A escolha do método se dará em função da pergunta avaliativa que se coloca, que sem dúvida é quem guiará a opção metodológica.

Torna-se um desafio avançar no sentido de institucionalizar a avaliação, visando integrá-la em um sistema organizacional no qual este seja capaz de influenciar o seu comportamento, ou seja, “*um modelo orientado para a ação ligando necessariamente as atividades analíticas às de gestão*” (Hartz, 2002).

## Referências Bibliográficas

BLACK, N. *Why we need observational studies to evaluate the effectiveness of health care*. BMJ. 1996, 312: 1215- 1218. 11 may.

CONTANDRIOPOULOS, A.P., CHAMPAGNE, F., DENIS, J.L., PINEAULT, R. *Avaliação na área de saúde: conceitos e métodos*. In: HARTZ, Z.M.A. (org.) Avaliação em Saúde. Dos Modelos conceituais à prática na Análise da implantação de programas. Rio de Janeiro: FIOCRUZ, 1997.

HABICHT J.P., VICTORA C.G., VAUGHAN J.P. *Evaluation designs for adequacy, plausibility and probability of public health programe performance and impact*. *Intern J. Epidemiology*. 1999, 28: 10-18.

HARTZ, Z.M.A *Institucionalizar e qualificar a avaliação: outros desafios para a atenção básica*. *Ciência e Saúde Coletiva*, 7(3): 419-421. 2002.

HARTZ, Z.M.A. Explorando novos caminhos na pesquisa avaliativa das ações de saúde. In: Hartz Z.M.A. et al. *Avaliação em Saúde. Dos Modelos conceituais à prática na Análise da implantação de programas*. Rio de Janeiro: FIOCRUZ, 1997. pág 19 – 28.

MALTA, D.C. *Buscando novas modelagens em saúde, as contribuições do Projeto Vida e Acolhimento para a mudança do processo de trabalho na rede pública de Belo*

*Horizonte, 1993-1996*. Tese Doutorado. Faculdade de Ciências Médicas, UNICAMP, Campinas, 2001.

NOVAES, H.M.D. *Avaliação de programas, serviços e tecnologias em saúde*. Rev Saúde Pública, 34(5):547-59, 2000.

OMS (Organização Mundial da Saúde). *Relatório Mundial da Saúde*. Genebra, 2000.

OPAS (Organização Panamericana de Saúde). *Observatorio de Políticas Para La Prevención De Enfermedades no Transmisibles en Latinoamerica y el Caribe* (2002). Division de Prevención y Control de Enfermedades. Programa de Enfermedades no Transmisibles. Washington, Diciembre.

OPAS (Organização Panamericana de Saúde) – Carmen – *Iniciativa para Conjunto de Ações para Redução Multifatorial de Enfermidades Não Transmissíveis* – outubro – 2003.

16 SANTOS, I.S.; VICTORA, C.G. *Serviços de Saúde: epidemiologia, pesquisa e avaliação*. Cad. Saúde Pública, v. 20, Sup. 2, pag 337-41, 2004.

VICTORA, C.G., HABICHT J.P., BRYCE, J. *Evidence-Based Public Health: Moving Beyond Randomized Trials*. *Americans Journals of Public Health*. March 2004, Vol. 94. Nº 3.

**Guia Metodológico de Avaliação e  
Definição de Indicadores –  
Doenças Crônicas Não Transmissíveis e  
Rede Carmen**

*Iná S. Santos*

O trabalho que se segue é uma proposta de avaliação e definição de indicadores da iniciativa Carmen para Doenças Crônicas Não Transmissíveis (DCNTs), a ser utilizada na América Latina, após revisão, críticas e aprovação dos demais países. Está estruturado em cinco partes. A Parte I contém a revisão da literatura sobre programas / intervenções de base comunitária para prevenção de DCNTs. A Parte II apresenta indicadores de avaliação e monitoramento. A Parte III discute os diferentes níveis de inferência possíveis de serem empregados nas avaliações de programas / intervenções. A Parte IV discute desenhos de estudos que podem ser usados para avaliação de desempenho e impacto de programas / intervenções. E a Parte V contém um guia de treinamento para avaliação. Com este trabalho, o Ministério da Saúde pretende induzir à prática da avaliação no Sistema único de Saúde (SUS) e contribuir com o debate dos países da América Latina no contexto da avaliação.

## A iniciativa Carmen

18

O objetivo da iniciativa Carmen é melhorar a saúde das populações nas Américas por meio da redução dos fatores de risco associados às DCNTs. Na América Latina e no Caribe, as DCNTs combinadas constituem a principal causa de morte antes dos 70 anos de idade. As DCNTs de maior importância em saúde pública na região são as cardiovasculares, os cânceres e o diabetes mellitus. Entre as doenças cardiovasculares, os acidentes vasculares cerebrais e a doença isquêmica do coração são as mais frequentes, em termos de mortalidade, e a hipertensão arterial sistêmica, em termos de prevalência. Dentre os cânceres, o de colo uterino e de mama, nas mulheres, e os de estômago, pulmão, colon e próstata, entre os homens, figuram entre os mais frequentes. O diabetes mellitus, por sua vez, contribui tanto para a mortalidade como para a incapacidade crônica.

Em 2000, a prevalência estimada de hipertensão arterial sistêmica na América Latina e Caribe variava de 14 a 40%, entre indivíduos de 35 a 64 anos, contabilizando cerca de 140 milhões de pessoas. Estimativas quanto ao diabetes mellitus indicavam quase 35 milhões de indivíduos acometidos. Ambas as enfermidades são fatores de risco para doenças cardiovasculares, além de poderem apresentar complicações, como necessidade de amputação de extremidades, insuficiência renal, cegueira, entre outras.

Os primeiros passos da Organização Pan-Americana de Saúde no sentido de apoiar programas de prevenção integrada das DCNTs ocorreram em 1996, tendo como

referência o Programa Cindi (*Countrywide Integrated Noncommunicable Diseases Intervention Program*) criado e desenvolvido pela Organização Mundial da Saúde (OMS) em países europeus, em 1985. Em setembro de 2002, a Conferência Pan-Americana de Saúde respaldou o enfoque de saúde pública para as DCNTs e aprovou a iniciativa Carmen como uma das principais estratégias de prevenção integrada das doenças crônicas a serem implementadas na região.

A iniciativa Carmen está comprometida com o desenvolvimento, implementação e avaliação de políticas que visem a redução simultânea de fatores de risco comuns a mais de uma das doenças crônicas; mobilização social e intervenções de base comunitária; vigilância epidemiológica dos fatores de risco; e atividades preventivas que possam contribuir com a redução das desigualdades em saúde.

Para a redução simultânea de fatores de risco comuns às DCNTs, a iniciativa Carmen utiliza três estratégias: prevenção integrada e promoção da saúde, efeito demonstrativo e promoção de equidade em saúde. A primeira estratégia compreende o desenvolvimento de intervenções que objetivem a prevenção e a redução, simultaneamente, de fatores de risco comuns às principais DCNTs. O tabagismo, a dieta inadequada e o sedentarismo são considerados fatores de risco estratégicos para essa ação preventiva.

O efeito demonstrativo compreende a estratégia de avaliar previamente, em áreas demonstrativas (áreas-piloto), a aceitabilidade, a segurança, a eficácia e a efetividade de intervenções a serem posteriormente introduzidas em maior escala. O efeito demonstrativo vai medir se houve modificação nos fatores de risco e na mortalidade por DCNTs correspondentes e se as mudanças observadas ocorreram dentro de uma seqüência temporal lógica em relação às intervenções. Para fins práticos, é necessário haver uma medida basal que permita comparações posteriores e um sistema de monitoramento dos fatores de risco e da mortalidade, bem como que coletas sistemáticas de informação sobre a intervenção sejam realizadas, a fim de avaliar seu desenvolvimento.

A estratégia de promoção de equidade em saúde visa identificar e atingir os grupos populacionais em maior desvantagem social, uma vez que as ações universais em saúde tendem a reforçar as desigualdades, beneficiando mais os menos necessitados (Gwatkin DR *et al*, 2004; Palmer N *et al*, 2004).

O estabelecimento de políticas de saúde, o desenvolvimento de ações de base comunitária e serviços de saúde receptivos à prevenção de doenças crônicas constituem as principais linhas de ação da iniciativa Carmen. Através de intervenções de base comunitária e estabelecimento de uma rede de apoio regional, a iniciativa Carmen pretende impactar positivamente a epidemia de DCNTs nas Américas.

O atual trabalho tem o objetivo de construir uma proposta didática e prática de um Guia Metodológico para avaliação de programas e intervenções destinados à prevenção e controle de DCNTs, a ser utilizado por gestores de saúde e de projetos dentro da iniciativa Carmen e contribuir para a capacitação técnica regional de avaliação em saúde Sública.

Especificamente, esse trabalho pretende: revisar a literatura disponível para descrever o desempenho e resultados de impacto de programas / intervenções preventivas de base comunitária para DCNTs; orientar na formulação de indicadores de monitoramento e avaliação para programas / intervenções desenvolvidos pela iniciativa Carmen; e subsidiar a construção de um Guia Metodológico de monitoramento e avaliação da iniciativa Carmen, em parceria com os demais países da América Latina e do Caribe.

# PARTE I

## Revisão da Literatura

*Iná S. Santos, Marcelo Capilheira  
e Carmen Moreira*

Desde a década de 70, vários projetos de intervenção populacional foram desenvolvidos com o objetivo de promover mudanças comportamentais, que reduzissem fatores de risco para doenças crônicas, inicialmente, as cardiovasculares. Tais projetos enfatizavam que intervenções realizadas apenas em indivíduos de alto risco, selecionados nos serviços de saúde, teriam impacto muito discreto no país como um todo. Por outro lado, se as intervenções fossem direcionadas para toda a população, pequenas mudanças em fatores de risco e hábitos de vida saudáveis teriam, potencialmente, grande impacto na saúde pública. Há vasta quantidade de publicações a cerca do tema. Entretanto, conforme apontado em revisões de literatura, apenas alguns estudos conseguiram apresentar conclusões válidas sobre o impacto dessas intervenções na saúde da população como um todo.

O primeiro grande projeto de intervenção populacional surgiu na Finlândia, com o *North Karelia Project*, em 1972. Posteriormente, foram implementados o *Stanford Three-Community Study* (1972-1974), o *Stanford Five-City Project* (1978-1980), o *Minnesota Heart Health Program* (1980-1993) e o *Pawtucket Heart Health Program* (1980-1991), todos nos EUA. Estão em andamento o *Countrywide Integrated Non-communicable Diseases Intervention – Cindi* –, na Europa e Canadá; e o *Conjunto de Acciones para la Reducción Multifactorial de Enfermedades No transmissibles – Carmen*, na América Latina (Nissinen A *et al*, 2001; PAHO, 2003).

22

## Metodologia da Revisão Bibliográfica

A revisão bibliográfica foi realizada através de pesquisa às bases de dados *Pubmed*, *Web of Science* e *Lilacs*. Considerando o grande número de intervenções para prevenção de doenças cardiovasculares, com as mais diversas abordagens e populações, optou-se por realizar a busca bibliográfica baseada em grandes projetos populacionais com alcance nacional e internacional (PAHO. *Carmen – An Initiative for Integrated Prevention of Noncommunicable Diseases in the Americas*, 2003). Selecionaram-se dois projetos pioneiros: o *North Karelia Project* e o *Stanford Three-Community Study*, seguidos pelos *Stanford Five-City Project*, *Minnesota Heart Health Program*, *Pawtucket Heart Health Program*, *Cindi* e *Carmen*.

Os descritores utilizados para a busca foram os nomes dos projetos supracitados nos títulos e resumos de publicações. A busca foi limitada a estudos em humanos e na área da saúde. Tanto na *Pubmed* quanto na base de dados *Web of Science*, o descritor “Carmen” identificou várias publicações de um estudo homônimo (*Carmen*

– *The Carbohydrate Ratio Management in European National diets*), as quais foram excluídas.

Foram encontrados 836 artigos e selecionados apenas os que apresentavam resultados de impacto populacional das intervenções propostas ou informações conceituais dos estudos em questão. Nesse processo, selecionaram-se 295 publicações, dentre as quais utilizaram-se 32 que descreviam os principais resultados de impacto populacional dos diversos estudos pesquisados.

Os artigos inicialmente selecionados e não utilizados foram excluídos por apresentarem resultados de grupos selecionados, discussões metodológicas já contidas em outras publicações, apenas aspectos financeiros, monitoramento com informações contempladas em publicações mais recentes ou ainda por apresentarem resultados já divulgados por outras fontes (livros, *guidelines*, etc). Ao final, foram também excluídos artigos em duplicata nas diferentes bases de dados. O Quadro 1 descreve o número de publicações localizadas e utilizadas nessa revisão.

Quadro 1. Resumo da revisão bibliográfica (2006)

Fonte	Descritores	Encontrados	Excluídos	Selecionados	Utilizados
Pubmed	Stanford Three-Community Study	8	1	7	3
	Stanford Five-City Program	57	14	43	11
	North Karelia Project	126	65	61	5
	Minnesota Heart-Health Program	94	50	44	8
	Pawtucket Heart Health Program	51	29	22	3
	Cindi	39	35	4	1
	Carmen	35	33	2	1
Web of Science	Stanford Three-Community Study	1	-	1	-
	Stanford Five-City Project	41	20	21	-
	North Karelia Project	148	95	53	-
	Minnesota Heart-Health Program	49	30	19	-
	Pawtucket Heart Health Program	-	-	-	-
	Cindi	20	18	2	-
	Carmen	148	146	2	-
Lilacs	Stanford Three-Community Study	-	-	-	-
	Stanford Five-City Project	-	-	-	-
	North Karelia Project	1	1	-	-
	Minnesota Heart-Health Program	-	-	-	-
	Pawtucket Heart Health Program	-	-	-	-
	Cindi	3	-	3	-
	Carmen	15	5	10	-
TOTAL		836	541	295	32

**Programas de Prevenção  
de Doenças Crônicas Não Transmissíveis  
em Países Desenvolvidos**

## North Karelia Project

O *Projeto North Karelia* teve início em 1972 como um projeto para prevenir a mortalidade por doenças cardiovasculares entre os moradores daquela província, no oeste da Finlândia. A Associação Finlandesa de Cardiologia coordenou inicialmente as discussões, reunindo especialistas, representantes da comunidade e representantes da Organização Mundial da Saúde (OMS). Mais tarde, o programa expandiu-se para contemplar outras doenças não transmissíveis. Tal projeto demonstrou que as DCNTs não são completamente evitáveis, mas que as intervenções podem reduzi-las drasticamente. A intervenção abrangeu ações para redução do tabagismo (campanhas na mídia) e do colesterol sérico (competições entre províncias) e projetos escolares, além de envolver várias instituições e estimular profissionais de saúde a enfatizar a importância desses temas aos pacientes.

Estudos de avaliação foram conduzidos a cada cinco anos, no período de 1972 a 1992, mostrando que, ao longo do tempo, o projeto foi bem sucedido. Como exemplo, as taxas de mortalidade por doença cardiovascular, em homens de 35 a 64 anos de idade, decresceram 57% no período de 1972 a 1992. O projeto contribuiu também para adoção de mudanças em políticas de saúde, agricultura e indústria, no intuito de estimular hábitos de vida e alimentares saudáveis. Em 1972, cerca de 90% das pessoas utilizavam na alimentação manteiga com pão e, em 1992, apenas 15% o faziam. No mesmo intervalo de tempo, o consumo anual de frutas e vegetais por pessoa aumentou de 20kg para 50kg.

Alguns resultados:

- 1) Abordagem preventiva entre estudantes de 12 – 13 anos de idade, por dois anos. Foram incluídos dois pares de escolas-intervenção: (1) uma em que as orientações aos estudantes eram dadas por profissionais da saúde; (2) outra em que os professores foram treinados para prevenção. Um par de escolas-controle, não recebeu intervenção.

As escolas-intervenção apresentaram menor prevalência de tabagistas em relação às escolas-controle, em todos os anos, exceto no último (Tabelas 1 e 2).

O número acumulado de carteiras de cigarro fumadas foi maior no grupo controle em relação ao grupo intervenção. Entre os homens, o grupo controle consumiu, até os 28 anos de idade, cerca de 1200 carteiras, enquanto que, no grupo intervenção,

900. Entre as mulheres, o grupo controle consumiu até os 28 anos de idade pouco menos de 500 cartelas e o intervenção, cerca de 350.

Tabela 1 – Prevalência de tabagistas nas escolas-intervenção e controle

Idade	Intervenção com profissionais de saúde (%)	Intervenção com educadores da própria escola (%)	Escola controle (%)	p*
13 anos	15,0	13,2	8,4	0,047
15 anos	25,3	24,5	34,8	0,11
16 anos	30,0	30,0	40,7	0,14
17 anos	33,6	30,7	44,4	0,003
21 anos	42,2	35,2	50,0	0,008
28 anos	34,6	34,3	42,8	0,115

\*Qui-Quadrado entre os três grupos

27

Tabela 2 – Prevalência de tabagismo nos acompanhamentos do estudo entre a população não-tabagista no início do período

Idade	Intervenção com profissionais de saúde (%)	Intervenção com educadores da própria escola (%)	Escola controle (%)	p*
15 anos	18,5	18,6	29,2	0,004
16 anos	23,5	22,2	37,7	<0,001
17 anos	28,3	25,8	41,2	0,001
21 anos	37,5	30,1	46,9	0,003
28 anos	30,8	29,3	41,2	0,026

\*Qui-Quadrado entre os três grupos

- 2) Intervenção para redução do tabagismo em North Karelia (área-modelo do projeto Cindi) iniciada em 1994, denominada “*Quit and Win*”. Acompanhamento de 2.114 indivíduos que pararam de fumar, com visitas no 1º, 6º e 12º mês de abstinência.

Tabela 3 – Prevalência de abstinência de tabagismo

Tempo de abstinência	Gênero	North Karelia (%)	Restante da Finlândia (%)	p
1 mês	Masculino	74,9	74,5	0,931
	Feminino	75,2	64,4	0,017
	Total	75,0	69,7	0,046
6 meses	Masculino	40,3	32,6	0,039
	Feminino	28,1	23,5	0,273
	Total	35,9	28,3	0,005
12 meses	Masculino	33,2	24,0	0,007
	Feminino	17,3	17,9	0,903
	Total	27,5	21,1	0,010

- 3) As Tabelas 4, 5 e 6 mostram resultados do projeto *North Karelia* sobre a mortalidade por doenças cardiovasculares na Finlândia. Na Tabela 4, as comparações são feitas entre os períodos de 1969 – 1971 e 1993 – 1995.

Tabela 4 – Taxas de mortalidade anual ajustada por idade nos períodos de 1969-1971 e 1993-1995 na Finlândia e em North Karelia com as respectivas diferenças percentuais de mudança na população masculina na faixa etária de 35 a 64 anos

Mortalidade	Taxa de mortalidade x 10 <sup>5</sup> (IC95%)	
	Finlândia	North Karelia
Todas as causas		
1969-71	1272 (1257-1287)	1509 (1431-1587)
1993-95	700 (690-710)	833 (779-887)
Diferença %	- 45,0	- 44,8
Doença cardiovascular		
1969-71	647 (637-657)	855 (796-914)
1993-95	261 (255-267)	338 (303-373)
Diferença %	- 59,7	- 60,5
Doença coronariana		
1969-71	465 (456-474)	672 (620-724)
1993-95	175 (170-180)	237 (208-266)
Diferença %	- 62,4	- 64,7
Doença cerebrovascular		
1969-71	93 (89-97)	90 (71-109)
1993-95	40 (38-42)	47 (34-60)
Diferença %	- 57,0	- 47,8

Tabela5 – Taxas de mortalidade por doenças coronarianas (DC) ajustadas por idade em North Karelia e toda Finlândia entre homens na faixa etária de 35 – 64 anos, entre 1969 a 1995

Local	Declínio da mortalidade por DC por ano (IC95%)		
	1969 – 1978	1979 – 1985	1986 – 1995
North Karelia	- 2,9 (-3,9; -1,8)	- 0,2 (-1,4; +0,9)	- 8,0 (- 9,8; - 6,1)
Finlândia	- 1,0 (-1,6; -0,4)	- 3,5 (-4,4; -2,6)	- 6,5 (- 7,1; - 5,9)

Tabela 6 – Taxas de mortalidade anual ajustada por idade nos períodos de 1969-1971 e 1993-1995 na Finlândia e em North Karelia com as respectivas diferenças percentuais de mudança na população feminina na faixa etária de 35 a 64 anos

Mortalidade	Taxa de mortalidade x 10 <sup>5</sup> (IC95%)	
	Finlândia	North Karelia
Todas as causas		
1969-71	475 (467-483)	501 (459-543)
1993-95	280 (274-286)	311 (277-345)
Diferença %	-41,1	-37,9
Doença cardiovascular		
1969-71	205 (200-210)	262 (232-292)
1993-95	65 (62-68)	93 (75-111)
Diferença %	-68,3	-64,5
Doença coronariana		
1969-71	82 (79-85)	118 (98-138)
1993-95	28 (26-30)	37 (26-48)
Diferença %	-65,9	-68,6
Doença cerebrovascular		
1969-71	69 (66-72)	65 (50-80)
1993-95	23 (21-25)	32 (21-43)
Diferença %	-66,7	-50,8

## The Stanford Community Projects

Os projetos *Three Community Study* (1972-1975) e *Five-City Multifactor Risk Reduction Project* (1978-1996) foram conduzidos pela Universidade de Stanford com objetivo de modificar fatores de risco para doenças cardiovasculares. Ambos os estudos envolveram amplas campanhas de educação comunitária sobre os fatores de risco, incluindo hipertensão arterial, altos níveis de colesterol sérico, tabagismo e excesso de peso. Um dos objetivos era desenvolver e avaliar métodos para atingir mudanças comportamentais que tivessem bom custo-efetividade, aplicáveis a grandes populações.

O projeto *Three Community Study* comparou uma comunidade controle com duas que receberam as intervenções na forma de campanhas, através dos meios de comunicação de massa. Tais campanhas, com duração de dois anos, ensinavam os prováveis fatores de risco para doenças cardiovasculares - e medidas para reduzi-los e adquirir hábitos de vida saudáveis. Em uma das comunidades, além disso eram dados conselhos individuais a pessoas com alto risco de doença cardiovascular.

O projeto *Five-City Multifactor Risk Reduction Project* utilizou a mesma metodologia, adaptada para um maior número de comunidades, e a intervenção tinha a duração de cinco anos. A hipótese era de que uma redução de 20% nos fatores de risco levaria a um declínio significativo nas doenças cardiovasculares nas cidades intervenção. Além disso, o programa se propunha a manter um sistema de vigilância de mortalidade por doença cardiovascular.

31

No *Three Community Study*, os indivíduos foram entrevistados antes, durante e depois da intervenção, com o objetivo de avaliar o conhecimento e os riscos para doenças cardiovasculares. As intervenções produziram mudanças no conhecimento e na prevalência dos fatores de risco no primeiro ano da intervenção, as quais se mantiveram até o segundo ano. A redução estimada do risco de doenças coronarianas e cerebrovasculares foi de 24%.

Os resultados do *Five-City Multifactor Risk Reduction Project* foram avaliados por vários estudos pontuais ao longo de 73 meses. Verificou-se que o conhecimento sobre fatores de risco aumentou em todas as comunidades, inclusive as controles, porém com maior expressão naquelas que receberam a intervenção. Com a intervenção, foram observadas reduções nos níveis de pressão arterial e colesterol

sérico, além de diminuição do tabagismo, representando um declínio geral de 16% no risco de desenvolver doenças cardiovasculares.

As mudanças geradas pelas intervenções mantiveram-se ao longo do tempo, sendo que alguns de seus componentes foram adotados e mantidos pela própria comunidade.

### Alguns resultados:

No *Three Community Study* participaram indivíduos entre 35 e 59 anos de idade, sendo submetidos ao estudo de base em 1972 e acompanhados em subseqüentes inquéritos até 1975. A Tabela 7 apresenta resultados observados entre 1972 e 1975.

Tabela 7 – Mudança percentual conforme as variáveis relacionadas à dieta nas cidades intervenção e controle (1972-1975)

Variável	Somatório percentual anual de mudança em média geométrica			p*
	Comunidades			
	Watsonville (intervenção)	Gilroy (intervenção)	Tracy (controle)	
Colesterol na dieta (mg/dia)	-11,2	-11,0	-2,8	0,01
Gordura saturada na dieta (g/dia)	-9,4	-10,3	-0,9	0,03
Colesterol plasmático (mg/100ml)	1,0	0,5	1,7	0,02
Peso relativo (peso atual/peso ideal)	0,0	0,0	0,3	0,04

\*Comparação da média de Watsonville e Gilroy vs Tracy (análise uni-caudal)

### The Minnesota Heart Health Program (MHHP)

O *Minnesota Heart Health Program* (1980-1993) foi planejado para testar o efeito da educação comunitária em saúde para prevenção de fatores de risco para doenças cardiovasculares. O programa foi um dos maiores ensaios comunitários realizados nos EUA. Os principais objetivos eram conscientizar a população quanto às doenças cardiovasculares, incentivar a participação de programas de saúde e estimular a adoção de comportamentos saudáveis, que reduzissem os fatores de risco para doenças cardiovasculares. Os principais fatores abordados foram tabagismo, níveis

altos de colesterol sérico, hipertensão arterial e inatividade física, sendo que o objetivo geral era adotar e manter as medidas de intervenção do programa.

O *MHHP* foi elaborado considerando a hipótese de que um programa educacional sistemático e com múltiplas estratégias é viável e muda a forma de pensar da população, em relação às doenças cardiovasculares e sua prevenção. Melhora, conseqüentemente, seus hábitos de vida, reduzindo os fatores de risco e, posteriormente, as taxas de doenças cardiovasculares. O programa consistiu em intervenções educacionais comunitárias em Dakota do Norte, Dakota do Sul e Minnesota. As principais intervenções foram: (1) rastreamento de fatores de risco para doenças cardiovasculares; (2) informações nutricionais em restaurantes; (3) campanhas educacionais anuais sobre prevenção de doenças cardiovasculares, desenvolvidas pela comunidade e sociedade civil organizada; (4) educação continuada para profissionais da saúde; (5) educação em veículos de comunicação de massa, como TV, jornais e rádios; (6) educação de adultos em locais de trabalho, igrejas e outras organizações; e (7) educação de jovens.

A avaliação do programa foi realizada para determinar o impacto e a sustentabilidade das atividades. As três principais medidas utilizadas para avaliar o impacto foram morbidade e mortalidade, fatores de risco e comportamentais associados e efeito de programas educacionais específicos. As informações eram obtidas através de registros oficiais e hospitalares, além de estudos populacionais anuais. Os resultados mostraram uma boa participação e conscientização da população, com diminuição dos fatores e comportamentos de risco e declínio das doenças cardíacas. Entretanto, não houve diferença significativa entre as áreas de intervenção e as de controle. Além disso, não houve redução nas taxas de morbidade ou mortalidade por doença cardiovasculares ou cerebrovascular. A maioria dos programas apresentou efeitos modestos e de curta duração.

O impacto do programa parece ter sido reduzido em função da forte tendência secular de aumento dos esforços para promoção da saúde durante o programa tanto nas áreas intervenção quanto nas controle. Os autores do *MHHP* sugerem que se deva considerar a avaliação de tendências seculares nas medidas de impacto e apontam a necessidade de outras estratégias que permitam determinar o verdadeiro papel das intervenções.

### Alguns resultados:

- 1) Coortes de estudantes adolescentes acompanhada por cinco anos (desde a 6ª série até a 12ª), com grupos-intervenção e controle
  - Entre os controles, 8,2% dos alunos da 6ª série tornaram-se fumantes semanais ou permaneceram nessa condição ao final de um ano. No grupo intervenção, a incidência correspondente foi de 2,5% ( $p < 0,001$ )

Análises das mesmas coortes com enfoque na intensidade do tabagismo mostraram que:

  - Ao final do acompanhamento, 14,6% da população era tabagista no grupo intervenção e 24,1% no grupo controle.
  - No grupo controle, os tabagistas consumiam cerca de 20 cigarros por semana e no grupo intervenção em torno de 10.
- 2) Comparação de três comunidades-intervenção com outras três comunidades-controle quanto a morbidade e mortalidade por doença cardiovasculares não demonstrou resultados estatisticamente significativos. As avaliações foram feitas até sete anos após a intervenção.
  - Entre os homens, houve uma tendência de redução do tabagismo, mas sem efeito adicional da intervenção. Entre as mulheres, a tendência secular foi pequena, com efeito significativo da intervenção na avaliação transversal. Na avaliação de coorte, observou-se um declínio na prevalência, porém com pouca evidência de efeito da intervenção.
  - Os níveis de colesterol, Índice de Massa Corporal (IMC) e pressão arterial mantiveram-se estáveis, sem evidência de mudança após a intervenção.
  - Quanto à atividade física, aparentemente, o programa contribuiu com a tendência secular de aumento da atividade física.

34

### The Pawtucket Heart Health Program (PHHP)

O Pawtucket *Heart Health Program* (1980 – 1991) consistiu em uma intervenção de base comunitária com foco na modificação de fatores de risco para doenças cardiovasculares entre os moradores de Pawtucket. Os principais objetivos eram reduzir a morbidade e a mortalidade cardiovasculares, utilizando estratégias que contassem com voluntários da própria comunidade, e desenvolver uma estrutura de organização comunitária capaz de sustentar programas de atenção à saúde. Houve envolvimento de toda a comunidade, incluindo grupos e organizações, no esforço de educar a população sobre fatores de risco, elaborar sistemas de suporte e

estratégias para iniciar e manter hábitos saudáveis de vida. Os fatores de risco alvos do projeto incluíam níveis altos de colesterol sérico, hipertensão arterial, tabagismo, sedentarismo e obesidade.

Formou-se uma unidade de intervenção para selecionar e treinar voluntários para desenvolver, implementar, avaliar e, eventualmente, coordenar um programa de mudanças de comportamentos de risco para doenças cardiovasculares. A unidade de intervenção era complementada por outra, de avaliação, que conduzia estudos domiciliares randomizados, desenhava e implementava sistemas de avaliação e monitorava mudanças relevantes em Pawtucket e cidades-controle. Uma estratégia de *marketing* foi desenvolvida para promover os programas de intervenção, além de um programa de educação popular, com material educativo próprio para pessoas com baixa escolaridade. Cada programa tinha como foco apenas um fator de risco: (1) *Lighten Up*, para controle de peso; (2) *Know Your Cholesterol*, que incluía rastreamento, aconselhamento e consultas para redução dos níveis de colesterol; e (3) *Quit and Win*, para eliminação do tabagismo. Foi desenvolvido ainda um software para registro dos participantes e manutenção de informações atualizadas sobre as intervenções.

35

Os resultados obtidos mostraram diferença significativa entre Pawtucket e a cidade-controle apenas quanto a uma atenuação no aumento do IMC ao longo da década. Os autores do programa enfatizam a importância da integração da comunidade, gestores regionais, estaduais e nacionais na implantação e, principalmente, na sustentação das ações educacionais de prevenção de doenças e fatores de risco para doenças cardiovasculares.

### **Alguns resultados:**

- 1) Comparação da cidade-intervenção (Pawtucket) com 71.204 habitantes e cidade-controle, com 98.478 habitantes. Foram realizados estudos transversais em cortes de indivíduos com idades entre 18 e 64 anos no início do estudo (seis inquéritos), durante e depois da intervenção educacional (Tabela 8).

Tabela 8 – Média do Índice de Massa Corporal (IMC) em Pawtucket (cidade-intervenção) e na cidade-controle, nos diferentes estágios do estudo

Fator de risco	Cidades e diferença entre estas	Estágios do estudo				
		Baseline (1981-84)	Intervenção (1987-91)	Pós-intervenção (1992-93)	Baseline x Intervenção	Baseline x Pós-intervenção
IMC (kg/m <sup>2</sup> )	Pawtucket	26,1	26,2	26,3	0,14	0,25
	Controle	26,4	26,7	27,3	0,25	0,88
	Diferença	-0,35	-0,46	-0,97	-0,11	-0,62
	p	0,043	0,006	<0,001	0,645	0,042

Os fatores de risco colesterol total, pressão arterial, tabagismo e a estimativa de morbidade por doenças cardiovasculares não sofreram mudanças significativas com a intervenção.

### Countrywide Integrated Noncommunicable Diseases Intervention (Cindi)

36

Entre 1978 e 1981, a OMS empreendeu iniciativa pioneira para desenvolver uma abordagem integrada de prevenção e controle de doenças crônicas não transmissíveis (DCNT). No ano de 1981, em um encontro convocado pela OMS e Escritório Regional da Europa, foi elaborado o conceito da abordagem integrada para prevenção e controle das DCNTs, bem como os princípios de colaboração internacional para implementação da proposta. Assim, em 1983, o programa *Countrywide Integrated Noncommunicable Diseases Intervention* (Cindi) foi estabelecido com o objetivo de dar suporte aos países-membros no desenvolvimento de políticas amplas e integradas para prevenção e redução das conseqüências das DCNTs. Para atingir tal objetivo, foram definidas como estratégias:

- combinar a promoção da saúde e a prevenção de doenças, desenvolvendo colaboração intersetorial e envolvimento da comunidade;
- destacar o papel dos profissionais de saúde; e
- otimizar a utilização de recursos existentes.

O desenvolvimento do projeto Cindi compreendeu quatro fases: (1) desenvolvimento do conceito, protocolos e desenho de um sistema de avaliação, entre 1981

e 1985; (2) desenvolvimento de programas nacionais e a criação de uma rede de trabalho internacional, entre 1982 e 1988; (3) implementação de programas nacionais para demonstração, entre 1985 e 1992; e (4) revisão e desenvolvimento da estrutura política do Cindi para atingir as metas estratégicas de saúde na Europa, entre 1990 e 1992.

A partir de 1990, em encontros com representantes dos países participantes, foi decidida a expansão do programa a toda Europa e Canadá, propondo assistir aos países no desenvolvimento de prevenção e controle das DCNT, auxiliar na definição de suas próprias estratégias e identificar áreas para ação em conjunto.

O programa Cindi tem cinco objetivos principais:

1. Realizar uma abordagem coordenada de prevenção ou redução das DCNTs. Para tal, é necessário o estabelecimento de uma estrutura organizacional adequada, para implementar o programa, e dispor de protocolos nacionais, para o avanço no desenvolvimento.
2. Desenvolver uma ampla abordagem de educação populacional, estrutura de serviços e outras estratégias, para prevenir e controlar as DCNTs. Devem ser identificados os principais grupos-alvo e os canais de intervenção, bem como estabelecer protocolos para alcance dos objetivos.
3. Desenvolver uma rede de informações para a implementação e monitorização dos programas. Tal item requer a elaboração e o uso de indicadores de resultado e processo, em relação aos vários fatores de risco, e a identificação dos tipos de DCNTs.
4. Os resultados dos programas devem ser avaliados. Para tal, são necessárias medidas e acompanhamento de indicadores adequados na população atingida, no intuito de avaliar o processo e a efetividade do programa, bem como transmitir tais resultados à população para reforçar a intervenção.
5. Promover a pesquisa em prevenção e controle das DCNTs e estabelecer métodos, meios e atividades para esse propósito.

O principal objetivo de cada país participante é planejar e conduzir o programa Cindi de forma que integre e coordene várias estratégias e atividades. A maioria dos países trabalhou em uma área-piloto para teste e treinamento, simultaneamente ao programa nacional. Todas as organizações importantes e estruturas administrativas devem estar envolvidas em uma única, intersetorial para implementação do programa, sendo que a coordenação central deve ser responsabilidade do ministério da saúde. No

Cindi, é de vital importância a existência de uma equipe executiva para elaboração de políticas, objetivos e metas do programa. Tal equipe pode estar alocada ou no ministério da saúde, em um centro universitário, algum instituto nacional ou outro local de natureza semelhante. Cada país deve elaborar um protocolo detalhado e um plano de ação para implementação nacional do programa, em parceria com a coordenação central do Cindi (internacional) e Escritório Regional da OMS. Tal processo deve ser realizado no início da implementação do programa e o plano reavaliado periodicamente. Em resumo, os estágios a ser seguidos compreendem:

- Análise de situação
- Estabelecimento de uma direção nacional do projeto
- Finalização do protocolo nacional e plano de ação do programa
- Desenvolvimento de diretrizes e métodos para intervenções em fatores de risco comuns
- Estudo de linha de base
- Início da intervenção
- Encontros de diretores nacionais do programa e participação destes nos encontros internacionais do Cindi
- Avaliações principais, em intervalos de cinco anos
- Avanço no desenvolvimento do programa Cindi nacional.

38

Os indicadores a serem monitorados referem-se tanto ao processo quanto ao resultado do programa, agrupados em dois níveis: (1) essenciais, ou mandatórios; e (2) recomendados, ou de considerável importância para o programa, sendo que os países que não os monitorar devem justificar as razões. Tais indicadores devem ser obtidos a partir de estatísticas oficiais e outras publicações, além de estudos populacionais.

### Indicadores Essenciais

- Estrutura etária e sexo da população
- Dados de mortalidade
- Dados sobre consumo alimentar
- Dados de estudo populacional quanto aos resultados (área-piloto)
- Dados de estudo populacional quanto ao processo (área-piloto)

Os estudos populacionais devem contemplar os seguintes indicadores:

- a) Fatores biológicos
  - Colesterol total
  - Pressão Arterial
  - Peso corporal
  - Altura
- b) Fatores comportamentais
  - Tabagismo
  - Consumo de álcool
  - Atividade física
- c) Escolaridade
  - Anos completos de estudo

### Indicadores Recomendados

Os indicadores recomendados referem-se a uma área-piloto e consiste em dados de morbidade e outras informações de estudos populacionais:

39

- Dados de Morbidade
  - a) Doença cardiovascular
    - Doença isquêmica do coração
    - Acidente vascular cerebral
  - b) Câncer
    - Sistema gastrointestinal
    - Estômago
    - Cólon e reto
    - Pulmão e laringe
    - Colo uterino
    - Mama
    - Próstata
  - c) Diabete Mellitus
    - Insulino-dependente
    - Não insulino-dependente
  - d) Doenças respiratórias crônicas

- e) Acidentes
  - Domésticos
  - Trânsito
  - Ocupacional
- Outras informações de estudos populacionais  
Fatores biológicos recomendados para inclusão:
  - HDL colesterol
  - Glicemia
  - Gama glutamil transferase
  - Medida laboratorial para validação de tabagismo auto-relatado

### **Multifactorial Monitoring of Trends and Determinants in Cardiovascular Disease (Monica)**

Embora as doenças cardiovasculares sejam responsáveis por cerca de metade das mortes em países desenvolvidos, há uma grande variabilidade de taxas de um país para outro, sem evidências suficientes para justificá-las.

40

Um dos objetivos do Projeto Monica é complementar antigos estudos transversais com uma investigação longitudinal da dinâmica de mudança das doenças. O projeto iniciaria o monitoramento simultâneo de mortalidade cardiovascular, morbidade, níveis de fatores de risco e tendências sociais e comportamentais, dentro de uma comunidade definida por um período de alguns anos.

Outro objetivo é medir as tendências de mortalidade cardiovascular, prevalência de doença cardíaca coronariana e cerebrovascular, além de avaliar a extensão com que estas tendências se relacionam com mudanças de fatores de risco conhecidos, hábitos de vida, cuidados em saúde e fatores socioeconômicos. O projeto trabalha com a hipótese de que as mudanças nas taxas de mortalidade por doenças cardiovasculares podem ser devidas a mudanças na incidência das doenças, na letalidade ou em ambas.

A estrutura geral do projeto utilizaria quatro fontes básicas de informação durante o período de dez anos:

- Dados administrativos rotineiramente disponíveis para o estudo populacional, do gestor local e fontes médicas locais. Alguns desses dados poderiam estar disponíveis anualmente e outros a cada década (censo populacional).
- Investigação de eventos cardiovasculares reconhecidos pela medicina como tal, fatais e não fatais, utilizando fontes médicas e médico-legais, validando o diagnóstico original de acordo com os critérios do Monica.
- Monitoramento contínuo ou intermitente dos cuidados imediatos dos eventos coronarianos e cerebrovasculares.
- Estudos populacionais para monitorar níveis de fatores de risco e comportamentais relacionados à saúde.

Considerando a população definida, cada núcleo do projeto deveria obter:

- Censo definindo o tamanho da população, distribuição por sexo e idade padronizada em grupos de cinco anos de idade.
- Estimativas anuais da população do meio do ano, considerando nascimentos, mortes, imigração e emigração ano a ano.
- Número de mortes estratificado por sexo e idade (grupos etários de cinco anos) na faixa de 25 a 64 ou 74 anos, para cada ano do estudo. As mortes deveriam ser classificadas conforme o Código Internacional de Doenças (CID).
- Mortes por doenças cardíacas coronarianas (incluindo morte súbita) na faixa etária de 25 a 64 anos.
- Mortes por acidente vascular cerebral na faixa etária de 25 a 64 anos.
- Outras doenças cardiovasculares (apenas diagnósticos de certeza) na faixa etária de 25 a 64 anos, revisados e validados, considerados potencialmente decorrentes de doença coronariana ou cerebrovascular.
- Infarto do miocárdio não-fatal na faixa etária de 25 a 64 anos
- Acidente vascular cerebral não-fatal na faixa etária de 25 a 64 anos.
- Estudos populacionais em pelo menos dois momentos no período de dez anos (início e final), com amostras diferentes. Esses estudos identificariam tabagismo, pressão arterial, colesterolemia, peso e altura.
- Dados adicionais do tratamento agudo dos eventos coronarianos e cerebrovasculares (quando relevante) em duas ou mais ocasiões durante o período do estudo ou monitoramento contínuo.
- Dados anuais dos serviços de saúde.

**Programas de Prevenção de Doenças  
Crônicas Não Transmissíveis em Países  
em Desenvolvimento**

## Who's Interhealth Programme

Embora a maioria dos programas de combate à epidemia de DCNTs tenham sido conduzidas em países desenvolvidos, o grande aumento dessas enfermidades em países em desenvolvimento levaram à adoção de programas de prevenção e controle. O eixo principal desse programa, desenvolvido em 1986 para ser implementado em países em desenvolvimento, consiste em intervenções com o objetivo de modificar os níveis dos principais fatores de risco para DCNTs na comunidade. Estas intervenções foram planejadas para construir capacitações regionais e trocar experiências sociais e tecnológicas na abordagem de controle e prevenção das DCNTs.

Utilizando a metodologia do Monica, compararam-se cinco áreas em desenvolvimento – Chile, Mauritània, Tanzània, Beijing e Tianjin (essas duas últimas na China), com cinco desenvolvidas – Finlândia, Lituània, Malta, Rússia e EUA. Como esperado, a prevalência dos fatores de risco foi maior nos países desenvolvidos. Entretanto, verificou-se a emergência do problema nos países em desenvolvimento. A situação encontrada nesses países foi semelhante a dos países desenvolvidos no passado.

43

## Programa Mirame!

Desde a década de 80, observa-se uma elevada prevalência de fatores de risco para uma vida saudável na população chilena. Com o objetivo de contribuir com informações sobre o tema, o Departamento de Saúde Pública da Universidade Católica do Chile iniciou, em 1986, uma linha de investigação chamada “Iniciativa Chilena para o Estudo das Enfermidades Crônicas do Adulto, seus Fatores de Risco e sua Prevenção”. Primeiramente, concentraram-se esforços em intervenções no nível primário de atenção à saúde, conforme recomendação da OMS, porém sem resultados satisfatórios. Assim, mudou-se o foco das ações para o nível educacional, resultando na elaboração de protocolos implementados em escolas de comunidades selecionadas conforme estudos anteriores de base. Entre as atividades desenvolvidas, organizou-se, em 1992, o programa Mirame!, com o propósito de promover hábitos de vida saudáveis na população escolar e grupos familiares. A hipótese adotada pelo programa foi a de que uma intervenção educacional na população escolar, baseada em conceitos sociais de educação, levaria à adoção de estilos de vida saudáveis.

As intervenções incluíam cursos e seminários para professores, pais e profissionais de saúde, incorporando três conceitos: (1) crianças aprendem observando e imitando ou-

tras pessoas, (2) a sociedade influencia o comportamento das crianças, e (3) as crianças necessitam de práticas positivas para desenvolver uma auto-imagem saudável.

Os resultados do estudo de base indicaram que as crianças adquirem comportamentos de risco para a saúde em idade precoce. As análises comparativas mostraram um declínio do consumo de álcool e fumo, além de redução da obesidade, hipertensão arterial e dislipidemia. Uma outra avaliação verificou que a comunidade, professores e pais permaneceram fortemente estimulados a continuar com as intervenções. Em 2001, o programa atingia 30.000 crianças escolares, com utilização de poucos recursos financeiros. Os autores concluíram que o programa poderia ser implementado com baixo custo e utilizado também para diversos temas, consistindo em uma boa opção para países em desenvolvimento.

### Tianjin Project

44

Iniciado em 1984, o Projeto *Tianjin* foi o primeiro grande projeto da China, com foco nas quatro principais doenças crônicas do país: doença cerebrovascular, doença coronariana, câncer e hipertensão arterial. O projeto consistia em áreas de intervenção e áreas-controle, selecionadas e randomizadas, no distrito urbano de *Tianjin*, com cerca de nove milhões de habitantes.

Os objetivos principais eram reduzir o consumo de sal na população, reduzir o tabagismo e prover cuidados para hipertensos, através da reorganização dos serviços de atenção primária à saúde. A viabilidade e o impacto do programa foram avaliados.

Os resultados mostraram redução no consumo de sal, especialmente entre os homens, em todos os estratos sociais, após três anos da intervenção. Após cinco anos, verificou-se redução na prevalência de hipertensão arterial e obesidade na faixa etária de 45 a 65 anos, porém aumento entre os mais jovens. Observou-se aumento do tabagismo entre os homens, principalmente naqueles com maior escolaridade

### Programa de Prevenção de Infartos na Argentina (Propia)

O *Programa de Prevenção de Infartos na Argentina (Propia)* é um programa nacional planejado para trabalhar em conjunto com serviços já existentes de prevenção de doença isquêmica do coração. O Programa propõe uma abordagem unificada de

promoção de saúde, educação e outras intervenções direcionadas a comunidades e indivíduos. O principal objetivo do programa é modificar comportamentos associados a fatores de risco para doença isquêmica do coração.

O programa foi criado em 1990, pela Universidade Nacional e La Plata, e adotado pelo ministério da saúde da Argentina em 1994, em reação aos altos índices de mortalidade por doença isquêmica do coração. A abordagem geral do programa consiste na identificação das necessidades da comunidade e planejamento de intervenções. Três intervenções principais foram propostas. A primeira tinha como foco modificar os fatores de risco na comunidade, através do envolvimento dos meios de comunicação, dos profissionais de saúde, educadores e legisladores. A segunda propunha o diagnóstico e manejo sistemático dos portadores de fatores de risco e daquelas que já haviam desenvolvido a doença. A terceira visava a adequação dos cuidados em saúde por parte dos serviços. Para tal fim, o *Propia* criou centros de assistência para as demandas locais, agências executivas e locais de informação.

A avaliação do programa incluiu aspectos internos, formativos e externos. Os aspectos internos e formativos foram avaliados paralelamente à implementação do programa. A avaliação interna fornecia rápidas informações de feedback aos trabalhadores e gestores. As avaliações formativas forneciam dados de vários componentes do programa, como objetivo de contribuir para o seu desenvolvimento. Além disso, foram realizadas avaliações do impacto do programa.

45

### **Conjunto de Acciones para la Reducción Multifactorial de Enfermedades No transmissibles (Carmen) – OPAS**

O projeto Carmen, criado em 1995, por iniciativa da Organização Pan-Americana de Saúde (OPAS) propunha-se como uma ferramenta prática para os países americanos atingirem o objetivo global da OMS de Saúde Para Todos no Ano 2000. O principal objetivo do projeto era promover coalizações locais e nacionais no intuito de estabelecer políticas e implementar intervenções para redução dos fatores de risco para doenças crônicas não transmissíveis. O projeto Carmen tem como foco o tabagismo, hipertensão arterial sistêmica, obesidade/sobrepeso, diabete mellitus e consumo excessivo de álcool. Porém, a abordagem específica de um fator depende da necessidade identificada em cada país. O projeto propõe uma abordagem integrada que combine serviços de cuidados preventivos em saúde para indivíduos com risco de desenvolver doenças crônicas e promoção de saúde para a população

em geral. Os alvos do projeto seriam atingidos por meio da comunidade, locais de trabalho, escolas e dos próprios serviços de saúde.

Embora tenha sido planejada após o projeto Cindi (Europa e Canadá), a iniciativa Carmen considerou características específicas da América Latina e países caribenhos. As intervenções são implementadas através do desenvolvimento de políticas e protocolos com maior custo-efetividade no manejo dos fatores de risco, educação de profissionais no sentido de reorientar os serviços de saúde para a prevenção, marketing para obter apoio político e social para o projeto, entre outras medidas. Os países interessados em desenvolver o projeto devem solicitar requerimento oficial, com o plano de ação para implementação das intervenções e para avaliação, que segue os protocolos do Cindi. Em cada país participante, o Carmen inicia com intervenções de demonstração que utilizam conhecimentos e serviços já existentes. Baseado nos resultados, tal experiência é expandida para todo o país.

46

As avaliações enfatizam a eficácia e efetividade das intervenções na mudança da morbi-mortalidade por DCNT, bem como da prevalência dos seus fatores de risco. O impacto das atividades do programa é avaliado através da análise de mudanças nos indicadores essenciais (mortalidade por doenças do sistema circulatório, diabetes, câncer e outras DCNT). Cada país participante coleta dados dos indicadores essenciais a cada 3 a 5 anos, embora que, para alguns, a avaliação de resultados necessitaria, pelo menos, 15 anos. Em curto prazo, avaliações de processo indicam como as intervenções funcionam, examinam o custo-efetividade das abordagens de implementação e documentam sua intensidade e alcance.

### **Estratégias da iniciativa Carmen**

A iniciativa Carmen apresenta dois componentes principais para atingir seus objetivos: o componente para redução de riscos e o componente para desenvolvimento de uma rede de trabalho. Cada componente apresenta estratégias e linhas de ação definidas para atingirem seus objetivos.

## A) Componente para redução de riscos

Este componente tem como objetivo promover e sustentar a redução dos fatores e condições de risco através de intervenções comunitárias. Para tanto, o programa propõe três estratégias principais: (1) promoção de saúde e prevenção de forma integrada, em que se aborde em conjunto os diversos fatores de risco, estimulando as comunidades a tornarem-se participantes ativas nas decisões sobre sua saúde, além de desenvolver estratégias para elaboração de consenso entre os diferentes setores da saúde envolvidos (governo, setor privado, etc.); (2) efeito demonstrativo, ou seja, intervenção inicial em uma área demonstrativa para se avaliar a aceitabilidade da intervenção, segurança e efetividade. Para tal, seriam necessários estudos de base, monitoramento dos fatores de risco e coleta sistemática de informações relacionadas à intervenção; e (3) promoção de equidade em saúde, através de intervenções que possam atingir os diversos grupos populacionais, com diferentes níveis de educação e renda.

Esse componente é guiado por três linhas de ação:

- *Elaboração de políticas.* Essa linha de ação é um processo cíclico envolvendo a definição de um problema, elaboração de políticas específicas, implementação da ação e avaliação. Tal processo deve ser adequado para cada país, frente a diversidade de fatores particulares de cada região. Deve contar ainda com a participação de diversos setores governamentais e não-governamentais, privados, formadores de opinião, entre outros.
- *Ações de base comunitária.* Essa linha de ação visa envolver famílias e comunidades nas intervenções, de forma a obter impacto populacional e sustentabilidade. Esta linha de ação propõe a identificação e formação de parcerias com líderes, grupos organizados e instituições, assegurando que todos tenham conhecimento e envolvimento com as intervenções. Nesse contexto, o setor de saúde é um parceiro no processo e não o centralizador das ações.
- *Serviços de saúde responsivos.* Os serviços de saúde devem mudar seu foco principal de atenção, passando a não somente atender as demandas agudas dos pacientes, mas também participar efetivamente de programas de prevenção, além de considerar as necessidades de indivíduos portadores de doenças crônicas.

## B) Componente para o desenvolvimento de uma rede de trabalho

Os principais objetivos desse componente seriam desenvolver, integrar e avançar a pauta de discussão para prevenção de DCNT nas Américas. Especificamente, esse componente pretende promover e sustentar intervenções com custo-efetividade favoráveis; desenvolver estratégias e metodologias adequadas para cada país; desenvolver estratégias para obter apoio de legisladores e outras pessoas-chave; promover e manter a participação da comunidade; e promover pautas regionais para enfrentar a epidemia de DCNT.

A principal estratégia para conduzir esse componente é a cooperação técnica entre países e regiões, através do desenvolvimento de redes de trabalho regionais e internacionais.

As linhas de ação adotadas nesse componente são as seguintes:

- *Advocacia*. Definida como a busca de desfechos influentes que afetam diretamente a vida das pessoas, incluindo políticas públicas e decisões no âmbito político, econômico e social.
- *Mobilização de recursos*. A obtenção de recursos para prevenção de DCNTs requer estrutura conceitual e operacional definidas, conhecimento e suporte técnico suficientes, além de instrumentos de avaliação adequados para assegurar o progresso e satisfazer a demanda da fonte dos recursos.
- *Monitoramento e avaliação*. Sistemas viáveis e eficientes de monitoramento e avaliação são elementos cruciais para mudanças dos processos em execução. Fornecem informações para o contínuo aperfeiçoamento dos programas.
- *Pesquisa*. Existe a necessidade de se fomentar pesquisas, especialmente nos países em desenvolvimento, para que se possa obter evidências importantes acerca da receptividade e efetividade das ações preventivas integradas para DCNT. A iniciativa Carmen mantém parcerias com instituições de pesquisa para suporte técnico no desenvolvimento de investigações.
- *Treinamento*. É necessária a formação de profissionais que se envolvam e sustentem a implementação de prevenção integrada de DCNT nas Américas. Tal necessidade deve ser alcançada com desenvolvimento de programas e treinamentos bem concebidos e de fácil acesso, sendo capazes de habilitar os profissionais a desenvolver e implementar intervenções preventivas.

## Carmen – Chile

O Chile foi o primeiro país latino-americano a integrar a rede Carmen, em 1996. O país possui um Plano Nacional de Promoção da Saúde, com metas para nutrição, atividade física e tabagismo, que se desenvolve de forma complementar à iniciativa Carmen. Tal ação tem como objetivo atingir populações de maior risco, desenvolvendo estratégias de educação em saúde, detecção e controle dos fatores de risco e vinculação das atividades com a comunidade.

Estudo conduzido em uma área demonstrativa (Valparaíso) mostrou que 73% da população na faixa etária de 25 a 64 anos tem pelo menos um fator de risco maior para DCNTs (tabagismo, hipertensão arterial, hipercolesterolemia ou diabetes). A prevalência de tabagismo na população foi de 40,6%, cerca de 85% eram sedentários, 20% tinham IMC acima de 30 kg/m<sup>2</sup> e quase metade (46,9%) apresentava níveis de colesterol altos. Tabagismo, inatividade física, obesidade e hipertensão foram significativamente mais prevalentes na população mais pobre.

O Chile possui cinco áreas demonstrativas do Carmen:

- a) *Área de Valparaíso*, com cerca de 430 mil habitantes. Área Carmen desde 1997.
- b) *Região Bio Bio*, com quase 2 milhões de habitantes. Área Carmen desde 2001.
- c) *Área Sul Oriente da Região Metropolitana*, com cerca de um milhão e 300 mil habitantes. Área Carmen desde 2001.
- d) *Área de Aconcagua*, com cerca de 600 mil habitantes. Área Carmen desde 2002.
- e) *Região de Maule*, com quase 1 milhão de habitantes. Área Carmen desde 2002.

As intervenções chilenas são executadas em dois níveis. Um deles tem por objetivo ampliar a cobertura dos programas, com metas temporais estabelecidas. O outro, restringe-se a projetos específicos demonstrativos, como saúde cardiovascular, educação a distância e capacitação de profissionais de saúde.

As parcerias firmadas no país para a condução da iniciativa Carmen incluem a Pontifícia Universidade Católica do Chile, por meio do Departamento de Saúde Pública; Instituto de Nutrição e Tecnologia de Alimentos (Inta), por meio do Departamento de Nutrição e Fisiologia; Cindi – Cataluña (Espanha); Sociedade Chilena de Cardiologia Preventiva; e Sociedade Chilena de Hipertensão.

O financiamento do Carmen – Chile é oriundo, fundamentalmente, de recursos ministeriais. Cerca de 80% do financiamento provém do setor público, 15% da OPAS/OMS e 5% do setor privado e organizações não-governamentais. Para o ano de 2003, o valor aprovado pelo Ministério da Saúde foi de US\$ 900,000. A OPAS/OMS subsidia a capacitação de recursos humanos e dá assessoria de técnicos e especialistas.

## Financiamento dos Programas

Os estudos norte-americanos (*Stanford Three-Community Study*, *Stanford Five-City Project*, *Pawtucket Heart Health Program* e *Minnesota Heart-Health Program*) foram financiados pelo *National Heart, Lung, and Blood Institute* dos EUA.

O *North Karelia Project*, na Finlândia, teve como idealizadores e incentivadores a OMS/Europa e órgãos do governo, sendo posteriormente firmadas parcerias com diversas entidades não-governamentais e comunitárias. No mesmo sentido, a OMS estendeu tal iniciativa para outras áreas da Europa (programa Cindi) e, recentemente, para as Américas, com a iniciativa Carmen. Da mesma forma, tais programas contam com financiamento oriundo da parceria entre governo e órgãos da sociedade organizada.

50

## Perspectivas

As principais conclusões do *III Fórum Global Mundial para Prevenção e Controle de Doenças Crônicas Não Transmissíveis* (OMS, Brasil, 2003) ratificam que o controle e prevenção das DCNT é um processo lento que requer uma liderança firme e constante. É necessário manter a capacidade crítica e enfrentar os obstáculos de maneira segura, visando a qualificação das iniciativas. O progresso do trabalho em prevenção está sendo visualizado, porém é necessário que os projetos tenham sustentabilidade. Foi enfatizado que resultados efetivos somente são alcançados com a formação de redes de trabalho, cujo fortalecimento foi um dos principais objetivos discutidos naquele fórum.

A proposta de atuar em áreas demonstrativas implica na necessidade de maior comprometimento na busca de evidências, devendo as intervenções ser rigorosamente avaliadas antes de aplicadas em larga escala.

Em novembro de 2004, no Canadá, no *IV Fórum Global sobre Prevenção e Controle de Doenças Crônicas*, representantes de diversos países participaram de ampla discussão sobre parcerias e planos de ação para o combate às doenças crônicas. Os principais temas de discussão consistiram em: políticas de ação, participação e planejamento.

Os principais pontos sugeridos em relação a políticas de ação foram: (1) o aumento da vontade política de mudança e comprometimento, em todos os níveis hierárquicos envolvidos, com decisões baseadas em evidências epidemiológicas; (2) as ações devem focar o sistema de saúde como um todo e não somente os serviços, de forma a incluir todos os setores e ter uma mudança sistêmica; (3) utilizar argumentos para o combate às doenças crônicas no contexto nacional e internacional, de forma a não perder sua importância política frente a outras prioridades existentes, como as doenças transmissíveis, por exemplo; (4) a comunidade internacional deve estar comprometida e cumprir sua agenda de trabalho; (5) desenvolver sólidas políticas de trabalho no combate às DCNTs; (6) atuar em parceria com coordenações nacionais e internacionais; e, (7) o desenvolvimento de estratégias de abordagens e ferramentas intersetoriais.

51

A participação engloba diversas instituições como governo, ONGs e outras parcerias. Os participantes do fórum sugeriram três etapas a serem seguidas para o trabalho em equipe: (1) identificar aspectos em comum, (2) concordância quanto a medidas operacionais, e (3) pactuar comprometimento e confiança mútuos.

O planejamento deve basear-se em uma estrutura de trabalho multisetorial, envolvendo medidas que mantenham o foco das intervenções, mobilizem a comunidade e seus vários setores, fortaleçam órgãos colaboradores (inclusive fora do setor de saúde), promovam mudanças comportamentais, com conseqüente envolvimento nas atividades propostas, com integração dos diferentes setores de prevenção (serviços de saúde, universidades, etc), elaboração de legislação específica, contínua avaliação e monitoramento.

As principais conclusões deste fórum consistiram na renovação do comprometimento em elaborar políticas de prevenção e controle das DCNTs e o reconhecimento da necessidade de urgência das ações e da importância em disponibilizar ferramentas, que ajudem os países e regiões no combate às enfermidades crônicas.

## Intervenções Atuais

Acompanhando o esforço global no combate às DCNTs, diversos países envolveram-se em vários programas e intervenções. Tome-se como exemplo alguns países membros do programa Cindi:

A Bielorrússia aderiu ao Cindi em 1994, implementando projetos demonstrativos. Em 2004, diversos protocolos sobre combate aos fatores de risco para DCNTs foram elaborados, direcionados à população, com recomendações baseadas em características individuais. No mesmo ano, foram conduzidos seminários de treinamento para cerca de 16 mil médicos e enfermeiros, envolvidos em uma intervenção antitabagismo (*International Quit & Win 2004*).

A Croácia desenvolveu, em 2004, ação preventiva em escolas através de medidas educacionais quanto à alimentação saudável, atividade física e riscos vinculados ao tabagismo e etilismo. Além disso, foram conduzidas intervenções para prevenção do diabetes e outras DCNTs. Foi desenvolvido no país um sistema de registro de diabéticos, contabilizando, até o final de 2004, 33 mil pacientes.

52

Foram implementadas estratégias e campanhas para modificar hábitos alimentares (*The Fibre for a Health Life Program*) e incentivar a população a participar de projetos envolvendo atividade física.

Na República Tcheca, no ano de 2004, os programas *Quit & Win* de combate ao tabagismo contou com a participação de duas mil pessoas. A campanha *Challenge your heart to move* contou com a participação de 2.100 pessoas, com as quais foram desenvolvidas ações de educação em saúde.

Na Finlândia, onde o programa Cindi foi pioneiro com o projeto *North Karelia*, as intervenções ainda são mantidas. Os programas mais importantes em andamento estão ligados a atividade física, sobrepeso, pressão arterial e tabagismo. O programa antitabagismo *Quit & Win 2004* contou com a participação de nove mil indivíduos e seis mil colaboradores, sendo que o número de participantes aumentou em 28% em relação ao ano de 2002.

Nos Estados Unidos, várias intervenções de combate às doenças cardiovasculares são desenvolvidas em parceria com o *Centers for Disease Control and Prevention* (CDC). Entre outros, o Wisconsin Cardiovascular Health Program, elaborado com base em

clientes de planos de saúde, aumentou o percentual de pacientes com a pressão arterial controlada, passando de 48% em 2000, para 58% em 2001. O programa identificava os pacientes e cada plano implantava estratégias para melhorar o controle pressórico de seus clientes.

O *New York Cardiovascular Health Program* avaliou os locais de trabalho através de instrumentos específicos e, posteriormente, orientou os empregadores a fazerem mudanças em prol da saúde cardiovascular dos seus trabalhadores. Após três anos de intervenção, houve um crescimento de 65% na participação em mudanças ambientais e políticas para promoção da saúde cardiovascular.

## Conclusões

O aumento global das DCNTs é resultado da mudança de vários aspectos na população ao longo do tempo. A transição epidemiológica dos países é um fator precursor da epidemia das DCNTs, em função da queda das taxas de natalidade, diminuição da mortalidade infantil, com conseqüente aumento da população e expectativa de vida. A evolução da sociedade levou a uma nova rotina cotidiana dos indivíduos, que é fruto da industrialização, urbanização, desenvolvimento econômico e globalização do mercado de alimentos, guiando a transição nutricional para uma dieta rica em gordura saturada e açúcares (OPAS, 2003).

53

No ano 2000, a doença isquêmica do coração foi responsável por 12,4% do total de óbitos no mundo, seguida das doenças cerebrovasculares (9,2%). No mesmo ano, 22,6% dos óbitos em países desenvolvidos tiveram a doença isquêmica do coração como causa, seguido das doenças cerebrovasculares (13,7%), cânceres do trato respiratório (4,5%) e infecções respiratórias baixas (3,7%). Nos países em desenvolvimento, a distribuição das causas de óbitos são mais homogêneas, sendo 9,1% devido a doença isquêmica do coração, 8,0% por doença cerebrovascular e 7,7% decorrentes de infecções respiratórias baixas (Beaglehole R & Yach D, 2003). Estimativas da OMS alertam que, para o ano de 2020, as mortes por doenças crônico-degenerativas representarão 73% dos óbitos no mundo (PAHO, 2003).

Desde a década de 70, vários projetos de intervenção populacionais foram desenvolvidos, com o objetivo de promover mudanças comportamentais que reduzissem fatores de risco para doenças crônicas. Tais projetos enfatizavam que intervenções realizadas apenas em indivíduos de alto risco, selecionados nos serviços de saúde,

teriam um impacto muito discreto no país como um todo. Por outro lado, se as intervenções fossem direcionadas para toda população, pequenas mudanças em fatores de risco e hábitos de vida saudáveis teriam, potencialmente, um grande impacto na saúde pública (Nissinen A et al, 2001). Esta estratégia de intervenção preventiva populacional envolve métodos de controle de massa e mudanças de normas comportamentais da sociedade. Tal estratégia possui algumas vantagens em função de ser radical, ou seja, tentar remover a causa básica da doença; possuir um grande potencial para a população, com reduções significativas na mortalidade; e ainda, ser apropriada do ponto de vista comportamental, procurando tornar o fator de risco socialmente inadequado. Este seria um ponto importante para manter os indivíduos sem os fatores de risco após a intervenção. Por outro lado, a estratégia populacional de intervenção possui algumas desvantagens. Esta oferece muitos benefícios para a população como um todo (diminuição de mortalidade e morbidade), porém pouco benefício aos indivíduos participantes, visto que a maioria viveria muitos anos sem qualquer problema de saúde. Tome-se como exemplo as imunizações, que diminuíram a incidência de doenças na população, porém sem um efeito visível aos indivíduos vacinados. Medidas que proporcionem recompensas sociais aos participantes de programas de mudança de estilo de vida, com reforço à auto-estima e aprovação social, devem ser utilizadas como motivadores para educação em saúde e sua manutenção (Rose G, 1985).

54

Grandes intervenções populacionais apresentaram resultados conflitantes. Após 25 anos de intervenção, a Finlândia reduziu a mortalidade por doenças cardiovasculares e aumentou a expectativa de vida de seus habitantes. Os autores atribuem os resultados favoráveis ao delineamento cuidadoso do programa, à abordagem utilizada e à organização da comunidade. Várias estratégias foram adotadas para a execução da intervenção, contando com a colaboração de órgãos do governo, profissionais de saúde e indústrias alimentícias, além de mudanças políticas (Puska P, 2002).

Estudos posteriores mostraram resultados pouco consistentes para determinados fatores de risco. O *Minnesota Heart Health Program* não conseguiu ter o alcance populacional desejado e com isso não pôde acelerar a tendência secular de diminuição da mortalidade por doenças cardiovasculares, embora o programa tenha obtido um efeito positivo na prevenção do tabagismo em jovens. Baseados nos programas *Minnesota* e *North Karelia*, alguns autores apontam como problemas maiores desse tipo de intervenção o aumento das ações preventivas nas comunidades controle, os poucos recursos para manter a intensidade das ações e o baixo poder estatístico das

avaliações de impacto (Murray DM, 1985; Luepker RV *et al*, 1994; Perry CL *et al*, 1992; Salonen JT *et al*, 1986).

O *Pawtucket Heart Health Program* apresentou resultados modestos quanto à redução de fatores de risco. Os programas *Stanford Three-Community Study* e *Five-City Project* mostraram reduções no colesterol, pressão arterial e tabagismo nos grupos intervenção (PAHO, 2003). Em uma análise conjunta das três intervenções norte-americanas (*Stanford Five-City Project*, *Minnesota Heart Health Program* e *Pawtucket Heart Health Program*) verificaram-se resultados modestos e, freqüentemente, estatisticamente não significativos. Os autores dessa análise atribuem tais resultados a três possibilidades: (1) as intervenções não são efetivas; (2) as intervenções não superam a forte tendência secular de redução natural dos fatores de risco; e (3) pouco poder dos estudos (Winkleby MA & Feldman HA, 1997). Em uma revisão sistemática sobre ensaios clínicos randomizados de intervenções sobre fatores de risco cardiovasculares, os autores apontaram resultados não significativos sobre a mortalidade e reduções modestas dos fatores de risco. As intervenções que utilizavam aconselhamento pessoal ou familiar, com ou sem tratamento farmacológico, tiveram maior efetividade (Ebrahim S & Smith GD, 1997).

55

Considerando a variabilidade de resultados dos programas de intervenção comunitária pode-se levantar dúvidas quanto a sua efetividade. Entretanto, a questão mais importante a ser apontada não está no julgamento da efetividade geral dos programas, mas sim na identificação de quais aspectos das intervenções são efetivas, em que populações e sob que condições (Sellers DE *et al*, 1997). Quando se avalia o resultado geral de um programa, admite-se que todos componentes necessários para sua execução funcionem de maneira sinérgica, o que pode não ocorrer. A efetividade de cada componente de uma intervenção varia conforme a população em questão e o contexto no qual ambos (intervenção e população) estão inseridos. Seria desejável que programas que se propõem primordialmente a trazer benefícios às comunidades, sobrepondo-se ao interesse científico, avaliassem com cuidado aspectos formativos e processuais, e não somente resultados (Mittelmark MB *et al*, 1993; Nissinen A *et al*, 2001). Tal idéia é reforçada por pesquisadores do *Stanford Five-City Project*, que sugerem a adoção de novas metodologias de abordagem comportamental, com testagem prévia, análises mais completas sobre os processos de mudanças comunitárias e avaliação do sucesso dos programas por meio de análises de processo, preferentemente às de resultado.

Finalmente, a acumulação em um mesmo indivíduo de mais de um fator de risco para DCNT é a regra mais do que a exceção. Da mesma forma, fatores de risco como o tabagismo, sedentarismo e dieta inadequada o são para mais de uma das DCNT. Assim, estratégias como a da iniciativa Carmen, que empenham-se em desenvolver, implementar e avaliar políticas voltadas para redução simultânea de uma série de fatores de risco comuns às DCNT, têm vantagem, em relação às intervenções isoladas, de maior benefício como forma de prevenção integrada.

## Quadro 2. Resultados gerais das principais intervenções

Projeto, País	Período	População	Resultados
<i>North Karelia Project (Finlândia)</i>	1972-97	> 250 mil em cada cidade	Nos primeiros 5 anos houve redução das doenças isquêmicas em North Karelia (área-intervenção) em relação à controle. Redução na prevalência de tabagismo, níveis de colesterol e pressão arterial ocorreram nas duas áreas.
<i>Stanford Three-Community Study (Estados Unidos)</i>	1972-74	> 12 mil em cada grupo	A redução da prevalência de tabagismo, níveis de colesterol e pressão arterial foram maiores no grupo intervenção.
<i>Stanford Five-City Project (Estados Unidos)</i>	1978-98	Entre 40 mil e 85 mil nos grupos intervenção, e 35 mil e 161 mil nos grupos controle	Significativa redução na prevalência dos fatores de risco no grupo intervenção. Não houve diferença quanto ao peso. Discreta diminuição nas taxas de mortalidade por doenças cardiovasculares, sem diferença entre os grupos.
<i>Pawtucket Heart Health Program (Estados Unidos)</i>	1980-91	> 70 mil em cada grupo	Redução na prevalência de fatores de risco em ambos grupos. Tendências estáveis de mortalidade por doença coronariana.
<i>Minnesota Heart-Health Program (Estados Unidos)</i>	1980-93	Entre 25 mil e 110 mil	Redução na prevalência dos fatores de risco nos grupos intervenção. Aumento da atividade física no grupo com intervenção educacional, porém sem diferença quanto ao peso. Redução das doenças coronarianas em todos os grupos. Não houve mudança na incidência de acidentes vasculares cerebrais.

Tabela Metodológica e Principais Resultados de Alguns Programas de Prevenção de DCNTs

Estudo	Teoria	Desfechos	Desenho	Metodologia	
				Representatividade da Amostra	Tamanho da Amostra/ Poder
<b>Pawtucket Heart Health Program</b>					
1	Teoria da aprendizagem social	Colesterol elevado, hipertensão arterial, tabagismo, obesidade e sedentarismo	Comparação de surveys	Randomização	Mínimo de 1000 indivíduos Alfa: 5%
2	-	Opção por alimentos mais saudáveis nas compras em supermercados	Surveys	Amostragem aleatória	N=1807 ao longo de 4 anos de estudo
3	Modelo transteórico para mudanças de comportamento	Atividade física	Surveys	Domicílios randomizados, e selecionados indivíduos conforme critérios definidos pelos autores	Surveys envolvendo 7529 indivíduos na cidade intervenção e 7732 na controle.
<b>North Karelia Project</b>					
1	-	Mortalidade	Comparação de surveys	Dados de mortalidade de surveys	Informações oficiais do país Alfa: 5%
2	-	Manutenção da abstinência ao tabagismo	Longitudinal, comparando intervenção e controle	Todos participantes em NK (n=618) e amostra randomizada da Finlândia (n=1500)	N=2114 (perdas 33 a 32%) Alfa: 5%
3	Abordagem de influência social	Prevenção do tabagismo	Longitudinal, comparando intervenção e controle	Estudantes entre 12 e 13 anos	4 escolas-intervenção e 2 escolas-controle
4	-	Hipertensão arterial, colesterol elevado, tabagismo	Comparação de surveys	Amostras randomizadas da população	Amostras grandes em cada survey e poucas recusas
<b>Minnesota Heart Health Program</b>					
1	Teoria do aprendizado social e outras não especificadas	Primeiro evento de doença coronariana definitiva e doença cerebrovascular	Longitudinal, acompanhando cada comunidade (intervenção e controle)	Comunidades não-randomizadas. Semelhantes em tamanho e estrutura.	Cerca de 230 mil habitantes
2	-	Mudança de hábitos quanto tabagismo, alimentação e exercício físico.	Longitudinal (coorte), comparando intervenção e controle	Todas escolas públicas das comunidades estudadas foram contempladas	N=2376 no baseline. No último follow-up houve 55% de não-respondentes

	Replicabilidade	Medidas	Intervalo entre as Medidas	Controle de Qualidade	Análise
	População acima de 18 anos	Amostras de sangue Peso e altura medidos Pressão arterial medida	1 ano	Revisita a 10% da amostra em cada survey	ANOVA
	Indivíduos acima de 18 anos de idade	Questionário	Anual	-	Descritiva
	Indivíduos com idade entre 18 e 64 anos	Questionário	Bianual	-	ANOVA
	População 35 a 64 anos	Bancos de dados oficiais	5 anos	-	Regressão linear Testes para comparar tendências
	Indivíduos que participaram do programa e pararam de fumar	Questionário	1º, 6º e 12º mês após ter parado de fumar	Perguntas sobre variáveis consideradas associadas à abstinência. Considerar não respondentes como fumantes	Qui-Quadrado, Exato de Fisher, Análise multivariável com Regressão logística
	Escolares de 12 e 13 anos que receberam intervenção por 2 anos	Questionários auto-aplicados	Anual	Perdas eram semelhantes aos respondentes	Qui-Quadrado (unidade: indivíduo e escola) e ANOVA (unidade: escola)
	População de 25 a 64 anos	Pressão arterial medida, colesterol dosado em amostra de sangue, tabagismo referido	5 anos	-	Diferenças e mudanças nas proporções no período, correlações ajustadas, análises de covariância
	Desfecho medido em indivíduos de 30 a 74 anos residentes em uma das cidades em estudo no momento do evento	Registros hospitalares de primeiro episódio de doença coronariana ou cerebrovascular	Contínuo por 9 a 11 anos	-	Tendências, incidências e regressão linear
	Estudantes da 6ª série	Questionário	Anual	-	Descritiva e comparação de tendências

Estudo	Teoria	Desfechos	Desenho	Metodologia		
				Representatividade da Amostra	Tamanho da Amostra/ Poder	
3	-	Prevenção do tabagismo	Longitudinal (coorte), comparando intervenção e controle	Todas escolas públicas das comunidades estudadas foram contempladas	N=2401 no baseline. No último follow-up houve 55% de não-respondentes.	
4	Teoria do aprendizado social	Grau de valorização dado a aspectos relacionados à saúde	Longitudinal (coorte), comparando intervenção e controle	Todas escolas públicas das comunidades estudadas foram contempladas	N=2406	
5	Teoria do aprendizado social	Grau de atividade física	Longitudinal (coorte), comparando intervenção e controle	Todas escolas públicas das comunidades estudadas foram contempladas	N=2376. Perdas ao final do estudo de 55% no grupo intervenção e 31% no controle	
6	Teoria do aprendizado social, teoria de comunicações persuasivas e modelos envolvendo líderes comunitários e instituições.		Longitudinal, acompanhando cada comunidade (intervenção e controle) e surveys	Comunidades não-randomizadas. Semelhantes em tamanho e estrutura. Amostras randomizadas para os surveys	Variando de 25 mil a 110 mil	
<b>Stanford Three-Community Study</b>						
1	-	Mudança de dieta	Longitudinal. Comparando comunidades intervenção e controle (surveys)	Amostragem em múltiplos estágios com randomização	Três comunidades (2 intervenção e 1 controle) com população entre 13 mil e 15 mil no baseline	
2	-	Mudança quanto a dieta, peso e colesterol, conforme nível socioeconômico e língua falada	Longitudinal. Comparando comunidades intervenção e controle (surveys)	Amostragem em múltiplos estágios com randomização	Três comunidades (2 intervenção e 1 controle) com população entre 13 mil e 15 mil no baseline	
3	-	Mudança quanto a dieta, peso e colesterol após 2 anos de intervenção	Longitudinal. Comparando comunidades intervenção e controle (surveys)	Amostragem em múltiplos estágios com randomização	Três comunidades (2 intervenção e 1 controle) com população entre 13 mil e 15 mil no baseline	
<b>Stanford Five-City Project</b>						
1	-	Morbidade e mortalidade por doença cardiovascular	Longitudinal. Acompanhamento da incidência de morbidade e mortalidade nas cidades intervenção e controle	Cidades semelhantes em tamanho e estrutura	Cidades com populações variando entre 34,300mil e 132,400mil	

	Replicabilidade	Medidas	Intervalo entre as Medidas	Controle de Qualidade	Análise
	Estudantes da 6ª série	Questionário e medida de tiocianato na saliva (detectar uso de tabaco)	Anual	Os falsos-negativos foram corrigidos (comparação da resposta no questionário com a medida na saliva)	Análise de coorte e tendências das surveys.
	Estudantes da 6ª série	Questionário	Anual	-	Comparação de tendências das surveys. Comparação por gênero.
	Estudantes da 6ª série	Questionário	Anual	-	Comparação de tendências das surveys. Comparação por gênero.
	Adultos 25 a 74 anos de idade	Questionário e medidas biológicas	Anual	-	Análises das coortes e tendências das surveys
	População com idade entre 35 e 59 anos	Questionário (histórico de dieta usual)	Anual	-	Utilizado transformação logarítmica. Análises de indivíduos e comunidades.
	População com idade entre 35 e 59 anos	Questionário, medidas de peso e colesterol sérico	Anual	-	Utilizado transformação logarítmica. Análises das surveys conforme nível socioeconômico e língua falada
	População com idade entre 35 e 59 anos	Questionário, medidas de peso e colesterol sérico	Anual	-	Comparação de surveys
	População de 30 a 74 anos	Bancos de dados oficiais. Dados de mortalidade incompletos eram investigados.	Contínuo durante 13 anos	Revisões bianuais do trabalho da equipe, revisão de 5% das revisões realizadas pela equipe, a equipe de validação de interpretações de exames eram avaliadas anualmente	Incidência, tendências. Análises das cidades como unidades.

Estudo	Teoria	Desfechos	Desenho	Metodologia	
				Representatividade da Amostra	Tamanho da Amostra/ Poder
2	Teoria do aprendizado e mudança comunitária	Redução no índice de massa corporal	Comparação de cidades intervenção e controle. Surveys transversais e em coortes.	Cidades semelhantes em tamanho e estrutura	Baseline com n=2504 e surveys entre 1800 e 2500 indivíduos. Perdas de 54% na coorte (6 anos). Percentual de respondentes nas surveys variando de 56 a 70%.
3	-	Análise dos efeitos da variabilidade do peso sobre os fatores de risco cardiovasculares	Coorte, com análise de surveys	Cidades semelhantes em tamanho e estrutura	269 homens e 361 mulheres. As perdas de acompanhamento chegaram a 63%, sendo estes indivíduos diferentes das restantes
4	Teoria do aprendizado e mudança comunitária	Diferença de gênero quanto a hipertensão arterial, tabagismo e hipercolesterolemia	Comparação de cidades intervenção e controle. Surveys transversais e em coortes	Cidades semelhantes em tamanho e estrutura	Baseline com 1713 indivíduos e 52% de perdas na coorte (6 anos). As surveys com amostra entre 1700 – 1800. O percentual de respondentes nas surveys variou de 61 a 76%.
5	Teoria do aprendizado social	Conhecimento sobre dieta e comportamento e nível de colesterol plasmático	Comparação de cidades intervenção e controle. Surveys transversais e em coortes	Cidades semelhantes em tamanho e estrutura	Coorte com n=777 e surveys incluíam cerca de 425 pessoas por cidade. Percentual de respondentes nas surveys variando entre 53 a 70%. Respondentes diferiam dos não-respondentes
6	Teoria do aprendizado e mudança comunitária	Pressão arterial e controle de hipertensão	Comparação de cidades intervenção e controle. Surveys transversais e em coortes	Cidades semelhantes em tamanho e estrutura	Baseline com n=2504 e surveys variando entre 1800-2500. Perdas de 47% nas coortes. Taxas de respondentes nas surveys variando entre 56 a 69%
7	-	Tabagismo	Comparação de cidades intervenção e controle. Surveys transversais e em coortes	Cidades semelhantes em tamanho e estrutura	Cerca de 440 indivíduos/cidade/survey. Coorte com n=805. Poder do estudo limitado. Percentual de respondentes nas surveys variando entre 59% e 76%. Não-respondentes eram diferentes. Coorte com cerca de 50% de perdas
8	-	Correlação entre atividade física e fatores de risco cardiovascular	Coorte, com avaliação no baseline e 5 anos após	Cidades semelhantes em tamanho e estrutura.	380 homens e 427 mulheres, das quatro cidades.

	Replicabilidade	Medidas	Intervalo entre as Medidas	Controle de Qualidade	Análise
	Indivíduos entre 12 e 74 anos de idade	Questionário. Medida de peso, altura e tiocianato sérico (tabagismo)	Anual	-	Análises das coortes e surveys transversais. Análises das cidades como unidades.
	Indivíduos entre 12 e 74 anos de idade	Questionário. Medida de peso, altura, pressão arterial, colesterol sérico e tiocianato sérico (tabagismo)	Anual	-	Comparação das surveys na coorte. Correlções e regressão linear múltipla
	Indivíduos entre 25 e 74 anos	Questionário. Medida de peso, altura, pressão arterial, colesterol sérico e tiocianato sérico (tabagismo)	Anual	Checagem de amplitude e lógica por computador, treinamento sistemático para os trabalhadores de campo	Regressão logística, análise de tendência temporal e fatores de interação
	Indivíduos entre 25 e 74 anos	Questionário e amostras de sangue para medir colesterol	Anual	-	Análise de covariância e regressão logística
	Indivíduos entre 12 e 74 anos de idade	Questionário, medidas de pressão arterial, peso e altura	Anual	-	Análises de covariância, tendências e regressão
	Indivíduos entre 25 e 74 anos	Questionário e medida de tiocianato sérico	Anual	-	Análise de variância e regressão logística múltipla
	Indivíduos entre 18 e 74 anos	Questionário, medidas de pressão, colesterol sérico, peso e altura	5 anos	-	Correlação

Estudo	Teoria	Desfechos	Desenho	Metodologia	
				Representatividade da Amostra	Tamanho da Amostra/ Poder
6	Teoria do aprendizado e mudança comunitária	Pressão arterial e controle de hipertensão	Comparação de cidades intervenção e controle. Surveys transversais e em coortes	Cidades semelhantes em tamanho e estrutura	Baseline com n=2504 e surveys variando entre 1800-2500. Perdas de 47% nas coortes. Taxas de respondentes nas surveys variando entre 56 a 69%
7	-	Tabagismo	Comparação de cidades intervenção e controle. Surveys transversais e em coortes	Cidades semelhantes em tamanho e estrutura	Cerca de 440 indivíduos/cidade/survey. Coorte com n=805. Poder do estudo limitado. Percentual de respondentes nas surveys variando entre 59%e 76%. Não-respondentes eram diferentes. Coorte com cerca de 50% de perdas
8	-	Correlação entre atividade física e fatores de risco cardiovascular	Coorte, com avaliação no baseline e 5 anos após	Cidades semelhantes em tamanho e estrutura.	380 homens e 427 mulheres, das quatro cidades.

	Replicabilidade	Medidas	Intervalo entre as Medidas	Controle de Qualidade	Análise
	Indivíduos entre 12 e 74 anos de idade	Questionário, medidas de pressão arterial, peso e altura	Anual	-	Análises de covariância, tendências e regressão
	Indivíduos entre 25 e 74 anos	Questionário e medida de tiocianato sérico	Anual	-	Análise de variância e regressão logística múltipla
	Indivíduos entre 18 e 74 anos	Questionário, medidas de pressão, colesterol sérico, peso e altura	5 anos	-	Correlação



**PARTE II**  
**Indicadores de Avaliação**  
*Iná S. Santos*

## Referencial Teórico

Este trabalho utiliza como referencial teórico o modelo desenvolvido por Habicht *et al*, 1999, de avaliação de programas / intervenções em saúde e nutrição. Esse modelo utiliza, fundamentalmente, dois eixos. O primeiro eixo (que será abordado nessa seção) refere-se ao tipo de indicadores necessários para a avaliação e o segundo, ao grau de inferência exigido para a tomada de decisões (o grau de inferência da avaliação).

A escolha do tipo de indicador, ou seja, do que deve ser medido, depende do que vai ser avaliado, se a oferta, a utilização, a cobertura ou o impacto do programa / intervenção. Há um *continuum* nessa seqüência. O programa / intervenção precisa ser oferecido e acessível à população-alvo, além de ter adequada qualidade. Uma vez disponível e adequado, é necessário que a população aceite o programa / intervenção e o utilize. Tal utilização resultará em uma dada cobertura da intervenção. Finalmente, a cobertura alcançada produzirá um impacto (resultado populacional) sobre um comportamento ou sobre a saúde. Quaisquer deficiências nos estágios iniciais dessa cadeia limitarão os resultados nos estágios finais.

68

O tipo de indicador a ser utilizado e a complexidade da avaliação dependem do uso que se pretende dar ao resultado da avaliação. Por exemplo, a agência financiadora de um programa poderá precisar demonstrar que o mesmo logrou um impacto positivo e estatisticamente significativo sobre a mortalidade por doenças cardiovasculares. O gestor local da intervenção, por sua vez, poderá estar interessado apenas em saber se uma determinada cobertura foi alcançada na população-alvo. O primeiro tipo de avaliação fornecerá evidências de efetividade, sendo relevante, por exemplo, na decisão de expandir ou não o programa. A segunda avalia o desempenho (processo) da intervenção e poderá apoiar a decisão, por exemplo, de que a intervenção não necessita de ajustes.

As avaliações de processo têm como objetivo produzir conhecimento para uso local sobre o que está sendo oferecido à população. Esse tipo de informação é muito útil como retro-alimentação do programa / intervenção, permitindo sua adaptação às necessidades da comunidade e, conseqüentemente, tornando mais provável o alcance dos resultados planejados. Avaliações de impacto, por outro lado, fornecem informações relevantes para uso local mas também com alguma capacidade de generalização.

Os indicadores de processo não são mais nem menos importantes do que os de impacto. Determinar como um programa / intervenção atua é tão importante quanto determinar seu impacto. Além disso, os diferentes tipos de indicadores não são excludentes. Aliás, como anteriormente enfatizado, em saúde pública, a regra é que as avaliações de impacto, mesmo que realizadas através de estudos randomizados, não dispensam a coleta de indicadores de processo (oferta, utilização e cobertura). A ausência de efeito sobre alguns indicadores de saúde (indicadores de impacto) como observado, por exemplo, no *Minnesota e no Pawtucket Heart Health Programs* pode ser atribuída à incapacidade desses programas em envolver um número suficiente de pessoas da população-alvo (indicadores de cobertura), necessário para a ocorrência de efeito em escala populacional (Rose G, 2001; Mittelmark MB, 1993).

Por outro lado, dependendo do objetivo da avaliação, indicadores somente de processo podem também não ser suficientes. Em programas / intervenções comunitários complexos, são esperadas diferentes interações em diferentes contextos sociais, culturais e econômicos. Diferentes interações, por sua vez, podem conduzir a resultados distintos, tornando necessário utilizar indicadores de impacto, mesmo para programas / intervenções já formalmente testadas (WHO, 2001; Victora CG, 2006).

A escolha do indicador depende ainda do montante de recursos disponíveis para a avaliação. Indicadores de oferta e utilização poderão ser obtidos através de registros disponíveis ou gerados pela própria intervenção. A cobertura e o impacto, por outro lado, requerem trabalho de campo, o que implica importante aumento de custo.

Idealmente, avaliadores, coordenadores e implementadores deveriam juntos, na fase de planejamento do programa / intervenção, construir o referencial teórico do mesmo e a cadeia causal através da qual seu efeito, se houver, ocorrerá. Nessa cadeia, os indicadores mensuráveis são mais facilmente identificados e os objetivos de futuras avaliações mais facilmente definidos.

O indicador específico a ser utilizado na avaliação depende das características do próprio programa / intervenção. A população a que o programa / intervenção se dirige gera indicadores de cobertura. A natureza do programa / intervenção, os instrumentos, equipamentos e recursos humanos utilizados, o método de veiculação à população-alvo, entre outros fornecem elementos para a formulação de indicadores

de oferta. Os registros sobre a implementação do programa construídos ao longo do tempo são úteis para a construção de indicadores de utilização e oferta. Os objetivos do programa / intervenção, por sua vez, permitem construir indicadores de impacto.

Nos próximos parágrafos, são apresentados alguns exemplos de indicadores extraídos de avaliações de programas / intervenções para redução dos fatores de risco cardiovasculares.

**Exemplos de Indicadores utilizados em  
Avaliações de Programas / Intervenções  
Preventivas de DCNTs**

## Indicadores de Oferta

Indicadores de oferta visam responder se ações ou atividades do programa / intervenção estão disponíveis para a população-alvo, se são acessíveis e se a qualidade é adequada.

O *Minnesota Heart Health Program* (Luepker RV *et al*, 1994) utilizou meios de comunicação em massa e, entre outras medidas, recrutou 100% dos médicos de atenção primária para programas de treinamento e envolveu organizações em programas de modificações ambientais.

O *Pawtucket Heart Health Program* (Carleton RA *et al*, 1995), por exemplo, oferecia programas formais de mudança de comportamento, como cursos de auto-ajuda, aulas para pequenos grupos, avaliações de risco e aconselhamento para maiores de 18 anos. Mais de 500 organizações comunitárias foram envolvidas no programa, incluindo todas as 27 escolas (públicas e privadas), a maioria das organizações sociais e religiosas, os maiores locais de trabalho, todos os supermercados e vários estabelecimentos menores, 19 restaurantes e a maioria das secretarias municipais.

72

Em 1995, Young DR *et al* publicaram uma análise sobre o efeito do *Stanford Five-City Project* no conhecimento, atitudes e comportamento em relação à atividade física. Durante o terceiro e o quarto ano do projeto, livretos com 16 páginas de exercícios foram distribuídos a 25.000 pessoas. Quatro chamadas de três minutos cada uma foram produzidas e apresentadas durante os noticiários noturnos de televisão. *Videotapes* com meia hora de duração foram distribuídos para grupos comunitários. No quinto ano, haviam sido computados mais de 20 contatos do projeto por pessoa (TV, rádio, jornal, material educativo impresso e diretamente pessoa a pessoa).

No *Pawtucket Heart Health Program* (Carleton RA *et al*, 1995), aproximadamente 60% dos 110.000 contatos foram de moradores de Pawtucket, a maioria mulheres e mais velhos.

O *Coronary Health Improvement Project (CHIP)* (Diehl HA, 1998) oferecia aos participantes reuniões diárias de segunda a quinta-feira, com 2h30min de duração, por quatro semanas em um programa educacional intensivo. Além disso, os participantes assistiam duas oficinas de nutrição aplicada, em domingos consecutivos, com duração de meio dia cada uma, seguidas de refeições.

## Indicadores de Utilização

Indicadores de utilização respondem se os indivíduos da população-alvo estão participando das atividades oferecidas pelo programa / intervenção.

No Pawtucket *Heart Health Program* (Carleton RA *et al*, 1995), durante os sete anos do programa (1984 a 1991), 42.000 pessoas participaram de uma ou mais atividades, resultando em 110.000 contatos documentados. As atividades de aconselhamento nutricional, controle de pressão arterial e peso foram as mais utilizadas pelos participantes.

Na avaliação de Elmer PJ *et al*, 1995, os participantes do estudo compareceram a uma média de 14,2 das 16 sessões de aconselhamento oferecidas pelo programa de aconselhamento.

O estudo de Young DR *et al*, em 1995, sobre o *Stanford Five-City Project* mostrou que, no quinto ano, 87 locais de trabalho estavam engajados nas atividades do projeto, totalizando cerca de 3.000 participantes.

73

## Indicadores de Cobertura

Os indicadores de cobertura permitem avaliar a proporção da população-alvo atingida pelo programa / intervenção, sendo necessário, portanto, conhecer o denominador (o número que se pretende alcançar com as atividades do programa / intervenção).

O *Minnesota Heart Health Program* (Luepker RV *et al*, 1994) teve uma cobertura de mais de 60% de todos os adultos com 25-74 anos de idade em atividades de rastreamento e programas educacionais e de mais de 30% de todos os adultos para intervenções face-a-face. A maioria dos jovens foi coberta através de atividades de promoção à saúde de base escolar.

## Indicadores de Impacto

Indicadores de impacto avaliam se os objetivos do programa / intervenção foram alcançados. A comparação da variação no consumo diário de colesterol e de gor-

dura saturada entre os três inquéritos das comunidades intervenção e controle do *Stanford Three Community Study* e da variação no colesterol plasmático foram os indicadores, respectivamente, de desfechos intermediário e final, da avaliação de Stern MP *et al*, 1976.

Na análise longitudinal de Fortmann SP *et al*, 1981, no *Stanford Three Community Study*, foram utilizadas como medidas de impacto as diferenças no consumo relatado de colesterol, gorduras saturadas e álcool, além da mudança relativa de peso e de colesterol plasmático, entre 1972 e 1975.

A avaliação de Luepker RV *et al*, 1994, sobre o impacto do *Minnesota Heart Health Program*, usou como indicadores as diferenças entre o grupo intervenção e controle quanto à prevalência ajustada de tabagismo e sedentarismo e quanto às médias ajustadas de escore de risco cardiovascular, colesterolemia, pressão arterial sistólica e diastólica e índice de massa corporal.

74

Em uma das avaliações do Pawtucket Heart Health Program, Carleton RA *et al*, 1995, utilizaram como indicadores de impacto as diferenças ajustadas entre Pawtucket e a cidade controle quanto à prevalência de tabagismo; taxas projetadas de doença cardiovascular para cada 10.000 habitantes de 35-64 anos de idade, para os 10 anos seguintes; e médias de colesterol total, pressão arterial sistólica e diastólica e índice de massa corporal.

Dunn AL *et al*, 1997, na avaliação randomizada do projeto *Active*, compararam o efeito no grupo de mudança de estilo de vida (grupo intervenção) com o grupo de exercícios estruturados (controle) quanto à variação média, em seis meses, no colesterol total e frações, pressão arterial sistólica e diastólica, peso, composição corporal, equivalentes metabólicos e escore de risco cardiovascular.

Puska P *et al*, 1998, na avaliação do North Karelia Project, utilizaram como indicadores os percentuais de mudança verificados de 1969-1971 a 1993-1995, nas taxas de mortalidade por todas as causas, por doenças cardiovasculares, doença coronariana, doença cerebrovascular, câncer e acidentes e violências, por 100.000 habitantes, na área demonstrativa do projeto (North Karelia) e em toda a Finlândia.

Cutter J *et al*, 2001, avaliando o impacto do *National Healthy Lifestyle Programme* em Singapura, mediram a diferença, de 1992 para 1998, nas prevalências de diabe-

tes mellitus, intolerância à glicose, hipertensão arterial, obesidade, hipercolesterolemia, tabagismo e sedentarismo.

O estudo randomizado de Korhonen M et al, 2003, comparando o efeito do aconselhamento nutricional intensivo com o aconselhamento habitual, sobre o controle da hipertensão arterial, entre pacientes de atenção primária, no North Karelia Project, utilizou como indicadores de impacto a diferença entre os dois grupos quanto a variação do peso corporal, colesterol total e frações e pressão arterial sistólica e diastólica. Indicadores de evidências de adesão ao aconselhamento (desfechos intermediários) foram obtidos e comparados entre o grupo intervenção e controle (consumo de sal, fibras, gorduras, entre outros, na dieta registrada por quatro dias consecutivos, e eliminação de sódio e potássio em amostras urinárias de 24 horas).

**Relação entre os Indicadores  
do Marco Referencial de Habitch *et al*  
e os de Avedis Donabedian**

Avedis Donabedian (Donabedian A, 1984) identifica três tipos de indicadores de avaliação: estrutura, processo e resultado. Indicadores de estrutura refletem as condições de área física, tecnologia apropriada, recursos humanos, medicamentos, acesso a normas de avaliação e manejo de pacientes etc. Indicadores de estrutura identificam as condições sob as quais o cuidado à saúde é oferecido aos usuários. Por exemplo, a atenção ao paciente hipertenso demanda a disponibilidade de balança de adultos, esfigmomanômetro, estetoscópio, medicamentos anti-hipertensivos, acesso a exames de laboratório para monitorização de órgãos-alvo, normas para investigação e manejo desses pacientes, além de profissionais adequadamente treinados para a tomada de decisões.

Indicadores de processo indicam o que é realmente oferecido aos usuários em termos de cuidado. Frente a um paciente com determinada queixa ou problema de saúde, indicadores de processo apontam o que os profissionais fazem, em termos de coleta de história, exame físico, exames complementares solicitados, tratamento e acompanhamento a curto, médio ou longo prazos. Geralmente, esses indicadores são comparados a padrões de qualidade previamente estabelecidos, como *guidelines* nacionais ou internacionais.

77

Os indicadores de resultado, por sua vez, refletem o quanto o usuário do serviço teve sua queixa ou problema resolvido, em decorrência da consulta, depois de decorrido um determinado período de tempo. Os indicadores de resultado precisam ser avaliados levando em conta o nível de complexidade do serviço e sua adequação para resolver o tipo de problema apresentado pelo paciente. Há problemas cuja complexidade pode ser atendida satisfatoriamente no nível primário de atenção. Outros exigirão tecnologia e recursos humanos de maior complexidade. Por exemplo, a pequena minoria de pacientes com hipertensão secundária, embora tendo tido seu problema detectado no nível primário, será mais bem manejada junto a especialistas, no nível secundário de atenção. A satisfação do paciente e a satisfação do profissional são duas outras dimensões do resultado definidas por Donabedian.

Não há conflitos entre os dois modelos. Os indicadores de estrutura e processo de Donabedian correspondem aos de oferta de Habitch *et al.* Há, no entanto, uma distinção a ser feita entre os dois referenciais teóricos. O referencial de Habitch *et al* tem enfoque mais voltado para a saúde pública e o de Donabedian mais para avaliação de serviços. Essa diferença conceitual se expressa na população-alvo das avaliações. No primeiro caso, a população-alvo é o grupo populacional a ser atingi-

do pelo programa / intervenção, enquanto que no segundo, é a população usuária do serviço.

## Importante

Para avaliações de programas / intervenções preventivas de fatores de risco cardiovasculares, o Manual Cindi (WHO, 1996) considera essencial que sejam coletados dados de mortalidade (nacionais e nas áreas demonstrativas), dados de consumo de alimentos (nacionais) e de desfechos esperados (nas áreas demonstrativas). Os desfechos esperados essenciais incluem: indicadores biológicos (colesterol total, pressão arterial sistólica e diastólica, peso e altura corporais), indicadores comportamentais (tabagismo, consumo de bebidas alcoólicas e atividade física), além de indicadores sócio-demográficos (sexo, idade e anos completos de escolaridade).

Indicadores recomendados (embora não essenciais) a serem coletados em áreas demonstrativas incluem: dados de morbidade (cardiopatia isquêmica, acidente vascular cerebral, diabetes mellitus, câncer, doenças respiratórias crônicas e acidentes).

78

Outros indicadores recomendados: HDL colesterol, glicemia, gama glutamil-transferase, tiocianato sérico, cotinina ou outra medida para validar tabagismo auto-referido. Padronização de medida desses indicadores é fornecida naquele mesmo documento.

Para monitoramento de processo e impacto de programas / intervenções sobre fatores de risco e morbi-mortalidade por DCNTs, em escala nacional ou regional, a bibliografia do Programa Cindi da Organização Mundial de Saúde (WHO, 1996) disponibiliza indicadores, bem como a metodologia de cálculo e seleção da amostra, intervalos de avaliação, metodologia da coleta de dados e instrumentos.

Importante destacar que, para fins de monitoramento, os **indicadores distais de impacto** (mortalidade por DANts), conforme apontado pelo referido documento, podem ser obtidos a partir de fontes secundárias dos sistemas de informação regularmente em curso nos países, como o Sistema de Informações de Mortalidade (SIM) do Brasil. Da mesma forma, **indicadores intermediários de impacto** (prevalência de DANts) podem ser obtidos a partir de bancos de dados de base hospitalar e de serviços de saúde. No Brasil, o Sistema de Internações Hospitalares do SUS (SIH/SUS), o Sistema de Informações Ambulatoriais (SIA/SUS), o Sistema de

Informação da Atenção Básica (SIAB), o Sistema de Informações sobre Orçamentos Públicos em Saúde (SIOPS) e a Rede Integrada de Informações para a Saúde (RIPSA) atenderiam a esse propósito.

O **monitoramento do processo** da iniciativa Carmen demanda inquéritos específicos de amostras populacionais. Esses inquéritos forneceriam indicadores de processo (oferta, utilização e cobertura de ações e atividades de prevenção e controle das DANTs), bem como indicadores de prevalência de fatores de risco para DANTs, prevalência referida de DANTs e mensurações de indicadores biológicos (peso, altura, pressão arterial, dosagens bioquímicas). O documento Cindi de protocolo e normas, acima mencionado (WHO, 1996), contém um questionário (*Methodology for Health Behaviour Survey – Cindi Health Monitor Questionnaire*) com quatro questões consideradas de aplicação obrigatória, que exploram processo para o período de 12 meses anteriores à entrevista: a 3.12 (sobre tabagismo), a 4.11 (sobre nutrição), a 5.7 (sobre consumo de bebidas alcoólicas) e a 6.14 (sobre atividade física); além de um questionário específico para processo, o *Process Evaluation Survey* (Anexo 10 daquele documento).



## **PARTE III**

# **Grau de Inferência da Avaliação**

*Iná S. Santos*

O segundo eixo do referencial teórico de Habicht *et al*, 1999, refere-se ao tipo de inferência causal (adequação, plausibilidade ou probabilidade) e ao grau de certeza necessária para a tomada de decisão. Tanto as avaliações de processo (oferta, utilização e cobertura) quanto as de impacto podem utilizar inferência de adequação, plausibilidade ou probabilidade.

## Avaliações de Adequação

Avaliações de adequação de um programa / intervenção respondem se as mudanças esperadas aconteceram ou não. Esse grau de inferência depende da comparação entre o desempenho (processo) ou o impacto do programa / intervenção e critérios de adequação estabelecidos *a priori*. Tais critérios poderão ser expressos em valores absolutos como, por exemplo, realização de duzentas oficinas sobre estratégias comportamentais para abandonar o tabagismo (indicador de oferta) ou 80% de adesão da população-alvo ao número mínimo de encontros de cada oficina para que a intervenção possa ser considerada realizada (utilização).

82 Os critérios de adequação poderão também expressar uma mudança, por exemplo, redução de 20% na prevalência de fumantes na área onde o programa / intervenção foi implementado. Mesmo quando objetivos específicos não tiverem sido estabelecidos *a priori*, o desempenho ou o impacto do programa / intervenção poderá ser avaliado através de séries temporais, tais como aumento da cobertura ou redução da mortalidade.

As avaliações de adequação não requerem um grupo controle se os resultados forem comparados com critérios préestabelecidos (por exemplo, 90% dos indivíduos engajados em atividade física regular). Para avaliar a adequação de mudança ao longo do tempo, serão necessárias pelo menos duas medidas, o que aumenta a complexidade da avaliação. De qualquer forma, avaliações de adequação são muito mais baratas do que os outros dois tipos de avaliação.

As avaliações de adequação do processo avaliam o quanto as atividades do programa / intervenção atenderam os objetivos esperados, por exemplo, quantos grupos de esporte foram criados, quantas vagas para atividades esportivas foram oferecidas, quantas pessoas se inscreveram para atividades esportivas ou qual a cobertura das atividades esportivas na população-alvo da intervenção. A avaliação poderá ser transversal, conduzida em uma única ocasião, durante ou ao final da intervenção.

Poderá também ser longitudinal, requerendo dados de base ou incluindo repetidas medidas para detectar tendências.

As avaliações de adequação de impacto avaliam se indicadores de saúde ou de comportamento melhoraram entre os indivíduos que receberam a intervenção ou na população-alvo como um todo. Novamente, a avaliação pode ser transversal ou longitudinal. Uma vantagem das avaliações de adequação é que podem, frequentemente, utilizar dados secundários, o que reduz os custos.

As avaliações de adequação limitam-se a descrever se as mudanças esperadas ocorreram. Uma avaliação da oferta ou utilização de um programa / intervenção, no entanto, pode dar uma idéia razoável sobre o sucesso do mesmo. Por exemplo, uma avaliação que demonstre que melhorou o aconselhamento para atividade física feito por médicos e que aumentou o número de hipertensos e diabéticos inscritos em grupos de atividade física, pode com segurança atribuir esse resultado a um programa de prevenção e controle da hipertensão arterial sistêmica e diabetes mellitus instituído nos serviços de saúde da comunidade.

Quando se mede cobertura ou impacto, por outro lado, é difícil inferir que qualquer melhora observada seja devida ao programa / intervenção, uma vez que a ausência de um grupo controle não garante que essas mudanças não teriam ocorrido de qualquer maneira, independentemente do programa. As melhoras observadas podem ter ocorrido por influências externas ao programa, como tendências seculares em mortalidade ou obesidade, melhorias socioeconômicas, presença de outras intervenções ou programas na mesma área etc.

As avaliações de adequação também poderão mostrar ausência de mudança no indicador medido. Sob condições normais, isso indicaria que a intervenção não foi efetiva. No entanto, em alguns casos, como, por exemplo, quando ocorre piora das condições socioeconômicas, a não mudança pode indicar que a intervenção foi efetiva, prevenindo a piora desses indicadores na população afetada.

Apesar da incapacidade das avaliações de adequação de afirmar, em uma relação de causa e efeito, a ligação entre as atividades do programa / intervenção e as mudanças observadas, essas avaliações podem fornecer elementos suficientes para analisar se os objetivos estão sendo atingidos. Se a avaliação mostrar que os objetivos não estão sendo atingidos, outras avaliações poderão ser necessárias, de forma a identificar

as causas do insucesso e orientar medidas saneadoras. O Quadro 1, traduzido de Habicht *et al*, 1999, apresenta as características das avaliações de adequação.

Quadro 1. Características das avaliações de adequação\*

Tipo de avaliação	Medidas	Em quem?	Comparado a quê?	Inferências
Adequação			Critérios de adequação previamente definidos	Objetivos alcançados
Desempenho (oferta, utilização, cobertura)	Atividades do programa	Profissionais que implementam o programa. Beneficiários do programa.		Atividades sendo realizadas conforme planejado no cronograma inicial
Transversal	Uma vez		Valor absoluto	
Longitudinal	Mudança		Valor absoluto e incremento	
Impacto	Indicadores de saúde e de comportamento	Beneficiários do programa ou população-alvo		A mudança observada na saúde ou no comportamento é na direção e magnitude esperadas
Transversal	Uma vez		Valor absoluto	
Longitudinal	Mudança		Valor absoluto e incremento	

\*Extraído e traduzido de Habicht et al, 1999

## Avaliações de Plausibilidade

Avaliações de plausibilidade vão além das avaliações de adequação porque tentam afastar fatores externos (fatores de confusão) que podem ter causado os resultados observados. As avaliações de plausibilidade tentam controlar o efeito de fatores de confusão selecionando um grupo controle antes do início da avaliação ou depois, durante a análise dos dados. Nessas avaliações, os grupos controles podem ser históricos, internos ou externos.

**Grupo controle histórico:** constituído pelas mesmas instituições ou população-alvo. Permite comparar a mudança de antes para depois do programa / intervenção, acompanhada de uma tentativa de descartar fatores externos.

**Grupo controle interno:** constituído por instituições, áreas geográficas ou indivíduos que deveriam ter recebido a intervenção mas não receberam, seja porque não puderam ou porque se recusaram a receber. Os indicadores podem então ser comparados entre três ou mais grupos de comunidades ou indivíduos com diferentes graus de exposição à intervenção. Uma associação tipo dose-resposta entre a intensidade da intervenção recebida e o desempenho ou impacto observado permite maior plausibilidade do que a comparação entre apenas grupos que receberam toda a intervenção e grupos que nada receberam. Esse tipo de avaliação requer comparação de dados transversais coletados ao final de um ciclo de intervenção.

**Grupo controle externo:** constituído por uma ou mais instituições ou áreas geográficas sem o programa. Nesse caso, a comparação pode ser transversal (intervenção *versus* controle ao final de um ciclo da intervenção) ou longitudinal controlada (comparando intervenção e controle no início e ao final de um ciclo).

85

O uso de qualquer tipo de grupo controle resulta em conclusões mais plausíveis do que as obtidas de avaliações sem grupo controle. A plausibilidade pode ser também bastante melhorada se esses tipos de grupos controle forem usados em combinação. Programas / intervenções que iniciam em momentos diferentes, em áreas separadas, por exemplo, permitem a combinação de controles históricos e controles externos. As áreas onde o programa / intervenção irá iniciar mais tarde constituem o grupo controle externo. As áreas que iniciarão mais tarde, por sua vez, terão controles históricos.

O pressuposto é que os grupos intervenção e controle serão semelhantes em todas as características relevantes, exceto a exposição à intervenção. Isso quase nunca é verdadeiro porque um dos grupos de comparação pode ser influenciado por um fator de confusão que não afeta tanto o outro grupo. Por exemplo, se um programa de atividade física for implementado em uma área com melhor sistema de saúde do que a área controle, a diferença em hospitalizações por hipertensão ou diabetes pode ser devida ao melhor sistema e não ao programa. Para se lidar com os fatores de confusão, há a necessidade de medir os prováveis confundidores e tratá-los estatisticamente através de emparelhamento, padronização, estratificação ou outra forma de análise multivariada (Rothman KJ & Greenland S, 1998). O controle de

fatores de confusão é particularmente importante quando estão sendo feitas comparações internas. Indivíduos que recusam a intervenção ou os que deveriam mas não recebem a intervenção, freqüentemente, diferem dos beneficiários em várias outras formas.

O desenho de avaliações com controles históricos é semelhante a avaliações de adequação em que uma série temporal é analisada sem comparações externas. Para que seja caracterizada como uma avaliação de plausibilidade, no entanto, é necessário tentar excluir outras causas possíveis para as tendências observadas, por exemplo, se a redução das hospitalizações por hipertensão ou diabetes não foi devida a mudanças no acesso aos serviços de saúde locais, à melhora na regularidade do suprimento de medicamentos ou ao treinamento dos profissionais de saúde para o manejo dessas doenças.

Em vários aspectos, as avaliações de plausibilidade assemelham-se aos “experimentos naturais” (Kleinbaum DG *et al*, 1982). O avaliador aproveita a existência oportuna de um grupo controle para examinar o efeito de um programa ou intervenção. Como o nome indica, uma conclusão de plausibilidade baseia-se largamente em julgamentos de valores de especialistas na área, incluindo os gestores e os avaliadores. Para obter um alto grau de plausibilidade nos resultados é necessário descartar todas as outras explicações prováveis, o que pode exigir várias comparações e simulações matemáticas.

A principal limitação das avaliações de plausibilidade é que não se pode descartar completamente todas as explicações alternativas para as diferenças observadas. No entanto, mesmo avaliações de plausibilidade menos rigorosas são freqüentemente suficientes para decidir-se sobre o futuro de um programa / intervenção. O Quadro 2, traduzido de Habicht *et al*, 1999, apresenta as características das avaliações de plausibilidade.

Quadro 2. Características das avaliações de plausibilidade

Tipo de avaliação	Medidas	Em quem?	Comparado a quê?	Inferências
Plausibilidade			Grupo controle oportunístico ou não-randomizado	O programa parecer ter um efeito maior e além do impacto de fatores externos ao programa
Desempenho (oferta, utilização, cobertura)	Atividades do programa	Profissionais que implementam o programa. Beneficiários do programa (dicotômico ou dose-resposta)		O grupo intervenção parece ter melhor desempenho do que o controle
Transversal	Uma vez		Grupo controle	
Longitudinal	Mudança		Antes-depois	
Longitudinal-controle	Mudança relativa		Comparação antes-depois entre grupo intervenção e controle	
Impacto	Indicadores de saúde e de comportamento	Beneficiários do programa ou população-alvo (dicotômico ou dose-resposta)		A mudança observada na saúde ou no comportamento parece ser melhor no grupo intervenção do que no controle
Transversal	Uma vez		Grupo controle	
Longitudinal	Mudança		Antes-depois	
Longitudinal-controle	Mudança relativa		Comparação antes-depois entre grupo intervenção e controle	
Caso-controle	Uma vez	População-alvo	Comparando exposição ao programa entre doentes (casos) e não-doentes (controles)	

Extraído e traduzido de Habicht et al, 1999

## Avaliações Probabilísticas

As avaliações de probabilidade têm como objetivo garantir que exista apenas uma pequena e conhecida probabilidade de que a diferença entre as áreas com programa / intervenção e as áreas controle seja devida ao azar, ao efeito de vieses ou de fatores de confusão. Essas avaliações requerem randomização dos grupos, sendo o padrão ouro das pesquisas de eficácia.

Embora a randomização não garanta que toda a possível confusão seja eliminada (uma crença errônea e comum), a probabilidade de confusão será mensurável, sendo parte do erro associado ao nível de significância utilizado ( $p < x\%$ ). O fator de confusão não precisa nem mesmo ser conhecido para esse procedimento funcionar. A randomização assegura que a associação estatisticamente significativa está relacionada diretamente e causalmente com a intervenção e não simplesmente afirma que os grupos de comparação são diferentes, como é o caso dos tipos anteriores de avaliação.

88

Há várias razões pelas quais as avaliações probabilísticas são frequentemente ineqüívocas para estimar a efetividade de programas / intervenções. Em primeiro lugar, o avaliador deve participar das fases iniciais de planejamento do programa / intervenção, para delinear a randomização. Serviços, comunidades ou indivíduos elegíveis têm que ser listados e randomizados para o grupo intervenção ou controle. Lamentavelmente, os avaliadores geralmente são chamados somente depois de o programa já haver sido implementado.

Também é necessário superar influências políticas que afetam a escolha do local onde o novo programa / intervenção será implementado. Intervenções são geralmente bem vistas e, em função disso, os planejadores sofrem pressão política para aplicá-las em comunidades mais influentes. Para garantir a alocação aleatória, o avaliador precisará influenciar diretamente o processo de implementação. Alternativas como desenhos escalonados (introdução por estágios) em que a intervenção é introduzida em uma seqüência randomizada, mas eventualmente estendida para todas as comunidades ou indivíduos elegíveis, têm sido propostas. Isso significa que os desenhos randomizados não são apropriados para examinar resultados com longos períodos de intervalo após o início da intervenção. O Quadro 3, traduzido e Habicht *et al*, 1999, apresenta as características das avaliações probabilísticas.

Quadro 3. Características das avaliações de probabilidade

Tipo de avaliação	Medidas	Em quem?	Comparado a quê?	Inferências
Probabilidade			Grupo(s) controle(s) randomizado(s)	O programa tem um efeito ( $p < 0,005$ )
Desempenho (oferta, utilização, cobertura)	Atividades do programa	Profissionais que implementam o programa. Beneficiários do programa		O grupo intervenção tem melhor desempenho do que o controle
Longitudinal-controle	Mudança relativa		Comparação antes-depois entre grupo intervenção e controle	
Impacto	Indicadores de saúde e de comportamento	Beneficiários do programa		Mudança na saúde ou comportamento melhor no grupo intervenção do que no controle
Longitudinal-controle	Mudança relativa	População-alvo	Comparação antes-depois entre grupo intervenção e controle	

Extraído e traduzido de Habicht et al, 1999

Em síntese, a força da inferência sobre a causalidade do efeito do programa / intervenção aumenta desde a descrição sem um grupo de comparação até a comparação com grupos controle possivelmente enviesados e finalmente com comparação com um grupo controle provavelmente não enviesado (por meio de ensaios randomizados).

Cada um dos quatro componentes do eixo de indicadores (oferta, utilização, cobertura e impacto) pode ser avaliado de acordo com os três graus de inferência (adequação, plausibilidade e probabilidade). Um exemplo é dado no Quadro 4 abaixo para um programa de promoção de atividade física.

Quadro 4. Exemplos de possíveis avaliações de Programas de Atividade Física

Tipo de avaliação	Oferta	Utilização	Cobertura	Impacto
Adequação	Mudanças no número de atividades disponíveis na comunidade	Mudança no número de pessoas participando de atividades	Medida da proporção dos indivíduos da população-alvo do programa participando das atividades	Medida da tendência de obesidade nas comunidades intervenção
Plausibilidade	Como acima, mas comparando comunidades intervenção com comunidades controle	Como acima, mas comparando comunidades intervenção com comunidades controle	Comparação da cobertura entre as comunidades intervenção e controle	Comparação da tendência de obesidade entre as comunidades intervenção e controle (ou dose-resposta)
Probabilidade	Como acima, mas as comunidades intervenção e controle devem ter sido randomizadas	Como acima, mas as comunidades intervenção e controle devem ter sido randomizadas	Como acima, mas com randomização prévia	Como acima, mas com randomização prévia

90

### Decidindo Quanto a Que Grau de Inferência Usar

Avaliações de adequação e de plausibilidade podem ser realizadas após o programa / intervenção haver sido lançado. As avaliações de adequação serão melhores se objetivos claros e exequíveis do programa tiverem sido preestabelecidos. As avaliações de plausibilidade, por sua vez, geralmente irão requerer informação de base do período pré-programático.

Avaliações complexas (por exemplo, com abordagem probabilística ou com avaliação de impacto) não devem ser realizadas antes de se demonstrar, através de avaliações mais simples, que o processo está ocorrendo na direção esperada. No caso do programa de atividade física do Quadro 4, por exemplo, dever-se-ia iniciar demonstrando que as atividades físicas estão sendo oferecidas nas comunidades e depois checar se a população está utilizando essas atividades. O terceiro estágio deveria incluir uma enquete domiciliar para avaliar se a cobertura do programa está sendo alcançada. Até aqui, as avaliações foram de adequação. Posteriormente, o tomador de decisão

poderá optar tanto por mostrar que a cobertura é maior nas áreas intervenção do que nas controle (uma afirmação de plausibilidade de que a maior cobertura é devida ao programa) ou talvez demonstrar uma redução da obesidade comparada ao período anterior à implementação do programa (uma avaliação de adequação).

A eficácia conhecida de um programa / intervenção é um fator importante na escolha do desenho da avaliação. Por exemplo, a eficácia da vacina contra o sarampo está bem documentada. Se uma avaliação de adequação mostrar que a cadeia de frio funciona bem e que a cobertura é alta, haverá pouca necessidade de avaliar o impacto dos programas de imunização sobre as taxas de sarampo ou mesmo sobre mudanças nas taxas de anticorpos contra o sarampo. Já no caso de programas para promover o consumo de fibras visando melhorar o controle da glicemia em pacientes diabéticos, a situação é bastante diferente. Como sua eficácia ainda não foi demonstrada, o aumento da ingestão não será suficiente para persuadir os financiadores sobre a utilidade dessa medida, sem que sejam utilizados fortes desenhos de plausibilidade. Nesse caso, mais desenhos probabilísticos serão necessários para persuadir os tomadores de decisão a implementar esse tipo de programa.

O momento em que a avaliação é planejada é fundamental. As avaliações probabilísticas, como visto, requerem que o avaliador esteja presente antes que o programa / intervenção inicie, de forma que as comunidades ou os indivíduos possam ser aleatoriamente alocados. Todos os métodos longitudinais, incluindo aqueles com grupo controle, requerem que sejam coletadas informações de base antes do programa / intervenção, ou que informações secundárias confiáveis para o período pré-programa estejam disponíveis. Em geral, as avaliações de oferta e utilização podem ser realizadas mais cedo e mais freqüentemente, uma vez que ajudam tomadores de decisão locais a melhorar as intervenções mais rapidamente do que esperando por resultados de longo prazo. Por outro lado, avaliações de cobertura e, particularmente, de impacto são em geral realizadas mais tarde.

Como regra geral, mesmo que a coleta de dados vá acontecer em uma fase mais tardia, as avaliações devem ser planejadas juntamente com o planejamento do próprio programa / intervenção.



**Exemplos de Estudos Utilizados  
para Avaliação de Acordo  
com o Grau de Inferência**

## Avaliação da Adequação do Desempenho (Processo) do Programa / Intervenção

Nesse grau de inferência causal, as atividades do programa / intervenção são medidas por profissionais que o implementam e dos usuários. Os resultados dessas medidas são comparados a critérios pré-definidos de adequação, com o objetivo de avaliar se as atividades estão sendo realizadas conforme planejado.

A análise de Fortmann SP *et al*, 1982, é um exemplo de avaliação longitudinal de adequação do desempenho do *Stanford Three Community Study*. A intervenção, uma campanha de educação alimentar em massa, bilíngüe (inglês e espanhol), era veiculada através de rádio, televisão, coluna de jornal escrita por médico ou nutricionista, correio (panfletos e livros de receitas), anúncios públicos e cartazes, visando atingir todos os segmentos da sociedade das duas cidades do grupo experimental. Com o objetivo de avaliar se o programa educacional era capaz de atingir não apenas os grupos socioeconômicos mais favorecidos (como geralmente ocorre) mas também os mais pobres, os autores analisaram o comportamento alimentar referido por uma amostra de participantes das cidades-alvo da intervenção (grupo experimental somente). Entrevistadores treinados aplicavam um questionário com 47 perguntas que permitiam estimar o consumo de colesterol, gorduras saturadas e insaturadas, açúcar refinado e álcool. Modelos de alimentos auxiliavam a avaliar o tamanho das porções consumidas. A mudança no consumo referido no ano dessa avaliação (1975), em relação à linha de base (pré-intervenção, em 1972), foi analisada para cada um de cinco estratos socioeconômicos da amostra. Essa foi uma análise interna do grupo intervenção que permitiu avaliar o processo da campanha entre os recipientes.

94

O desempenho de um programa de cuidados a pacientes diabéticos foi avaliado (Assunção MC *et al*, 2001; Assunção MC *et al*, 2002), na cidade de Pelotas, RS. Os autores avaliaram indicadores de oferta e de processo de cuidado aos pacientes, tendo como referencial as recomendações de organismos internacionais em manejo de diabetes mellitus.

## Avaliação da Adequação do Impacto do Programa ou Intervenção

Duas das seis comunidades participantes do *Minnesota Heart Health Program* foram incluídas no *Class of 1989 Study* (Kelder SH *et al*, 1994). Todos os alunos da sexta série das escolas públicas, em ambas as comunidades, foram convidados a

participar de uma avaliação de linha de base, em 1983, e anualmente até a formatura, em 1989. Os alunos da comunidade do grupo intervenção receberam dupla intervenção: a do próprio *Minnesota Heart Health Program* e a específica para os escolares. Os desfechos investigados foram mudança na prevalência e no padrão de tabagismo, mudança nas preferências alimentares, no escore de atividade física e em horas de exercício físico por semana. Na análise, os autores descreveram se os alunos do grupo intervenção e controle, separadamente, mantinham-se ou mudavam o tipo de comportamento que apresentavam na sexta série. Embora o estudo fosse um quase-experimento com grupo controle externo, a análise apresentada, por não controlar o efeito de fatores de confusão, caracteriza uma avaliação de adequação de impacto.

A intervenção *Coronary Health Improvement Project (CHIP)* conduzida em Kalamazoo, Michigan, incluiu 304 indivíduos com alto risco coronariano (70% estavam 10% ou mais acima do peso ideal, 14% eram diabéticos, 47% eram hipertensos e 32% tinham história de doença coronariana) (Diehl HA, 1998). O objetivo da intervenção era desenvolver e testar a efetividade de um modelo para programas de redução de risco de base comunitária. Nesse programa educacional de 40 horas, desenvolvido durante quatro semanas, os participantes, que pagavam para participar do estudo, eram encorajados, em reuniões diárias, a caminhar ou a fazer exercícios durante 30 minutos por dia e aderir a uma dieta (*Optimal Diet*) baseada em vegetais. Os fatores de risco foram avaliados para todos os participantes antes da intervenção e 30 e 120 dias após o início do programa. O impacto do programa foi medido através da mudança nas médias dos fatores de risco (peso, pressão arterial sistólica e diastólica, frequência cardíaca, colesterol plasmático, entre outros) antes e depois do programa, sem explorar a presença de confusão. A ausência de um grupo controle e o não controle de fatores de confusão permitiram que os autores avaliassem apenas se a meta de redução dos fatores de risco era factível com a metodologia empregada. O resultado observado superestima o efeito real em decorrência do efeito placebo e do efeito Hawthorne (Fletcher, RH, 1996). O efeito placebo é aquele percebido pelo indivíduo devido à expectativa positiva de que uma medicação ou tratamento tenha um efeito, isto é, pelo poder da sugestão. Pelo efeito Hawthorne, os indivíduos tendem a mudar o comportamento, para melhor, quando são alvos de interesse e atenção especial em um estudo, independentemente da natureza específica da intervenção que possam estar recebendo.

Uma avaliação conduzida na cidade de Pelotas, RS, analisou o resultado do cuidado a pacientes diabéticos em uma abordagem de adequação (Araújo RB *et al*, 1999).

Os pacientes diabéticos atendidos pelo menos uma vez nos cinco anos anteriores à avaliação, em uma unidade de atenção básica da cidade, foram visitados em casa, examinados e encaminhados para realização de exames laboratoriais complementares. Os resultados (glicemia e hemoglobina glicosilada) foram comparados com as recomendações da literatura.

O estudo de Assunção MC *et al*, 2001, foi uma avaliação da adequação de um programa de manejo do diabetes mellitus realizada na rede de postos de saúde de atenção básica na cidade de Pelotas, RS. Através de um estudo transversal, analisando somente os beneficiários do programa, os autores compararam a prevalência de pacientes com níveis glicêmicos aceitáveis em relação aos critérios recomendados pelo Ministério da Saúde do Brasil e da Associação Latino-Americana de Diabetes (ALAD).

### **Avaliação de Plausibilidade do Desempenho (Processo) do Programa ou Intervenção**

96

Stern MP *et al*, 1976, analisaram o comportamento alimentar de uma amostra de moradores das três cidades norte-americanas participantes do *Stanford Three Community Study*, dois anos após o início da intervenção. Os objetivos da campanha eram informar, persuadir e ensinar como reduzir o consumo de colesterol e gorduras saturadas e aumentar o de gorduras poliinsaturadas. Uma amostra de indivíduos com alto risco cardiovascular de somente uma das comunidades intervenção (Watsonville) foi aleatoriamente selecionada para receber a intervenção de forma intensiva. Utilizando um questionário com 47 questões, que permitiam estimar o consumo habitual diário de colesterol e gorduras, os autores compararam a mudança alimentar entre as comunidades intervenção (Watsonville e Gilroy), o grupo de alto risco de Watsonville com intervenção intensiva e o grupo controle (Tracy), antes e dois anos depois de iniciada a campanha.

Fortmann SP *et al*, em 1981, utilizando metodologia semelhante, analisaram a mudança de consumo referido nas três comunidades, no terceiro ano após o início da campanha.

## Avaliação de Plausibilidade do Impacto do Programa ou Intervenção

### Com Grupo Controle Histórico

Elmer PJ *et al*, 1995, avaliaram o efeito de um programa de mudança de estilo de vida aplicado a ambos os grupos (intervenção e controle) de um estudo randomizado, cujo objetivo era medir o efeito de cada um de cinco diferentes tipos de medicamentos anti-hipertensivos, contra um placebo. O programa educacional recomendava a redução do peso corporal, a diminuição do consumo de sal e bebidas alcoólicas e o aumento da atividade física. O efeito do programa educacional sobre a pressão arterial e os lipídeos sanguíneos foi medido no grupo placebo somente, caracterizando uma avaliação tipo antes-e-depois sem grupo controle.

Da mesma forma, o estudo de Diehl HA, 1998, citado anteriormente, avaliando o impacto do projeto *CHIP* sobre fatores de risco cardiovasculares, foi do tipo antes-e-depois, sem grupo controle.

A avaliação de impacto do *National Health Lifestyle Programme*, um programa de prevenção dos principais fatores de risco cardiovasculares, implementado pelo Ministério da Saúde, em 1992, em Singapura, foi uma avaliação longitudinal de plausibilidade. Os autores (Cutter J *et al*, 2001) utilizaram dados secundários de inquéritos nacionais de saúde, de base populacional, de 1992 (linha de base) e 1998. Foram medidas as prevalências de diabetes mellitus, hipertensão arterial, obesidade, tabagismo, sedentarismo e hipercolesterolemia, na população de 18-69 anos de idade. O programa incluía ampla utilização da mídia para promoção de estilos de vida saudáveis, medidas legais para coibir o tabagismo e programas de saúde em escolas, locais de trabalho e comunidades. Foram comparadas as prevalências padronizadas por idade e estratificadas por sexo e etnia, antes e depois de iniciado o programa.

97

### Com Grupo Controle Interno

A variação no colesterol plasmático em uma mesma comunidade (Watsonville) de acordo com graus diferentes de exposição à campanha, como investigado no estudo de Stern MP *et al*, 1976, no *Stanford Three Community Study*, é um exemplo de avaliação da plausibilidade de impacto do programa. Uma relação de dose-resposta entre a

intensidade da intervenção e o impacto permite maior plausibilidade do que o resultado da comparação apenas entre os grupos que receberam ou não a intervenção.

O estudo de Puska P *et al*, 1998, comparando os coeficientes de mortalidade por doença cardiovascular e câncer de pulmão ajustados para sexo e idade, entre North Karelia (área demonstrativa pioneira na implementação do programa) e a Finlândia como um todo, antes e depois da intervenção, é outro exemplo de avaliação de plausibilidade do impacto de uma intervenção. O tempo de exposição e a intensidade da intervenção em North Karelia era maior do que no restante do país (início em 1972 e 1977, respectivamente).

### Com Grupo Controle Externo

A avaliação do *Stanford Three Community Study*, publicada em 1981, por Fortmann SP *et al*, foi um avaliação quase-experimental do efeito de uma intervenção educacional sobre os hábitos alimentares nas duas comunidades, comparativamente à comunidade controle. A comunidade controle (Tracy) foi definida como tal antes da aplicação da intervenção nas duas comunidades experimentais.

98

A análise do impacto do *Minnesota Heart Health Program*, conduzida por Luepker RV *et al*, 1994, comparando transversalmente e longitudinalmente as mudanças nos fatores de risco cardiovasculares, entre a população-alvo dessa intervenção educacional (uma de cada um dos três pares de comunidades selecionadas) e a população das três comunidades controle, é um exemplo desse tipo de avaliação. As medidas de efeito foram ajustadas para idade, escolaridade e estado marital dos entrevistados.

A análise de Carleton RA *et al*, 1995, sobre a mudança nos fatores de risco cardiovasculares decorrente da intervenção educacional em Pawtucket (*The Pawtucket Heart Health Program*), comparativamente à observada em cidade controle, foi uma avaliação de plausibilidade com grupo controle externo. A cada dois anos, entre 1981 e 1993, eram realizados inquéritos de amostras populacionais das duas cidades e comparadas as prevalências e médias de fatores de risco antes, durante e após a intervenção, ajustadas para sexo, idade, escolaridade e país de nascimento.

O estudo de Young DR *et al*, em 1995, sobre o *Stanford Five-City Project*, comparou as comunidades intervenção e controle, de 1979 a 1980, quanto a médias em gasto

energético diário, número de atividades habituais e proporção de atividades vigorosas. Como um desfecho intermediário, foi também comparado o conhecimento dos dois grupos sobre exercício.

Em 1978, os alunos da sétima série (12 a 13 anos de idade) de North Karelia, Finlândia, foram alvo de um programa de prevenção do tabagismo (Vartiainen *et al*, 1998). O programa que teve a duração de dois anos, ocorreu na mesma área e durante a implementação do North Karelia *Project*, um programa de prevenção cardiovascular, de base comunitária, direcionado a adultos e iniciado em 1972. Três pares de escolas foram selecionados. Em cada par, uma escola localizava-se na capital da província e outra no interior. O primeiro par recebia o programa veiculado por educadores da área da saúde e o segundo, pelos próprios professores. O terceiro par foi selecionado de uma província vizinha e consistia no grupo controle. Em 1978, 1980, 1981, 1982, 1986 e 1993 (aos 28 anos de idade) os alunos responderam um questionário auto-aplicado e uma entrevista sobre fatores de risco cardiovascular. A prevalência de fumantes e a intensidade do hábito entre o grupo intervenção e controle foram comparadas nos diferentes inquéritos realizados. A existência de um grupo controle permitiu aos autores controlar o efeito da tendência secular e da intervenção mais ampla que ocorria no país como um todo.

99

Na intervenção randomizada do projeto *Active* (Dann *et al*, 1997), além das análises convencionais dos ensaios randomizados, os autores abdicaram do princípio da intenção-de-tratar e usaram abordagens de plausibilidade para aumentar a credibilidade dos resultados. Assim, uma análise apenas dos indivíduos que lograram atingir o critério estabelecido pelo *Centers for Disease Control and Prevention* (CDC) de atividade física (pelo menos três dias por semana, 20 minutos ou mais de cada vez), dos dois grupos, os autores investigaram a presença de estratégias comportamentais e cognitivas, entre outras, veiculadas pela intervenção. Conforme esperado, o uso dessas estratégias foi mais freqüente no grupo experimental.

Em uma avaliação do *Pawtucket Heart Health Program*, publicada por Gans *et al*, 1999, os autores exploraram desfechos intermediários do programa. Em um estudo transversal, uma amostra da população-alvo de Pawtucket e da área controle foi entrevistada e comparada quanto ao conhecimento sobre fatores de risco e sobre medidas preventivas cardiovasculares.

Uma intervenção comunitária e multicêntrica foi conduzida pelo *German Cardiovascular Prevention Study*, cujo objetivo era reduzir a prevalência e as mé-

dias populacionais de fatores de risco cardiovasculares relacionados com o estilo de vida (hipertensão arterial, tabagismo, hiperlipidemia e obesidade) (Scheuermann W *et al*, 2000). Nessa intervenção, indivíduos da comunidade e clínicos gerais de duas cidades alemãs eram estimulados a desenvolver, cooperativamente, atividades promotoras de um estilo de vida saudável. Três inquéritos populacionais entre indivíduos de 25 a 69 anos de idade (1985/86, 1988/1989 e 1991/1992), com amostras independentes, foram realizados para avaliar as mudanças ocorridas após o início do estudo. Uma amostra de indivíduos obtida a partir de inquéritos populacionais de toda a Alemanha Ocidental foi utilizada como controle (dados secundários). Os autores controlaram o efeito da tendência secular nas mudanças dos fatores de risco estudados através de um grupo controle externo ao estudo, caracterizando uma avaliação de plausibilidade do impacto da intervenção.

### **Avaliação Probabilística do Desempenho (Processo) do Programa ou Intervenção**

100

O projeto *Active* foi um ensaio clínico randomizado, conduzido entre adultos sedentários, cujo objetivo era comparar o efeito de uma intervenção para promover um estilo de vida voltado para atividade física (30 minutos de atividade física moderada na maioria dos dias da semana) e uma intervenção tradicional com exercícios estruturados (20-60 minutos por dia, por 3-5 dias/semana) sobre fatores de risco cardiovasculares (Dunn AL, 1997). Como avaliação do processo, os autores analisaram as estratégias psicológicas usadas em cada grupo para atingir o nível de atividade física recomendada pelo *Centers of Disease Control and Prevention* (CDC, EUA). Os participantes foram avaliados quanto à prevalência das dez estratégias normalmente usadas para progredir de um estado de baixa para alta motivação para mudar o comportamento em relação à atividade física. De acordo com a Teoria Cognitiva Social, as cinco estratégias cognitivas (aumentar conhecimento sobre atividade física, alerta sobre os riscos do sedentarismo à saúde, conseqüências desfavoráveis do sedentarismo à família e amigos, compreensão dos benefícios da atividade física e utilização das oportunidades para atividade física) e as cinco comportamentais (substituição de alternativas através da opção pela atividade física em momentos de cansaço ou stress, estimulando a busca de apoio de familiar ou amigo, premiando a si mesmo por ser ativo, comprometendo-se em ser ativo e utilizando formas de lembrar a si mesmo que precisa ser ativo - por exemplo, tendo sempre à mão um par de tênis) – foram comparadas entre os dois grupos. A observação de que o grupo que recebeu aconselhamento utilizava mais as estratégias cognitivas e

comportamentais do que o grupo dos exercícios estruturados é uma evidência probabilística de que a intervenção educacional estava atuando na forma desejada.

### **Avaliação Probabilística do Impacto do Programa ou Intervenção**

No estudo de Dann AL *et al*, 1997, descrito anteriormente, os grupos de aconselhamento (experimental) e de exercícios estruturados (controle) foram comparados quanto ao alcance do objetivo de reduzir os fatores de risco cardiovasculares. Tal como com a avaliação de processo, a alocação aleatória dos indivíduos a cada um dos grupos, após o emprego de critérios estritos de seleção, tornou os dois grupos teoricamente semelhantes entre si, exceto pelo tipo de intervenção que receberam. Os resultados obtidos a partir da randomização de um grupo homogêneo de participantes permitiu inferir que a diferença observada entre os dois grupos, em termos de atingir os critérios recomendados de atividade física, foi devida à própria intervenção.

O estudo randomizado de Korhonen M *et al*, 2003, no *North Karelia Project* comparou o efeito do aconselhamento nutricional intensivo com o aconselhamento habitual, sobre o controle da hipertensão arterial, um e dois anos após iniciada a intervenção, entre pacientes de atenção primária.



# **PARTE IV**

## **Planejando a Avaliação de Programas / Intervenções**

*Iná S. Santos*

A avaliação é um processo que procura determinar, tão sistematicamente e objetivamente quanto possível, a relevância, efetividade, eficiência ou impacto de atividades em relação a objetivos específicos pré-definidos (Last, 2001). Avaliar significa estimar o quanto um programa / intervenção logrou alcançar os objetivos a que se propunha, além de ser um instrumento para subsidiar a implementação de medidas que visem melhorar a qualidade das atividades atuais do programa / intervenção e o futuro planejamento e tomadas de decisão a ele relacionadas. Idealmente, a avaliação deveria fazer parte do planejamento dos programas / intervenções. Momentos para realizar avaliações sobre o processo e o impacto do programa / intervenção, bem como a alocação de recursos para conduzir a avaliação, deveriam ser planejados antes da implementação do mesmo. Lamentavelmente, na maioria das vezes, tal não é o caso e, como será discutido mais adiante, em decorrência disso, criam-se algumas limitações à avaliação.

104

Há várias maneiras de avaliar um programa. A melhor maneira, no entanto, é aquela que responde de forma apropriada a pergunta formulada por aqueles que, de posse dos resultados da avaliação, precisarão tomar decisões sobre o futuro ou a implementação do programa. Algumas perguntas, portanto, precisam ser respondidas antes de planejar uma avaliação. Qual o objetivo da avaliação? A avaliação pretende demonstrar que o programa / intervenção está atingindo seus objetivos ou que está funcionando conforme planejado? Quem solicita a avaliação e para tomar qual tipo de decisão? A decisão do que medir dependerá de quem utilizará os resultados e das decisões que precisarão ser tomadas em decorrência da avaliação. Assim, na avaliação de um programa nacional de promoção da atividade física, por exemplo, o coordenador das atividades de um bairro precisará saber apenas se o número de instrutores treinados será suficiente para atender a demanda da comunidade. Para o coordenador municipal poderá ser importante saber quantas pessoas estão inscritas e freqüentam as atividades do programa. Para a coordenação estadual, o mais importante poderá ser identificar as características do programa nos municípios que alcançaram maiores coberturas na população-alvo. E, finalmente, ao Ministério da Saúde, se o investimento está reduzindo a mortalidade por doenças cardiovasculares nas regiões beneficiadas pelo programa.

Além do objetivo da avaliação, o montante de recursos disponíveis é um outro fator que auxilia na definição do estudo necessário para a avaliação. Os custos de avaliar aumentam com a complexidade da avaliação. Estudos simples, conduzidos com informações geradas pelo próprio programa e envolvendo somente beneficiários são mais baratos. Avaliações que envolvam amostragem populacional e que necessitem

um grande número de participantes (como as avaliações de impacto) são mais caras. Da mesma forma, estudos que exijam controle rigoroso de cada passo da avaliação, além de algumas centenas de participantes, como os ensaios randomizados para demonstrar eficácia, são mais demorados e mais caros.

Quanto ao desenho do estudo, uma mesma pergunta poderá ser respondida utilizando-se metodologias diferentes, mais ou menos complexas. Idealmente, nenhuma intervenção ou programa deveria ser aplicado em escala populacional sem que sua eficácia houvesse sido previamente determinada. Tal não é o caso, no entanto, na maioria das intervenções realizadas em saúde pública. Determinar a eficácia de um programa ou intervenção significa testar se, em condições ideais de execução, o programa ou intervenção logra obter o efeito esperado. Condições ideais de execução significam que a avaliação planejou e implementou medidas que garantam que cada passo do programa / intervenção sejam realizados plenamente ou na intensidade mínima necessária para que o efeito esperado possa ser obtido ou verificado.

O melhor modelo de estudo, dentre os disponíveis até o momento, para demonstrar eficácia são os ensaios controlados randomizados, que geralmente são montados para testar medicamentos. A adesão dos participantes é um aspecto essencial nessa metodologia e os pesquisadores empregam substancial esforço na implementação de estratégias que promovam (termo de concordância assinado, contatos telefônicos e por correio, visitas em domicílio, pagamento de passagens, acesso facilitado etc) e demonstrem adesão ao tratamento (contagem de comprimidos, dosagens sanguíneas ou urinárias de marcadores biológicos etc).

No contexto dos ensaios clínicos com medicamentos, a intervenção é uma pílula ou uma vacina, ou seja, a “intervenção” cujo efeito se quer medir é veiculada compactamente, por exemplo, por via oral ou parenteral. O mecanismo de ação da intervenção é bioquímico e previamente testado em modelos laboratoriais e em cobaias. Em intervenções visando mudança de hábitos e de comportamentos, como é o caso dos fatores de risco para DCNTs, no entanto, a “pílula” é um pacote de medidas que podem incluir amplo leque de atividades, variando desde sensibilização dos potenciais beneficiários até reuniões sistemáticas de grupo, sessões de atividades físicas etc. Diferentemente dos ensaios com drogas, para ser veiculada, a intervenção depende de técnicas que trabalhem com estratégias cognitivas e comportamentais e seu mecanismo de ação segue uma rede causal complexa. Nessa realidade, os ensaios randomizados descaracterizam-se e precisam abrir espaço para outras metodologias de avaliação que melhor atendam as especificidades da saúde pública.

Embora os ensaios randomizados sejam considerados o padrão ouro em inferência causal, dentro da pesquisa epidemiológica, e de haverem qualificado a atenção à saúde individual, através do movimento da medicina baseada em evidências (do qual é uma das pedras fundamentais), a tentativa de replicar seus princípios na área de saúde pública tem encontrado inúmeras limitações.

Há atualmente uma corrente de epidemiologistas que, embora postulando rigor nas investigações, propõe a adaptação dos princípios dos ensaios randomizados ao contexto da vida real, quando se trata de avaliações na esfera da saúde pública, mesmo quando o objetivo é medir eficácia (Black N, 1996; Habicht JP *et al*, 1999; Rychetnik *et al*, 2002; McQueen DV, 2002; Victora CG *et al*, 2004).

Inúmeras situações previnem a realização de ensaios randomizados em saúde pública (Santos & Victora, 2004). Como regra, a rede causal ligando o programa / intervenção ao desfecho é tão longa e complexa que o tamanho de amostra necessário para a realização de um estudo randomizado é praticamente proibitivo. Além disso, as fontes possíveis de confusão entre o programa / intervenção e o desfecho são tão numerosas que, por mais cuidadosos que tenham sido os critérios de inclusão, é muito provável que diferenças de linha de base entre o grupo intervenção e controle sejam detectadas e necessitem de controle na análise, violando dessa forma o princípio da randomização.

106

A não-adesão, as perdas de acompanhamento e o cruzamento de indivíduos de um para outro grupo, todas características esperadas em estudos populacionais de larga escala, previnem que o princípio da análise por intenção-de-tratar seja aplicado rigorosamente. O mais razoável é que as análises por intenção-de-tratar sejam obrigatoriamente complementadas por análises de subgrupos ou de dose-resposta.

Dessa forma, idealmente, apenas novos programas / intervenções devem ser testados através de ensaios randomizados. Uma vez demonstrada a eficácia, no entanto, e o programa / intervenção aplicado em larga escala, a medida de impacto ou efetividade deve dar lugar a metodologias menos complexas e mais adequadas ao contexto da saúde pública.

Finalmente, convém enfatizar que nenhuma avaliação é perfeita. Todos os métodos de avaliação atualmente disponíveis têm limitações. O ato de avaliar, no entanto, traz vantagens que extrapolam o conhecimento dos resultados propriamente ditos. O resultado de qualquer avaliação, por mais simples, desde que rigorosamente

planejada e conduzida, produzirá um impacto sobre o programa / intervenção que será seguramente melhor do que não avaliar.

Nas próximas seções são descritas as etapas de planejamento de cada um dos tipos de estudos epidemiológicos especificamente voltados para avaliação de programas / intervenções de prevenção e controle de DCNTs. Inicialmente, são descritos os delineamentos baseados em dados primários, coletados pela própria avaliação e, posteriormente, os estudos oriundos de dados secundários.

Três etapas iniciais são comuns a todos os tipos de estudo e devem ser realizadas antes do planejamento metodológico propriamente dito: a revisão bibliográfica, o conhecimento detalhado do programa e a definição da cadeia causal que liga o programa / intervenção ao resultado esperado.

## ETAPA 1: Revisar a literatura

É importante aprender com a experiência de outros estudos. Outras intervenções, de mesma natureza ou de natureza semelhante, devem ser buscadas na literatura. O rastreamento de bancos informatizados de dados como *Pubmed*, *Lilacs*, *Cochrane*, entre outros, utilizando como descritores, por exemplo, a combinação de termos como “avaliação”, “intervenção comunitária”, “programa”, “<nome de programas>”, entre outros, e os seus correspondentes em inglês, dá acesso a inúmeras publicações avaliativas.

107

A seção de Material e Métodos dos artigos científicos localizados deve ser lida com especial atenção porque é nela que o autor relata como o estudo foi planejado e conduzido, podendo ser inspirador para o novo estudo. Além disso, a validade interna dos resultados obtidos pelas investigações depende da forma como os dados foram coletados e analisados. Uma maneira prática de analisar de forma sistemática a qualidade de estudos experimentais e observacionais é utilizar critérios de qualidade preestabelecidos, como os de Downs SH & Black N, 1998; Altman DG *et al*, 2001; e Des Jarlais DC *et al*, 2004.

## ETAPA 2: Conhecer o programa

Para que a avaliação seja bem planejada é necessário que os avaliadores, se forem externos, busquem conhecer o programa / intervenção (objetivos, metas, conteúdo, metodologia, instrumentos, atividades, logística etc) e que, se necessário, visitem os locais e as atividades promovidas pelo mesmo antes de desenhar a metodologia de avaliação. A participação dos coordenadores e implementadores do programa / intervenção é muito importante a partir dessa fase da avaliação. O envolvimento precoce de representantes do programa no processo de avaliação é importante não apenas para que detalhes não aparentes possam ser considerados, mas também para diminuir a tensão que uma avaliação, geralmente, provoca entre os gestores e implementadores, além de garantir comprometimento com as decisões que venham a ser tomadas em função dos resultados obtidos.

O conhecimento detalhado do programa auxiliará na definição do desenho mais adequado de avaliação e guiará a construção dos indicadores. Por exemplo, é improvável que um programa de atividade física, planejado e dirigido exclusivamente para indivíduos hipertensos usuários de unidades básicas de saúde, possa resultar em diminuição do sedentarismo em toda a comunidade. Depreende-se daí que a melhor estratégia de avaliação será estimar o efeito entre os usuários do programa, ao invés de na comunidade como um todo. Indicadores que meçam a oferta de material e situações promotoras da atividade física (*folders* educativos, equipamento de ginástica, áreas planejadas de lazer, instrutores treinados dentro das equipes, atividades programadas, entre outros), a aceitação e utilização dos hipertensos desses equipamentos e serviços, a cobertura obtida dessas ações entre a população hipertensa em tratamento nas unidades básicas de saúde e o benefício detectável em termos de condicionamento físico, redução da pressão arterial, diminuição do número e da dose de drogas anti-hipertensivas, são alguns dos exemplos de indicadores que poderiam ser gerados na avaliação de um programa dessa natureza.

108

## ETAPA 3: Identificar a rede causal que liga o programa / intervenção ao impacto esperado

Antes de serem implementados, os ensaios randomizados com medicamentos (também chamados de Ensaios de Fase III) são precedidos por ensaios de laboratório com cobaias e pelos ensaios de Fase I e II, com seres humanos. Nesses ensaios,

o mecanismo de ação do novo medicamento e a tolerância em seres humanos são demonstrados.

As intervenções na área comportamental humana também precisam estar apoiadas em concepções prévias de mecanismo de ação. Um programa desenvolvido na Alemanha (Scheuermann W *et al*, 2000), por exemplo, foi estruturado com base em uma teoria denominada “*Community Related Behavioural Medicine*” testada no *German Cardiovascular Prevention Study*” (GCP Study Group, 1988). O programa *Active* (Dunn AL *et al*, 1997) baseou-se na “*Social Cognitive Theory*” (Bandura A, 1986) e no “*Stages of Change Model*” (Prochaska JO *et al*, 1983). O *Minnesota Heart Health Program* (Murray DM, 1995) envolveu várias estratégias e teorias, incluindo a “*Social Cognitive Theory*” (Bandura A, 1986), “*Persuasive Communications Theory*” (Hovland CI *et al*, 1953; McGuire WJ, 1973) e modelos para envolvimento de líderes comunitários e instituições.

Além da teoria que fundamenta o programa / intervenção, é importante a construção de um modelo lógico que mostre o mecanismo provável de ação, através do qual o programa / intervenção deve operar para o alcance de seus objetivos. Se o programa / intervenção não tem (que é o que acontece na maioria das vezes), o avaliador deve construir, juntamente com os gestores e coordenadores do programa, um marco conceitual em que sejam identificados, de forma esquemática, os passos através dos quais se presume que o programa / intervenção deva alcançar o desfecho esperado. Com um modelo desse tipo, é possível especificar as etapas essenciais na implementação do programa, estabelecer o desenho da avaliação, as variáveis a serem medidas e a estratégia para coleta de dados, além de auxiliar na interpretação dos achados.

Essa materialização conceitual é muito importante para a avaliação do programa por duas razões. Primeiro, porque auxilia na definição das variáveis que deverão ser investigadas. Segundo, porque ajuda na distinção entre as variáveis mediadoras e variáveis confundidoras do efeito do programa / intervenção, o que é muito importante na fase de análise e interpretação dos resultados da avaliação (Victora CG *et al*, 2006).

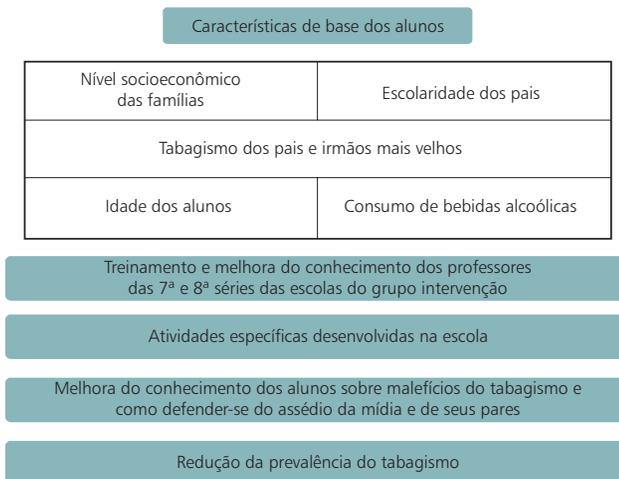
A Figura 1 é um exemplo simplificado de um marco conceitual que poderia ser construído para uma intervenção visando a redução do tabagismo entre jovens escolares. Com base nesse marco conceitual, as variáveis “treinamento e melhora do conhecimento dos professores sobre efeitos danosos do tabagismo à saúde”, “atividades específicas desenvolvidas pela escola” e “aumento do conhecimento dos estudantes nesse

assunto” seriam exemplos de mediadores entre a intervenção propriamente dita e o desfecho esperado (redução do tabagismo). As variáveis “escolaridade dos pais”, “tabagismo dos pais e irmãos mais velhos” e “idade dos alunos”, por exemplo, seriam potenciais modificadores do efeito da intervenção sobre o desfecho. Uma variável modificadora de efeito, por definição, é aquele em cujos níveis varia a magnitude de efeito de uma dada exposição (Rothman & Greenland, 1998).

O efeito de confusão também poderá ser identificado. Por definição, uma variável confundidora é aquela que se associa independentemente com o desfecho e com a intervenção que está sendo avaliada e não faz parte da cadeia pela qual a intervenção produz o desfecho (Rothman & Greenland, 1998). Algumas variáveis poderão ter um efeito confundidor positivo, outras, negativo. Se, concomitantemente, fossem desenvolvidas campanhas anti-tabágicas na comunidade geral (direcionadas à população como um todo), essas seriam potenciais fatores de confusão positivos, porque seu efeito, se presente, ocorreria no mesmo sentido do da intervenção com os escolares, superestimando seu efeito. As atividades específicas desenvolvidas na escola seriam fatores mediadores mais ou menos facilitadores, conforme o tipo, através dos quais a intervenção atingiria o efeito desejado.

110

Figura 1. Marco Conceitual da Rede Causal Ligando a Intervenção ao Desfecho



# **Avaliações com Dados Primários**

## Usando Estudos Transversais

Nas avaliações transversais, a amostra pode ser selecionada a partir de duas fontes diferentes: da população-alvo a que se dirige o programa / intervenção ou dos beneficiários. Em avaliações conduzidas entre a população-alvo, o *status* quanto à exposição (beneficiários e não beneficiários do programa / intervenção) e quanto à presença do desfecho (indivíduos apresentando ou não o desfecho de interesse) são obtidos em uma única entrevista e exame, ao final da intervenção ou após decorrido um determinado tempo de intervenção. A hipótese sendo testada é que, se a intervenção funcionar, a prevalência do desfecho (por exemplo, a prevalência de indivíduos fisicamente ativos) entre os expostos (beneficiários do programa) será maior do que a prevalência entre os não expostos. Se o programa / intervenção for protetor (por exemplo, promover a redução do tabagismo), a prevalência esperada de fumantes entre os participantes do programa será menor do que entre os não participantes.

112

Em seu desenho, análise e interpretação, as avaliações transversais conduzidas entre a população-alvo são semelhantes aos quase experimentos com grupo controle interno, exceto pelos critérios de seleção da amostra. Nos estudos transversais, a amostra da avaliação é selecionada a partir dos mesmos critérios que habilitam para a participação no programa / intervenção. Nos quase experimentos com grupo controle interno, os participantes da avaliação são selecionados com base na sua participação (grupo intervenção) ou não (grupo controle) no programa / intervenção.

Um exemplo de avaliação transversal, com amostras independentes da população-alvo, foi o estudo de Cutter *J et al*, 2001, que avaliaram o impacto sobre os fatores de risco cardiovasculares de um programa nacional de promoção de estilo de vida saudável, seis anos após sua implementação, em Singapura.

Avaliações transversais podem ser conduzidas entre apenas os beneficiários do programa. Nesse caso, o *status* dos participantes é comparado à meta estipulada pelo programa / intervenção ou baseada em parâmetros locais ou internacionais. Por exemplo, na avaliação de um programa de controle de pacientes diabéticos em unidades básicas de saúde, conduzida por Assunção MC *et al*, 2005, cerca de 50% dos diabéticos acompanhados apresentavam glicemia controlada. Tal achado estava em concordância com a regra das metades descrita na literatura que diz que, metade dos doentes crônicos sabem ser portadores da doença; desses, metade faz

o tratamento; e, dos que se tratam, metade está compensada (Wilber JA & Barrow JG, 1972; McKinlay J & Marceau L, 2000).

Uma outra hipótese que poderá ser testada em amostras constituídas apenas por beneficiários é se graus diferentes de exposição ao programa estão associados a diferentes intensidades de desfecho. Na Finlândia, o impacto do programa foi maior na área demonstrativa, onde as atividades do programa eram mais intensas, do que no restante do país.

**Vantagens das avaliações transversais:** a principal vantagem das avaliações transversais é sua simplicidade e conseqüente menor custo, comparativamente às outras avaliações. Avaliações transversais respondem perguntas simples, mas necessárias, antes que avaliações mais complexas (e mais caras) sejam conduzidas. Sua realização deveria ser prática rotineira dentro dos serviços de saúde ou de qualquer programa / intervenção que vise melhorar a saúde das populações humanas.

**Limitações das avaliações transversais:** as avaliações transversais devem ser realizadas sempre que o grau de inferência causal entre o programa / intervenção e os resultados obtidos não for uma questão crucial para o tomador de decisões. Se for suficiente saber o *status* da população-alvo ou dos participantes do programa / intervenção em relação ao desfecho, sem precisar atribuir necessariamente ao programa / intervenção esse resultado, então as avaliações transversais são suficientes e nenhuma desvantagem deve ser registrada. Se, por outro lado, houver necessidade de atribuir o efeito detectado ao programa / intervenção, a causalidade reversa será um potencial viés para essa inferência.

As etapas para o planejamento e condução de estudos transversais com vistas à avaliação de resultados de programas / intervenções estão descritas nos próximos parágrafos.

### **ETAPAS 1-3: proceder como especificado no início da PARTE IV, nas páginas 107 a 110.**

#### **ETAPA 4. Definir os objetivos da avaliação**

A avaliação poderá se limitar a medir o alcance dos objetivos do programa / intervenção (avaliação de adequação) ou expandir essa avaliação de forma a afastar o efeito de fatores externos sobre os resultados observados (avaliação de plausibilidade). Os resultados obtidos após controle de fatores de confusão serão mais plausíveis do que os de análises não ajustadas.

O estudo de Cutter J *et al*, 2001, objetivou analisar o impacto de um programa nacional de promoção de estilos de vida saudáveis sobre fatores de risco cardiovasculares em amostra representativa da população de Singapura.

O estudo de Assunção MC *et al*, 2005, foi planejado para avaliar a proporção de diabéticos com controle adequado da glicemia e identificar os fatores associados ao controle glicêmico entre os atendidos na rede de unidades básicas de saúde da cidade de Pelotas, RS.

114

#### **ETAPA 5. Definir os indicadores de desfecho, os fatores de confusão e os mediadores de efeito entre o programa / intervenção e o desfecho**

Definições operacionais do desfecho e das variáveis mediadoras e confundidoras deverão ser feitas *a priori*. Se for objetivo da avaliação descartar o efeito de fatores de confusão, os avaliadores deverão planejar a busca de informações sobre potenciais confundidores a partir da própria amostra e/ou de fontes externas.

A avaliação na Finlândia utilizou os códigos da Classificação Internacional de Doenças, nona revisão, para classificar as categorias do desfecho: doenças do sistema circulatório (390-459), doença isquêmica do coração (410-414) etc.

Os desfechos investigados na avaliação de Cutter et al, em Singapura, foram definidos conforme critérios diagnósticos recomendados pela OMS (WHO, 1994; WHO, 1996; WHO 1998a; WHO, 1998b), pelo Colégio Americano de Medicina Esportiva (American College of Sports Medicine, 1998) e pelo Instituto Nacional de Saúde

dos Estados Unidos (National Institutes of Health Publication, 1993). Hipertensão arterial, por exemplo, foi definida como pressão sistólica (média de duas aferições, por observadores treinados, em sala silenciosa, em repouso)  $\geq 140$  mmHg ou pressão diastólica (fase V de Korotkoff)  $\geq 90$  mmHg ou ambas ou uso atual auto-relatado de medicação anti-hipertensiva.

Os critérios de adequação do controle do diabetes mellitus empregados por Assunção MC et al, 2005, para definir o desfecho, foram os da Associação Latino-Americana de Diabetes (ALAD, 2000), da Associação Americana de Diabetes (ADA, 2004) e do Ministério da Saúde do Brasil (Ministério da Saúde, 1997), que estabeleceram valores de glicemia iguais ou inferiores a 10 mmol/L (180mg/dl), em sangue capilar, como indicativos de controle glicêmico “aceitável”. Informações para definição de variáveis confundidoras, modificadoras ou mediadoras entre o programa e o desfecho (idade, escolaridade, sexo, peso e altura corporais, tempo de diagnóstico, co-morbidade, atividade física, dieta, tabagismo, uso de hipoglicemiantes orais e de insulina, satisfação com o serviço, infra-estrutura das unidades básicas para o atendimento de diabéticos, especialidade do médico, presença de nutricionista na equipe, entre outras) foram também coletadas.

115

Se a avaliação incluir apenas beneficiários do programa / intervenção poderá ser importante coletar informações que indiquem o grau de exposição dos participantes às atividades oferecidas. A definição do ponto de corte para esses graus de exposição poderá se basear em critérios de adequação no âmbito clínico ou comportamental, estabelecidos por organismos reconhecidos. O grau de exposição poderá ser expresso através do tempo de participação nas atividades do programa / intervenção, número de sessões assistidas, número de adesivos de nicotina retirados e utilizados etc. A avaliação do programa de diabetes (Assunção MC et al, 2005) investigou o número de consultas médicas nos últimos seis meses (uma, duas e três ou mais) e a correspondente prevalência de controle da glicemia.

Independentemente do objetivo da avaliação e do tipo de estudo, sempre que possível, planejar a coleta de informações com metodologia tanto quantitativa quanto qualitativa. Com a colaboração de profissionais da área qualitativa (antropólogos, psicólogos, sociólogos, pedagogos), a metodologia qualitativa (como observação estruturada, *exit interviews*, grupos focais, entrevistas com informantes chaves, por exemplo) poderá ser usada. Informações obtidas através de metodologia qualitativa complementam os achados quantitativos. A metodologia qualitativa facilita a compreensão dos motivos de eventual fracasso ou sucesso dos programas / inter-

venções, uma vez que identifica com maior precisão falhas presentes na implementação e utilização dos mesmos.

## ETAPA 6. Desenhar o estudo

Uma amostra de indivíduos beneficiários do programa / intervenção ou uma amostra da população onde o mesmo se desenvolve será selecionada e, em um único contato, a exposição (ou o grau de exposição ao programa) e a presença do desfecho serão coletadas.

Essa metodologia de avaliação é planejada para ter uma única medida de prevalência do desfecho, por exemplo, proporção de fumantes entre beneficiários e não beneficiários do programa ou proporção de fumantes conforme o tempo de participação no programa ou, simplesmente, proporção de fumantes na população-alvo, independentemente da participação no programa. Uma única medida transversal na população-alvo ou no grupo intervenção é suficiente e as prevalências dos desfechos ou o efeito dose-resposta medido.

116

As Figuras 2, 3 e 4 mostram a estrutura simplificada de avaliações transversais realizadas, respectivamente, com amostra da população-alvo do programa / intervenção (Figura 2, como, por exemplo, o estudo de Cutter J *et al*, 2001), apenas com beneficiários (Figura 3, como, por exemplo, o estudo de Assunção MC *et al*, 2005) e entre beneficiários com análise tipo dose-resposta (Figura 4, como, por exemplo, o estudo de Puska P *et al*, 1998).

O planejamento do estudo deve tentar prevenir os principais tipos de vieses que costumam afetar esse tipo de delineamento, entre eles a taxa de perdas e recusas. Essas, preferentemente, não devem ultrapassar os 10%, sob pena de ameaçarem a validade interna da avaliação. O treinamento cuidadoso dos entrevistadores em técnicas de entrevista e em abordagem dos respondentes, o planejamento de, no mínimo, três revisitas, em dias e horários diferentes, são estratégias a serem adotadas para prevenir esse potencial viés.

A literatura epidemiológica em estudos etiológicos costuma referir a causalidade reversa como um tipo de viés que pode afetar os resultados dos estudos transversais. Pela natureza do desenho transversal em que, em uma mesma ocasião, são medidos a exposição e o desfecho, a temporalidade entre ambos não pode ser ga-

rantida. A temporalidade é um critério imprescindível de causalidade, afirmando que a causa, para ser tal, deve, necessariamente, preceder temporalmente o efeito (critérios de causalidade de Bradford Hill)(Rothman & Greenland, 1998). Esse é o caso quando o objetivo do estudo é investigar “causalidade” entre o programa / intervenção e o desfecho. Na maioria das situações de avaliação em que o desenho selecionado foi o transversal, como nos estudos de Puska P *et al* e de Cutter J *et al*, acima mencionados, o objetivo não era responder se o desfecho foi “causado” pelo programa / intervenção, mas sim, e simplesmente, se o efeito esperado foi alcançado, não se aplicando questionar a presença do viés de causalidade reversa.

No estudo de Assunção MC *et al*, por outro lado, características internas do programa (indicadores de infra-estrutura e recursos humanos dos serviços) e características individuais dos pacientes foram ajustadas umas para as outras, visando identificar os fatores independentemente associados ao resultado encontrado. Nessa análise, entre os diabéticos que consultaram um número maior de vezes (três ou mais), nos seis meses anteriores à entrevista, a proporção com hiperglicemia foi maior do que entre os que consultaram apenas uma vez, ficando definida uma associação inversa entre a intensidade de exposição ao programa (medida através do número de consultas) e o desfecho. Nesse caso, o viés de causalidade reversa, provavelmente, esteve presente. Não seria plausível supor que os pacientes apresentavam pobre controle glicêmico *porque* consultaram mais vezes. O mais provável é que pacientes com hiperglicemia tornam-se mais sintomáticos, necessitando, portanto, de cuidados mais frequentes.

117

## ETAPA 7: Definir os critérios para seleção dos participantes

Os critérios de seleção para a avaliação serão os mesmos que habilitam os indivíduos a participar do programa, (amostras de beneficiários somente ou amostras da população-alvo). O estudo de Cutter J *et al*, em Singapura, tinha como população de referência todos os residentes em Singapura com idade entre 18 e 69 anos.

No estudo de Assunção MC *et al*, eram candidatos a participar da avaliação os diabéticos que houvessem consultado nos sete dias anteriores à visita dos entrevistadores às unidades de saúde, em duas semanas não consecutivas, de acordo com registros do serviço. O uso de medicamentos hipoglicemiantes orais ou insulina e a confirmação do paciente de ter recebido o diagnóstico de diabetes por um médico foram critérios para assegurar que os entrevistados apresentavam a doença. Dos 461 pacientes identificados, 44 não se sabiam diabéticos. Seus registros foram

revisados na unidade de saúde, sua condição de não-diabéticos confirmada sendo, conseqüentemente, excluídos do estudo.

## ETAPA 8: Escolher o processo de amostragem

Se a avaliação pretender medir um efeito populacional, os participantes deverão ser selecionados, idealmente, por processo aleatório simples. Quando possível, membros da população-alvo do estudo são listados e, por sorteio simples, um número suficiente é selecionado para compor o estudo. Porém, geralmente não é esse o caso, porque essa listagem não existe para a maioria das comunidades, principalmente nos países em desenvolvimento. Há duas alternativas: fazer um censo populacional (se a população total não for muito grande e relativamente concentrada em uma área geográfica) ou utilizar outras abordagens de amostragem, que é o que geralmente é feito. Estratos populacionais identificados por outras fontes, com outros objetivos, poderão ser usados para a seleção dos participantes. Amostragem por múltiplos estágios, iniciando por conglomerados, tendo como base os setores censitários do Instituto Brasileiro de Geografia e Estatística (IBGE), seguida por amostragem aleatória simples, são estratégias bastante utilizadas em estudos populacionais que visam amostras representativas. Em vários estudos transversais realizados em cidades de médio porte, essa metodologia tem-se mostrado factível (Hallal PC *et al*, 2003; Domingues MR *et al*, 2004; Hallal PC *et al*, 2005; Dias-da Costa JS *et al*, 2005, entre outros).

118

Figura 2. Estrutura simplificada de avaliação transversal com amostra da população-alvo do programa

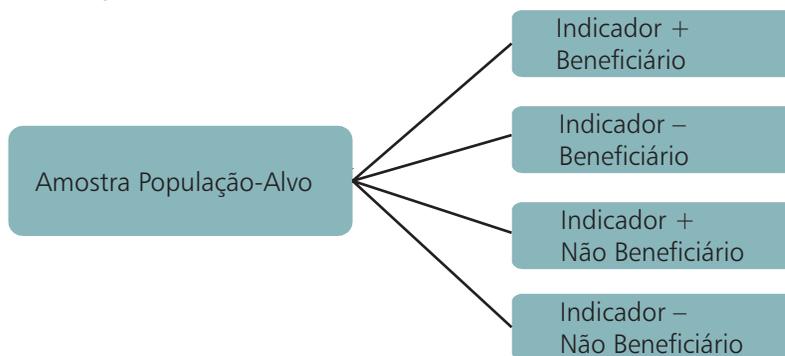


Figura 3. Estrutura simplificada de avaliação transversal entre usuários do programa (estudo descritivo)

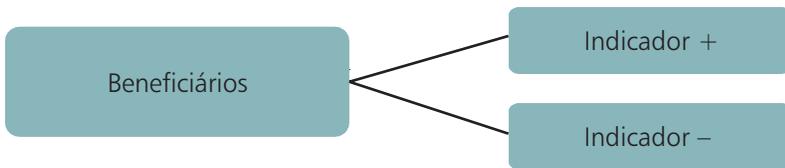
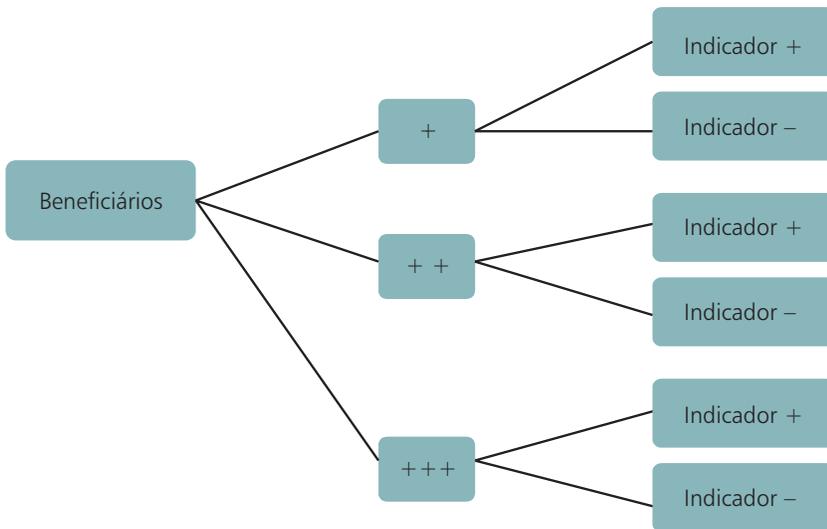


Figura 4. Estrutura simplificada de avaliação transversal entre usuários do programa com análise tipo dose-resposta



Na avaliação em Singapura, seis centros geográficos da ilha foram selecionados e os participantes localizados em duas fases. Inicialmente, as áreas em torno dos seis centros foram divididas conforme o tipo de moradia, como um *proxis* para nível socioeconômico, e selecionadas sistematicamente. Na primeira fase, utilizando uma base de dados nacional, os autores sortearam 11.200 endereços dentro das seis áreas. Na segunda fase, entre os moradores das residências incluídas na primeira, uma amostra aleatória de 7.500 pessoas de 18-69 anos de idade foi selecionada. A seleção dos indivíduos foi estratificada de forma a conter número representativo de todos os grupos étnicos residentes no país (chineses, malásios e indianos).

## ETAPA 9: Definir a unidade amostral e calcular o tamanho da amostra

Nas avaliações transversais, quanto nos demais desenhos de estudo, a unidade amostral poderá ser indivíduos ou grupos de indivíduos (salas de aula, escolas, locais de trabalho, bairros etc). Embora a unidade de estudo e observação sejam os indivíduos, o uso de conglomerados como unidade amostral tem fortes implicações no poder do estudo e na metodologia de análise dos dados. Mesmo que sejam incluídas todas as, digamos, 300 crianças de cada uma de duas escolas, totalizando 600 crianças, o uso de escolas como unidades amostrais implica uma alta correlação entre as unidades de observação (correlação intraclasse), uma vez que alunos de uma mesma escola são mais semelhantes entre si em relação a uma série de outras características, também importantes para o desfecho de interesse, do que os de outra escola. O limitado número de graus de liberdade para estimar essa correlação implica que, ignorar o agrupamento da unidade amostral na fase de análise, conduzirá a aumento espúrio do poder do estudo (Cornfield J, 1978; Murray DM, 1995; Murray DM, 1998).

120

No estudo de Assunção MC *et al*, por exemplo, a unidade amostral foram as unidades básicas de saúde da cidade e a unidade de estudo foram os pacientes diabéticos consultando nesses serviços no período de duas semanas.

Os parâmetros requeridos para o cálculo do tamanho amostral dependem se foram previstas análises de associação ou não. Se análises de associação foram planejadas, os parâmetros utilizados são os mesmos requeridos para os demais estudos transversais. Em geral são estabelecidos como aceitáveis erro alfa de 5% e erro beta de 20% (poder de 80%). Os outros parâmetros incluem a razão expostos:não expostos, a prevalência do desfecho entre não expostos e o tamanho da razão de prevalências que se quer ser capaz de estimar. A razão expostos:não expostos pode ser estimada a partir dos registros do próprio programa / intervenção, se existirem, ou através de um estudo piloto que estime a cobertura do programa na população-alvo. A prevalência do desfecho entre não expostos poderá ser obtida de outros estudos populacionais realizados na mesma população ou estimado a partir de estudos em populações semelhantes ou através de estudo-piloto. Quanto menor a razão de prevalências estatisticamente significativa que se desejar ser capaz de estimar, maior será o tamanho amostral requerido. Um acréscimo de 10% para corrigir as perdas e recusas e, sobre esse, de 15% para controle de fatores de confusão, é geralmente requerido.

Se o objetivo do estudo for descritivo (análises de adequação), o tamanho da amostra deverá ser calculado tendo como parâmetros o erro alfa (em geral de 5%), a prevalência do desfecho que se quer verificar (tendo como base, por exemplo, a meta do programa) e o tamanho do erro (em pontos percentuais) que se aceita incorrer nessa estimativa. O cálculo do tamanho da amostra para a avaliação em Singapura, por exemplo, estimou que seriam necessários 5.000 respondentes para detectar uma queda na prevalência dos fatores de risco de 10-15%, em comparação aos achados em estudo realizado seis anos antes da implementação da intervenção. Para localizar esse número de respondentes, seria necessário visitar 10.000 domicílios.

Independentemente do objetivo do estudo, se a amostra populacional for obtida por conglomerados (domicílios, escolas, postos de saúde), é aconselhável usar um fator de correção para lidar com o efeito de delineamento. O efeito de delineamento é decorrente da variabilidade da prevalência que está sendo medida dentro e entre os aglomerados estudados. Quanto mais próximas as prevalências entre os aglomerados, menor o efeito de delineamento. Se o efeito de delineamento não for conhecido para a população estudada (geralmente não é), uma dica prática para esse cálculo é multiplicar por dois o tamanho da amostra obtida tendo como pressuposto a amostragem aleatória simples.

121

No caso de estudos realizados através de visitas domiciliares, o cálculo do número de domicílios a serem visitados pode ser obtido dividindo-se o tamanho da amostra requerido (no numerador) pelo produto (denominador) da multiplicação entre número de pessoas por domicílio e a proporção de pessoas na população na faixa etária desejada. Os fatores do denominador poderão ser obtidos de dados censitários da região a ser estudada.

## **ETAPA 10: Planejar a logística do estudo**

Como a avaliação será implementada no campo? Como os entrevistadores se deslocarão para localizar os participantes da avaliação? Onde serão feitas as entrevistas? Como e onde serão coletadas amostras biológicas? Esses e outros são passos do estudo que precisam ser planejados com antecipação. Aspectos práticos do trabalho, desde contatos e autorização de autoridades locais e líderes comunitários, antes e durante o campo, distribuição e coleta de material, controle de qualidade, entre outros, precisam ser planejados com antecedência. Como regra, vale a pena gastar todo o tempo necessário na fase de planejamento e com isso antecipar os problemas

que poderão surgir ao longo do estudo. Embora essa previsão nunca seja completa (a vida real é cheia de surpresas!), quanto mais energia for despendida na fase de planejamento, menor a probabilidade de surpresas desagradáveis ao longo do estudo.

### **ETAPA 11: Estimar os custos e o tempo necessário para fazer a avaliação**

Listar todo o material necessário (questionários, formulários, balanças, antropômetros, esfigmomanômetros, vidraria de laboratório, anticoagulantes, agulhas, seringas etc) para a coleta de dados. Definir o valor a ser pago a cada entrevista completa. Considerar gastos com deslocamento da equipe de campo (vales transporte, gasolina ou motorista) e com a comunicação dos entrevistadores com a equipe central do estudo (créditos pré-pagos de telefone celular ou cartão telefônico ou chamadas a cobrar). Custos com coletadores de sangue, digitadores e equipe de controle de qualidade devem ser somados aos itens anteriores. Gastos com treinamento da equipe devem também ser computados entre as despesas.

122

O tempo para a realização da avaliação também precisa ser planejado e, preferentemente, respeitado. As avaliações são feitas para subsidiar decisões e, portanto, devem ser fornecidas com a maior brevidade possível. A velocidade do trabalho depende da complexidade dos dados coletados, do número de trabalhadores de campo e da distância a ser percorrida para localização dos entrevistados. O volume de recursos disponíveis poderá modificar essa velocidade para mais ou para menos.

### **ETAPA 12: Construir e reunir os instrumentos para a coleta de dados e controle de qualidade**

Boa parte da informação será coletada na forma de questionários para serem, em geral, aplicados por entrevistadores treinados. A construção do questionário requer um trabalho cuidadoso e vários testes antes que possa ser aplicado no estudo. As perguntas devem ser claras, objetivas, cada uma contendo um único assunto, formuladas de acordo com a realidade cultural e a escolaridade da população na qual vão ser aplicadas e, preferentemente, pré-codificadas, evitando-se perguntas abertas. Para a pré-codificação, é necessário testar as perguntas em pequenas amostras, antes de considerá-las prontas para inclusão no questionário.

Um manual de instruções especificando o objetivo de cada pergunta, bem como os procedimentos em possíveis situações ou respostas, deve ser elaborado para acompanhar os entrevistadores durante todo o trabalho de campo.

Instrumentos para coleta de informações qualitativas como *checklists* para observações e entrevistas para *exit interviews* devem ser elaboradas com o auxílio dos profissionais da área.

O uso de procedimentos padronizados por organismos internacionais para mensuração de parâmetros biológicos como pressão arterial, peso, altura, pregas cutâneas é altamente recomendado. A padronização repetida das medidas entre os trabalhadores de campo é a melhor estratégia para garantir a qualidade da coleta de dados e a comparabilidade de achados de um mesmo estudo ou de estudos diferentes.

Um aspecto importante a considerar é que, quando o objetivo é comparar prevalências obtidas de estudos realizados no mesmo local, em períodos diferentes, ou em locais diferentes, a padronização dos instrumentos e procedimentos para coleta de dados é crucial. No estudo de Cutter J *et al*, em 1998, em Singapura, por exemplo, foi seguido o mesmo protocolo usado no estudo realizado em 1992, o qual foi baseado no modelo da OMS para trabalhos de campo sobre diabetes e outras DCNTs (Dowse G & Zimmet P, 1992) e no protocolo Monica para amostras populacionais (WHO, 1990)

123

### **ETAPA 13: Treinar a equipe de campo e testar a logística e os instrumentos**

O tempo investido no treinamento exaustivo da equipe e na realização de estudo-piloto, que teste a logística e os instrumentos, será revertido em qualidade dos dados coletados. Várias técnicas podem ser usadas para o treinamento dos trabalhadores de campo: leitura passo a passo do questionário e do manual de instruções, simulações de entrevistas entre os próprios treinandos, entrevistas com pessoas em salas de espera etc. Uma vez treinados e selecionados os entrevistadores com desempenho ótimo e entendido que a avaliação está pronta para ser implementada, é importante realizar um piloto do estudo. Comparativamente ao campo das artes cênicas, o estudo piloto é o ensaio geral (com guarda-roupa e maquiagem) que precede a estréia da peça teatral e onde tudo deve funcionar perfeitamente. A escolha do grupo de indivíduos que serão entrevistados no estudo piloto depende da con-

veniência do pesquisador. Preferentemente, os incluídos devem ter características semelhantes aos que serão efetivamente avaliados e não deverão ser incluídos na avaliação propriamente dita. Por exemplo, se uma intervenção foi planejada para atingir bairros periféricos de uma grande cidade e a avaliação para ser implementada em parte desses bairros, então o estudo-piloto deverá ser conduzido entre moradores de bairros não selecionados para a avaliação.

## **ETAPA 14: Iniciar o trabalho de campo**

Como com os demais desenhos, o monitoramento do trabalho da equipe de campo deve ser permanente. O contato facilitado dos entrevistadores com o coordenador de campo e reuniões semanais de equipe são imprescindíveis para o bom andamento do estudo. As reuniões de equipe têm com principal objetivo “afinar” a coleta de dados e manter a qualidade de trabalho do pessoal de campo. O coordenador de campo sana dúvidas que surgirem no decorrer do trabalho, revisa os questionários preenchidos pelos entrevistadores, aponta inconsistências no preenchimento, pede esclarecimentos ou solicita que o entrevistador retorne ao entrevistado para clarear dúvidas, sempre que necessário. Se o estudo vai se manter por vários meses no campo, novas sessões de treinamento para padronização de medidas devem ser planejadas e implementadas ao longo do tempo.

124

## **ETAPA 15: Digitar os dados em um programa de computador, proceder a limpeza e edição dos dados**

Uma opção bastante usada é armazenar o banco de dados em computador através da criação de estrutura com a função *Enter Data* do programa Epi Info (Epi Info, 2005). Duas digitações, geralmente por dois digitadores diferentes, são realizadas. A dupla digitação e comparação dos bancos permite detecção e correção de inconsistência. A estrutura do banco de dados pode ser programada definindo limites possíveis para as respostas, consistência entre respostas e pulos entre campos, dependendo de códigos preestabelecidos.

Como dados numéricos são coletados, em geral, em forma contínua, muitas vezes é necessário editá-los de uma outra forma, tanto para fins de análise como para facilitar a apresentação e a compreensão dos mesmos. É sempre preferível coletar informações de variáveis contínuas de forma não agrupada (por exemplo, idade em

anos completos, escolaridade em anos escolares com aprovação, número de cigarros fumados por dia, número de porções de legumes / verduras por semana) e só posteriormente agrupar essa informação em uma nova variável (idade em grupos de 5 ou 10 anos, escolaridade em graus, número de cigarros em maços e  $\geq 5$  porções de verduras / legumes por dia). A vantagem de coletar informações de variáveis contínuas de forma não agrupada decorre do fato de ser possível dessa forma calcular médias e desvios padrão dessas variáveis. Além disso, em termos de análise, medidas contínuas em geral conferem maior poder de detectar associações que de fato existam, comparativamente aos seus correspondentes em medidas categóricas (sim ou não) (Altman DG, 1991).

Procedimentos detalhados para digitação, comparação, limpeza e edição dos dados, bem como sobre o planejamento da logística de inquéritos de base populacional, podem ser obtidos no livro de Barros FC & Victora CG, 1998.

## **ETAPA 16: Analisar e interpretar os resultados**

Análises descritivas de avaliações transversais consistem em descrever os indivíduos que ingressaram na avaliação e em apresentar as medidas de tendência central e de dispersão (desfechos contínuos) ou a prevalência (desfechos categóricos) dos indicadores, comparando-as ao previamente estipulado como adequado pelo programa / intervenção ou a parâmetros nacionais ou internacionais recomendados. Não haverá, nesse tipo de avaliação, controle para fatores de confusão, uma vez que não estará caracterizada a presença de um grupo controle.

125

Se o estudo for feito com uma amostra da população-alvo, a análise e apresentação dos dados consiste inicialmente em descrever a distribuição da amostra e a prevalência de cada um dos desfechos conforme as variáveis independentes estudadas. A participação no programa / intervenção (a exposição de interesse) será uma das variáveis independentes analisadas e apresentadas. A associação entre o programa e as demais variáveis independentes com o desfecho poderá ser estimada através de testes *t* de Student, para desfechos contínuos ou qui-quadrado, para desfechos categóricos. A força da associação e o controle do efeito de variáveis de confusão poderão ser obtidos através de análise de regressão linear múltipla, para desfechos contínuos, ou regressão logística (desfechos dicotômicos com baixa prevalência na amostra estudada) ou de Poisson, para altas prevalências (Barros AJ & Hirakata VN, 2003).

Como já mencionado, em estudos cuja unidade amostral foi um grupo de pessoas, a análise deverá levar em conta esse agrupamento. Uma estratégia apropriada é a análise por modelos de multiníveis (Goldstein H, 2003), com aplicação de modelos lineares hierárquicos. Tipicamente, os programas / intervenções comunitárias têm pelo menos dois níveis de variáveis: variáveis do nível individual, que compreendem as características individuais do respondente, e variáveis do nível da comunidade, que descrevem as características da unidade amostral (bairros, comunidades, escolas, serviços de saúde etc).

No estudo de Assunção MC *et al*, 2005, por exemplo, em que as unidades amostrais foram as unidades básicas de saúde, os dados foram analisados por regressão logística, com modelagem de multiníveis, de acordo com um modelo hierárquico de determinação da glicemia. O nível hierárquico mais alto eram as unidades de saúde, o segundo, os médicos e, mais caudalmente, as variáveis do paciente.

126

Se a avaliação for conduzida apenas entre beneficiários do programa, a análise poderá ser do tipo dose-resposta, com controle para fatores de confusão, se necessário (geralmente, é). A magnitude da associação entre os indicadores de desfecho entre os participantes fortemente expostos à intervenção e os moderadamente expostos poderá ser comparada à observada entre os fracamente expostos tomados como controles. A existência de grupo controle não contra-indica análise tipo dose-resposta. Nesse caso, indivíduos com diferentes graus de exposição são comparados aos não expostos.

As avaliações transversais descritivas, tanto de processo quanto de impacto, são as mais simples e econômicas de serem conduzidas. Em princípio, nenhuma avaliação mais complexa deveria ser realizada antes que as mais simples tenham demonstrado indicadores favoráveis.

**Nota:**

A metodologia de avaliação e monitoramento de ações de prevenção e controle de DCNTs proposta pelo Cindi (WHO, 1996) enquadra-se nesse tipo de desenho com amostras populacionais independentes. A avaliação e o monitoramento Cindi têm como objetivo comparar indicadores e tendências (mudanças) entre os indicadores (fatores de risco biológicos e comportamentais) de diferentes países (dados nacionais) e entre os indicadores de cada país com os das áreas demonstrativas, ao

longo do tempo, através de estudos transversais. Conforme anteriormente mencionado, para que essa comparabilidade seja garantida, a metodologia de avaliação e os indicadores utilizados precisam ser padronizados. O Cindi recomenda que a amostra de cada avaliação deverá ser independente uma da outra e selecionada, preferencialmente, de forma aleatória simples entre os residentes na área de estudo. A amostragem aleatória simples é importante para garantir os parâmetros pré-definidos para o cálculo do tamanho amostral (mínimo de 200 homens e 200 mulheres de cada uma das seguintes faixas etárias: 15-24, 25-34, 35-44, 45-54 e 55-65, totalizando 2000 pessoas). A avaliação dos fatores de risco (avaliação de impacto) deverá ser realizada no início do programa e após, a cada cinco anos, no mesmo período do ano, para prevenir variações sazonais de fatores comportamentais, como alimentação, por exemplo.

No Cindi, as avaliações de desempenho (processo) são recomendadas para as áreas demonstrativas, embora possam também ser realizadas nacionalmente. Informações sobre implementação do programa, exposição à intervenção e processo de mudança do comportamento de risco deverão ser obtidas através de estudo transversal envolvendo pelo menos 800 pessoas (400 homens e 400 mulheres).

127

## Usando Estudos Quase-experimentais

Nas avaliações localizadas pela revisão bibliográfica e apresentadas na Parte I do atual trabalho, a metodologia quase-experimental foi a mais empregada. Na maioria das vezes, o programa / intervenção está em curso há algum tempo, quando os avaliadores são chamados para analisar se está sendo desenvolvido como planejado e alcançando os objetivos pretendidos. Outras vezes, o programa / intervenção foi originalmente planejado para contar com a participação de um grupo controle externo, sendo programadas avaliações quase-experimentais.

Os estudos quase-experimentais assemelham-se aos estudos randomizados pela presença de um grupo controle e por seu caráter longitudinal (acompanhamento dos grupos intervenção e controle ao longo do tempo). Distinguem-se dos ensaios randomizados porque a alocação dos grupos intervenção e controle é feita de forma não aleatória.

**Vantagens das avaliações quase experimentais:** a principal vantagem das avaliações quase-experimentais, comparativamente às avaliações transversais, é decor-

rente da presença de um grupo controle, que permite afastar o efeito de tendências seculares, do fenômeno de regressão à média e outros vieses, que podem comprometer a validade dos resultados encontrados.

**Limitações das avaliações quase-experimentais:** a principal desvantagem é a dificuldade em afastar o efeito de variáveis não medidas, o que, acrescido da ausência de randomização para alocação dos grupos intervenção e controle, implica o risco de atribuir-se ao programa / intervenção um sucesso ou fracasso imerecidos. Para aumentar a validade interna desse tipo de avaliação, pode ser necessário utilizar simulações matemáticas complexas que estimem a melhor / pior situação dos indicadores na população-alvo, caso o programa / intervenção não tivesse acontecido.

O planejamento de avaliações quase-experimentais pode variar, fundamentalmente, em função do tipo de grupo controle que será utilizado. A seguir são apresentadas as etapas do planejamento de avaliações quase-experimentais.

### **ETAPAS 1-3: proceder como especificado no início da PARTE IV, nas páginas 107 a 110.**

128

#### **ETAPA 4. Definir os objetivos da avaliação**

Uma vez revisada a literatura, conhecido o programa / intervenção e definida a rede causal supostamente responsável pelo efeito esperado, é necessário definir os objetivos da avaliação.

A formulação dos objetivos será decisiva para o desenho da avaliação. Os objetivos poderão ser de três tipos:

- a) medir a diferença entre o grupo intervenção e o controle, após decorrido um determinado tempo de programa na área intervenção, (como, por exemplo, a avaliação do *Minnesota Heart Health Program*, publicada por Luepker RV *et al*, 1994; a avaliação de Winkleby MA *et al*, 1993, do *Stanford Five-City Project*; e a de Jousilahti P *et al*, 1994, no *North Karelia Project*); ou
- b) comparar a mudança verificada ao longo do tempo no grupo intervenção com a ocorrida no grupo controle (como o estudo de Fortmann SP *et al*, 1981, do

- Stanford Three Community Study*; de Stern MP *et al*, 1976 e de Carleton RA *et al*, 1995, do *Pawtucket Heart Health Program*); ou
- c) medir a mudança ao longo do tempo no grupo intervenção apenas, tomado como seu próprio controle (Diehl HA, 1998, no *Coronary Health Improvement Project*; Elmer PJ *et al*, 1995, no *Treatment of Mild Hypertension Study*; e os estudos de Nothwehr F *et al*, 1994, e de Arnett DK *et al*, 2002, no *Minnesota Heart Health Program*).

Cada um desses objetivos tem implicações sobre o desenho da avaliação e, particularmente, sobre a escolha do grupo controle.

## **ETAPA 5. Definir os indicadores de desfecho e os potenciais confundidores e mediadores de efeito entre o desfecho e a intervenção**

Definições operacionais dos indicadores de desfecho e das demais variáveis são necessárias para a adequada coleta e análise dos dados. Diferentes estudos podem utilizar diferentes definições de uma mesma variável. Por exemplo, Luepker RV *et al*, 1994, no *Minnesota Heart Health Program*, definiram como tabagista atual os participantes que relataram haver fumado pelo menos 100 cigarros ao longo da vida e que fumassem atualmente. Winkleby MA *et al*, 1993, avaliando o impacto do *Stanford Five-City Project*, definiram como tabagista indivíduos que relataram ser fumantes diários e que houvessem fumado um ou mais cigarros na semana anterior. Em uma intervenção de base escolar aninhada ao *Minnesota Heart Health Program*, Kelder SH *et al*, 1994, alunos da 6<sup>a</sup> a 12<sup>a</sup> séries foram classificados como fumantes se referissem tabagismo semanal. Jousilahti P *et al*, 1994, no *North Karelia Project* consideraram fumantes os que referiram tabagismo semanal. Jousilahti P *et al*, 1994, no *North Karelia Project* consideraram fumantes os indivíduos que relataram tabagismo por pelo menos um ano e que fumaram, em média, mais de uma vez por dia no mês anterior à entrevista. Para que medidas repetidas em uma mesma população ou que resultados de avaliações diferentes possam ser comparadas, é muito importante que os mesmos critérios sejam utilizados em todas as ocasiões.

Particularmente importante é planejar a coleta e definir indicadores de processo e variáveis mediadoras. Métodos qualitativos de pesquisa poderão auxiliar na obtenção de alguns desses dados. Tais informações serão muito importantes para compreender razões de sucesso ou de fracasso das intervenções e para respaldar a

plausibilidade dos resultados que virem a ser detectados. Ao definir a rede causal (Etapa 3), ficam claros os mediadores através dos quais a intervenção, se funcionar, deverá atuar. Demonstrar que esses passos intermediários foram alcançados (ou que fracassaram) podem ser a chave da resposta necessária às futuras tomadas de decisão.

## ETAPA 6. Desenhar o estudo

O desenho do estudo envolve, fundamentalmente, a decisão de quem será o grupo controle. Os indivíduos beneficiados ou a populações-alvo do programa / intervenção constituem o grupo intervenção (experimental). O grupo controle, por sua vez, pode ser de três tipos: externo, interno e histórico. Algumas intervenções são implementadas tendo uma comunidade ou área controle pré-definida. Nesse caso, o grupo controle é formado por indivíduos ou comunidades de fora da área de cobertura do programa (grupo controle externo). Diferentemente das intervenções randomizadas, no entanto, o grupo controle externo e o grupo intervenção foram alocados a essa condição de forma não aleatória. Nesse caso, os indivíduos ou grupos de indivíduos são distribuídos pelos grupos intervenção ou controle pela vontade dos responsáveis pelo programa / intervenção, por contingências políticas ou por razões de gerenciamento ou administrativas do programa. Por exemplo, antes de sua implementação, o *Minnesota Heart Health Program* selecionou três pares de comunidades e uma comunidade de cada par, sem sorteio, foi selecionada para receber a intervenção (Luepker RV *et al*, 1994). Mesmo que o programa / intervenção não tenha elegido antecipadamente um grupo controle, a avaliação poderá fazê-lo, em geral emparelhando os dois grupos com base em características potencialmente confundidoras.

130

Por outro lado, se o grupo controle for constituído por indivíduos elegíveis para o programa / intervenção (porque atendem os critérios de seleção), mas não beneficiados (porque não aceitaram ou porque não houve suficientes recursos para abranger a totalidade dos potenciais beneficiários) ou se for constituído por indivíduos que receberam menores doses da intervenção, então o quase-experimento terá um grupo controle interno. Parte de uma das avaliações do *Stanford Three Community Study* publicada por Stern MP *et al*, 1976, por exemplo, analisou o efeito da intervenção sobre indivíduos de alto risco cardiovascular que participaram de nove sessões de aconselhamento em grupo, com duração de 1 hora e meia a 3

horas e meia cada um, comparativamente aos que receberam apenas a intervenção veiculada pelos meios de comunicação de massa e por correio.

Embora as opções de grupo controle possam ser várias para estudos quase-experimentais, muitas vezes a disponibilidade de um grupo controle pode ser um problema. Em programas / intervenções de cobertura universal, dependendo do tempo desde o lançamento do programa, pode não haver um grupo controle externo ou interno disponível (áreas ou comunidades totalmente virgens do programa). Essas situações exigem muita criatividade do avaliador para tornar possível uma avaliação controlada. Áreas com coberturas parciais do programa (não por vontade do avaliador, mas por circunstâncias locais ou de gerenciamento do programa) poderão ser utilizadas como controles internos e análises comparadas e do tipo dose-resposta conduzidas.

Finalmente, os próprios beneficiários ou a população-alvo do programa poderá ser seu próprio controle, através de uma análise tipo antes-e-depois. A análise de Arnett DK *et al*, 2002, por exemplo, mostra o impacto do *Minnesota Heart Health Program* sobre os fatores de risco cardiovasculares através de três inquéritos populacionais (1985-1987, 1990-1992, 1995-1997) realizados após o primeiro, em 1980-1982. Como o objetivo da intervenção era obter impacto sobre toda a população, todos os habitantes das áreas intervenção eram potenciais beneficiários. Dessa forma, foi obtida uma amostra de base populacional para examinar o efeito antes-e-depois da intervenção em somente comunidades alvo.

131

Para que o grupo intervenção possa ser seu próprio controle através de uma comparação do tipo antes-e-depois ou que mudanças no tempo  $t_i - t_0$  possam ser comparadas entre o grupo experimental e o controle (Figuras 5 e 6), é necessário que uma medida de linha de base tenha sido realizada, antes do início do programa, sendo por isso chamadas de avaliações longitudinais.

Independentemente do tipo de grupo controle utilizado, se avaliações de linha de base não tiverem sido conduzidas ou, se conduzidas, não tiverem sido analisadas, estará caracterizado um estudo do tipo transversal e não um quase-experimento.

Para que medidas tipo antes-e-depois sejam feitas, é necessário definir o momento para a coleta de informações de linha de base de ambos os grupos. Normalmente, a avaliação de linha de base é realizada antes que a intervenção seja implementada no grupo intervenção. É possível também incluir indivíduos que estejam recém-in-

gressando nas atividades do programa ou que hajam ingressado há um tempo inferior ao necessário para que mudanças dele decorrentes possam ser verificáveis.

As Figuras 5 e 6 apresentam as possíveis estruturas de avaliações com delineamento quase-experimental, respectivamente, com grupo controle interno (como o estudo de Stern MP *et al*, 1976) e externo (como o estudo de Luepker RV *et al*, 1994). Dadas as especificidades metodológicas dos estudos longitudinais, sem grupo controle, o planejamento desse tipo de avaliação quase-experimental será apresentado separadamente.

O Quadro 1 mostra um resumo dos tipos de grupos controles (internos ou externos) que podem ser utilizados nas avaliações quase-experimentais.

**Quadro 1. Tipos de grupos controles que podem ser utilizados nos diferentes desenhos quase-experimentais**

Tipo de avaliação	Medidas	Em quem?	Comparado a quê?	Tipo de grupo controle
Longitudinal	Mudança	População-alvo ou beneficiários	Antes-depois	Interno ou beneficiários somente (dose-resposta)
Longitudinal-controle	Mudança relativa	População-alvo ou beneficiários	Comparação antes-depois entre grupo intervenção e controle	Interno ou externo
Caso-controle	Uma vez	População-alvo	Comparando exposição ao programa entre doentes (casos) e não-doentes (controles)	Interno

Extraído e traduzido de Habicht et al, 1999.

Figura 5. Estrutura simplificada de avaliação quase experimental com grupo controle interno

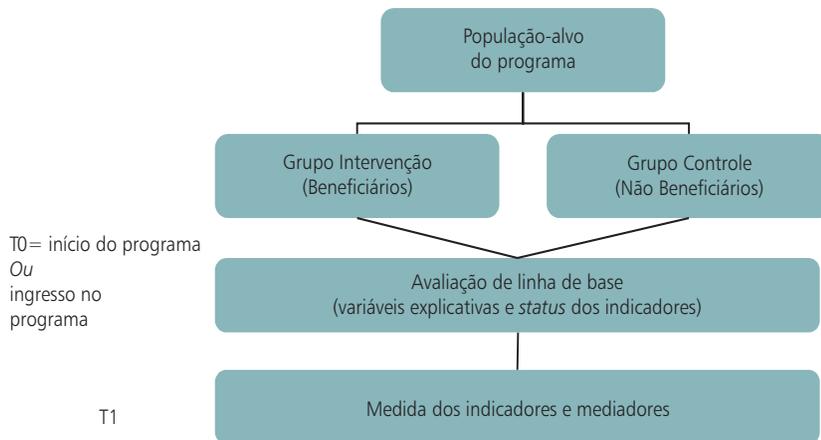
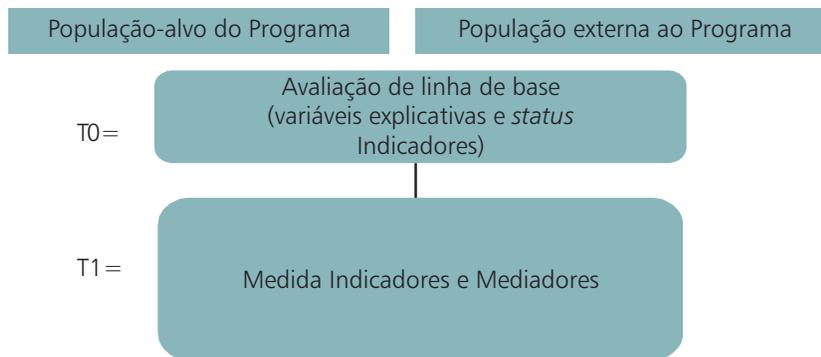


Figura 6. Estrutura simplificada de avaliação quase experimental com grupo controle externo

133



O planejamento dos estudos quase-experimentais deve tentar prevenir os principais tipos de vieses que costumam afetar esse tipo de delineamento:

**Viés de perda de acompanhamento:** sempre que o desenho do estudo implicar o acompanhamento do mesmo indivíduo ao longo do tempo, é imprescindível que

o pesquisador estabeleça cuidados que previnam as perdas de acompanhamento. Esses cuidados incluem, além do consentimento informado, a tomada de endereço e telefone pessoais, de residência e de trabalho para contato, não apenas do próprio participante da avaliação, mas também de um ou mais amigos ou familiares que estejam em contato permanente com o mesmo. As perdas, se numerosas (digamos, se superiores a 10%), podem introduzir um tipo de viés de seleção, uma vez que indivíduos perdidos costumam diferir dos demais em outras características importantes para a determinação do desfecho. Como regra, independentemente do tipo de estudo epidemiológico, a melhor maneira de lidar com vieses é preveni-los. Não há método estatístico que compense um estudo cujos participantes tenham sido selecionados de forma enviesada ou cujas informações tenham sido obtidas de forma tendenciosa.

134

**Viés de informação:** esse é um tipo comum de viés nos estudos que envolvem estratégias educacionais. Os desfechos proximais (mudanças em hábitos e comportamentos), quando obtidos de maneira referida pelo participante, são particularmente suscetíveis a esse tipo de viés. Nem sempre é possível mas, se for e os recursos permitirem, o relato deve ser validado, pelo menos em uma subamostra, através de medidas objetivas. No *Stanford Five-City Project*, por exemplo, a informação sobre tabagismo foi validada através de dosagens bioquímicas (monóxido de carbono no ar expirado e de tiocinato sérico) (Winkleby MA *et al*, 1993).

Em geral, em programas / intervenções cujo conteúdo é educacional, os participantes tendem a responder sobre a mudança de hábitos ou comportamentos de forma idealizada (como deveria ser ao invés de como realmente é). O viés de informação, se presente em ambos os grupos ou se for diferencial e presente, por exemplo, mais frequentemente em um dos grupos (intervenção ou controle), tenderá a desviar a medida de efeito. No primeiro caso, em direção à unidade, enfraquecendo o efeito protetor, por exemplo, quando esse de fato existe. No segundo caso, a direção do viés variará conforme o grupo em que o erro for mais frequente: exacerbará ou diminuirá o efeito protetor da intervenção se for mais frequente, respectivamente, no grupo intervenção ou controle.

## ETAPA 7: Definir os critérios para seleção dos participantes

Idealmente, os grupos controles externos devem ser constituídos por comunidades o mais semelhantes possível às comunidades intervenção quanto a caracterís-

ticas que poderão confundir o efeito real do programa / intervenção. População total, estrutura populacional sócio-demográfica, coeficientes de mortalidade geral, específicos ou proporcionais, índices de desenvolvimento humano, aspectos sócio-culturais, são algumas das características que podem confundir os resultados observados de intervenções comunitárias. Da mesma forma, se uma intervenção tiver como população-alvo apenas indivíduos de estratos socioeconômicos menos favorecidos, o grupo controle, seja interno ou externo, deverá ser selecionado apenas entre indivíduos com essa característica.

Os três pares de comunidades selecionadas no *Minnesota Heart Health Program* eram emparelhadas quanto ao tamanho (25.000-110.000 habitantes), tipo de comunidade (agrícola, urbana ou metropolitana) e distância em relação à área metropolitana de Mineápolis – St. Paul (< 250 milhas).

A definição dos critérios de seleção depende da população-alvo a que se destina a intervenção. Se o objetivo do programa / intervenção for atingir uma cobertura universal, independente da idade, nível de risco ou de outras características dos potenciais beneficiários, então a seleção dos indivíduos para a avaliação deverá atender essa característica.

135

## **ETAPA 8: Definir a unidade amostral e calcular o tamanho da amostra**

Tanto quanto para os estudos transversais e para os demais desenhos, um aspecto importante a considerar, no planejamento dos estudos quase-experimentais, é a unidade através da qual os participantes serão selecionados. Embora a unidade de estudo e observação sejam os indivíduos, a unidade amostral para seleção poderão ser grupos, aglomerados de indivíduos. Tal método de seleção dos sujeitos tem implicações no cálculo do tamanho da amostra, no poder do estudo e na metodologia de análise, a qual deve levar em conta a aglomeração das unidades amostrais (Cornfield J, 1978; Murray DM, 1998; Goldstein H, 2003).

Para o cálculo do tamanho amostral, há diversos livros textos (como Pocock SJ, 1983), que fornecem fórmulas, e vários pacotes estatísticos que estimam o número de indivíduos ou grupos de indivíduos (conforme a unidade amostral) necessários para o estudo. Os parâmetros requeridos para o cálculo manual ou por computador são os mesmos. A prática habitual na pesquisa científica é de aceitar em 5% como

verdadeiro que um tratamento, programa ou intervenção funciona, quando de fato não funciona (erro alfa de 5% ou nível de significância de 5%). A prática habitual na pesquisa acadêmica é declarar ausência de efeito 20% das vezes em que, de fato, há efeito (erro beta de 20%), geralmente referido como um poder de 80%. Quanto menores os erros alfa e beta, maior será o tamanho da amostra requerido.

O valor do  $p$  também poderá ser decidido entre testes bicaudais ou unicaudais. Se a avaliação puder assumir que o efeito do programa / intervenção só poderá ocorrer em uma única direção, então  $p$  unicaudais poderão ser estabelecidos para os cálculos amostrais. Nesse caso, o tamanho de amostra requerido é menor do que para testes bicaudais, mantidas os demais parâmetros de cálculo amostral.

O tamanho da diferença do desfecho, seja em médias com os respectivos desvios padrão (por exemplo, níveis plasmáticos de colesterol total ou de pressão arterial) ou em proporções (por exemplo, taxas de adoção de atividade física) entre os grupos intervenção e controle constituem um outro parâmetro a ser estabelecido para o cálculo do tamanho amostral. A magnitude alcançável da diferença entre o grupo intervenção e controle (em médias ou em proporções) pode ser estimada através de revisão da literatura e da relevância dessa diferença para a saúde individual ou populacional.

136

## **ETAPA 9: Planejar a logística do estudo**

Como já enfatizado, se análises tipo antes-e-depois tiverem sido planejadas, é necessário que uma medida de linha de base seja realizada, em ambos os grupos, antes que o grupo experimental receba o programa / intervenção. Para que as informações sejam comparáveis, os instrumentos e procedimentos de coleta de dados devem ser idênticos aos que serão aplicados depois de decorrido um período de intervenção.

## **ETAPA 10: Definir o tempo necessário de intervenção para que o desfecho esperado possa ser detectado**

A revisão bibliográfica sobre a experiência de outros programas de prevenção de DCNTs poderá auxiliar na definição do tempo mínimo necessário para que um resultado mensurável possa ser detectado. Em geral, o tempo entre a medida de linha de base e a avaliação final deve ser de pelo menos seis meses (Kleinbaum DG *et al*, 1982).

O estudo de Vartiainen R et al, 1998, para avaliar o efeito sobre o tabagismo do *North Karelia Youth Project* realizado entre escolares da 7ª série do grupo experimental, em 1978, re-entrevistou os estudantes pela primeira vez dois anos após a intervenção (1980) quando eles estavam em média com 15 anos de idade. Esses estudantes foram novamente reavaliados aos 16, 17, 21 e 28 anos de idade.

O Manual Cindi (WHO, 1996) recomenda que avaliações de impacto sejam realizadas a cada cinco anos e avaliações de monitoramento, a cada dois.

## **ETAPA 11: Construir e reunir os instrumentos para a coleta de dados e controle de qualidade**

As recomendações tecidas para os estudos transversais, quanto aos instrumentos de coleta de dados, aplicam-se da mesma forma aos quase-experimentos. Como esse tipo de avaliação implica, necessariamente, em pelo menos duas coletas de informações quanto ao status dos indicadores de desfecho, é muito importante, para garantir a comparabilidade, que a metodologia de coleta e a padronização das medidas realizadas sejam rigorosamente as mesmas.

137

Como são feitas pelo menos duas medidas em cada participante do estudo, é necessário construir instrumentos para cada uma dessas visitas. A quantidade de informação e o instrumento para a segunda visita, em geral são menores do que para a primeira. Não é necessário na segunda visita coletar informações sobre variáveis confundidoras da potencial associação entre o programa / intervenção e o desfecho. Ao contrário, precisarão ser coletadas informações sobre variáveis mediadoras entre a intervenção e o desfecho e repetidas as perguntas e medidas definidoras do desfecho.

Os instrumentos de coleta devem ser iguais para o grupo intervenção e controle. Questões sobre variáveis mediadoras e de processo e que, portanto, somente serão respondidas pelos indivíduos do grupo intervenção, deverão ser aplicadas aos dois grupos. Perguntas filtro poderão auxiliar nos pulos dentro do questionário, o que será bastante útil quando a entrevista for realizada com indivíduos do grupo controle, para os quais as perguntas dessa natureza não se aplicarão. Tal procedimento tem por objetivo manter os entrevistadores tão cegos quanto possível sobre o *status* dos entrevistados quanto à exposição ao programa / intervenção. O cegamento dos entrevistadores nos quase-experimentos, tanto quanto nos estudos randomizados, é importante para evitar o viés do entrevistador. De acordo com esse viés, o entre-

vistador poderá, consciente ou inconscientemente, insistir mais ou menos com um dos grupos do estudo com relação a variáveis desfecho ou intermediárias, desviando assim a medida de efeito. O cegamento dos entrevistadores, juntamente com a padronização de coleta e de medidas, constituem as principais estratégias para prevenção desse tipo de viés.

## **ETAPA 12: Treinar a equipe de campo e testar a logística e os instrumentos**

Nos quase-experimentos, tanto quanto nos demais tipos de estudos, o treinamento dos trabalhadores de campo deve priorizar o cegamento quanto aos objetivos do estudo. No caso dos quase-experimentos, especificamente, é interessante que os entrevistadores entrevistem e examinem indivíduos dos dois grupos (intervenção e controle), nas duas ocasiões. Tal procedimento tem por finalidade evitar que um erro sistemático introduzido por um entrevistador e não detectado pelos coordenadores do estudo, seja distribuído desigualmente entre os dois grupos, introduzindo com isso desvios das estimativas.

138

## **ETAPA 13: Iniciar o trabalho de campo**

Diferentemente dos estudos transversais, o trabalho de campo dos estudos quase-experimentais será realizado em pelo menos duas fases (pré e pós-intervenção). As recomendações feitas para os estudos transversais, no entanto, aplicam-se igualmente para cada uma das fases dos quase experimentos.

## **ETAPA 14: Digitar os dados em um programa de computador, proceder a limpeza e edição dos dados**

Proceder conforme recomendado para os estudos transversais. É importante, no entanto, não esquecer de criar uma variável que identifique o grupo a que o indivíduo pertence (intervenção ou controle). Se a unidade amostral for um agrupamento de indivíduos, também é importante que essa unidade tenha uma variável que a identifique.

## ETAPA 15: Analisar e interpretar os resultados

A análise dos estudos quase-experimentais com grupo controle externo ou interno consiste, inicialmente, em comparar os grupos experimental (intervenção) e controle, em relação aos potenciais fatores de confusão. Essa análise fornece resposta quanto ao primeiro critério que define uma variável como potencial confundidora da associação entre a intervenção (exposição de interesse) e o desfecho: estar associada com a exposição de interesse.

Análises brutas de desfechos contínuos (glicemia, colesterolemia, pressão arterial, peso) podem ser realizadas através de testes *t* de Student com cada uma das variáveis explicativas (independentes). Análises de desfechos categóricos, em duas ou mais categorias, podem ser conduzidas através de testes qui-quadrado.

Se houver necessidade de controle de fatores de confusão (que, em geral, há, uma vez que é muito provável que, sendo coletadas informações suficientes sobre os dois grupos, sejam detectadas diferenças estatisticamente significativas quanto a potenciais confundidores) análises multivariadas devem ser realizadas. Como regra geral, se o desfecho for contínuo, realizar regressão linear múltipla; se categórico, regressão logística para desfechos raros (prevalência de, no máximo, 10%) ou regressão de Poisson, para desfechos frequentes (Barros AJ *et al*, 2003).

139

O principal aspecto a ser considerado na análise de avaliações quase-experimentais é a exclusão do efeito de confusão. As avaliações quase-experimentais com grupo controle externo ou interno podem utilizar diversas estratégias não-excludentes para controle de confusão. Tais estratégias podem ser aplicadas desde a seleção do grupo controle (através de emparelhamento) até o tratamento estatístico dos dados (estratificação e análise multivariada). Modelagens de multiníveis para estudos com unidades amostrais agrupadas (Goldstein H, 2003), conforme abordado nos parágrafos acima para estudos transversais, aplicam-se aos quase-experimentos tanto quanto aos demais desenhos de avaliações.

São consideradas potenciais confundidoras as variáveis que se associem estatisticamente com o desfecho (indicadores que o programa pretende modificar), com a exposição (o programa) e que não fazem parte da cadeia que liga a exposição de interesse ao desfecho (Rothman KJ, 1998). Para aumentar a sensibilidade do controle de fatores de confusão, podem ser mantidas nos modelos multivariados as variáveis que se associem estatisticamente ao desfecho e ao programa com  $p \leq 0,20$ .

Descartar o efeito de confusão o mais exaustivamente possível é a melhor estratégia para aumentar a plausibilidade de que o resultado encontrado seja de fato decorrente do programa / intervenção.

### Usando Estudos Longitudinais com Análise Tipo Antes-e-depois, sem Grupo Controle

Nesse tipo de avaliação, há apenas o grupo intervenção e a mudança verificada entre a linha de base (o “antes” da intervenção) e depois de decorrido um determinado tempo de exposição ao programa / intervenção (o “depois”) é tomada como efeito. Esse tipo de avaliação tende a superestimar o real efeito do programa / intervenção. O efeito placebo e o efeito Hawthorne (Fletcher RH *et al*, 1996), além do fenômeno de regressão à média (Murray, 1998), na ausência de um grupo controle, são responsáveis por essa superestimação.

140

Um grupo de indivíduos ingressando em um programa / intervenção ou uma amostra da população-alvo onde um programa / intervenção será implantado tem suas características de base medidas. Decorrido um intervalo de tempo necessário para que possam ocorrer mudanças, os mesmos indivíduos ou uma amostra da população-alvo são novamente entrevistados e examinados para verificação do desfecho. A diferença verificada entre as duas medidas constitui o desfecho (por exemplo, mudança na prevalência de tabagismo do início para após um ano de participação ou implementação do programa / intervenção). Esse tipo de grupo controle é chamado de histórico e o desenho do estudo pode ser considerado um quase-experimento em que o grupo controle é o próprio grupo intervenção antes da implementação do programa / intervenção.

O estudo de Diehl HA, 1998, sobre o efeito de um programa educacional de base hospitalar na diminuição de fatores de risco cardiovasculares, o *Coronary Health Improvement Project* (CHIP), desenvolvido em Kalamazoo, Michigan, foi uma avaliação tipo antes-e-depois, sem grupo controle. Para cada participante, os fatores de risco coronarianos foram medidos antes do início do programa, 30 dias após completado o programa e novamente após 12 meses.

A avaliação publicada por Elmer PJ *et al*, 1995, sobre o resultado de mudanças no estilo de vida no tratamento da hipertensão leve, o *Treatment of Mild Hypertension Study* (TOMHS), mostra, entre outras, a variação média da pressão arterial e dos

lipídeos sanguíneos, após 1 e 4 anos de participação no programa (grupo intervenção somente).

**Vantagens das avaliações tipo antes-e-depois:** comparativamente aos demais quase-experimentos, esse tipo de avaliação tem a vantagem do menor custo e maior simplicidade em seu planejamento e condução.

**Limitações das avaliações tipo antes-e-depois:** são as mesmas das demais avaliações quase-experimentais, acrescidas da possibilidade de outros vieses que tendem a afastar as medidas de efeito da unidade. Embora o fato de os participantes da avaliação poderem ser seus próprios controles, o que controla uma série de características confundidoras individuais, a ausência de um grupo controle propriamente dito torna esse desenho suscetível ao fenômeno de regressão à média e vulnerável à tendência secular, além dos demais vieses acima mencionados.

A possibilidade de vieses é maior em análises tipo coorte, quando um único grupo de participantes (ao invés de duas amostras independentes da população-alvo) é selecionado para ser avaliado antes e depois. Nesse caso, as perdas de acompanhamento e o abandono do programa / intervenção são alguns dos exemplos de vieses que se somam aos anteriores.

141

**ETAPAS 1-3: proceder como especificado no início da PARTE IV, nas páginas 107 a 110.**

#### **ETAPA 4. Definir os objetivos da avaliação**

Avaliações com esse delineamento permitem estimar se as mudanças ocorreram no sentido e na intensidade esperados. Por exemplo, o *Treatment of Mild Hypertension Study* (TOMHS) tinha como objetivos: reduzir em 5% o peso corporal médio do grupo intervenção e em 25% a excreção urinária de sódio, além de perda de peso de 10 libras (1 libra= 0,453 kg) ou alcançar o peso ideal, o que fosse menor (objetivo individual), entre outros. A avaliação publicada por Elmer PJ *et al*, 1995, teve por objetivo descrever o alcance desses objetivos.

A avaliação poderá se limitar a medir a diferença no desfecho entre antes e depois da intervenção (como no estudo de Diehl HA, 1998) ou poderá pretender afastar

o efeito de fatores externos ao programa / intervenção (como a avaliação de Arnett Dk *et al*, 2002, no *Minnesota Heart Health Program*). Nesse último caso, os resultados encontrados serão mais plausíveis do que no primeiro.

## **ETAPA 5. Definir os indicadores de desfecho, os fatores de confusão e os mediadores de efeito entre o programa / intervenção e o desfecho**

Definições operacionais do desfecho e das variáveis mediadoras deverão ser definidas de forma semelhante como para os demais desenhos de estudos. Se for objetivo da avaliação descartar o efeito de fatores de confusão, os avaliadores deverão planejar a busca de informações sobre potenciais confundidores a partir do próprio grupo em estudo e/ou de fontes externas. Na avaliação de Arnett Dk *et al*, 2002, o efeito do *Minnesota Heart Health Program* sobre os fatores de risco cardiovasculares foi ajustado para idade, sexo, escolaridade e grupo étnico dos participantes.

142

Se a avaliação incluir apenas beneficiários do programa / intervenção poderá ser importante coletar informações que indiquem o grau de exposição do participante às atividades do programa / intervenção. Nesse caso, deve-se proceder como para os desenhos anteriores. A coleta de informações que indiquem a intensidade da exposição deverá ser planejada para a segunda entrevista ou para a entrevista da segunda amostra.

## **ETAPA 6. Desenhar o estudo**

Uma coorte de indivíduos beneficiários do programa / intervenção ou uma amostra da comunidade que receberá o programa / intervenção será selecionada e características de base serão medidas. As características a serem medidas incluem, obrigatoriamente, indicadores de resultado, cuja mudança se deseja mensurar (por exemplo, médias basais de peso corporal ou prevalência basal de sedentarismo), além de outras características dos participantes.

Opcionalmente, se for objetivo do estudo descartar efeito de confusão, características modificáveis ao longo do tempo e potenciais confundidores poderão ser coletadas na linha de base e na aferição final do estudo. A Figura 7 apresenta a estrutura esquemática de um estudo longitudinal, sem grupo controle, com cronologia que

permite descartar, embora não perfeitamente, o efeito de variáveis externas ao longo da implementação do programa / intervenção. A primeira coleta de dados deverá ser planejada para antes da implementação em campo do programa / intervenção ou para o momento de ingresso dos participantes no programa (a linha de base).

Figura 7. Estrutura simplificada de avaliação tipo antes-e-depois, sem grupo controle



143

O planejamento do estudo deve tentar prevenir os principais tipos de vieses que costumam afetar esse tipo de delineamento. A regressão à média e a tendência secular são duas fontes potenciais de vieses a que estão sujeitas as avaliações tipo antes-e-depois. Pela regressão à média, indivíduos que em uma primeira medida apresentam valores extremos de qualquer parâmetro biológico, como pressão arterial, colesterolemia, glicemia etc, tendem, em uma segunda oportunidade, independentemente de qualquer tratamento ou intervenção, apresentar valores mais próximos à média do grupo. Nesse sentido, o viés de regressão à média, afasta as medidas de efeito da unidade, atribuindo ao programa / intervenção um impacto superior ao real. Os estudos que medem os mesmos indivíduos antes-e-depois são particularmente suscetíveis a esse tipo de viés. Na ausência de um grupo controle propriamente dito, como é o caso, não há estratégia capaz de prevenir sua ocorrência.

A tendência secular caracteriza-se pelas modificações nas incidências e prevalências de desfechos que ocorrem ao longo do tempo em decorrência de fatores extrínsecos ao programa / intervenção. Sem a presença de um grupo controle, tal viés

não é passível de prevenção, mesmo que o estudo avalie amostras de indivíduos diferentes antes-e-depois.

O efeito placebo e o efeito Hawthorne também contribuem para a superestimação de efeito das avaliações tipo antes-e-depois que incluem os mesmos indivíduos nas duas medidas. Indivíduos que são alvo de qualquer programa ou estudo apresentam melhores resultados para um mesmo “tratamento” do que seus pares na vida real. Isso se deve a dois fenômenos: ao fato de receber um “tratamento” (efeito placebo, que também está presente fora do contexto das investigações) e ao fato de ser alvo de atenção diferenciada, no caso, a atenção da equipe do estudo (efeito Hawthorne).

Além desses, nos estudos tipo antes-e-depois em que o mesmo grupo de indivíduos é seguido ao longo do tempo, a perda de acompanhamento e o abandono, como anteriormente mencionado para os demais tipos de quase-experimentos, são vieses que podem e devem ser prevenidos.

## **ETAPA 7: Definir os critérios para seleção dos participantes.**

A definição dos critérios de seleção depende da população-alvo a que se destina a intervenção e do objetivo do programa / intervenção. Se o objetivo do programa / intervenção for obter um impacto populacional, então qualquer indivíduo que preencher os critérios de inclusão para ser beneficiário do programa (mesmo não sendo) poderá fazer parte da amostra de avaliação. No *Minnesota Heart Health Program* (Arnett DK *et al*, 2002), eram selecionados adultos com idade entre 25 e 74 anos.

Se o programa pretender modificar o risco apenas dos beneficiários, bastará uma amostra representativa desses para que esse efeito possa ser estimado. Elmer PJ *et al*, 1995, por exemplo, analisaram a mudança de peso, pressão arterial, colesterol e triglicerídeos, bem como de consumo de sal e bebidas alcoólicas e de taxas de sedentarismo, ao longo de quatro anos do TOMHS. Avaliações de caráter populacional, como no primeiro caso, são mais caras do que aquelas em que os indivíduos podem ser localizados a partir dos registros do próprio programa.

## ETAPA 8: Escolher o processo de amostragem

Quando o programa / intervenção é direcionado a toda a população de um área, é necessário selecionar uma amostra representativa dessa população. No *Minnesota Heart Health Program*, por exemplo, a avaliação de Arnett DK *et al*, 2002 selecionou uma amostra da população-alvo do programa. Adultos com idade entre 25 e 74 anos eram selecionados por amostragem probabilística em dois estágios. Inicialmente, de forma aleatória, foram selecionados 40 grupos de 1.000 domicílios cada um. No segundo estágio, por amostragem aleatória simples, em cada grupo, era selecionado um número de domicílios de forma a alcançar os 5.000 adultos requeridos para o tamanho da amostra. Somente domicílios em que residisse pelo menos uma pessoa da faixa etária de interesse eram incluídos no sorteio.

## ETAPAS 9 a 16:

Proceder como para os demais quase-experimentos.

## ETAPA 17: Analisar e interpretar os resultados

Como regra geral, se o estudo for feito com duas amostras independentes da população-alvo, a análise é semelhante a de um quase-experimento com grupo controle interno. A diferença entre os estudos reside nos critérios de seleção da amostra. Nos quase experimentos com grupo controle interno, o grupo controle foi selecionado com base na sua não exposição ao programa / intervenção (semelhante às coortes observacionais controladas). Os passos da análise e descrição dos resultados são os mesmos descritos para os quase-experimentos com grupo controle interno.

Se o estudo incluir duas amostras de beneficiários do programa, a análise poderá ser do tipo dose-resposta, com controle para fatores de confusão, se necessário (geralmente, é). A mudança nos indicadores de desfecho entre os participantes fortemente expostos à intervenção e os moderadamente expostos poderá ser comparada à observada entre os menos expostos, tomados como controles.

A análise tipo dose-resposta dos estudos longitudinais consiste, inicialmente, em descrever os indivíduos que ingressaram na avaliação. Essa fase da análise inclui a verificação da distribuição das variáveis confundidoras (por exemplo, idade, sexo,

escolaridade e nível socioeconômico) e dos indicadores que o estudo pretende modificar (prevalência de tabagismo, de sedentarismo e de sobrepeso, ou média de pressão arterial, peso corporal, colesterolemia, HDL colesterol e de glicemia). Dados sobre o processo do programa (por exemplo, número de sessões oferecidas e número médio de participantes por sessão, número médio de sessões assistidas pelos participantes, número de adesivos com nicotina distribuídos, número médio de adesivos utilizados pelos participantes) devem ser analisados e apresentados.

O desfecho deve ser calculado através da diferença entre os níveis dos indicadores observados na segunda e na primeira medida. Diferença observada na taxa de fumantes ou de sedentários ou mudança na concentração plasmática média de HDL colesterol são exemplos de como os desfechos poderão ser estimados. Quando os participantes do estudo forem os mesmos na primeira e na segunda medida, a significância estatística de mudanças em desfechos contínuos deverá ser estimada através de testes *t* emparelhados.

146

Como nos exemplos anteriores, análises brutas entre desfechos contínuos (glicemia, colesterolemia, pressão arterial, peso) e cada uma das demais variáveis, além da intensidade da intervenção, podem ser realizadas através de testes *t*. Análises de desfechos categóricos, em duas ou mais categorias, devem ser conduzidas através de testes qui-quadrado.

Conforme o grau desejado de inferência da avaliação (ver Parte III desse documento), os estudos longitudinais tipo antes-e-depois poderão ser analisados sem controle de fatores de confusão (avaliações de adequação). Nesse caso, depois da descrição da amostra e das variáveis de processo, apresenta-se a diferença absoluta ou proporcional verificada nos indicadores de desfecho. Por exemplo, o estudo de Diehl HA, 1998, encontrou uma redução no colesterol total de  $41,3 \pm 33$  mg/dl, entre os homens, e de  $21,6 \pm 28$  mg/dl, entre as mulheres, após quatro semanas de participação no CHIP.

Análises mais sofisticadas, não apenas dos estudos tipo antes-e-depois sem grupo controle, mas também dos quase-experimentos de modo geral, incluem uma gama variada de tratamento estatístico dos dados. Na ausência de randomização, o afastamento de outros fatores que poderiam confundir a associação verificada pode tornar-se uma tarefa complexa. Aumentará a credibilidade dos resultados da avaliação se puder ser demonstrado, por exemplo, que:

- a) as tendências esperadas aconteceram: queda relativamente rápida da mortalidade cardiovascular em áreas com o programa;
- b) houve efeito tipo dose-resposta: associação inversa entre intensidade da intervenção e mortalidade cardiovascular;
- c) mudanças em variáveis mediadoras associaram-se a mudanças no desfecho: queda na mortalidade cardiovascular entre indivíduos, por exemplo, que pararam de fumar em comparação aos que continuaram fumando;
- d) não verificação de impacto na ausência do programa: mortalidade cardiovascular entre os que recusaram participar do programa semelhante à verificada em áreas sem o programa e maior do que na área coberta;
- e) a magnitude do efeito é compatível com mudanças em variáveis mediadoras: o aumento na participação das atividades oferecidas pelo programa é compatível com o grau de redução da mortalidade cardiovascular, como estimado por simulações matemáticas.

## Usando Estudos Tipo Casos e Controles

Brevemente, a utilização de um estudo de casos e controles para avaliar o impacto ou desempenho de um programa / intervenção será planejada tendo como casos indivíduos que logrem obter um determinado desfecho. Em uma avaliação do impacto de um programa para redução das taxas de infarto, por exemplo, a exposição ao programa é comparada entre os casos (indivíduos infartados) e os controles (uma amostra de indivíduos sem infarto da população de onde se originaram os casos). A hipótese é que, se o programa / intervenção proteger os beneficiários, o odds de exposição entre os casos (número de casos expostos dividido pelo número de casos não expostos) será menor que o odds de exposição entre os controles (número de controles expostos dividido pelo número de controles não-expostos) e a razão de odds, inferior à unidade.

Aspectos metodológicos sobre o planejamento, organização, implementação e análise de estudos de casos e controles podem ser obtidos em obras específicas (Schlesselman JJ, 1982). Na revisão da literatura realizada, não foi localizada nenhuma avaliação de programas / intervenções de prevenção de DCNTs que tivesse utilizado a metodologia de casos e controles. Seu uso, no entanto, deveria ser incentivado, pelas vantagens metodológicas desse delineamento.

**Vantagens das avaliações caso-controle:** comparativamente às avaliações quase-experimentais, as avaliações caso-controle têm a vantagem de fornecer resultados mais rapidamente e de poderem ser iniciadas precocemente à implementação do programa / intervenção.

**Limitações das avaliações caso-controle:** são as mesmas que limitam os resultados das avaliações quase-experimentais. A falta de controle de todos os fatores que podem influir sobre o desfecho limita a conclusão de que os resultados alcançados sejam devidos exclusivamente ao programa / intervenção.

As próximas seções apresentam as etapas de planejamento, implementação e análise de estudos de casos e controles que visem avaliar o impacto de programas / intervenções.

**ETAPAS 1-3: proceder como especificado no início da PARTE IV, nas páginas 107 a 110.**

148

#### **ETAPA 4. Definir os objetivos da avaliação**

A avaliação poderá ter como objetivo investigar o impacto do programa / intervenção sobre a frequência do desfecho que o mesmo pretende modificar.

#### **ETAPA 5. Desenhar o estudo**

Os estudos de casos e controles poderão ser de três tipos, dependendo do tipo de grupo controle utilizado: estudos de casos e controles tradicionais, em que os controles estão livres do desfecho até o final do estudo; estudos de casos e controles caso-concomitantes, em que os controles são selecionados entre a população em risco de apresentar o desfecho e, uma vez experimentado o desfecho, não se tornam mais elegíveis para o grupo controle; e estudos de casos e controles inclusivos, em que os controles são selecionados da população em estudo, independentemente de seu *status* quanto à ocorrência do desfecho, e no qual indivíduos com o desfecho farão parte, indistintamente, do grupo de casos e do grupo controle (Rodrigues & Kirkwood, 1990).

Além das estratégias e oportunidades de seleção dos controles, cada um dos três tipos difere quanto a interpretação das razões de odds. A Figura 8 mostra um diagrama simplificado da estrutura de avaliações com desenho de casos e controles.

Dependendo da medida de desfecho, cada um desses desenhos de casos e controles poderá ser mais apropriado. Se a medida de desfecho for prevalência, o estudo de casos e controles tipo inclusivo será mais indicado, porque a razão de odds desse tipo de desenho corresponde à razão de prevalências dos estudos transversais. Se a medida de efeito for mais bem expressa em termos de incidência, os estudos de casos e controles tipo caso-concomitante serão mais adequados, porque a razão de odds deles obtidas corresponde à densidade de incidência (pessoas-ano-em risco) dos estudos de coortes. Os estudos de casos e controles tradicionais fornecem razão de odds que correspondem ao risco relativo (razão de incidências cumulativas).

Dos três tipos, os estudos de casos e controles tradicionais são os mais sensíveis para detectar associações, podendo, no caso de desfechos frequentes, superestimar a magnitude da associação entre o programa / intervenção e o desfecho.

O planejamento do estudo deve tentar prevenir os principais vieses que costumam afetar esse tipo de delineamento:

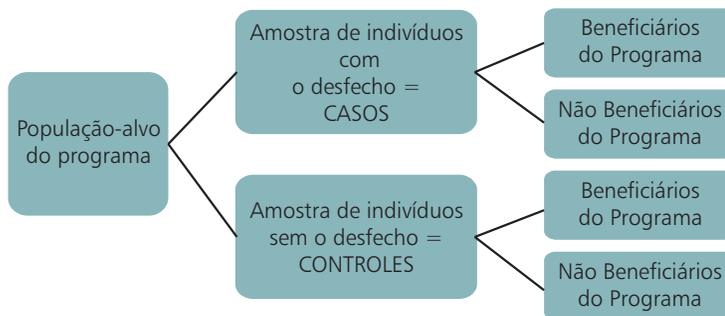
- *Viés de memória:* é um tipo de viés de informação em que, tipicamente, as pessoas que experimentam desfechos desfavoráveis tendem a lembrar mais facilmente de exposições ocorridas no passado do que seus contrapartes, levando a erro de classificação da exposição. Nesse sentido, quando a intervenção é uma exposição protetora, indivíduos parcialmente ou temporariamente expostos no passado, se forem do grupo de casos lembrarão mais facilmente de terem participado do programa do que os controles, enfraquecendo dessa forma (ou mesmo mostrando como risco) o efeito protetor do programa / intervenção. Técnicas de entrevista, a formulação da pergunta, a idade e a escolaridade do entrevistado, o tempo desde a exposição ao programa e o grau de detalhamento das questões investigadas são fatores que devem ser levados em conta de forma a evitar que essa fonte de viés afete a validade das respostas.
- *Viés de sobrevivência:* para prevenir esse tipo de viés de seleção, é importante selecionar os casos entre indivíduos recém-experimentando o desfecho (casos incidentes), ao invés de selecionar casos prevalentes. A seleção de casos prevalentes poderá levar à falsa conclusão de que um fator de fato protetor, de bom

prognóstico (tanto que garantiu a sobrevivência após a ocorrência do evento), seja interpretado como fator de risco.

- *Viés de causalidade reversa*: como o desfecho é registrado no presente e a exposição, necessariamente, aconteceu no passado, pode ser difícil demonstrar que o processo mórbido que desencadeou o desfecho não estava presente desde antes do ingresso do sujeito ao programa / intervenção. Nesse caso, indivíduos expostos ao programa eram já portadores da doença em estágios mais precoces da história natural e o estudo demonstrará uma associação espúria de aumento de risco entre os participantes do programa. Esse é o caso, por exemplo, em avaliações de programas de prevenção do câncer do colo uterino em áreas ou grupos em que o programa foi recentemente implementado. Como muitas mulheres estarão em fase adiantada de câncer quando se submeterem aos primeiros exames citopatológicos, parecerá a princípio que o programa estará associado a um risco aumentado de mortalidade por câncer invasivo. A medida que passar o tempo e que a cobertura do programa atingir níveis suficientes, o efeito protetor do programa poderá ser demonstrado através de um odds de exposição ao programa entre mulheres casos (câncer invasivo) menor do que o odds de exposição entre as controles.

150

Figura 8. Estrutura simplificada de avaliação com desenho de casos e controles



## ETAPA 6: Definir os critérios para seleção dos casos

Diferentemente dos demais desenhos, os estudos de casos e controles têm os participantes selecionados em função da presença ou ausência do desfecho. Os casos são indivíduos que apresentam o desfecho que o programa / intervenção quer prevenir.

Definições operacionais dos desfechos são usadas para identificar o grupo de casos. Programas de prevenção de DCNTs poderão definir como casos indivíduos nos quais seja feito o diagnóstico de infarto agudo do miocárdio, acidente vascular cerebral ou doenças intermediárias como diabetes mellitus e hipertensão arterial sistêmica.

Tanto quanto para os demais tipos de estudos, a definição desses desfechos deverá, preferencialmente, seguir recomendações de organismos nacionais ou internacionais reconhecidos. A padronização dos procedimentos diagnósticos é tão importante quanto para os demais desenhos. Ressalte-se uma vez mais que, para prevenir o viés de sobrevivência, os casos devem ser selecionados de entre os novos portadores do desfecho (casos incidentes).

## **ETAPA 7: Definir os critérios para seleção dos controles**

A definição dos controles constitui-se na etapa mais delicada do planejamento dos estudos de casos e controles. Por definição, os controles devem ser selecionados em uma população tal que, se viessem a se tornar casos, seriam captados pelo estudo.

A maioria dos livros textos costuma dividir os estudos de casos e controles em estudos de base hospitalar e estudos de base comunitária. Em princípio, se os casos forem selecionados de pacientes hospitalizados, os controles poderão ser selecionados de entre outros pacientes dos mesmos hospitais, cuja hospitalização foi decorrente de outras enfermidades não sabidamente relacionadas com a exposição de interesse. No contexto dos programas / intervenções de prevenção das DCNTs (a exposição de interesse), os controles deveriam ser selecionados de entre pacientes hospitalizados por outras doenças não relacionadas ao tabagismo ou ao sedentarismo. Preferencialmente, para evitar super-representação de enfermidades que possam inadvertidamente estar associadas com a exposição de interesse, é aconselhável que o grupo controle contenha uma gama variada de diagnósticos.

Nos estudos de casos e controles de base comunitária, os controles poderão ser selecionados a partir dos endereços dos casos ou, se possível, de entre os demais indivíduos que freqüentam o local ou o serviço onde foram localizados os casos. Indivíduos livres dos sintomas ou sinais que caracterizam os casos são selecionados para controles.

Uma prática bastante freqüente nos estudos de casos e controles é a seleção dos controles emparelhada aos casos, de acordo com algumas características previamente definidas, por exemplo, idade e sexo. O emparelhamento tem como objetivo controlar o efeito de confusão já na fase de seleção da amostra. Como regra, portanto, somente vale a pena emparelhar com base em fortes fatores de confusão, do contrário o estudo perderá em precisão das estimativas de risco. Assim, por exemplo, se o caso for um homem de 55 anos, o controle deverá ser do mesmo sexo e com idade variando entre 52-57 anos (um pequeno intervalo de tolerância facilitará a localização dos controles). O emparelhamento tem conseqüências na metodologia de análise dos estudos de casos e controles, conforme será abordado mais adiante.

Quando o número de casos é limitado, uma forma de otimizar o tamanho da amostra é selecionar-se mais de um controle para cada caso. Teoricamente, pode-se selecionar livremente o número de controles por caso. Estudos com mais do que quatro controles por caso, no entanto, não acrescentam eficiência ao tamanho da amostra, podendo, inclusive, implicar em aumento improdutivo de custos do estudo.

152

## **ETAPA 8: Definir os potenciais confundidores e mediadores de efeito entre o desfecho e a intervenção**

Demonstrar associações entre o *status* de caso ou controle e variáveis mediadoras e confundidoras do efeito do programa / intervenção é tão importante quanto para os demais estudos. Obter razões de odds de exposição ao programa entre os casos inferior às obtidas entre os controles aumenta a plausibilidade de um efeito protetor do programa / intervenção. Definições operacionais dessas variáveis são necessárias e seguem o recomendado para os desenhos anteriores.

## **ETAPA 9: Calcular o tamanho da amostra**

Além dos erros alfa (geralmente de 5%) e beta (geralmente de 20%), o cálculo do tamanho amostral utiliza como parâmetros a razão doentes:não doentes (“doença” querendo representar o desfecho dicotômico de interesse), a freqüência de exposição ao programa / intervenção entre os não doentes e a razão de odds que se deseja ser capaz de estimar. Pequenos estudos-pilotos podem ser necessários para que alguns desses parâmetros possam ser estimados.

## **ETAPA 10: Planejar a logística do estudo**

Tal como com os demais delineamentos, a estratégia para localização dos casos e dos controles precisa ser definida previamente. Em estudos de casos e controles de base comunitária, por exemplo, os casos precisam ser tais que motivem a busca de consulta médica ou que sejam diagnosticados em consulta médica, sendo conseqüentemente localizados em serviços ambulatoriais. Academias de ginástica, locais públicos de lazer seriam outras fontes possíveis de casos.

## **ETAPA 11: Definir o tempo necessário de intervenção para que o desfecho esperado possa ser detectado**

Essa é uma particular vantagem dos estudos de casos e controles sobre os desenhos quase-experimentais. À medida que os casos forem sendo registrados a avaliação de casos e controles pode ser iniciada, não havendo necessidade de esperar por um tempo mínimo para que o programa / intervenção atue como acontece com os demais desenhos de avaliação.

153

## **ETAPAS 12 a 16**

Seguem as mesmas recomendações feitas para os desenhos anteriores.

## **ETAPA 17: Analisar e interpretar os resultados**

A análise dos estudos de casos e controles inclui, inicialmente, a descrição, separadamente, de casos e controles, conforme variáveis independentes, inclusive a participação no programa /intervenção que está sendo avaliado. Se o programa tiver um efeito protetor sobre o desenvolvimento do desfecho, a prevalência da exposição (participação no programa) será menor no grupo de casos do que no de controles.

A força da associação entre a exposição e o desfecho tanto bruta quanto ajustada será avaliada através de regressão logística e a medida de efeito obtida, a razão de odds. Razões de odds menores do que a unidade significam que a exposição é protetora contra o desfecho, enquanto que as maiores do que um indicam o contrário.

Estudos de casos e controles cujos controles tenham sido selecionados conforme algum tipo de emparelhamento devem ser analisados levando isso em conta, do contrário as razões de odds serão desviadas para a unidade. Em estudos de casos e controles emparelhados, o método adequado de análise é a regressão logística condicional. Esse tipo de análise baseia-se nos pares discordantes. Em estudos com um controle por caso, a razão de odds para exposições dicotômicas é obtida dividindo o número de pares em que o caso é exposto e o controle não o é, pelo número de pares em que o caso não é exposto e o controle o é.

Relativamente aos demais desenhos, os estudos de casos e controles são pouco usados nas avaliações. Para avaliações que desejem demonstrar plausibilidade entre o desfecho e o programa / intervenção, o uso desse tipo de delineamento deveria ser estimulado.

## Usando Ensaios Randomizados

154

Como mencionado, os ensaios controlados randomizados são o padrão ouro para demonstrar eficácia. Tais estudos não são exclusivos de eficácia, podendo ser utilizados igualmente para avaliar efetividade (efeito da intervenção em condições reais de aplicação, sem que medidas de promoção da adesão dos participantes sejam implementadas ao longo do estudo). Para que ensaios randomizados sejam usados para demonstrar eficácia ou efetividade de programas / intervenções de larga escala populacional é necessário que a avaliação seja planejada juntamente com o planejamento do programa / intervenção, de forma a garantir a alocação aleatória dos grupos. Em programas / intervenções planejadas para terem cobertura universal, esse tipo de avaliação precisa contar com a possibilidade de atrasar o início do programa / intervenção em áreas ou comunidades sorteadas para o grupo controle.

Há uma ampla bibliografia em forma de livros texto em epidemiologia abordando o planejamento e execução de ensaios controlados randomizados. Há igualmente, na literatura de periódicos, ampla divulgação de normas e recomendações sobre a apresentação dos resultados dos estudos randomizados, apontando itens essenciais a serem mencionados na sessão de metodologia e análise, de forma a assegurar a validade interna das pesquisas. A Figura 9 mostra a estrutura de um estudo controlado randomizado para avaliação de um programa / intervenção. Os estudos de Dunn AL *et al*, 1997, sobre o projeto *Active* e de Korhonen M *et al*, 2003, em North Karelia são exemplos de avaliações randomizadas.

O planejamento do estudo deve tentar prevenir os principais vieses que costumam afetar esse tipo de delineamento:

- *Viés de perda de acompanhamento*: todos os estudos cujo delineamento envolve o acompanhamento de indivíduos ao longo do tempo são suscetíveis a esse tipo de viés. Indivíduos que são perdidos ou que abandonam o “tratamento” devem ser considerados diferentes dos que se mantêm aderidos quanto a outros aspectos prognósticos. Estratégias que otimizem a localização dos participantes e que promovam a adesão ao programa / intervenção devem ser planejadas e implementadas ao longo dos estudos randomizados que visem medir eficácia.

Em forma esquemática, o planejamento dos ensaios randomizados envolve as seguintes etapas:

**ETAPAS 1-3: proceder como especificado no início da PARTE IV, nas páginas 107 a 110.**

#### **ETAPA 4: Definir os objetivos**

Os estudos randomizados permitem avaliar a eficácia e a efetividade de programas / intervenções. Como para os demais delineamentos, os objetivos devem ser claros, preferentemente em pequeno número, e formulados de tal forma que permitam uma medida quantitativa de seu alcance.

Na avaliação do *Project Active*, acima mencionado, foi medida a eficácia de uma estratégia de aconselhamento educacional para promover a atividade física, comparativamente a exercícios estruturados tradicionais.

A avaliação do *North Karelia Program* publicada por Korhonen M *et al*, 2003, foi 149 uma avaliação da efetividade do aconselhamento nutricional intensivo para pacientes com hipertensão arterial, em tratamento não farmacológico, em atenção primária à saúde.

## ETAPA 5: Definir os indicadores de desfecho, os fatores de confusão e os mediadores de efeito entre o programa / intervenção e o desfecho

O desfecho é o objetivo que o programa / intervenção pretende alcançar. Operacionalmente, a maioria dos desfechos podem ser definidos em mais de uma maneira (apenas o desfecho morte é mais conservador nesse sentido!). Ser tabagista, por exemplo, poderá ser definido como fumar pelo menos um cigarro por dia, todos os dias, sendo excluídos os fumantes ocasionais e os fumantes de final de semana. A escolha da definição pode estar baseada na possibilidade de comparar os achados do atual estudo com os de outros já existentes na literatura (para o que, o uso de definições aplicadas por outros autores é a melhor estratégia) ou na possibilidade logística ou financeira da definição em questão. Nesse último caso, para definir tabagista através de dosagem de cotinina na urina, por exemplo, é necessário que o estudo disponha de orçamento que permita a realização de dosagens laboratoriais.

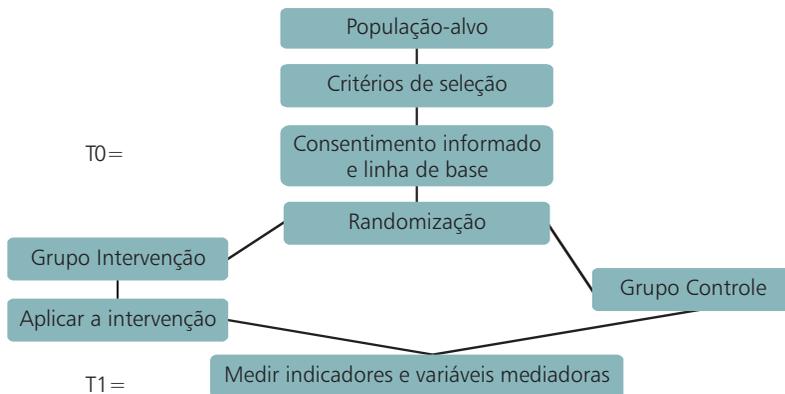
156

No projeto *Active*, como desfecho, foram comparadas as variações médias de peso, índice de massa corporal, colesterol total e frações, pressão arterial, gordura corporal, perfil de risco e de equivalentes metabólicos, além dos traçados de eletrocardiogramas de repouso e de esforço, entre os grupos intervenção e controle. No *North Karelia*, foram comparadas, nos dois grupos, as variações médias em peso corporal, colesterol total e frações e em pressão arterial.

No *Active*, as variáveis de confusão investigadas foram a idade, o sexo e o tabagismo; e as mediadoras, os desfechos intermediários (adoção de estratégias comportamentais, auto-eficácia e benefícios / barreiras) identificados nos dois grupos.

No *North Karelia*, os fatores de confusão investigados foram idade, sexo, tabagismo e tratamento medicamentoso para hipertensão ao ingressar no estudo. As variáveis mediadoras foram a variação na excreção urinária de 24 horas de sódio e potássio.

Figura 9. Estrutura simplificada de estudo randomizado para avaliação de programas



## ETAPA 6. Desenhar a interven o

O grupo interven o do projeto *Active* era aconselhado a acumular pelo menos 30 minutos de atividade f sica de moderada intensidade, na maioria dos dias da semana, sendo cada indiv duo aconselhado conforme o est gio de prontid o para mudan a em que se encontrasse. Durante seis meses, os participantes se reuniam em pequenos grupos: uma hora por semana,   noite, nas 16 primeiras semanas, e a cada duas semanas nas semanas 17 a 24. Nessas reuni es, os participantes aprendiam estrat gias cognitivas e comportamentais apropriadas para seu grau de prontid o motivacional. Os facilitadores usavam a t cnica de solu o de problemas para ajudar os participantes a iniciar, adotar e manter um programa de atividade f sica. Nenhum exerc cio estruturado era realizado durante essas sess es, exceto para explicar, por exemplo, o que seria considerado atividade f sica moderada.

157

O grupo controle do *Active* recebia exerc cios tradicionais (exerc cios com intensidade de 50-85% da capacidade aer bica m xima, com dura o de 20-60 minutos por sess o, tr s a cinco dias por semana. Sess es individuais supervisionadas eram oferecidas cinco dias por semana por seis meses. L deres de grupo ajudavam os participantes a estabelecer metas real sticas, monitoravam a atividade f sica e incentivavam verbalmente os participantes. Ap s tr s semanas de exerc cios supervisionados, os participantes escolhiam as atividades aer bicas que mais gostavam e individualizavam seus exerc cios.

O aconselhamento nutricional do North Karelia consistiu em, com a ajuda de um *folder* elaborado por nutricionista clínica, transmitir informação detalhada sobre alimentos e dicas para mudar a alimentação. Enfermeiras aconselhavam os pacientes do grupo intervenção a comer alimentos sem gordura ou pobres em gordura ou que contivessem gordura de origem vegetal. Os participantes eram também aconselhados a diminuir o sal da alimentação e a usar outros temperos na comida. Diários para registrar os alimentos consumidos eram revisados e comentados por escrito por nutricionista clínica. Além disso, o grupo intervenção tinha duas reuniões de grupo com o médico e a nutricionista clínica, aos seis e dezoito meses de estudo sobre o papel do estilo de vida no tratamento da hipertensão. O grupo controle por sua vez recebeu o tratamento de rotina para hipertensão.

## ETAPA 7: Definir os critérios para seleção dos participantes

158

Os critérios de seleção geralmente incluem a idade, o sexo e outros atributos que qualificam os indivíduos a participar do programa / intervenção (critérios de inclusão) ou que os excluem (critérios de exclusão). Em condições ideais, devem receber a intervenção aqueles com maior probabilidade de se beneficiar da mesma. Nos ensaios randomizados com medicamentos, um grupo de indivíduos em uma determinada fase da doença serão adequados para a intervenção, por exemplo, se diagnósticos concorrentes com o desfecho que se quer medir puderem ser excluídos. Por exemplo, entre indivíduos com cardiopatia isquêmica candidatos a um ensaio para promover atividade física, seria interessante excluir os portadores de neoplasias fatais. De forma comparável, nos ensaios comportamentais para adoção de estilos de vida saudáveis, indivíduos dependentes de álcool e outras drogas poderiam ser excluídos.

Um outro aspecto a considerar é que os ensaios comunitários, por definição, (Kleinbaum DG *et al*, 1982) são aqueles aplicados a indivíduos sadios mas em risco de contrair a doença que se quer evitar. No caso particular da iniciativa Carmen, cujo objetivo mais distal é reduzir a mortalidade por doenças cardiovasculares, cânceres e acidentes, praticamente toda a população adulta encontra-se em potencial risco, uma vez que são essas as principais causas de morte, nessa faixa etária, em áreas em transição epidemiológica, como é o caso da maioria dos países latinos e do Caribe.

Os critérios de inclusão do *Active* foram: voluntários sadios que residissem ou trabalhassem em um raio de 10 milhas do centro onde o estudo era realizado, tivessem

idade entre 35 e 60 anos e que fossem sedentários. Eram excluídos os voluntários que tivessem história de infarto, derrame, diabetes mellitus insulino-dependente, osteoporose ou artrite; que estivessem acima de 140% do peso corporal ideal; que planejassem mudar de endereço no período de realização do estudo; que consumissem três ou mais drinques de bebidas alcoólicas por dia; que se exercitassem 20 minutos ou mais por dia, pelo menos três dias da semana ou que tivessem um gasto energético total superior a 36 kcal/kg/dia, para os homens, ou 34 kcal/kg/dia, para as mulheres; que tivessem pressão arterial sistólica  $\geq 160$  mmHg ou diastólica  $\geq 100$  mmHg; que usassem medicamentos como  $\beta$ -bloqueadores que pudessem limitar o desempenho no exercício; e, para as mulheres, as que planejassem engravidar nos próximos dois anos. Os voluntários foram convidados por meios locais de comunicação em massa, igrejas, grupos comunitários e por pessoa a pessoa.

No North Karelia, os critérios de seleção foram: idade de 25 anos ou mais, pressão sistólica entre 140-179 mmHg e/ou diastólica 90-109 mmHg e/ou que estivessem recebendo medicação anti-hipertensiva, em nove centros de saúde de North Karelia e no centro de atenção primária de Kuopio.

## **ETAPA 8. Definir a unidade amostral e calcular o tamanho da amostra**

Como já mencionado anteriormente, um aspecto importante a considerar no planejamento das avaliações, inclusive naquelas que usam desenhos experimentais randomizados, é a unidade através da qual os participantes serão selecionados. Embora a unidade de estudo e observação sejam os indivíduos, a unidade amostral para seleção poderão ser grupos, aglomerados de indivíduos. Esse foi o caso, por exemplo, da intervenção em escolares feita pelo North Karelia (Vartiainen *E et al*, 1998) e pelo *Mirame!* (Berríos X *et al*, 2004). Como já enfatizado, tal método de seleção dos sujeitos tem implicações no cálculo do tamanho da amostra, no poder do estudo e na metodologia de análise, a qual deve levar em conta a aglomeração das unidades amostrais (Murray DM, 1998; Goldstein H, 2003).

Os parâmetros requeridos para o cálculo amostral são os mesmos apresentados para as avaliações quase-experimentais: erro alfa de 5%, poder de pelo menos 80%, tamanho da diferença do desfecho que se deseja ser capaz de detectar e que tenha relevância do ponto de vista da redução de risco. Nesse particular, diferenças estatisticamente significativas não são necessariamente relevantes do ponto de vista da

redução do risco individual ou populacional. Por último, convém lembrar uma vez mais que o número de pessoas a serem incluídas no estudo e a distância entre elas são os principais determinantes do custo de uma avaliação.

O *Active* calculou o tamanho da amostra em base individual utilizando os seguintes parâmetros: erro alfa de 5%, erro beta de 10%, diferença de 3 ml/kg/min em consumo de oxigênio entre os grupos intervenção e controle, após seis meses de participação no projeto, e taxa de abandono de 15%.

## ETAPAS 9 a 12

Correspondem às etapas 10 a 13 das avaliações transversais e seguem as mesmas recomendações tecidas para aquele tipo de delineamento.

## ETAPA 13. Obter o consentimento informado

160

É uma exigência dos códigos de ética em pesquisa experimental envolvendo seres humanos (*International Ethical Guidelines for Biomedical Research Involving Human Subjects, Genebra: CIOMS, WHO, 1993*; Diretrizes e Normas Regulamentadoras de Pesquisa Envolvendo Seres Humanos do Conselho Nacional de Saúde do Ministério da Saúde do Brasil, Resolução 196/96) a obtenção do consentimento livre e esclarecido dos indivíduos elegíveis sobre o protocolo do estudo, antes de sua inclusão. Dos que aceitarem participar são coletadas as informações de linha de base.

## ETAPA 14: Coletar informações de linha de base

Na linha de base (antes da aplicação da intervenção no grupo experimental), são coletadas informações de ambos os grupos (experimental e controle) quanto aos parâmetros que se pretende modificar (peso, colesterol total, pressão arterial, glicemia etc) e quanto aos potenciais confundidores da associação entre a intervenção e o desfecho. Espera-se que o passo seguinte, a randomização, distribua igualmente essas características entre os dois grupos (intervenção e controle) e que apenas a intervenção sendo testada seja a única diferença a incidir entre os dois grupos. Essas

informações, em geral, são aquelas apresentadas na primeira tabela das publicações dos estudos randomizados.

### **ETAPA 15: Proceder à randomização**

A randomização é um procedimento de sorteio através do qual os participantes têm chances iguais de pertencer a um grupo ou outro (intervenção ou controle). A alocação randomizada deve ser feita de acordo com as unidades consideradas para o cálculo de tamanho amostral. Se foi calculado número de escolas, por exemplo, essas deverão ser randomizadas. Nas escolas selecionadas para o grupo experimental, por exemplo, todos os alunos que atenderem os critérios de seleção deverão receber a intervenção.

Existem vários procedimentos de randomização, desde os mais simples (randomização aleatória simples) até formas mais ou menos complexas (randomização estratificada e randomização em blocos). Vários livros texto de epidemiologia abordam com detalhe essas técnicas.

Embora mais simples, a randomização aleatória simples pode ter o inconveniente de distribuir casos e controles de forma não balanceada quanto a características de linha de base (por exemplo, o número de mulheres pode ser muito maior no grupo intervenção do que no controle ou vice-versa). A randomização estratificada, como o nome indica, distribui os participantes uniformemente quanto a estratos de uma ou mais características cujo efeito se deseja controlar. A randomização por blocos é uma extensão da estratificada. Para cada estrato de duas ou mais variáveis agrupadas, um bloco de tamanho pré-fixado de participantes é dividido igualmente entre os dois grupos.

O *Active* utilizou randomização estratificada de forma a garantir um número equilibrado de homens e mulheres nos dois grupos.

### **ETAPA 16: Aplicar a intervenção no grupo experimental**

Aplicar a intervenção ao grupo aleatoriamente selecionado como experimental. As ações de intervenção devem ser aplicadas conforme planejado. O conteúdo, duração e técnica de reuniões de grupo, por exemplo, devem ser implementadas dentro

do esquema planejado e previamente acordado com os participantes. Se o ensaio randomizado for de eficácia, as estratégias planejadas para garantir a adesão mínima considerada suficiente para a obtenção de efeito devem ser acionadas regularmente. No estudo de Dunn AL *et al*, 1997, por exemplo, os participantes que não compareciam às sessões semanais da intervenção eram contatados e encorajados a retornar às atividades o mais breve possível.

Um aspecto importante a destacar é com relação ao cegamento. Os ensaios randomizados ideais são duplo cegos, significando que os participantes e os membros da equipe que medem as variáveis de desfecho não são sabedores a qual grupo os primeiros pertencem. O duplo cegamento garante as vantagens metodológicas de validade interna obtidas através da randomização. Nos ensaios randomizados com drogas, o cegamento de ambas as partes é obtido através do mascaramento de cápsulas, comprimidos ou injeções. Características de cor, sabor, odor, formato etc são mantidas o mais semelhantes possíveis, de forma a dificultar a identificação do produto ativo e do placebo. Nos ensaios com aconselhamento, no entanto, o cegamento do participante não é possível, restando apenas a possibilidade de realização de estudos uni-cegos.

162

No estudo de avaliação da efetividade do aconselhamento nutricional do North Karelia, a intervenção era realizada pelos profissionais do próprio centro de saúde, mas as medidas necessárias para a determinação do desfecho (linha de base e aos 12 e 24 meses) foram feitas por uma enfermeira do estudo, que não participava das atividades da pesquisa nos centros de saúde. Embora não fosse possível cegar os participantes do estudo (ensaio “aberto”), esse procedimento permitiu que a avaliação fosse uni-cega.

## **ETAPA 17: Acompanhar os participantes ao longo do tempo**

O número de contatos da equipe do estudo com os participantes depende da natureza da intervenção e do tempo necessário para que ocorra impacto sobre os desfechos estudados. O tempo de acompanhamento tem como base a resposta à pergunta: quanto tempo é necessário para que a intervenção se traduza em efeito mensurável sobre a variável desfecho? A resposta a essa pergunta pode ser obtida através da revisão bibliográfica, em estudos de outros autores, ou através da realização de um pequeno estudo-piloto experimental com apenas o grupo intervenção. Estudos piloto são também úteis para estimar o tamanho do efeito esperado da

intervenção, além das médias e prevalências de linha de base das variáveis, necessários para o cálculo do tamanho amostral.

No *Active*, os participantes do grupo intervenção foram acompanhados semanalmente ou quinzenalmente ao longo dos seis meses de estudo. Os do grupo controle foram acompanhados 3-5 dias por semana, semanalmente no mesmo período. As medidas para o desfecho foram realizadas na linha de base e após seis meses.

No *North Karelia*, os pacientes do grupo intervenção eram agendados para consultar com a enfermeira do centro de saúde quatro vezes durante o primeiro ano (1, 3, 6 e 9 meses) e quatro vezes no segundo ano (15, 18, 21 e 24 meses). A cada visita, os pacientes do grupo intervenção eram aconselhados quanto à dieta. Os participantes foram avaliados ao ingressarem no estudo (linha de base) e decorridos 12 e 24 meses. Essas avaliações ocorriam no próprio centro de saúde onde os pacientes recebiam o tratamento intervenção ou controle.

Especial cuidado deve ser tomado para evitar perdas de acompanhamento. Se muito numerosas (digamos, se superiores a 10%, em quaisquer dos grupos) poderão ameaçar a validade interna do estudo. Os indivíduos perdidos nos estudos prospectivos costumam diferir dos demais em outras características geralmente relacionadas com desfechos de saúde, introduzindo um potencial viés. No projeto *Active*, dos 121 participantes do grupo intervenção e dos 114 do grupo controle, respectivamente, 116 e 106, completaram a avaliação de seis meses, não havendo diferença estatisticamente significativa entre os grupos em termos de adesão (Dunn *AL et al*, 1997).

No *North Karelia*, embora fossem programadas várias visitas para o grupo intervenção, não há relato de que tenham sido usadas estratégias para promover a adesão dos participantes (Korhonen *M et al*, 2003). No grupo intervenção, as taxas de adesão foram maiores do que no controle, respectivamente, 88% *versus* 78% ( $p < 0,0005$ ) no primeiro acompanhamento (um ano) e 77% *versus* 66% ( $p = 0,002$ ) no segundo ano. Indivíduos aderentes apresentam, em geral, melhores respostas aos tratamentos de qualquer natureza do que os não aderentes.

## ETAPA 18: Medir o desfecho em ambos os grupos

Além da padronização na aplicação da entrevista e nas medidas de parâmetros biológicos (como frequência cardíaca e pressão arterial, por exemplo), é importante manter cego, tanto quanto possível, o membro do estudo que coleta as informações quanto ao *status* de intervenção ou controle de cada participante. Como acima mencionado, o cegamento do avaliador das medidas de desfecho é muito importante para preservar a validade interna da avaliação.

## ETAPA 19: Analisar os dados por intenção-de-tratar

Para análise, os dados do estudo devem ser armazenados em formato eletrônico. Os procedimentos recomendados para digitação, comparação, limpeza e edição dos dados podem ser consultados no livro de Barros & Victora, como anteriormente mencionado. Não havendo diferenças entre os dois grupos quanto a outras características que não a intervenção propriamente dita (o que, espera-se tenha sido logrado com êxito pela randomização), a análise dos ensaios randomizados é relativamente simples. Médias ou proporções das variáveis desfecho são comparadas estatisticamente entre o grupo intervenção e controle. Diferenças inter-grupos (por exemplo, diferença de perda de peso desde a linha de base até o momento do acompanhamento no grupo intervenção e no grupo controle) podem ser calculadas e sua significância estatística estimada. Nesse último caso, por exemplo, a intervenção terá sido bem sucedida na redução de peso se a média de perda de peso no grupo intervenção for maior do que a média de perda verificada no grupo controle.

Se a randomização tiver falhado na garantia do equilíbrio entre os grupos quanto a potenciais variáveis de confusão, medidas estatísticas de análise multivariada poderão ser empregadas. Especial atenção na análise deverá ser empregada quando a amostra estudada for do tipo aglomerada. Modelos de multiníveis poderão ser empregados para que a análise seja pertinente à forma de seleção dos sujeitos e para que as medidas de desfecho não sejam espuriamente superestimadas (Cornfield J, 1978; Murray DM, 1995).

Análise por intenção-de-tratar significa que os participantes deverão ser analisados nos grupos aos quais foram alocados através da randomização. Indivíduos randomizados para o grupo experimental, mas que não aderiram à intervenção, devem ser analisados como fazendo parte do grupo intervenção, sendo incorreto analisá-

los como se pertencessem ao grupo controle. Análises de sub-grupos deverão ser realizadas preferentemente se tiverem sido programadas na fase de planejamento do estudo.

Um aspecto importante nos estudos randomizados para avaliação de programas / intervenções de cadeias longas é analisar se os desfechos intermediários aconteceram conforme esperado. As variáveis mediadoras são então tomadas como desfechos intermediários e sua associação com o status intervenção ou controle analisada. No projeto *Active*, o alcance dos critérios de mudança de comportamento (acumular 30 minutos ou mais de atividade física pelo menos moderada, na maioria dos dias da semana) e de sustentabilidade da mudança de comportamento ao longo do tempo foram comparados entre os dois grupos. O alcance desses critérios fazia parte da rede causal que ligava a intervenção à mudança no perfil de risco cardiovascular dos participantes do estudo.

Da mesma forma, no estudo em North Karelia, a rede causal ligando o aconselhamento nutricional e a mudança no risco cardiovascular passava pela mudança no teor de gordura, fibras e sal da alimentação dos participantes. Demonstrar que as mudanças intermediárias aconteceram conforme estimado, aumentou a plausibilidade do estudo.

## **Avaliações com Dados Secundários**

## Usando Sistemas de Vigilância de Fatores de Risco e Doenças Crônicas como Instrumento de Avaliação

Os tipos de estudos de avaliação até aqui descritos tinham em comum o fato de a unidade de observação serem indivíduos. É possível, como já anteriormente mencionado, fazer estudos em que a unidade de observação sejam grupos de indivíduos, os chamados ESTUDOS ECOLÓGICOS. Os grupos geralmente são definidos por critérios geográficos (cidades, estados ou países) mas podem ser grupos menores como salas de aula, fábricas etc. Os estudos ecológicos são mais usados para combinar grandes bancos de dados (como censos, estatísticas vitais etc) como fonte de informação para ambos, o fator em estudo e a doença, na mesma população. Uma característica dos estudos ecológicos é que não se tem informação sobre a exposição e a doença dentro de cada grupo, isto é, conhece-se o número de pessoas expostas e o número de casos dentro de cada grupo, mas não se conhece o número de casos expostos. Os sistemas de vigilância dos ministérios da saúde dos países são importantes fontes de dados para medir diretamente o efeito de intervenções de larga escala.

Os estudos ecológicos podem ser de dois tipos: uma comparação de medidas entre um número determinado de grupos (em geral dez grupos ou mais), em um dado ponto no tempo (análise transversal), ou uma avaliação da tendência, em um ou mais grupos, ao longo de um determinado período de tempo (**SÉRIES TEMPORAIS**). No primeiro caso, compara-se a frequência da exposição (existência ou cobertura ou intensidade do programa / intervenção) e frequência do desfecho de interesse. Nas séries temporais estima-se as mudanças ocorridas ao longo do tempo entre a exposição e o desfecho. Os desfechos podem ser prevalência, incidência ou mortalidade mas, geralmente, os dois últimos são mais empregados, pela maior disponibilidade dessa informação.

Através da análise da tendência temporal do número ou taxas de eventos (como morte por DCNTs) ou fatores de risco (como tabagismo, obesidade, hipercolesterolemia e sedentarismo) obtidos de dados secundários, coletados pelos sistemas de vigilância, constituem uma estratégia de avaliação do impacto de programas / intervenções relativamente barata e suficiente. Em alguns casos, a associação temporal entre mudanças na tendência de ocorrência de casos da doença ou evento e a intervenção é tão óbvia, que a vigilância por si só pode fornecer evidências suficientes de associação. Tal foi o caso, por exemplo, da avaliação do efeito da legislação de

uso obrigatório de capacetes em alguns estados norte-americanos sobre a mortalidade por causas externas (Fleming NS, 1992; Kraus JF, 1994).

As informações publicadas pelo Ministério da Saúde do Brasil, em 2005, são um exemplo de avaliação ecológica baseada em dados de rotina da vigilância epidemiológica no país.

Um exemplo de avaliação através de séries temporais foi o estudo de Puska P *et al*, 1998, na Finlândia, para avaliar o impacto sobre a mortalidade por doenças cardiovasculares na área demonstrativa do North Karelia *Project* e no restante do país.

**Vantagens das avaliações ecológicas:** a principal vantagem das avaliações ecológicas é o baixo custo e a conveniência do uso de dados já coletados, oriundos de diferentes bancos de dados, desde que possam ser facilmente agregados. Os estudos ecológicos podem ser a única forma factível de estudar uma exposição. Por exemplo, quando a exposição varia pouco na área estudada (intervenções feitas através de meios de comunicação de massa; intervenções universais feitas sobre toda a população, como legislações) os delineamentos ecológicos, cobrindo áreas maiores, são capazes de detectar variação na exposição entre diferentes grupos.

168

**Limitações das avaliações ecológicas:** embora os estudos ecológicos sejam de fácil análise e apresentação, problemas metodológicos são muito frequentes e podem limitar a inferência causal dos resultados obtidos. A limitação mais conhecida é a da falácia ecológica em que a correlação observada entre duas variáveis ecológicas é, frequentemente, muito diferente da correlação no nível individual, dentro da mesma população. O exemplo clássico do estudo do sociólogo Emil Durkheim, em 1953, que coletou dados sobre frequência de suicídio e tipo de religião em diversas províncias do oeste europeu, concluindo que as províncias predominantemente protestantes tinham maiores taxas de suicídio do que as predominantemente católicas. A falácia ecológica, no entanto, não permitia identificar que eram geralmente os católicos, residentes em províncias predominantemente protestantes, os que se suicidavam.

Uma outra limitação dos estudos ecológicos é a alta colinearidade dos dados. Como já mencionado anteriormente, a magnitude das associações oriundas de análises ecológicas é maior do que as de análises individuais.

As etapas para o planejamento e análise de avaliações ecológicas transversais assemelham-se às das avaliações transversais feitas em indivíduos. As avaliações através de séries temporais assemelham-se às individuais do tipo antes e depois, com ou sem grupo controle. Alguns aspectos, no entanto, merecem destaque e serão discutidos a seguir.

### **ETAPAS 1-3: proceder como especificado no início da PARTE IV, nas páginas 107 a 110.**

#### **ETAPA 4. Definir os objetivos da avaliação**

A avaliação poderá ter como objetivo medir o alcance dos objetivos do programa / intervenção ou medir a correlação temporal entre o programa / intervenção e o desfecho. A análise de Puska *et al*, mencionada, é um exemplo de análise de correlação.

169

#### **ETAPA 5. Identificar possíveis indicadores de desfecho e da exposição**

Definido o objetivo, o próximo passo é construir uma lista de possíveis variáveis de desfecho e variáveis independentes que poderão responder, mesmo que só parcialmente, ao objetivo do estudo. De posse dessa lista, o próximo passo será identificar as bases de dados que contenham as informações necessárias. Identificadas as bases, autorização para o uso precisa ser formalmente obtida. Bancos de dados de acesso público, em geral, podem ser facilmente utilizados e analisados.

Os dados de mortalidade de North Karelia e de toda a Finlândia (Puska P *et al*, 1998) foram obtidos da *Statistics Finland*, uma organização que revisa o diagnóstico dado pelos médicos nos atestados de óbito, de acordo com os códigos da Classificação Internacional de Doenças.

## ETAPA 6. Desenhar o estudo

Dados sobre o número de beneficiários do programa / intervenção serão selecionados e, em um único momento, a exposição (ou o grau de exposição ao programa) e a presença do desfecho serão correlacionadas.

Essa metodologia de avaliação é planejada para analisar uma única ou mais de uma medida de prevalência do desfecho. No primeiro caso, por exemplo, volume de impostos sobre a venda de cigarros, um *proxis* da prevalência de tabagismo (desfecho), em áreas beneficiárias e não beneficiárias de um programa de combate ao fumo. No segundo, a mudança no montante de impostos arrecadados sobre a venda de cigarros, à medida em que passa o tempo de introdução do programa na área.

Uma vez que os dados nos estudos ecológicos são médias de todos os indivíduos, a associação entre exposição e desfecho não necessariamente reflete a exposição individual, o substrato da falácia ecológica. Também, o uso de variáveis *proxis* distorce ainda mais o significado da associação. Além disso, na maioria das vezes, os estudos ecológicos não dispõem de todos os dados necessários para o controle de fatores de confusão. Todos esses problemas contribuem para tornar questionável a validade dos resultados desses estudos. Mais recentemente, tem sido dada ênfase a estudos que congreguem, ao mesmo tempo, variáveis ecológicas (de grupos) e variáveis individuais (Rothman KJ, 1998).

170

O planejamento do estudo deve tentar prevenir, tanto quanto possível, esses principais vieses. No entanto, mesmo se afetados por variáveis de confusão, os estudos ecológicos serão suficientes para estabelecer se os objetivos de um programa / intervenção estão sendo alcançados (ecológicos de análise transversal) ou ainda, se é plausível supor que as mudanças observadas no desfecho são devidas ao programa / intervenção (séries temporais). No estudo de Puska P *et al*, mencionado, por exemplo, o objetivo era responder se o efeito esperado sobre a mortalidade por causas cardiovasculares havia sido alcançado ou não.

## ETAPA 7. Juntar bancos de dados

Se as informações necessárias provêm de bancos diferentes, será necessário fazer um *merge* entre os dados. Para tanto, é necessário que os bancos a serem unidos tenham uma variável identificadora em comum, como municípios, estados, países

etc. Fontes de dados individuais, mesmo que tenham identificadores pessoais, podem ter mais difícil acesso por razões de confidencialidade.

## ETAPA 8. Analisar e interpretar os resultados

Análises descritivas de avaliações ecológicas consistem em descrever, em forma tabular, gráfica ou através de mapas, as frequências dos desfechos de interesse, de acordo com o *status* de exposição (com ou sem o programa / intervenção). Resultados de análises realizadas pelo Ministério da Saúde do Brasil, em 2005, por exemplo, mostraram a variação, entre 1996 e 2002, nas taxas de mortalidade por acidente vascular cerebral, infarto agudo do miocárdio e diabetes mellitus, padronizadas pela população de 2000, no país como um todo e em cada uma das cinco regiões. As análises da situação de saúde do Brasil, publicadas em 2004 e 2005, são outros exemplos de análises ecológicas que mostram tendências temporais em desfechos e fatores de risco no país e regiões.

Se estiverem disponíveis variáveis confundidoras da associação, técnicas como padronização por idade, como empregada por Puska *et al*, e estratificação por níveis da variável confundidora poderão ser utilizadas.

171

Se estiverem disponíveis tanto variáveis de nível coletivo quanto individual, uma estratégia de análise é através de modelos de multiníveis (Goldstein H, 2003), como anteriormente descrito para os demais desenhos de avaliações.

## Conclusões e Recomendações

A decisão quanto ao tipo de estudo que deverá ser conduzido depende de vários fatores: da pergunta que a avaliação precisa responder, de quem será o tomador de decisão, do marco teórico do programa / intervenção, do momento em que a avaliação será conduzida, da existência de estudos demonstrando eficácia da intervenção e do montante de recursos disponível.

Definir os objetivos da avaliação com base no conhecimento do referencial teórico do programa / intervenção é um dos primeiros passos necessários para a decisão de qual desenho utilizar para a avaliação. O desejo de descartar o papel de variáveis de confusão sobre o efeito verificado entre o programa / intervenção e o impacto

é o segundo elemento importante nessa decisão. Se for desejado afastar esse efeito, então o estudo deverá ser tal que informações sobre um grupo controle sejam coletadas.

Um aspecto importante a ressaltar na análise dos dados das avaliações é que, independentemente do delineamento do estudo, a unidade de análise deve ser a mesma unidade utilizada para seleção dos participantes. Seleção por aglomerados (grupos) implica análise que leve em conta essa aglomeração.

As avaliações devem fornecer as respostas que os tomadores de decisão necessitam e em tempo para que esses possam levar esses resultados em conta em suas decisões. Informações perfeitas e uma avaliação ideal serão inúteis se estiverem disponíveis somente depois que as decisões tiverem sido tomadas, uma situação muito freqüente. Portanto, os avaliadores precisarão conhecer não somente o tipo de decisão que terá que ser tomada mas também quando essas decisões serão tomadas. O desenho e a condução da avaliação devem então ser tais que permitam atender esses prazos.

172

Os custos são geralmente o principal fator na escolha do desenho da avaliação. Os tomadores de decisão são particularmente sensíveis a esse aspecto porque terão que subtrair do orçamento do programa os fundos necessários para a avaliação. Os avaliadores devem discutir com os tomadores de decisão as implicações econômicas de diferentes desenhos, incluindo os seguintes aspectos: a) vale a pena fazer uma avaliação de impacto? b) é necessário coletar novos dados? Se sim, em que nível? c) o desenho inclui uma comparação entre a intervenção e o controle do tipo quase experimental? d) quão raro é o evento a ser medido e quão pequena é a diferença a ser detectada? e) quão complexa será a análise dos dados?

Finalmente, uma palavra a mais sobre o rigor das avaliações nos programas / intervenções em saúde pública. Vinte anos de avaliações randomizadas e quase-experimentais foram incapazes de demonstrar resultados positivos da maioria dos programas / intervenções comunitários (Winkleby M, 1994). É possível que os programas / intervenções comunitários sejam incapazes de impactar a população-alvo mais do que o atingível através da tendência secular isoladamente. Antes de chegar a essa conclusão, no entanto, convém perguntar se os métodos de avaliação utilizados são apropriados para detectar a verdade. Essa crítica é especialmente aplicável aos estudos randomizados que foram desenvolvidos para serem aplicados no âmbito da clínica, particularmente na área de medicamentos e que funcionam melhor

quando o pesquisador consegue obter um rigoroso controle das variáveis extrínsecas à intervenção propriamente dita. Por definição, os programas / intervenções em saúde pública operam por cadeias complexas e por mecanismos às vezes imprevisíveis, através de interações com características contextuais do local e da população onde está sendo aplicado. Dessa forma, cabe o cepticismo de que os mais rigorosos métodos de investigação atualmente disponíveis são incapazes de avaliar a efetividade dos programas / intervenções, mesmo quando os resultados mostram que o programa / intervenção não funciona. Novos programas / intervenções devem ser testadas formalmente através de ensaios randomizados. Áreas demonstrativas seriam os locais ideais para essas avaliações. Uma vez demonstrada a eficácia, no entanto, outros desenhos metodológicos, controlados ou não, deveriam ser empregados com segurança e suficiência para fins de avaliação.

Dado o estado atual da arte em termos de metodologia para avaliação de programas, a recomendação mais sensata parece ser de que a avaliação tenha claros os objetivos, que as informações coletadas sejam tão abrangentes e rigorosas quanto possível (de forma a tornar plausível os resultados que virem a ser detectados) e que os delineamentos sejam simples.



# PARTE V

## Manual de treinamento

*Iná S. Santos, Deborah Carvalho Malta, Otaliba Libânio de Moraes  
Neto, Lenildo de Moura,  
Luciana Sardinha, Elisabeth Duarte e Denise Bontempo*

## Introdução

A criação de um Guia Metodológico para avaliação de processo e impacto de programas / intervenções comunitárias que visem à prevenção de doenças crônicas não transmissíveis (DCNTs) é uma iniciativa da Organização Mundial da Saúde (OMS)/ Organização Pan-Americana de Saúde (OPAS), em parceria com o Ministério da Saúde do Brasil e a. Na América Latina e no Caribe, as DCNTs combinadas constituem a principal causa de morte antes dos 70 anos de idade. As DCNTs de maior importância em saúde pública na região são as doenças cardiovasculares, os cânceres e o diabetes mellitus.

Os primeiros passos da OPAS, no sentido de apoiar programas de prevenção integrada das DCNTs, ocorreram em 1996, tendo como referência o Programa *Cindi* (*Countrywide Integrated Noncommunicable Diseases Intervention Program*) criado e desenvolvido pela OMS, em países europeus, em 1985. Em setembro de 2002, a Conferência Pan-Americana de Saúde respaldou o enfoque de saúde pública para as DCNTs e aprovou a Iniciativa *Carmen* como uma das principais estratégias de prevenção integrada das doenças crônicas a serem implementadas na região.

176

A Iniciativa *Carmen* está comprometida com o desenvolvimento, implementação e avaliação de políticas que visem a redução simultânea de fatores de risco comuns a mais de uma das doenças crônicas; mobilização social e intervenções de base comunitária; vigilância epidemiológica dos fatores de risco; e atividades preventivas que possam contribuir com a redução das desigualdades em saúde.

Para a redução simultânea de fatores de risco comuns às DCNTs, a Iniciativa *Carmen* utiliza três estratégias: prevenção integrada e promoção da saúde, efeito demonstrativo e promoção de equidade em saúde. A primeira estratégia compreende o desenvolvimento de intervenções que objetivem a prevenção e a redução, simultaneamente, de um conjunto de fatores de risco comuns às principais DCNTs. O tabagismo, a dieta inadequada e o sedentarismo são considerados fatores de risco estratégicos para essa ação preventiva. A elaboração de guias que instrumentalizem os planejadores e avaliadores de programas e intervenções destinadas à prevenção e controle dos fatores de risco e à promoção da saúde fazem parte da estratégia de apoiar ações que possam provar efetividade nos diferentes contextos onde são aplicadas.

Avaliar significa investigar, de forma sistemática, o mérito, valor ou significado de um objeto (Scriven M. Minimalist theory of evaluation: the least theory that practice requires. *American Journal of Evaluation* 1998;19:57-70). A importância e aplicação da avaliação em programas e serviços de saúde têm crescido nas últimas décadas, ao mesmo tempo em que se constroem instrumentos que facilitem sua realização. Têm esse caráter o *Manual para Avaliação de Atividade Física*, elaborado pelo Centro de Controle e Prevenção de Doenças (CDC) do Departamento de Saúde e Serviços Humanos dos Estados Unidos (U.S. Department of Health and Human Services. Centers for Disease Control and Prevention. *Physical Activity Evaluation Handbook*. Atlanta, GA. EUA, 2002) e a *Avaliação Participativa de Municípios, Comunidades e Ambientes Saudáveis*, organizado por Akerman & Mendes (Akerman M, Mendes R. *Avaliação Participativa de Municípios, Comunidades e Ambientes Saudáveis: a trajetória brasileira – memória, reflexões e experiências*. São Paulo: Mídia Alternativa Comunicação e Editora, 2006).

O atual *Manual para Avaliação de Desempenho e Impacto de Programas de Prevenção e Controle de DCNTs* tem como objetivo principal auxiliar na definição do tipo de avaliação necessária para um projeto ou programa específico. Seu conteúdo foi adaptado do trabalho de Victora, 1996, em documento do Unicef (Victora CG. *What type of evaluation is needed? Using evaluation needs to design choices. Training materials*. New York: Unicef Evaluation and Research Office. FRMWK.LEC, 1996).

177

## Sete Passos para Avaliação de Programas de Prevenção e Controle de DCNTs

Ao longo do texto, o termo “programa” é usado para representar o objeto da avaliação e pode consistir em qualquer tipo de ação organizada de saúde pública, incluindo programas de saúde, projetos comunitários, intervenções em serviços de saúde, mobilizações comunitárias, pesquisas, sistemas de vigilância, atividades de implementação de políticas públicas e não necessariamente um programa propriamente dito.

Esse guia é um instrumento prático, que organiza e resume os principais elementos envolvidos na avaliação de programas. Os seguintes passos são recomendados para o planejamento e implementação de uma avaliação de programas ou intervenção:

PASSO 1	Que decisões precisam ser tomadas e quem serão os tomadores de decisão?
PASSO 2	Definir quem vai fazer a avaliação
PASSO 3	Definir que indicadores serão medidos
	Que tipo de inferência é requerida?
	Dadas as respostas anteriores que tipo de estudo é mais apropriado e factível?
PASSO 4	Escrever o protocolo da avaliação
PASSO 5	Preparar os instrumentos de pesquisa, coleta e análise
PASSO 6	Elaborar relatório final
PASSO 7	Disseminar os resultados e as recomendações da avaliação

### **PASSO 1: Definir que decisões serão tomadas em decorrência da avaliação e quem serão os tomadores de decisão**

178

Em um mundo ideal, a avaliação estaria presente desde o planejamento de novos programas / intervenções. Avaliações de processo e de resultados seriam planejadas com o objetivo de responder perguntas importantes a respeito da implementação e, posteriormente, do impacto alcançado sobre a saúde das populações-alvo. Esse, no entanto, não é o caso na esmagadora maioria dos programas / intervenções. Além disso, quando realizadas, inúmeras avaliações não têm uma definição clara de a que decisões poderão levar ou quem tomará essas decisões. Não surpreende que pouca, se é que alguma ação resulte dessas avaliações.

Considere, por exemplo, um programa de controle de hipertensão arterial sistêmica. Que perguntas poderiam ser feitas e que decisões poderiam ser tomadas por diferentes atores?

Diferentes tomadores de decisão requerem diferentes tipos de informações e diferentes avaliações para tomarem suas decisões. A menos que esses aspectos estejam claros no início da avaliação, os resultados poderão ter nenhuma utilidade.

### Exemplo de perguntas que podem fazer diferentes tomadores de decisão em um Programa de Controle da Hipertensão Arterial conduzido em serviços de saúde

Tomador de decisão	Possível pergunta	Possível decisão
Chefe de Posto de Saúde	Meu estoque de medicamentos anti-hipertensivos é suficiente?	Solicitar envio regular de maior quantidade de anti-hipertensivos
Secretário Municipal de Saúde	A maioria dos hipertensos recebe os medicamentos regularmente?	Expandir o Programa Promover a utilização dos serviços
Secretário Estadual de Saúde	Quais as características do programa nos municípios com maiores coberturas?	Melhorar/expandir o Programa para todos os municípios Explorar possíveis causas de falha
Ministro da Saúde	O investimento está reduzindo a mortalidade por doenças cardiovasculares?	Continuar/interromper o Programa Expandir para outros estados/regiões
Cientista	O tratamento com anti-hipertensivos reduz a mortalidade entre maiores de 50 anos de idade?	Recomendar o Programa como uma intervenção efetiva

179

#### Perguntas para as partes interessadas na avaliação

Antes de planejar a avaliação, faça as seguintes perguntas aos atores interessados:

O que você gostaria de conseguir com este programa?

Quais são as perguntas fundamentais que esta avaliação deve responder?

Como você vai usar os resultados desta avaliação?

Que recursos você dispõe para a realização desta avaliação?

## PASSO 2. Definir quem vai fazer a avaliação

A escolha de avaliadores inapropriados pode ser frustrante. Competência técnica é essencial mas outras características também devem ser consideradas. Um excessivo “cientificismo” poderá ser uma desvantagem.

Características de um bom avaliador:
Ser tecnicamente competente
Desejar trabalhar junto com as pessoas do programa
Ser sensível às preocupações dos tomadores de decisão
Aceitar que a avaliação não é uma ciência exata (embora deva ser rigorosa)
Ser capaz de apresentar resultados no tempo previsto

Há prós e contras de se ter avaliadores INTERNOS (pessoas envolvidas no programa) versus EXTERNOS

180

AVALIADORES INTERNOS
<b>Vantagens</b>
Conhece a organização, o programa e o funcionamento
Compreende e pode interpretar comportamentos pessoais
Conhecido da equipe, não representa ameaça
Tem maior chance de adotar as recomendações
Barato, rápido
<b>Desvantagens</b>
É mais difícil ser objetivo, constrangimentos
Pode evitar chegar a conclusões negativas
Tende a aceitar os pressupostos da organização
Geralmente tem pouca disponibilidade de tempo
Falta de capacidade técnica e treinamento para avaliar

AVALIADORES EXTERNOS
<b>Vantagens</b>
Mais objetivo, não pessoalmente envolvido
Livre do viés da organização
Pode trazer novas perspectivas
Pode ter experiência em avaliação
Maior disponibilidade
Pode funcionar como facilitador entre as partes
Pode articular o contato da organização com outras fontes de recursos
<b>Desvantagens</b>
Pode não conhecer a organização, suas políticas, procedimentos e pessoal
Pode ser percebido como ameaça
Pode ser caro
Não vai acompanhar as recomendações dadas
Pode desconhecer a realidade local
Pode desconhecer as limitações à realização das recomendações

### PASSO 3. Definir quais indicadores serão medidos

Muitas avaliações tentam medir resultados complexos, de longo prazo, como o impacto do programa, sem antes avaliar se o programa alcançou seus objetivos mais imediatos. Isso resulta em um grande desperdício de recursos.

Um bom indicador deve ser:
<i>Quantitativo</i> : mensurável (percentagem, razão, número)
<i>Factual</i> : ter o mesmo significado para diferentes pessoas
<i>Válido</i> : mede o que se propõe a medir
<i>Confiável</i> : pode ser checado
<i>Sensível</i> : reflete as mudanças na situação

Um indicador é uma medida usada para demonstrar uma mudança ou o resultado de uma atividade, projeto ou programa. Um indicador deve ser mensurável, factual, válido, confiável e sensível.

Os indicadores devem ser construídos a partir dos objetivos, estrutura e funcionamento do programa ou intervenção.

O primeiro eixo do marco teórico (Habicht *et al*, 1999) define os tipos de indicadores, ou seja, quão longe se deseja ir na avaliação: indicadores de oferta, de utilização, de cobertura e de impacto. A definição desse tipos de indicadores encontra-se no quadro abaixo.

PRIMEIRO EIXO: TIPOS DE INDICADORES		
INDICADOR	PERGUNTA	DEFINIÇÕES
Oferta (“inputs”)	Os serviços ou atividades estão disponíveis? São acessíveis? A qualidade é adequada?	Primeiro os serviços devem ser oferecidos, disponíveis, acessíveis à população alvo e de qualidade adequada.
Utilização (“outputs”)	Os serviços estão sendo utilizados?	Segundo, a população precisa aceitar os serviços e utilizá-los.
Cobertura (“outcome”)	A população alvo está sendo atingida?	Terceiro, essa utilização resultará em uma dada cobertura populacional. A cobertura é uma medida particularmente útil, representando a interface entre serviços ou oferta de atividades (o processo gerencial) com a população (o quadro epidemiológico).
Impacto	Houve um resultado favorável?	Finalmente, a cobertura alcançada pode levar a um impacto sobre o desempenho, o comportamento ou a saúde.

## Exercício

Preencher o quadro abaixo com possíveis exemplos de indicadores para avaliação de:

- a) programa para promover atividade física entre adultos
- b) programa nutricional conduzido em escolas com o objetivo de prevenir a obesidade infantil

INDICADOR	PROJETO/PROGRAMA	
	Atividade física	Prevenção de obesidade infantil
Oferta		
Utilização		
Cobertura		
Impacto		

## Tipos de Indicadores: Possíveis Respostas ao Exercício

INDICADOR	PROJETO/PROGRAMA	
	Atividade física	Prevenção de obesidade infantil
Oferta	Número de grupos oferecidos por semana para os interessados em participar. Treinamento dos facilitadores e qualidade do aconselhamento (uso de referencial teórico na abordagem para mudar comportamento, conteúdos das reuniões etc).	Número de escolas oferecendo merenda escolar nutricionalmente balanceada para crianças de 6-10 anos de idade. Características nutricionais da merenda (quantidade de fibras, gordura, açúcar etc., por porção).
Utilização	Número de participantes das reuniões.	Número de escolas oferecendo merenda. Número de crianças participando do programa.
Cobertura	Percentual da população alvo que compareceu a 75% ou mais das seções.	Percentual de escolas oferecendo merenda do programa para 90% ou mais das crianças. Percentual de crianças que receberam pelo menos 80% das merendas oferecidas pelo programa.
Impacto	Diminuição da prevalência do sedentarismo na população alvo.	Redução na prevalência de obesidade infantil entre escolares.

184

### Observe que:

- A ordem lógica vai de oferta a impacto
- Problemas nos estágios iniciais da cadeia resultarão em falhas nos estágios mais tardios
- Diferentes “tomadores de decisão” estarão interessados em diferentes indicadores. Não existem indicadores “melhores” ou “piores”. Medir impacto não é necessariamente melhor do que medir se os serviços ou atividades foram oferecidos e utilizados.
- A escolha dos indicadores depende também de quanto se deseja gastar com a avaliação. Oferta e utilização podem ser avaliadas nos próprios serviços ou

usando os sistemas de informação de rotina. Cobertura ou impacto, no entanto, quase sempre requerem coleta de dados de campo, com importantes implicações de custo.

### Dúvidas comuns sobre o marco teórico:

Onde estão as avaliações de “processo” nesse marco teórico?

- As avaliações de processo cobrem dados de oferta, utilização e cobertura, diferente das avaliações de impacto.

### Dúvidas comuns...

Porque usar a terminologia “oferta → utilização → cobertura → impacto” invés de “input → output → outcome → impacto”?

- Oferta, utilização são às vezes chamadas de “inputs” e “outputs” e cobertura, de “outcome”. Não importa a terminologia desde que haja uma progressão lógica desde indicadores mais imediatos para os mais distais.

185

### Dúvidas comuns...

Como entram as avaliações “formativas” e “somativas” no marco teórico?

- O termo “formativa” refere-se a avaliações realizadas durante o desenvolvimento de uma intervenção, com o objetivo de afiná-la. Avaliações “somativas” referem-se àquelas realizadas depois que o programa ou projeto estiver sendo implementado. O marco teórico está primariamente preocupado com avaliações somativas.

## Dúvidas comuns...

### As avaliações qualitativas não estão sendo deixadas de lado?

- Há frequentemente uma certa confusão entre métodos qualitativos de pesquisa (como entrevistas aprofundadas, grupos focais etc) e avaliação da qualidade das atividades de um programa. Essa última pode ser avaliada através de métodos quantitativos (por ex: avaliação do conhecimento dos profissionais de saúde) ou qualitativos. A avaliação da qualidade dos serviços ou das atividades oferecidas é parte essencial das avaliações de “oferta”.

Embora o marco teórico se baseie em indicadores quantitativos, os métodos qualitativos são essenciais para complementar os dados quantitativos, especialmente quando as intervenções falham. Por exemplo, se a oferta de serviços ou de atividades for inapropriada, discussões com os profissionais do projeto e com os usuários podem ajudar a identificar os fatores limitantes e a sugerir alternativas.

186

## Dúvidas comuns...

### E sobre as análises de custo?

- Análises de custo são uma parte essencial de todas as avaliações. Os dados sobre custo podem ser divididos por indicadores de oferta, utilização, cobertura ou impacto para avaliar a eficiência e o custo-benefício da intervenção.

## Dúvidas comuns...

O marco teórico é apropriado somente para programas envolvendo serviços?

- Embora o marco teórico tenha sido originalmente formulado para tais programas, esse se adapta bem a outros tipos de intervenções. Por exemplo, o quadro abaixo mostra como o marco teórico pode ser usado para avaliar um projeto para promover atividade física.

PRIMEIRO EIXO: TIPOS DE INDICADORES		
INDICADOR	PERGUNTA	DEFINIÇÕES
Oferta (“inputs”)	Os serviços ou atividades estão disponíveis?	Quantos workshops ou cursos de treinamento/ conscientização/ sensibilização foram organizados?
	São acessíveis?	Essas atividades eram acessíveis a todas as pessoas, inclusive as que trabalham?
	A qualidade é adequada?	Os conteúdos e os facilitadores eram adequados?
Utilização (“outputs”)	Os serviços estão sendo utilizados?	Quantas pessoas participaram dessas atividades de promoção da atividade física?
Cobertura (“outcome”)	A população alvo está sendo atingida?	Qual a proporção de indivíduos da população alvo participando das atividades?
Impacto	Houve um resultado favorável?	Qual foi a tendência na proporção de indivíduos engajados em atividades físicas?

O tipo de indicador a ser usado depende também do estágio de desenvolvimento em que se encontra o programa. Quando um programa está em fase inicial de implementação, digamos, durante os dois primeiros anos de funcionamento, avaliações de processo são mais úteis, indicando se o planejado está sendo realizado a contento. Para avaliações de processo, poderão ser construídos indicadores de oferta, utilização e cobertura. Depois que o programa estiver maduro, operando por alguns anos, pode-se realizar avaliações somativas, de impacto. Os indicadores nesse tipo de avaliação deverão incluir oferta, utilização, cobertura e impacto. A coleta de informação sobre indicadores de processo em avaliações de impacto é importante para aumentar a plausibilidade de um resultado favorável ao programa. Mesmo nas avaliações de eficácia feitas em áreas demonstrativas, como recomendado pela estratégia CARMEN, as avaliações de impacto devem estar apoiadas em evidências de que o processo é satisfatório.

É importante lembrar que, embora as avaliações de impacto devam ser realizadas após um período de implementação, alguns delineamentos necessitam coleta de indicadores de linha de base, antes do início do programa. Tais delineamentos incluem os estudos tipo antes-e-depois, sem grupo controle, os quase-experimentos e os estudos randomizados que pretendam comparar mudanças ocorridas ao longo de um período entre o grupo intervenção e controle.

## Dúvidas comuns...

### Quais as diferenças entre utilização e cobertura?

- O quadro a seguir resume as diferenças entre indicadores de utilização e de cobertura.

	UTILIZAÇÃO	COBERTURA
Fonte de dados	Dados de rotina do programa/ projeto	Inquérito ( <i>survey</i> )
Custo da coleta de dados	Nenhum-Baixo	Intermediário-Alto
Frequência da coleta de dados	Pode ser freqüente	Pouco freqüente
Fornecer informação de base populacional	Somente se: a) a população alvo for bem definida b) a intervenção é única ou é uma meta* c) todos os provedores são monitorados d) natureza e qualidade dos dados de rotina são adequadas	Sempre
Fornecer informação sobre equidade (taxas para subpopulações)	Não	Sim
Exemplos	No. de cursos de aconselhamento nutricional realizados	% de agentes de saúde treinados para aconselhamento nutricional
	No. de sessões de atividade física realizadas	% de adultos que freqüentaram 3-5 sessões por semana durante 6 meses
	No. de obesos inscritos no programa de atividade física	% de obesos inscritos no programa de atividade física

(\*) Intervenção única é aquela que deve ocorrer somente uma vez para cada indivíduo, por exemplo, parto por parteira, dose única de BCG, matrícula no programa etc. Alguns sistemas podem também coletar dados sobre metas específicas, por ex., conclusão do primeiro grau, terceira dose de DTP, o sexto mês consecutivo de abstinência do fumo etc. Os dados de utilização podem ser extrapolados para uma população alvo somente se essa informação estiver registrada.

## Que Tipo de Inferência é Requerida?

- Sempre Reinventar a Roda?
- Frequentemente muitos avaliadores acham que devem demonstrar que uma intervenção cuja eficácia já foi comprovada teve um impacto, enquanto que seria suficiente apenas demonstrar que a intervenção foi conduzida adequadamente e que atingiu a população alvo.

## Qual o grau de certeza que você deseja ter?

- Segundo Eixo do Marco Teórico – Tipo de Inferência

INFERÊNCIA	PERGUNTA	COMENTÁRIOS
Adequação	Os objetivos do programa foram alcançados? Os resultados apontam para a direção esperada?	Uma simples avaliação se os objetivos iniciais, por ex., uma redução de um terço na prevalência de tabagismo, foram atingidos.
Plausibilidade	É provável que os objetivos tenham sido alcançados devido ao programa?	Pode-se razoavelmente afastar influências externas?
Probabilidade	Qual a probabilidade estatística de que o programa tenha tido um efeito?	Estritamente falando, uma análise de probabilidade requer um desenho experimental com alguns indivíduos ou comunidades sendo aleatoriamente alocados para receber a intervenção, enquanto indivíduos e comunidades controle não recebem.

189

## Exercício

Classifique as seguintes afirmativas conforme o tipo de inferência (adequação, plausibilidade, probabilidade) e o tipo de indicador medido (oferta, utilização, cobertura, impacto):

1. O objetivo de aumentar em 20% a participação de maiores de 50 anos nos grupos comunitários de promoção de atividade física foi alcançado.

190

2. O número de participantes foi substancialmente maior nos distritos onde houve programa de treinamento de facilitadores do que em distritos similares sem o programa.
3. O número de restaurantes oferecendo cardápios *light* aumentou em 15% no 5º ano de funcionamento do programa.
4. Foram construídas áreas de lazer em 100 comunidades. Em 50 delas, aleatoriamente selecionadas, um programa de atividade física foi implementado. As áreas de lazer nessas 50 comunidades foram usadas significativamente mais frequentemente, comparativamente às outras 50.
5. A prevalência de hipertensão em comunidades com programas de estilo de vida saudável foi 12% menor que em outras comunidades emparelhadas por tamanho, localização e tipo de clientela.
6. O número de adolescentes fumantes foi reduzido à metade em comparação a antes da implementação do programa.
7. Nas vilas em que o agente comunitário de saúde (ACS) participou do curso de treinamento sobre DCNTs houve 50-70% mais encontros de obesos sobre estratégias para mudança de hábitos alimentares do que em vilas similares em que o ACS não foi treinado.
8. Uma amostra representativa de domicílios de uma cidade foi aleatoriamente selecionada e o chefe da família entrevistado sobre o recebimento, pelo correio, de um livro de receitas nutricionalmente balanceadas. Uma vez que 20% dos entrevistados não receberam o livro, o objetivo de divulgação universal do programa não foi alcançado.
9. Todas as crianças da segunda série de uma escola pública foram listadas e metade delas, aleatoriamente selecionadas, receberam aulas de culinária para uma alimentação saudável. A prevalência de sobrepeso, ao final de um ano, foi 14% ( $p < 0,05$ ) menor do que o das crianças que não receberam as aulas.
10. Nos estados em que o programa foi implementado mais intensamente – conforme demonstrado pelo número de participantes nos encontros e pelo número de adesivos de nicotina distribuídos – houve um número substancialmente menor de mortes por câncer de pulmão do que nos demais estados do país.

## Respostas do Exercício

1. Participação de maiores de 50 anos: ADEQUAÇÃO, IMPACTO
2. No. de participantes: PLAUSIBILIDADE, COBERTURA
3. Restaurantes: ADEQUAÇÃO, OFERTA

4. Áreas de lazer e educação: PROBABILIDADE, UTILIZAÇÃO
5. Comunidades com programa: PLAUSIBILIDADE, UTILIZAÇÃO
6. Adolescentes fumantes: ADEQUAÇÃO, IMPACTO
7. Treinamento dos ACS: PLAUSIBILIDADE, OFERTA
8. Livros distribuídos: ADEQUAÇÃO, COBERTURA
9. Aulas de culinária: PROBABILIDADE, IMPACTO
10. Participantes e adesivos de nicotina: PLAUSIBILIDADE, IMPACTO

## Análises de Adequação

- São relativamente simples e devem ser feitas em todas as avaliações, mesmo naquelas com desenhos de plausibilidade ou probabilidade. Primeiro deve-se avaliar se os objetivos foram alcançados antes de tentar atribuí-los ao programa.
- É necessário haver objetivos claramente definidos, se possível quantitativamente, para realizar análises de adequação. Intervenções bem formuladas deverão conter esses objetivos. Se não disponíveis, uma discussão com os responsáveis pelo programa poderá ajudar a definir quais eram as expectativas e definir esses objetivos.
- A adequação pode ser avaliada em termos de se um objetivo absoluto (por ex., educação básica universal) ou um objetivo relativo (por ex., uma redução de 33% no tabagismo) foi alcançado. Objetivos relativos requerem a existência de dados de base.
- Se a eficácia da intervenção já foi comprovada, uma avaliação de adequação da cobertura é suficiente. Por exemplo, uma alta cobertura de vacina (com uma cadeia de frio funcional) ou uma alta cobertura de um projeto de fornecimento de água e esgoto terão quase que certamente um impacto sobre a saúde.
- As avaliações de adequação respondem à maioria das perguntas dos gerentes de programas em nível local, municipal, estadual e nacional.

Avaliações feitas através de estudos descritivos são as mais usadas, tanto por serem mais simples como por serem menos dispendiosas. As avaliações descritivas, como o nome indica, mostram associações entre o programa e os resultados observados, sem provar relação causa-efeito. Incluem-se nesse tipo de avaliação:

- **estudos de casos** (descrição do programa, dos participantes e dos resultados em um ponto no tempo ou em uma série temporal);

- **estudos transversais** (prevalências de fatores de risco e exposição ao programa medidos em um mesmo momento);
- **estudos ecológicos** (são estudos transversais cuja unidade de observação são grupos de indivíduos, como escolas, comunidades, serviços de saúde etc); e
- **séries temporais** (tendências seculares de prevalências e incidências de agravos à saúde, geralmente obtidos a partir de dados de vigilância epidemiológica).

## Análises de Plausibilidade

- Análises de plausibilidade são requeridas quando o tomador de decisões precisa se sentir seguro de que os efeitos observados foram devidos ao programa e não a fatores externos.
- Por exemplo, a mortalidade pode cair ou a prevalência de hipertensão diminuir devido a mudanças na sociedade que não eram relacionadas ao programa. Seria injusto atribuir ao programa essas conquistas que eram devidas a fatores externos.
- O oposto pode também acontecer. Mesmo que nenhuma mudança pudesse ser detectada nos indicadores, fatores externos poderiam ter levado a uma pior situação que o programa preveniu. Análises de plausibilidade podem ajudar a detectar um efeito tipo “barreira de proteção”. Em um exemplo hipotético, um programa educacional para promover alimentação saudável foi desenvolvido em uma região com alta prevalência de sobrepeso/obesidade. Na região controle, a prevalência de sobrepeso/obesidade e a oferta no mercado de alimentos eram semelhantes às da região intervenção antes da implementação do programa. Decorridos dois anos, uma avaliação do programa mostrou que as prevalências de sobrepeso/obesidade nas duas regiões mantiveram-se inalteradas, embora o mercado de alimentos na região intervenção tivesse piorado no último ano, com forte ingresso de alimentos industrializados. O efeito esperado da mudança no tipo de alimentos disponíveis na região intervenção seria de agravamento nas taxas de sobrepeso/obesidade. A intervenção, no entanto, criou uma barreira de proteção, evitando a piora das taxas de sobrepeso/obesidade.
- Plausibilidade requer algum tipo de comparação ou grupo controle. Esse pode ser:
  - **Grupo controle histórico:** as mesmas instituições ou populações são comparadas antes e depois do programa.

- **Grupo controle interno:** as instituições, áreas geográficas ou indivíduos atingidos pelo programa são comparados a outros que deveriam ter recebido o programa mas não o receberam porque não puderam ou porque recusaram.
- **Grupo controle externo:** as instituições ou áreas geográficas atingidas pelo programa são comparadas a outras que não foram planejadas para recebê-lo.
- Independente do tipo de grupo controle, o avaliador deve tentar descartar diferenças externas entre o grupo controle e o grupo beneficiário do programa.

Avaliações de plausibilidade podem ser feitas através do emprego de estudos do tipo:

- ***antes-e-depois*** (a mesma amostra ou uma amostra diferente da população alvo é avaliada antes e depois da implementação de um programa, servindo como seu próprio grupo controle);
- ***quase-experimentos*** (grupo intervenção e grupo controle não definidos por randomização e comparados quanto à medida do desfecho).

## Exercício

- Dê exemplos de possíveis fatores externos que poderiam influir sobre um programa destinado a promover uma alimentação saudável visando a redução da obesidade.

Fator Externo	Mudanças Esperadas Sobre a Prevalência de Obesidade

194

## Respostas do Exercício

### Possíveis fatores externos a serem descartados em uma análise de plausibilidade

Ao analisar o possível impacto de um programa destinado a promover uma alimentação saudável, o efeito de mudanças nos seguintes fatores externos devem ser afastados sempre que possível:

Fator externo	Mudanças esperadas sobre a prevalência de obesidade
Fatores socioeconômicos (renda <i>per capita</i> , renda familiar, escolaridade média, taxa de desemprego etc.)	Mudanças socioeconômicos estão estatisticamente associadas a taxas de obesidade
Idade	A taxa de obesidade tende a aumentar com o aumentar da idade
Sexo	De acordo com os fatores socioeconômicos, mulheres e homens de uma mesma faixa etária podem apresentar diferentes prevalências de obesidade
Atividade física	Mudanças nos níveis de atividade física podem reduzir a prevalência de obesidade
Melhora do aconselhamento nutricional em serviços de saúde	A melhora do aconselhamento nutricional em serviços de saúde pode reduzir a taxa de obesidade
Disponibilidade de alimentos saudáveis e oportunidades para optar por alimentos saudáveis	A sazonalidade de colheitas e a oferta de produtos saudáveis a preços aceitáveis podem reduzir a prevalência de obesidade

## Análises de Plausibilidade

### Observe que:

Os métodos para análises de plausibilidade incluem simulações matemáticas e análises de regressão multivariada.

Veja alguns exemplos de afirmações de plausibilidade de uma avaliação sobre um programa de combate ao tabagismo:

- A mortalidade por infarto caiu mais rapidamente durante o período do programa do que a mortalidade por outras causas.
- Uma simulação matemática mostrou que a diminuição observada na mortalidade por infarto não poderia ser explicada por mudanças nos fatores demográficos, socioeconômicos ou nutricionais, nem por outras intervenções.
- Uma comparação entre diferentes estados mostrou que aqueles com maiores coberturas do programa tiveram as menores mortalidades proporcionais por infarto.

## **Análises de Probabilidade**

Requerem locação aleatória das unidades de estudo para o grupo controle e intervenção. Por exemplo, todos os municípios ou todas as escolas em uma região seriam listadas e metade deles, aleatoriamente selecionados, receberiam o programa enquanto que a outra metade não. Testes estatísticos comparam então os indicadores de oferta, utilização, cobertura e impacto nos dois grupos, sendo calculada a probabilidade (valor  $p$ ) de que as diferenças observadas sejam devidas ao acaso.

Avaliações de probabilidade requerem estudos controlados randomizados (alocação aleatória dos grupos intervenção e controle).

### **Observe que:**

- Análises de probabilidade são importantes no desenvolvimento de novas intervenções. Por exemplo, para estabelecer que a redução no consumo de sal reduz a incidência de hipertensão arterial, ou no desenvolvimento de novos métodos de aconselhamento.
- Uma vez provada a eficácia da intervenção através desses estudos, as avaliações de campo usam normalmente desenhos de plausibilidade ou adequação.

196

A maioria dos cientistas são treinados em avaliações de probabilidade, podendo não ser sensíveis à necessidade de desenhos de adequação ou de plausibilidade.

## **Outras dúvidas comuns sobre o marco teórico:**

### **Testes estatísticos podem ser usados somente em avaliações de probabilidade?**

Não. Avaliações de plausibilidade frequentemente utilizam métodos estatísticos, incluindo análises multivariadas. A diferença é que as avaliações de probabilidade requerem randomização prévia para decidir onde os serviços ou atividades serão empreendidos.

## Sobre amostras aleatoriamente selecionadas?

Amostras aleatoriamente selecionadas (também chamadas amostras probabilísticas) são essenciais em várias avaliações de adequação e plausibilidade da cobertura e do impacto. Amostras aleatoriamente selecionadas não devem ser confundidas com alocação aleatória de unidades da intervenção e do grupo controle, que é a base dos desenhos probabilísticos.

## Que tipo de estudo é mais apropriado e factível?

### Observe que:

- ANÁLISES DE ADEQUAÇÃO na maioria das vezes requerem uma única medida e a comparação dos indicadores com metas preestabelecidas.
- ANÁLISES DE PLAUSIBILIDADE requerem grupo controle histórico, interno ou externo. Fatores externos também precisam ser descartados.
- ANÁLISES DE PROBABILIDADE requerem alocação aleatória das unidades de estudo para receberem ou não a intervenção.

197

## Considerações de custo na escolha do desenho da avaliação

Frequentemente, o montante disponível para avaliação é limitado.

### Observe que:

- Avaliações de oferta e utilização são as mais baratas porque podem utilizar dados de rotina.
- Avaliações de cobertura geralmente requerem pesquisas de campo mas os tamanhos de amostra são relativamente razoáveis
- Avaliações de impacto também requerem estudos de campo mas os tamanhos de amostra são frequentemente maiores que os anteriores
- Avaliações de adequação geralmente são mais baratas do que avaliações de plausibilidade porque não necessitam grupo controle nem análise de dados complexa.

## Combinando Diferentes Delineamentos

Eixos	Oferta	Utilização	Cobertura	Impacto
Adequação	1º	2º	3º	4º (b)
Plausibilidade			4º (a)	5º
Probabilidade				

Deve-se iniciar avaliando a adequação da oferta, depois a adequação da utilização e da cobertura. O quarto nível poderia ser tanto a avaliação de adequação do impacto quanto de plausibilidade da cobertura e, finalmente, poderia ser avaliada a plausibilidade de um impacto. De acordo com o marco teórico, avaliações de probabilidade devem ser realizadas menos freqüente, ficando reservadas para, quando possível, determinar a eficácia de novas intervenções, antes de sua implementação em larga escala. Para a maioria das tomadas de decisão, são suficientes os resultados de avaliações não probabilísticas.

### Observe que:

198

As avaliações mais baratas devem ser feitas primeiro e que as mais complexas não devem ser realizadas sem que as avaliações anteriores mostrem que o programa está funcionando bem.

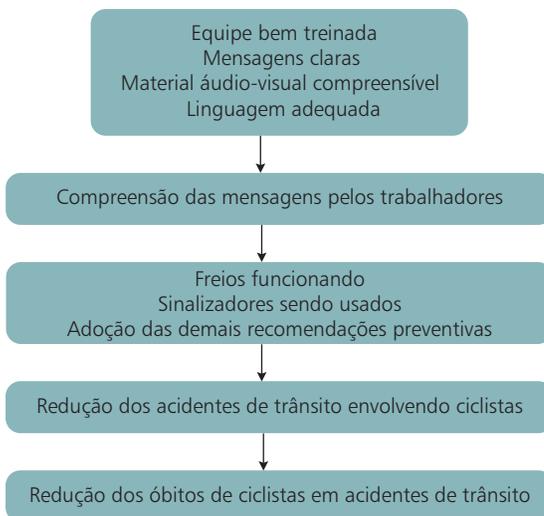
## PASSO 4. Com base nos dados anteriores, escrever o protocolo da avaliação

Um protocolo de avaliação, como um projeto de pesquisa, deve ser detalhado, buscando antever todas as necessidades que poderão surgir ao longo do estudo. A revisão da literatura disponível sobre avaliações anteriores de programas e intervenções pode ser inspiradora e antecipar problemas encontrados por outros avaliadores. Estude bem o programa a ser avaliado para conhecer os objetivos, metas, métodos, recursos materiais e humanos, sistema e tipo de registro.

Em seguida, junto com os gestores e executores do programa, construa a cadeia causal (lógica) através da qual espera-se que o programa alcance as metas planejadas. A construção da cadeia lógica será de grande auxílio para que você crie os indicadores da avaliação e identifique outros fatores que poderão atuar como confundidores ou

modificadores de efeito do programa. A identificação de potenciais confundidores e modificadores de efeito são importantes para que se planeje a coleta de informações e para a fase de análise e interpretação dos resultados. Considere, por exemplo, uma intervenção educacional planejada para reduzir acidentes de trânsito, dirigida a trabalhadores que se deslocam de bicicleta. A intervenção controlada consistia em reunir grupos de cerca de vinte ciclistas de cada vez, apresentar um filme rodado na cidade do estudo, mostrando situações de perigo e como preveni-las. Deslocamento em fila indiana e na mesma mão que os automóveis na via pública, sinalização manual antes de virar as esquinas, respeito aos sinais de trânsito e semáforos, uso de campanha e espelho retrovisor, entre outras, eram mensagens veiculadas. Os trabalhadores eram informados sobre equipamentos de segurança para bicicletas e ciclistas e recebiam revisão grátis ou aplicação de freios nas próprias bicicletas e sinalizadores para os aros. Um exemplo de cadeia causal ligando a intervenção ao resultado esperado (redução de acidentes de trânsito e óbitos entre ciclistas) é mostrado na figura a seguir.

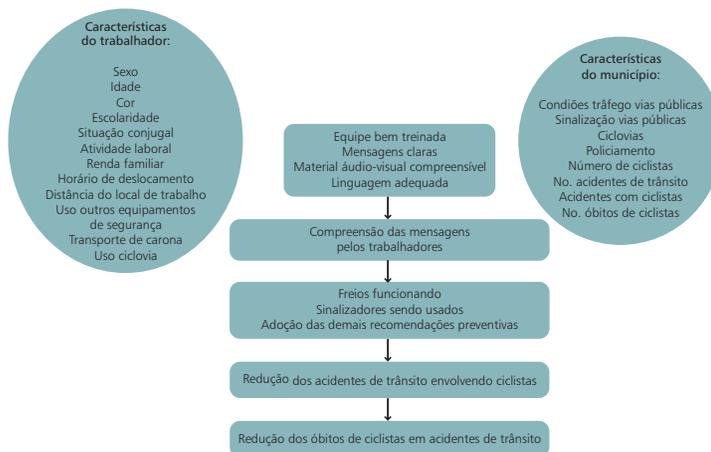
### Cadeia causal ligando a intervenção ao resultado esperado



Para que a intervenção seja bem sucedida, é necessário que a equipe que fará a intervenção esteja bem treinada, as mensagens sejam claras e objetivas, o material áudio-visual seja de boa qualidade e compreensível e que a linguagem empregada seja apropriada ao grupo de trabalhadores. Atendidas essas premissas, os trabalhadores deverão compreender o conteúdo das mensagens e, conseqüentemente aumentar seu nível de conhecimento sobre prevenção de acidentes de trânsito. O aumento do conhecimento deverá promover a adesão ao comportamento preventivo e uso dos sinalizadores, o que, juntamente, com a adequação de funcionamento dos freios, deverá reduzir o número de acidentes envolvendo ciclistas e a conseqüente diminuição de óbitos por essa causa no município.

Mas, a ocorrência de acidentes de trânsito tem causalidade multifatorial e o efeito da intervenção é apenas um dentre os fatores implicados em sua ocorrência. A figura a seguir apresenta outras variáveis do próprio indivíduo e do município que poderão confundir ou modificar o efeito da intervenção.

200



Essa rede de causalidade será muito útil no momento de dar o PASSO 5 desse manual. Deverão ser coletadas informações que possam identificar se a cadeia causal da intervenção ocorreu e sobre características dos indivíduos e da comunidade. Na análise, os grupos intervenção e controle serão comparados quanto a essas características. Se ambos forem semelhantes quanto a essas variáveis e apenas a in-

tervenção for o fator que distinga os dois grupos, então as diferenças de incidência de acidentes e óbitos de ciclistas poderá ser atribuída em maior grau de certeza à própria intervenção.

Defina os objetivos da avaliação e decida com os tomadores de decisão o delineamento do estudo (considerando objetivos da avaliação, tempo e recursos disponíveis) que melhor se adéque às necessidades. Todos os delineamentos utilizados na epidemiologia são aplicáveis às avaliações de desempenho e impacto de programas. Uma descrição detalhada do planejamento de estudos especificamente voltados para avaliação de programas pode ser encontrados nos capítulos anteriores deste documento.

### **Um bom protocolo de avaliação deve incluir:**

- Uma breve descrição do projeto ou programa: O que está sendo avaliado?
- Os objetivos da avaliação: O que vai ser feito?
- A metodologia: Como será feito?
- O grupo de avaliação: Quem vai fazer?
- O cronograma: Quando será feito?
- O orçamento: Quanto custará?
- O relatório final: Como serão apresentados os resultados e para quem?

201

O protocolo deve explicitar quem são o coordenador e a equipe responsável pela avaliação, com a definição de papéis e atribuições. O cronograma de execução da avaliação deve ser previsto e, posteriormente, respeitado. Quanto custará a avaliação é um outro item do protocolo, lembrando que estudos com delineamentos mais complexos e indicadores mais distais costumam ser mais dispendiosos. A complexidade dos delineamentos cresce desde estudos transversais, passando por estudos de casos e controles, estudos tipo antes-e-depois sem grupo controle, quase-experimentos e ensaios randomizados. Estudos a partir de dados secundários são em geral, menos dispendiosos do que aqueles com dados primários. Dados secundários são informações originalmente coletadas para outros fins e que passam a ser utilizados na avaliação. Séries temporais que utilizam informações de vigilância em fatores de risco e doenças cardiovasculares e cânceres, medindo a tendência secular, são exemplos de avaliações com dados secundários. Indicadores de oferta e utilização são mais baratos de serem obtidos do que indicadores de cobertura e impacto, que, via de regra, exigem a coleta de dados populacionais e primários.

O cronograma e o orçamento devem contemplar os seguintes elementos: planejamento da avaliação, desenvolvimento dos instrumentos de pesquisa, teste piloto dos instrumentos de pesquisa, coleta de dados, processamento de dados (codificação, entrada de dados), análise dos dados, redação do relatório final e disseminação dos resultados.

## **PASSO 5. Preparar os instrumentos de pesquisa, coleta e análise**

A execução do trabalho propriamente dito precisa ser planejada em detalhes. Se a avaliação incluir a coleta de dados populacionais e uma amostra tiver que ser selecionada, os procedimentos necessários para que essa seja representativa da população de origem devem ser fixados antes do início do trabalho de campo. Os instrumentos a serem utilizados deverão ser preparados com antecedência, cobrir todas as questões importantes para a pesquisa e testados quanto a sua adequação e compreensão. A elaboração de manuais de instrução para preenchimento de questionários e uso de equipamentos da pesquisa são de grande auxílio durante o trabalho de campo.

202

Os equipamentos de campo deverão ser listados, adquiridos e testados com antecipação. Pode ser necessário utilizar balanças, antropômetros, fitas inextensíveis, estetoscópios, esfigmomanômetros, glicosímetros, tiras reagentes para glicose capilar, seringas e agulhas descartáveis etc.

- Planejar procedimentos de amostragem
- Preparar questionários e formulários
- Reunir equipamento para o campo
- Treinar a equipe de pesquisa
- Fazer os arranjos para a logística de campo
- Verificar os arranjos para entrada de dados e análise

O adequado treinamento do coordenador, dos entrevistadores e dos supervisores de campo é vital para obtenção de dados de qualidade. A leitura de questionários e de manuais, bem como simulações de entrevistas devem ser feitas durante o treinamento. O uso adequado dos equipamentos e a leitura de seus resultados devem também realizados durante o treinamento. Todo o investimento necessário de tempo e esforço deve ser feito para que, ao final do treinamento, a equipe de campo esteja adequada e homoganeamente treinada. A seleção definitiva dos trabalhadores

de campo poderá ser feita ao final do treinamento entre aqueles que mostrarem desempenho adequado para a realização do trabalho de campo.

- Os arranjos para a logística de campo (se aplicável), incluindo a obtenção de autorização das autoridades locais, preparação de cartas de apresentação e crachás, distribuição dos trabalhadores de campo, transporte etc, devem estar prontos antes do início do campo.
- O coordenador deve estar preparado para revisar os questionários e os procedimentos realizados no campo e reunir a equipe semanalmente para corrigir e prevenir possíveis erros. A entrada de dados deve ser planejada e, se possível, iniciada logo após o início do campo.
- Uma descrição detalhada dos procedimentos necessários para a organização do trabalho de campo e outros aspectos práticos de estudos de base populacional pode ser encontrado em *Barros FC, Victora CG. Epidemiologia da saúde Infantil: um manual para diagnósticos comunitários. São Paulo: Hucitec-Unicef, 1998.* Como o indica o título, trata-se de um manual para estudos em saúde infantil mas os princípios subjacentes são aplicáveis a estudos com outros grupos populacionais e podem ser de grande auxílio para os avaliadores.

## **PASSO 6. Escrever o relatório final**

A redação do relatório final deve ser feita de forma clara e objetiva de forma a consistir em um real auxílio para o tomador de decisões. Deve conter um Resumo Executivo, que é um resumo mais extenso do que os escritos para artigos publicados nas revistas científicas; a metodologia deve ser descrita detalhadamente para que o leitor possa compreender exatamente como a avaliação foi feita; a seção de Resultados deve conter análises principais e de subgrupos e, sempre que possível, apresentar os achados na forma de gráficos que facilitem a compreensão do leitor; e as recomendações devem ser factíveis e apresentadas em ordem de prioridade.

O quadro a seguir apresenta um *checklist* a ser seguido para a elaboração de um relatório final.

<b>CHECKLIST PARA A ELABORAÇÃO DE RELATÓRIO DE AVALIAÇÃO</b>	
<b>Resumo Executivo</b>	
Extensão	2-3 páginas, espaço simples.
Conteúdo	Deve incluir contexto, problema, metodologia, resultados e recomendações.
Coerência	Todos os elementos devem ser consistentes com o conteúdo do relatório.
Abrangência	Todos os aspectos e resultados importantes do relatório devem ser incluídos no Resumo.
<b>INTRODUÇÃO</b>	
Background	Deve incluir informações relevantes sobre geografia, demografia, economia e política (se aplicáveis), bem como a descrição detalhada do projeto ou programa avaliado.
Colocação do problema	Deve apresentar a justificativa específica para a realização da avaliação.
Objetivos	Os objetivos devem ser claramente apresentados e ser relacionados com o problema acima colocado.
Factibilidade	Os objetivos devem ser realísticos realizáveis.
Estudo proposto	As ações específicas para alcançar cada um dos objetivos devem ser claramente descritas.
Organização	Todos os aspectos introdutórios devem ser colocados na seção de Introdução e não ao longo do relatório.
<b>Metodologia</b>	
Coerência	Os métodos devem ser coerentes com os objetivos apresentados na Introdução.
Escolha de indicadores e delineamento	Os indicadores usados devem ser claramente descritos e justificados à luz dos objetivos da avaliação. O delineamento do estudo também deve ser justificado.
Amostragem	Os critérios específicos de seleção da amostra devem ser descritos, incluindo a adequação do tamanho da amostra e o processo de amostragem.

Instrumentos	Todos os instrumentos (questionários, formulários de coleta de dados, equipamentos) devem ser claramente descritos e, se possível, incluídos como apêndice. A validade e confiabilidade dos instrumentos deve ser discutida.
Equipe de pesquisa	A formação, seleção e o treinamento dos entrevistadores e outros membros da equipe devem ser descritos.
Procedimentos	O relatório deve incluir uma descrição detalhada dos procedimentos de campo incluindo entrevistas, mensurações etc.
Vieses e controle de qualidade	As potenciais fontes de vieses (incluindo vieses de seleção, informação, do entrevistador, de medidas etc) devem ser reconhecidas e descritas as ações tomadas para minimizá-los.
Limitações	Todas as limitações importantes na implementação da avaliação, relativas ao planejamento inicial, devem ser discutidas abertamente.
<b>Resultados</b>	
Coerência	A fonte de cada grupo de resultados e como foram obtidos a partir da metodologia usada devem ser claramente apresentados.
Compleição	Os dados apresentados devem fazer uso completo da amostragem, instrumentos e procedimentos empregados. Sempre que disponíveis, dados de subgrupos devem ser apresentados.
Relevância da interpretação	A explicação dos resultados da avaliação deve ser consistente com os dados reais e com os métodos.
Relevância programática	Deve ser feita uma clara descrição dos insights importantes para o projeto ou programa oriundos da avaliação.
Importância potencial	A relevância potencial local e, se aplicável, para outros locais e até mesmo internacional deve ser descrita.
<b>Recomendações</b>	
Coerência	As recomendações devem ser baseadas em dados reais a ser coerentes com os objetivos da avaliação.
Foco	Devem ser explícitas, concretas e enfocadas no tomador de decisões específico.
Prioridade	O relatório deve definir a prioridade relativa das diferentes recomendações.
Factibilidade	Tempo e custo devem ser considerados entre as recomendações.
<b>ESTILO</b>	
Extensão	A extensão ideal é de não mais de 50 páginas, espaço simples, não incluindo os anexos.

Formato	Títulos, subtítulos e gráficos devem ser usados. O relatório deve ser bem organizado e ter bom impacto visual.
Clareza	As palavras devem ser simples e as frases curtas. Uma estrutura lógica racional deve ser seguida.
<b>Aspectos Gerais</b>	
Conteúdo	Todas as seções devem ser igualmente desenvolvidas, sem estender uma seção em detrimento de outras.
Abrangência	Todos os objetivos apresentados inicialmente devem levar aos correspondentes métodos, resultados e recomendações.
Acurácia	O relatório deve conter apenas informações, resultados e afirmativas válidas. Vieses pessoais não devem afetar a apresentação ou interpretação dos dados.
Relevância	Insights reais para a identificação e/ou solução de problemas importantes devem ser buscados.
Criatividade	Sempre que possível, o relatório deve ser criativo quanto a abordagens técnicas, interpretação de resultados (esperados e não esperados) e recomendações.

206

## **PASSO 7. Disseminar os resultados e as recomendações da avaliação**

Os resultados e recomendações deverão ser divulgados para os seguintes grupos:

**Tomadores de decisão:** aqueles que tomarão decisões em consequência dos resultados e recomendações da avaliação constituem o grupo primário a quem o relatório deve ser dirigido. Esses deverão receber uma cópia impressa do relatório completo e os resultados discutidos individualmente e em grupos de trabalho.

**Comunidades:** as comunidades envolvidas na avaliação devem receber os resultados através da mídia, material impresso ou reuniões locais. Se relevantes, os resultados devem ser disseminados para toda a população não apenas para as comunidades envolvidas.

**Organizações profissionais:** publicações em revistas científicas ou de categorias profissionais e apresentações em congressos e reuniões científicas são estratégias para atingir esses grupos.

# Referências Bibliográficas

1. ALAD 2000 (Guias). *For diagnosis and management of type-2 diabetes mellitus with evidence-based medicine*. Available at <http://www.alad.org/gulasalad.html> [9ago2004].
2. ALTMAN D, SCHULZ KF, MOHER D, EGGER M, DAVIDOFF F, ELBOURNE D, GÖTZSCHE PC, LANG T. *The revised CONSORT Statement for reporting randomized trials: explanation and elaboration*. *Ann Intern Med* 2001;134:663-694.
3. ALTMAN DG. *Practical statistics for medical research*. London: Chapman & Hall, 1991.
4. AMERICAN COLLEGE OF SPORTS MEDICINE. *Position stand: the recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness and flexibility in healthy adults*. *Med Sci Sports Exerc* 1998;30:975-991.
5. AMERICAN DIABETES ASSOCIATION. *Standards in medical care in diabetes*. *Diabetes Care* 2004;27(S1):S15-S35.
6. ARAÚJO RB, SANTOS I, CAVALETI MA, COSTA JSD, BÉRIA JU. *Avaliação do cuidado prestado a pacientes diabéticos em nível primário*. *Rev Saúde Pública* 1999;33:24-32.
7. ARNETT DK, McGOVERN PG, JACOBS DR, SHAHAR E, DUVAL S, BLACKBURN H, LUEPKER RV. *Fifteen-year trends in cardiovascular risk factors (1980-1982 through 1995-1997)*. The Minnesota Heart Health Program. *Am J Epidemiol* 2002;156:929-935.
8. ASSUNÇÃO MC, SANTOS IS, Costa JSD. *Avaliação do processo da atenção médica: adequação do tratamento de pacientes com diabetes mellitus*, Pelotas, Rio Grande do Sul, Brasil. *Cad Saúde Pública* 2002;18:205-211.
9. ASSUNÇÃO MC, SANTOS IS, GIGANTE DP. *Atenção primária em diabetes no sul do Brasil: estrutura, processo e resultado*. *Rev Saúde Pública* 2001;35:88-95.

10. ASSUNÇÃO MC, SANTOS IS, VALLE NCJ. *Blood glucose control in diabetes patients seen in primary health care centers*. Rev Saúde Pública 2005;39:183-190.
11. BANDURA A. *Social foundations of thought and action: a social cognitive theory*. Englewood Cliffs: Prentice Hall, 1986.
12. BARROS AJ & HIRAKATA VN. *Alternatives for logistic regression in cross-sectional studies: na empirical comparison of models that directly estimate the prevalence ratio*. BMC Med Res Methodol 2003;3:21.
13. BARROS FC, VICTORA CG. *Epidemiologia da saúde Infantil: um manual para diagnósticos comunitários*. São Paulo: Hucitec-Unicef, 1998.
14. BEAGLEHOLE R, YACH D. *Globalisation and the prevention and control of non-communicable disease: the neglected chronic diseases of adults*. Lancet, 2003;362:903-8.
15. BERRÍOS X, BEDREGAL GP, GUZMÁN AB. *Costo-efectividad de la promoción de la salud en Chile. Experiencia Del programa "Mirame!"*. Rev Méd Chile, 2004;132:361-70.
16. BERRIOS X, KOPONEN T, HUIGUANG T, KHALTAEV N, PUSKA P. *Distribution and prevalence of major risk factors of noncommunicable diseases in selected countries: the WHO Inter-Health Programme*. Bull WHO 1997, 75(2):99-108.
17. BLACK N. *Why we need observational studies to evaluate the effectiveness of health care*. BMJ 1996;312:1215-1218.
18. CARLETON RA, LASATER TM, ASSAF AR, FELDMAN HA, MCKINLAY S, Pawtucket Heart Health Program Writing Group. *The Pawtucket Heart Health Program: community changes in cardiovascular risk factors and projected disease risk*. Am J Public Health 1995;85:777-785.
19. CENTER OF DISEASE CONTROL AND PREVENT (CDC). Available at: <http://www.cdc.gov/nccdphp/publications/aag/cvh.htm> [3mar2006]

20. COUNTRYWIDE INTEGRATED NONCOMMUNICABLE DISEASE INTERVENTION PROGRAM. Cindi Highlights 2004. Available at [http://www.euro.who.int/healthy-cities/publications/20050201\\_1](http://www.euro.who.int/healthy-cities/publications/20050201_1) [3mar2006].
21. CORNFIELD J. *Randomization by group: a formal analysis*. Am J Epidemiol 1978;108:100-102.
22. CUTTER J, TAN BY, CHEW SK. *Levels of cardiovascular disease risk factors in Singapore following a national intervention programme*. Bull WHO 2001;79:908-915.
23. DES JARLAIS DC, LYLES C, CREPAZ N, TREND GROUP. *Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND Statement*. Am J Public Health 2004;94:361-366.
24. DIAS-DA-COSTA JS, HALLAL PC, WELLS JC, DALTOE T, FUCHS SC, MENEZES AM, OLINTO MT. *Epidemiology of leisure-time physical activity: a population-based study in southern Brazil*. Cad Saúde Pública 2005;21:275-282.
25. DIEHL HA. *Coronary risk reduction through intensive community-based lifestyle intervention: the Coronary Health Improvement Project (CHIP) experience*. Am J Cardiol 1998;82:83T-87T.
26. DOMINGUES MR, ARAÚJO CL, GIGANTE DP. *Knowledge and perceptions of physical exercise in an adult urban population in southern Brazil*. Cad Saúde Pública 2004; 20:204-215.
27. DONABEDIAN A. *La Calidad de la Atención Médica: definición y métodos de evaluación*. México: La Prensa Médica Mexicana, S.A., 1984.
28. DOWNS SH, BLACK N. *The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomised studies of health care interventions*. J Epidemiol Community Health 1998;52:377-384.
29. DOWSE G, ZIMMET P. *A model protocol for diabetes and other noncommunicable disease field surveys*. Health Stat Q 1992;45:360-372.

30. DUNN AL, MARCUS BH, KAMPERT JB, GARCIA ME, KOHL HW, BLAIR SN. *Reduction in cardiovascular disease risk factors: 6-month results from Project Active*. Preventive Medicine 1997;26:883-892.
31. EATON CB, LAPANE KL, GARBER CE, GANS KM, LASATER TM, CARLETON RA. *Effects of a community-based intervention on physical activity: The Pawtucket Heart Health Program*. Am J Public Health 1999;89(11):1741-1744.
32. EBRAHIM S, SMITH GD. *Systematic review of randomised controlled trials of multiple risk factor interventions for preventing coronary heart disease*. BMJ, 1997;314:1666-1669.
33. ELMER PJ, GRIMM R, LAING B, GRANDITS G, SVENDSEN K, VAN HEEL N, BETZ E, RAINES J, LINK M, STAMLER J, NEATON J. *Lifestyle intervention: results of the Treatment Mild Hypertension Study (TOMHS)*. Prev Med 1995;24:378-388.
34. EPI INFO. *A word processing, database, and statistics system for epidemiology on computers*. Centers for Disease Control: Division of surveillance and epidemiologic studies, Epidemiology program office. Atlanta, Georgia, 2005 (versão3.3.2).
35. FARQUHAR JW, FORTMANN SP, MACCOBY N, HASKELL WL, WILLIAMS PT, FLORA JA, TAYLOR CB, BROWN, SOLOMON DS, HULLEY SB. *The Stanford Five-City Project: design and methods*. Am J Epidemiol 1985;122:323-324.
36. FLEMING NS, BECKER ER. *The impact of the Texas 1989 motorcycle helmet law on total and head-related fatalities, severe injuries, and overall injuries*. Med Care 1992;30:832-845.
37. FLETCHER RH, FLETCHER SW, WAGNER EH. *Epidemiologia Clínica: elementos essenciais*, 3a. edição. Porto Alegre: Artes Médicas, 1996.
38. FORTMANN SP, FLORA JA, WINKLEBY MA, SCHOOLER C, TAYLOR CB, FARQUHAR JW. *Community intervention trials: reflections on the Stanford Five-City Project experience*. Am J Epidemiol 1995;142(6):576-586.

39. FORTMANN SP, HASKEL WL, WILLIAMS PT, VARADY AN, HULLEY SB, FARQUHAR JW. *Community surveillance of cardiovascular diseases in the Stanford Five-City Project: Methods and initial experience*. Am J Epidemiol 1986;123(4):656-669.
40. FORTMANN SP, TAYLOR CB, FLORA JA, JATULIS DE. *Changes in adult cigarette smoking prevalence after 5 years of community health education: The Stanford Five-City Project*. Am J Epidemiol 1993;137:82-96.
41. FORTMANN SP, TAYLOR CB, FLORA JA, WINKLEBY MA. *Effect of community health education on plasma cholesterol levels and diet: The Stanford Five-City Project*. Am J Epidemiol 1993;137:1039-55.
42. FORTMANN SP, VARADY NA. *Effects of a community-wide health education program on cardiovascular disease morbidity and mortality. The Stanford Five-City Project*. Am J Epidemiol 2000;152(4):316-322.
43. FORTMANN SP, WILLIAMS PT, HULLEY SB, HASKELL WL, FARQUHAR JW. *Effect of health education on dietary behavior: the Stanford Three Community Study*. Am J Clin Nutr 1981;34:2030-2038.
44. FORTMANN SP, WILLIAMS PT, HULLEY SB, MACCOBY N, FARQUHAR JW. *Does dietary health education reach only the privileged?* Circulation 1982;66:77-82.
45. FORTMANN SP, WINKLEBY MA, FLORA JA, HASKELL WL, TAYLOR CB. *Effect of long-term community health education on blood pressure and hypertension control*. Am J Epidemiol 1990;132:629-46.
46. GANS KM, ASSMANN SF, SALLAR A, LASATER TM. *Knowledge of cardiovascular disease prevention: an analysis from two New England communities*. Prev Med 1999;29:229-237.
47. GCP STUDY GROUP. *The German Cardiovascular Prevention Study (GCP): design and methods*. Eur Heart J 1988;9:1056-1066.
48. GOLDSTEIN H. *Multilevel statistical models*. London: Hodder Arnold, 2003.

49. GWATKIN DR, BHUIYA A, VICTORA C. *Making health systems more equitable*. Lancet 2004;364:1273-1280.
50. HABICHT JP, VICTORA CG, VAUGHAN JP. *Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact*. Intern J Epidemiology 1999;28:10-18.
51. HALLAL PC, AZEVEDO MR, REICHERT FF, SIQUEIRA FV, ARAÚJO CL, VICTORA CG. *Who, when, and how much? Epidemiology of walking in a middle-income country*. Am J Prev Med 2005;28:156-161.
52. HALLAL PC, VICTORA CG, WELLS JC, LIMA RC. *Physical inactivity: prevalence and associated variables in Brazilian adults*. Med Sci Sports Exerc 2003;35:1894-1900.
53. HOVLAND CI, JANIS IL, KELLEY NH. *Communication and persuasion*. New Haven, CT: Yale University Press, 1953.
54. HUNT MK, LEFEBVRE C, HIXSON ML, BANSPACH SW, ASSAF AR, CARLETON RA. *Pawtucket Heart Health Program point-of-purchase nutrition education program in supermarkets*. Am J Public Health 1990;80(6):730-732.
55. JADUE L, VEGA J, ESCOBAR MC, DELGADO I, GARRIDO C, LASTRA P, ESPEJO F, PERUGA A. *Factores de riesgo para las enfermedades no transmisibles: metodología y resultados globales de la encuesta de base del programa Carmen (Conjunto de Acciones para la Reducción Multifactorial de las Enfermedades No transmisibles)*. Rev Med Chile 1999;127(8):1004-13.
56. JOUSILAHTI P, TUOMILEHTO J, KORHONEN HJ, VARTIAINEN E, PUSKA P, NISSINEN A. *Trends in cardiovascular disease risk factor clustering in eastern Finland: results of 15-year follow-up of the North Karelia Project*. Prev Med 1994;23:6-14.
57. KELDER SH, PERRY C, KLEPP KI, LYTTLE LL. *Longitudinal tracking of adolescent smoking, physical activity, and food choice behaviors*. Am J Public Health 1994;84:1121-1126.

58. KELDER SH, PERRY CL, KLEPP KI. *Community-wide youth exercise promotion: long-term outcomes of the Minnesota Heart Health Program and the Class of 1989 study*. J School Health 1993;63(5):218(6).
59. KLEINBAUM DG, KUPPERLL, MORGENSTERN H. *Epidemiologic Research: principles and quantitative methods*. New York: van Nostrand Reinhold, 1982.
60. KORHONEN M, KASTARINEN M, UUSITUPA M, PUSKA P, NISSINEN A. *The effect of intensified diet counseling on the diet of hypertensive subjects in primary health care: a 2-year open randomized controlled trial of lifestyle intervention against hypertension in eastern Finland*. Prev Med 2003;36:8-16.
61. KRAUS JF, PEEK C, MCARTHUR DL, WILLIAMS A. *The effect of the 1992 California motorcycle helmet use law on motorcycle crash fatalities and injuries*. JAMA 1994; 272:1506-1511.
62. KORHONEN T, SU S, KORHONEN HJ, UUTELA A, PUSKA P. *Evaluation of a national Quit and Win Contest: determinants for successful quitting*. Prev Med 1997;26:556-564.
63. KORHONEN M, KASTARINEN M, UUSITUPA M, PUSKA P, NISSINEN A. *The effect of intensified diet counseling on the diet of hypertensive subjects in primary health care: a 2-year open randomized controlled trial of lifestyle intervention against hypertension in Eastern Finland*. Prev Med 2003;36:8-16.
64. LANDO HA, PECHACEK TF, PIRIE PL, MURRAY DM, MITTELMARK MB, LICHTENSTEIN E, NOTHWEHR F, GRAY C. *Changes in adult cigarette smoking in the Minnesota Heart Health Program*. Am J Public Health 1995;85(2):201-208.
65. LAST JM. *A dictionary of epidemiology*. 4th edition. New York: Oxford University Press, 2001.
66. LUEPKER RV, MURRAY DM, JACOBS DR, MITTELMARK MB, BRACHT N, CARLAW R, CROW R, ELMER P, FINNEGAN J, FOLSOM AR, GRIMM R, HANNAN PJ, JEFFREY R, LANDO H, MCGOVERN P, MULLIS R, PERRY CL, PECHACECK T, PIRIE P, SPRAFKA M, WEISBROD R, BLACKBURN H. *Community education for cardiovascular disease prevention: risk factor changes*

- in the Minnesota Heart Health Program. Am J Public Health 1994;84:1383-1393.*
67. LUEPKER RV, RASTAM L, HANNAN PJ, MURRAY DM, GRAY C, BAKER WL, CROW R, JACOBS DR, PIRIE PL, MASCIOLI SR, MITTELMARK MB, BLACKBURN H. *Community education for cardiovascular disease prevention: morbidity and mortality for cardiovascular disease prevention results from the Minnesota Heart Health Program. Am J Epidemiol 1996;144(4):351-362.*
  68. McGUIRE WJ. *Persuasion, resistance, and attitude change.* In: DeSola I, Schramm W, eds. *Handbook of communication.* Chicago: Rand McNally & Company, 1973.
  69. McKINLAY J, MARCEAU L; *US public health and the 21st century: diabetes mellitus. Lancet 2000;356:757-761.*
  70. McQUEEN DV. *The evidence debate* (editorial). *J Epidemiol Community Health 2002;56:83-84.*
  71. MINISTÉRIO DA SAÚDE. *Diabetes mellitus: guia básico para diagnóstico e tratamento.* Brasília (DF): Ministério da Saúde, 1997.
  72. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE. DEPARTAMENTO DE ANÁLISE DE SITUAÇÃO DE SAÚDE. OPAS/OMS. *A vigilância, o controle e a prevenção das doenças crônicas não transmissíveis. DCNT no contexto do Sistema Único de Saúde brasileiro. Situação e desafios atuais.* Brasília: Ministério da Saúde, 2005.
  73. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE. DEPARTAMENTO DE ANÁLISE DE SITUAÇÃO DE SAÚDE. *Saúde Brasil 2004: uma análise da situação de saúde. Série G. Estatística e Informação em Saúde.* Brasília: Ministério da Saúde, 2004.
  74. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE. DEPARTAMENTO DE ANÁLISE DE SITUAÇÃO DE SAÚDE. *Saúde Brasil 2005: uma análise da situação de saúde. Série C. Projetos, Programas e Relatórios.* Brasília: Ministério da Saúde, 2005.

75. MITTELMARK MB, HUNT MK, HEATH GW, SCHMID TL. *Realistic outcomes: lessons from community-based research and demonstration programs for the prevention of cardiovascular diseases*. J Public Health Policy 1993;14:437-462.
76. MONICA Manual. Geneva, World Health Organization, 1990.
77. MURRAY DM, HANNAN PJ, JACOBS DR, MCGOVERN PJ, SCHMID L, BAKER WL, GRAY C. *Assessing intervention effects in the Minnesota Heart Health Program*. Am J Epidemiol 1994;139(1):91-103.
78. MURRAY DM. *Design and analysis of community trials: lessons from the Minnesota Heart Health Program*. Am J Epidemiol 1995;142:569-575.
79. MURRAY DM. *Design and analysis of group-randomized trials*. New York: Oxford University Press, 1998.
80. NISSINEN A, BERRIOS X, PUSKA P. *Community-based noncommunicable disease interventions: lessons from developed countries for developing ones*. Bull WHO 2001;79(10):963-970.
81. NOTHWEHR F, ELMER P, HANNAN P. *Prevalence of health behaviors related to hypertension in three blood pressure treatment groups: the Minnesota Heart Health Program*. Prev Med 1994;23:362-368.
82. ORGANIZAÇÃO PAN AMERICANA DE SAÚDE (OPAS). *Doenças crônico-degenerativas e obesidade: estratégia mundial sobre alimentação saudável, atividade física e saúde*. OPAS, 2003.
83. PALMER N, MUELLER D, GILSON L, MILLS A, HAINES A. *Health financing and equity of use in low income settings*. Lancet 2004;364:1365-1370.
84. PAN AMERICAN HEALTH ORGANIZATION. *Carmen – An Initiative for Integrated Prevention of Noncommunicable Diseases in the Américas*. PAHO, 2003.
85. PAN AMERICAN HEALTH ORGANIZATION. *New non-communicable diseases program at PAHO*. The global and regional predominance of non-communicable diseases. *Epidemiol Bull* 1995;16(4):6-8.

86. PERRY CL, KELDER SH, MURRAY DM, KLEPP KI. *Communitywide smoking prevention: long-term outcomes of the Minnesota Heart Health Program and class of 1989 study*. Am J Public Health 1992;82:1210-1216.
87. POCOCK SJ. *CLINICAL TRIALS. A practical approach*. New York: John Wiley & Sons, 1983.
88. PROCHASKA JO, DICLEMENTE CC. *The stages and processes of self-change in smoking: towards an integrative model of change*. J Consult Clin Psychol 1983;51:390-395.
89. PROKHOROV A, PERRY CL, KELDER SH, KLEPP KI. *Lifestyle values of adolescents: results from Minnesota Heart Health Youth Program*. Fall 1993;28:637.
90. PUSKA P, VARTIAINEN E, TUOMILEHTO J, SALOMAA V, NISSINEN A. *Changes in premature deaths in Finland: successful long-term prevention of cardiovascular diseases*. Bull WHO 1978;76:419-425.
91. PUSKA P. *Successful prevention of non-communicable diseases: 25 year experiences with North Karelia Project in Finland*. Public Health Med 2002;4:5-7.
92. QUADROS CA, VICTORA CG, COSTA JS. *Coverage and focus of a cervical cancer prevention program in southern Brazil*. Pan Am J Public Health 2004;16:223-232.
93. RODRIGUES L, KIRKWOOD BR. *Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls*. Int J Epidemiol 1990;19:205-213.
94. ROSE G. *Sick individuals and sick populations*. Int J Epidemiol 2001;30:427-432.
95. ROTHMAN KJ, GREENLAND S. *Modern Epidemiology*, 2nd edition. Washington: Lippincott-Raven Publishers, 1998.
96. SALONEN JT, KOTTKE TE, JACOBS DR, HANNAN PJ. *Analysis of community-based cardiovascular disease prevention studies – evaluation issues*

*in the North Karelia Project and the Minnesota Heart Health Program. Int J Epidemiol* 1986;15:176-182.

97. SANTOS IS, VICTORA CG. *Serviços de saúde: epidemiologia, pesquisa e avaliação. Cad Saúde Pública* 2004;20:S337-S341.
98. SCHEUERMANN W, RAZUM O, SCHEIDT R, WIESEMANN A, VON FRANKENBERG H, TOPF G, NÜSSEL E. *Effectiveness of a decentralized, community-related approach to reduce cardiovascular disease risk levels in Germany. Eur Heart J* 2000;21:1591-1597.
99. SCHLESSELMAN JJ. *Case-control studies. Design, Coduct, Analysis.* New York: Oxford University Press, 1982.
100. SELLERS DE, CRAWFORD SL, *et al.* *Understanding the variability in the effectiveness of community heart health programs: a meta-analysis. Soc Sci Med* 1997;44:1325-1339.
- 218 101. STERN MP, FARQUHAR JW, MACCOBY N, RUSSEL SH. *Results of a two-year health education campaign on dietary behavior. Circulation* 1976;54:826-833.
102. TAYLOR CB, FOTMANN SP, FLORA J, KAYMAN S, BARRET DC, JATULIS D, FARQUHAR JW. *Effect of long-term community health education on Body Mass Index. The Stanford Five-City Project. Am J Epidemiol* 1991;134:235-249.
103. TAYLOR CB, JATULIS DE, FORTMANN SP, KRAEMER HC. *Weight variability effects: a prospective analysis from the Stanford Five-City Project. Am J Epidemiol* 1995;141:461-465.
104. U.S. Department of health and human services. *Worldwide efforts to improve heart health: a follow up to the Catalonia Declaration – Selected Program Descriptions.* U.S. Department of health and human services: June, 1997.
105. UNITED STATES NATIONAL INSTITUTES OF HEALTH. *Second report of the expert panel on detection, evaluation and treatment of high blood cholesterol in adults.* National Institutes of Health Publication no. 93-3096. Washington, DC: US Department of Health and Human Services, 1993.

106. VARTIAINEN E, PAAVOLA M, MCALISTER A, PUSKA P. *Fifteen-year follow-up of smoking prevention effects in the North Karelia Youth Project*. Am J Public Health 1998;88:81-85.
107. VICTORA CG, HABICHT JP, BRYCE J. *Evidence-based public health: moving beyond randomized trials*. Am J Public Health 2004;94:400-405.
108. VICTORA CG, HUTTLY SR, FUCHS SC, OLINTO MTA. *The role of conceptual frameworks in epidemiological analysis: a hierarchical approach*. Int J Epidemiol 1997;26(1):224-227.
109. VICTORA CG, OLINTO MA, BARROS FC, NOBRE LC. *The recent fall in diarrhoea mortality in Northeastern Brazil: did ORT play a role?* Health Policy Plan 1996;11:132-141.
110. VICTORA CG, SCHELLENBERG JA, HUICHO L, AMARAL J, ARIFEEN S, PARIYO G, MANZI F, SCHERPBIER RW, BRYCE J, HABICHT JP. *Context matters: interpreting impact findings in child survival evaluations*. Health Policy Plann 2006;20(S1):i18-i31.
111. WILBER JA, BARROW JG. *Hypertension – a community problem*. Am J Med 1972;52:653-663.
112. WILLIAMS EL, WINKLEBY MA, FORTMANN SP. *Changes in coronary heart disease risk factors in the 1980s: evidence of a male-female crossover effect with age*. Am J Epidemiol 1993;137:1056-67.
113. WINKLEBY M. *The future of community-based cardiovascular disease intervention studies*. Am J Public Health 1994;84:1369-1372.
114. WINKLEBY MA, FELDMAN HA, MURRAY DM. *Joint analysis of three U.S. community intervention trials for reduction of cardiovascular disease risk*. J Clin Epidemiol 1997;50:645-658.
115. WINKLEBY MA, FORTMANN SP, ROCKHILL B. *Cigarette smoking trends in adolescents and young adults: the Stanford Five-City Project*. Prev Med 1993;22:325-334.

116. WORLD HEALTH ORGANIZATION. *Countrywide Integrated Noncommunicable Diseases Intervention (Cindi) Programme. Protocol and Guidelines*. EUR/ICP/CIND 94 02/PB04. Copenhagen, 1996.
117. WORLD HEALTH ORGANIZATION. *Evaluation in health promotion. Principles and perspectives*. Geneva: WHO Regional Publications, European series, No. 92, 2001.
118. WORLD HEALTH ORGANIZATION. *Guidelines for controlling and monitoring the tobacco epidemic*. Geneva, 1998b.
119. WORLD HEALTH ORGANIZATION. *Handbook for Process Evaluation in Noncommunicable Disease Prevention (Cindi)*. WHO, 1999.
120. WORLD HEALTH ORGANIZATION. *Obesity – Preventing and managing the global epidemic: report of a WHO consultation on obesity*. Geneva, 1998a.
121. WORLD HEALTH ORGANIZATION. *Policy and Development and Noncommunicable disease prevention: The road from Kaunas. Policy development and implementation processes in the Cindi and Carmen Noncommunicable disease intervention programmes – A comparative study*. WHO/Europe, 2004.
122. WORLD HEALTH ORGANIZATION. *Prevention of diabetes mellitus: report of a WHO study group*. WHO Technical Report Series no. 862. Geneva, 1994.
123. WORLD HEALTH ORGANIZATION. *WHO Global Forum IV on Chronic Disease Prevention and Control – Final Report*. Ottawa, Canada, 2004.
124. YOUNG DR, HASKELL WL, JATULIS DE, FORTMANN SP. *Associations between changes in physical activity and risk factors for coronary heart disease in a community-based sample of men and women: The Stanford Five-City Project*. Am J Epidemiol 1993;138:205-16.
125. YOUNG DR, HASKELL WL, TAYLOR CB, FORTMANN SP. *Effect of community health education on physical activity knowledge, attitudes, and behavior*. Am J Epidemiol 1996;144:264-274.

# **Anexo**

## **Glossário de Termos**

**Adequação (Tipo de inferência):** Tipo de inferência que avalia o quanto as atividades dos programas atingiram seus objetivos.

**Adesão:** Comportamento relacionado a saúde que obedece às recomendações de um médico, outro profissional de saúde ou um investigador em um projeto de pesquisa.

**Amostragem:** Processo de seleção de um número de indivíduos pertencentes a um grupo ou universo.

**Amostragem por múltiplos estágios:** Seleção aleatória de regiões, escolas, locais de trabalho, seguida de amostragem aleatória de pessoas dentro de cada um destes grupos.

**Análise bi-caudal:** Teste de significância estatística baseado na premissa que os dados são distribuídos em ambas as direções a partir de um mesmo valor central.

222

**Análise de variância (ANOVA):** Técnica estatística que isola e avalia a contribuição de variáveis categóricas independentes para a variação da média de uma variável dependente contínua.

**Análise multivariável:** Conjunto de técnicas usadas quando a variação em diversas variáveis tem que ser estudadas simultaneamente.

**Análise por intenção de tratar:** Procedimento na condução e análise de ensaios clínicos randomizados. Todos os pacientes alocados em cada braço do regime de tratamento são analisados juntos, representando aquele braço do tratamento, tendo eles recebido / completado o tratamento prescrito ou não.

**Análise uni-caudal:** Teste de significância estatística baseado na premissa que os dados são distribuídos em apenas uma das direções a partir de um valor central.

**Área-piloto:** Área geográfica utilizada para um ensaio do estudo. Reproduz todas as estratégias e métodos a serem utilizados.

**Avaliação:** Processo que procura determinar de forma mais sistemática e objetiva possível a importância, efetividade e impacto das atividades propostas, conforme os seus objetivos.

**Cadeia causal:** Metáfora para a complexa e multifatorial causalidade das doenças (ou eventos). Por exemplo, indivíduos têm uma alimentação rica em gordura, como consequência desenvolvem hipercolesterolemia, que causa obstrução nas artérias coronárias, levando a infarto agudo do miocárdio. Além disso, diversos outros fatores contribuem para tal evento, como: hipertensão arterial, diabetes mellitus, obesidade etc., formando uma cadeia causal de eventos.

**Campanha:** É uma estratégia com abrangência limitada no tempo, que visa o alcance de um objetivo previamente determinado. Por exemplo, campanhas antitabagismo conduzidas através de alertas na mídia, medidas educacionais, etc.

**Causalidade reversa:** Em estudos transversais, nem sempre é possível estabelecer com segurança a precedência temporal da exposição sobre a doença. Por exemplo, pode-se não ter clara a relação entre obesidade e sedentarismo: as pessoas são obesas porque são sedentárias (a falta de exercício é a causa da obesidade), ou são sedentárias porque não têm disposição física para praticarem exercícios devido a obesidade (a obesidade é a causa do sedentarismo).

**Cegamento (estudos cegos):** Estudo em que o observador/pesquisador e/ou os indivíduos participantes desconhecem o grupo ao qual pertencem (controle ou intervenção), em estudos experimentais. Em estudos transversais o cegamento pode ser dado pelo desconhecimento do pesquisador, na análise de dados, sobre a origem da população em estudo, por exemplo.

223

**Censo:** Contagem de uma população. Identificam-se todos os indivíduos em todos os locais de habitação, com informações de idade, data de nascimento, sexo, ocupação, nacionalidade, estado civil, renda, etc.

**Cobertura:** Medida da extensão do alcance dos serviços de saúde na cobertura das potenciais necessidades da comunidade. É expressa pela proporção em que o numerador é o número de pessoas que receberam o serviço e o denominador é o número que deveria ter sido alcançado.

**Coefficiente de mortalidade geral:** Coeficiente que se expressa dividindo-se o número de óbitos por todas as causas em um determinado período pela população ajustada para o meio do período, relativo a uma determinada área. O valor final é multiplicado por um múltiplo de dez, o qual é a base de referência do denominador do coeficiente, isto é, os habitantes.

**Coeficiente de mortalidade por causas específicas:** Coeficiente que indica a mortalidade por uma causa específica, ou um grupo de causas, ao qual esteve exposta uma determinada população, durante um certo período. É dado pelo número de óbitos por aquela causa dividido pelo número de pessoas em risco no meio do período.

**Coeficiente de mortalidade proporcional:** Coeficiente que resulta da divisão do número de óbitos ocorridos por determinada causa ou grupo de causas, em relação ao total de óbitos.

**Coorte:** Termo utilizado para descrever um grupo de pessoas com uma característica em comum que são acompanhadas durante um período de tempo, como em estudos de coorte.

**Correlação:** Grau em que as variáveis mudam juntas.

**Correlação intraclasse:** Representa o quanto membros de um grupo se parecem entre si, mais do que se parecem com membros de outros grupos.

**Covariância:** A covariável é uma das variáveis preditivas do desfecho em estudo.

**Custo-efetividade:** É um método de avaliação em que os resultados ou conseqüências dos programas de saúde são medidos em unidades físicas (por exemplo, anos de vida ganhos ou casos detectados). Tal método justifica-se quando os benefícios são dificilmente monetarizáveis.

**DCNT (Doenças crônicas não transmissíveis):** Hipertensão arterial, diabetes mellitus, dislipidemia, cânceres, entre outras.

**Desfecho:** Todo resultado possível decorrente de um ou mais fatores de exposição.

**Desfechos categóricos:** Desfechos expressos em categorias. Por exemplo: o índice de massa corporal (IMC) pode ser expresso nas seguintes categorias: indivíduos obeso, com sobrepeso, normais e desnutridos.

**Desfechos contínuos:** Desfechos expressos numa escala numérica. Por exemplo: houve uma diminuição de 10 mmHg na pressão arterial sistólica da população estudada.

**Desvio-padrão:** medida de dispersão ou variação. É um resumo do quanto estão dispersos os valores em torno do centro.

**Duplo cego:** Medida de cegamento de um estudo em que os observadores e os indivíduos estudados desconhecem a que grupo estes pertencem.

**Efeito de delineamento:** Utilizado para cálculo de tamanho de amostra quando esta é selecionada por conglomerados. É a razão entre as imprecisões associadas à estimação de um parâmetro sob dois desenhos amostrais. O *deff* funciona como um “preço” a ser pago pela escolha de amostragem por conglomerados, que aumenta a imprecisão devido às possíveis correlações das unidades amostrais dentro e entre os conglomerados.

**Efeito dose-resposta:** Relação do desfecho observado (resposta) numa população conforme a variação do nível de exposição (dose).

**Efeito Hawthorne:** Tendência que os indivíduos têm de mudar o comportamento quando são alvos de interesse e atenção especial em um estudo, independentemente da natureza específica da intervenção que possam estar recebendo.

225

**Efeito placebo:** O efeito atribuído por pacientes a uma medicação ou tratamento que foi prescrito por seu médico. O efeito percebido é devido à expectativa que uma medicação ou tratamento tenha um efeito, isto é, pelo poder da sugestão.

**Efetividade:** Medida de uma intervenção quando aplicada em circunstâncias rotineiras, reais.

**Eficácia:** Medida de uma intervenção quando aplicada em circunstâncias ideais.

**Eficiência:** Efeitos ou resultados alcançados em relação ao custo em termos de dinheiro, recursos e tempo.

**Emparelhamento:** Processo de construir um grupo para o estudo e o grupo comparação, comparáveis em relação a fatores externos.

**Ensaios comunitários:** Experimento em que a unidade alocada para receber uma intervenção (terapêutica ou preventiva) é uma comunidade inteira ou uma subdivisão política.

**Ensaio randomizado:** Estudo epidemiológico em que os indivíduos de uma população são alocados randomicamente em grupos, geralmente chamados de grupos estudo (experimental ou intervenção) e controle, para receber ou não uma intervenção.

**Eqüidade:** Distribuição justa de determinado atributo populacional junto com a eficiência, a liberdade de escolha pelo consumidor e a maximização da saúde. O conceito de eqüidade não é necessariamente equivalente à igualdade, embora os dois termos sejam por vezes usados como sinônimos. Quando se define eqüidade na prestação de saúde em termos de igualdade, o conceito envolve duas dimensões importantes: a eqüidade horizontal – tratamento igual de indivíduos que se encontram numa situação de saúde igual – e a eqüidade vertical – tratamento apropriadamente desigual de indivíduos em situações de saúde distintas.

**Erro alfa (erro tipo 1):** Declarar que uma diferença existe quando, na verdade, não existe.

**Erro beta (erro tipo 2):** Declarar que uma diferença não existe quando, na verdade, existe.

**Estratificação:** O processo ou o resultado de separar a amostra em várias subamostras, de acordo com critérios específicos.

**Estudo de linha de base (dados de baseline):** Fornece um conjunto de informações no início do estudo que, posteriormente, serão utilizados para comparações.

**Estudos de casos e controles:** Estudo epidemiológico observacional de pessoas selecionadas com base na presença da doença (ou outro desfecho) de interesse e um grupo adequado de indivíduos (controle, referência) sem a doença.

**Estudos experimentais:** Estudos em que as condições estão sob controle direto do investigador. O pesquisador altera intencionalmente um ou mais fatores com o objetivo de avaliar seu efeito.

**Estudos longitudinais tipo antes e depois:** Estudo de intervenção em que todos os pacientes recebem o mesmo tratamento e sua condição é verificada antes do início e em um ou vários momentos após o tratamento.

**Estudos observacionais:** Estudos epidemiológicos que não envolvem qualquer intervenção ou experimento.

**Estudos transversais:** Estudo observacional que analisa a relação entre desfechos e outras variáveis de interesse que existam em uma população definida em um dado período.

**Exposição:** Proximidade e/ou contato com a fonte potencialmente causadora ou protetora de um determinado desfecho.

**Fator de risco:** Aspecto do comportamento pessoal ou estilo de vida, exposição ambiental, característica intra-útero ou genética, que com evidências epidemiológicas é sabido estar associado com a condição de saúde.

**Força de associação:** Definida como a magnitude e significância estatística do risco medido.

**Grupo controle externo:** Um ou mais grupos, comunidades, instituições ou áreas geográficas que não receberam a intervenção.

227

**Grupo controle histórico:** Pessoas ou pacientes utilizados para comparações que possuam a condição ou tratamento em estudo em épocas diferentes, geralmente no início do período.

**Grupo controle interno:** Instituições, áreas geográficas ou indivíduos que poderiam ter recebido a intervenção completa, porém não receberam porque não puderam ou se recusaram.

**História natural:** O curso de uma doença do seu início à sua resolução.

**Incidência:** Número de novos eventos que ocorrem em uma população, em um determinado período.

**Indicador:** Parâmetro utilizado com o objetivo de avaliar e fornecer subsídios para planejamento e acompanhamento de ações específicas.

**Índice de Massa Corporal (IMC):** Medida antropométrica definida como peso em quilogramas dividido pelo quadrado da altura em metros

**Inferência:** Processo em que se usa a informação de uma amostra de indivíduos para fazer alguma inferência sobre a população-alvo.

**Intervenção:** Mudança intencional em algum aspecto dos indivíduos em estudo. Intervenção em saúde pública pode ser definida como o fato que é aplicado a vários, maioria ou todos os membros de uma comunidade, com o objetivo de levar benefícios à comunidade como um todo, bem como aos indivíduos.

**Marco conceitual:** Estrutura teórica que fundamenta a organização e hierarquização dos diversos fatores (variáveis) pertencentes a uma cadeia causal. A construção de um marco conceitual requer conhecimento sobre determinantes sociais e biológicos do desfecho em estudo, bem como a relação temporal entre elas.

**Medicina baseada em evidências:** Uso consistente de evidências atuais derivadas de boas publicações de pesquisas clínicas e epidemiológicas, considerando os riscos e benefícios para os pacientes.

228

**Medidas de dispersão:** Termo geral para denominar as medidas que descrevem a variabilidade numérica, ou seja, a medida do afastamento dos dados em relação a um valor central. Por exemplo, a amplitude e o desvio-padrão.

**Medidas de efeito:** Medida que expressa o efeito de um fator na frequência ou risco de um desfecho em saúde. Ver risco relativo, razão de prevalências e razão de odds.

**Medidas de tendência central:** Termo geral para diversos valores da distribuição de um conjunto de valores ou medidas localizados na média ou em torno desta. As mais usadas são a média, mediana e moda.

**Metodologia qualitativa:** Metodologia de pesquisa baseada em observações e coleta de informações caracterizadas por medidas em uma escala categórica, envolvendo aspectos sociológicos e antropológicos. Utiliza métodos de observação de participantes ou informantes-chave, entrevistas, entre outros.

**Metodologia quantitativa:** Metodologia de pesquisa que coleta informações numericamente mensuráveis.

**Modelos multiníveis:** A análise de modelos multinível é a integração de um grupo ou macroníveis de variáveis em estudos epidemiológicos no intuito de incorporar múltiplos níveis de determinação no estudo de desfechos de saúde.

**Modificador de efeito:** Modificação de efeito ou interação é a variação da magnitude do efeito de uma exposição em diferentes níveis de uma outra variável. Essa última variável é chamada de modificador de efeito.

**Monitoramento:** Desempenho intermitente de medidas de análise rotineiras com objetivo de detectar mudanças no ambiente ou estado de saúde das populações.

**Morbidade:** Qualquer afastamento, subjetivo ou objetivo do estado de bem-estar fisiológico ou psicológico.

**Mortalidade:** Estimativa da porção da população que morre durante um período específico de tempo.

**Nível de significância:** Probabilidade de se achar uma diferença tão grande ou maior quanto aquela observada.

229

**Nível primário de atenção à saúde:** Atenção essencial à saúde, baseada em métodos práticos, cientificamente evidentes e socialmente aceitos e em tecnologias tornadas acessíveis a indivíduos e famílias na comunidade por meios aceitáveis e a um custo que as comunidade e países possam suportar, independentemente de seu estágio de desenvolvimento, em um espírito de autoconfiança e autodeterminação. Constitui a porta de entrada, o primeiro contato de indivíduos, famílias e comunidades com o sistema nacional de saúde.

**Odds:** Razão da probabilidade da ocorrência de um evento para a não ocorrência. Por exemplo: o número de casos de uma doença dividido pelo número de pessoas em risco (não casos) ao final de um período.

**Oferta (indicador):** Indicador que avalia se os serviços estão disponíveis e acessíveis e se a qualidade é adequada.

**Padrão-ouro:** Método, procedimento ou medida que é amplamente aceita como sendo a melhor avaliação. Geralmente usada para comparação de novos métodos.

**Plausibilidade (Tipo de inferência):** Permite avaliar se a intervenção proposta pelo programa apresenta efeitos a ele atribuídos, apesar de fatores externos.

**Poder:** Capacidade do estudo em demonstrar uma associação existente.

**População-alvo:** Grupo do qual a população em estudo é selecionada.

**Prevalência:** Número de eventos em uma população em um determinado momento.

**Probabilidade (Tipo de inferência):** Permite avaliar se o programa possui um efeito mensurável estatisticamente significativo.

**Processo:** Maneira como os profissionais ou outros membros de uma equipe atendem os pacientes ou participantes de um programa. Inclui interesse, cortesia, respeito, além de adesão a normas definidas de boa prática profissional.

230

**Processo aleatório simples:** Processo de amostragem em que todos os indivíduos elegíveis possuem a mesma probabilidade de ser selecionados. Por exemplo: listasse todos indivíduos de uma sala de aula e, aleatoriamente, sorteiam-se números de uma urna opaca que correspondem a cada indivíduo previamente numerado.

**Programa:** Conjunto de procedimentos para conduzir uma atividade com um objetivo previamente definido.

**Protocolo:** Plano ou conjunto de etapas a serem seguidos em um estudo, investigação ou em programas de intervenção.

**Quase-experimentos:** Situação na qual o investigador não tem o controle total sobre a alocação e/ou o tempo da intervenção, mas conduz o estudo como um experimento alocando indivíduos em grupos, sem randomização.

**Randomização:** Alocação de indivíduos em grupos a partir de um processo aleatório de escolha.

**Rastreamento:** Identificação presuntiva de doença não identificada ou defeito, pela aplicação de testes, exames ou outros procedimentos que possam ser aplicados rapidamente.

**Razão de odds:** Medida de efeito dada pela razão de dois odds.

**Razão de prevalências:** Medida de efeito dada pela razão entre a prevalência no grupo exposto e a prevalência no grupo não exposto.

**Regressão a média:** Fenômeno através do qual, se indivíduos com valores extremos de um teste forem selecionados e o teste repetido, é provável que o segundo conjunto de valores seja mais próximo da parte central (estatisticamente normal) da distribuição de frequência.

**Regressão de Poisson:** Análise de regressão de dados utilizando distribuição de Poisson.

**Regressão linear:** Análise de regressão de dados utilizando modelos lineares.

**Regressão logística:** Análise de regressão de dados utilizando modelos logísticos.

**Risco relativo:** Medida de efeito dada pela razão de riscos de um desfecho entre o grupo exposto e o não exposto.

231

**Saúde pública:** É a ciência e a arte de prevenir doenças, prolongar a vida e promover saúde através de esforços organizados pela sociedade.

**Temporalidade:** Postula que a exposição sempre precede o efeito.

**Teste exato de Fisher:** Teste de associação para tabelas 2x2 baseado na distribuição exata, hipergeométrica, das frequências da tabela.

**Teste Qui-quadrado:** Qualquer teste estatístico baseia-se na comparação de um teste estatístico com a distribuição qui-quadrada. Os testes qui-quadrados mais antigos e utilizados são aqueles que detectam se duas ou mais distribuições de populações diferem uma da outra. Distribuição Qui-quadrada: uma variável é dita como tendo uma distribuição qui-quadrada com  $k$  graus de liberdade se está distribuída como a soma dos quadrados de  $k$  variáveis aleatórias independentes, cada qual com uma distribuição normal com média zero e variância um.

**Teste t de Student:** Teste estatístico que usa a distribuição t para determinar se duas médias são significativamente diferentes ou ainda, para testar coeficientes de correlação ou regressão linear.

**Uni cego:** Os participantes do estudo não têm conhecimento de sua pertinência aos grupos de pesquisa. Por exemplo, pelo uso de placebos em ensaios clínicos.

**Unidade amostral:** Indivíduos ou grupos de indivíduos (salas de aula, postos de saúde, escolas) que serão selecionados para compor um estudo.

**Utilização (indicador):** Avalia se a população-alvo do programa utiliza os serviços disponibilizados.

**Validade externa:** Um estudo tem validade externa ou é generalizável, se puder fazer inferências sem vieses à população-alvo.

**Validade interna:** Os grupos são selecionados e comparados de forma que as diferenças observadas entre eles, quanto ao desfecho, só possam ser atribuídas ao efeito da exposição sob investigação. A validade interna depende da qualidade metodológica do estudo.

**Variáveis confundidoras:** Variáveis que podem distorcer a associação entre uma exposição e o desfecho de interesse; não são variáveis mediadoras e estão associadas a ambos, o desfecho e a exposição de interesse.

**Variáveis independentes:** Fatores relacionados ao desfecho em estudo na composição da cadeia causal de determinação.

**Variáveis mediadoras:** Variáveis que aparecem no meio da cadeia causal entre uma variável independente e uma dependente.

**Variável:** É qualquer valor que varie, qualquer atributo, fenômeno ou evento que possa ter valores diferentes.

**Viés:** Desvio dos resultados ou inferências sobre a verdade; processo que induz a tal desvio.

**Viés de informação:** Relato seletivo ou omissão de informações sobre a saúde.

**Viés de perda de acompanhamento:** Em um estudo de acompanhamento (coortes, por exemplo), indivíduos que não são mais acompanhados (mortes, não-cooperação, migração, falta de registros adequados, etc) podem possuir características diferentes daqueles que permanecem em acompanhamento.

**Viés de sobrevivência:** Os estudos, particularmente os transversais, excluem indivíduos que já foram a óbito em decorrência da doença. Tal fato gera uma diminuição enviesada da prevalência da exposição.

**Viés do entrevistador:** Erro sistemático na coleta de dados feita pelo entrevistador, consciente ou inconscientemente.

**Vigilância epidemiológica:** Coleta, organização e análise sistemática e contínua de dados e disseminação da informação em tempo para tomar decisões.

ISBN 853341316-5



9 788533 413160



disque saúde: 0800 61 1997

[www.saude.gov.br/svs](http://www.saude.gov.br/svs)

[www.saude.gov.br/bvs](http://www.saude.gov.br/bvs)

Secretaria de  
Vigilância em Saúde

Ministério  
da Saúde

