

CONCEBER E ENCOMENDAR AVALIAÇÕES DE IMPACTO CONTRAFACTUAIS

UM GUIA PRÁTICO PARA AS AUTORIDADES DE GESTÃO DO FSE



Comissão Europeia

Direção-Geral do Emprego, dos Assuntos Sociais e da Inclusão
Unidade A3

Manuscrito concluído em outubro de 2012

Nem a Comissão Europeia nem qualquer pessoa agindo em seu nome são responsáveis pelo uso que possa ser dado às informações contidas nesta publicação.

AGRADECIMENTOS

Este guia prático inspira-se no trabalho fornecido, a pedido da Direção-Geral do Emprego da Comissão Europeia, pelos seguintes especialistas:

Stephen Morris, NatCen Social Research and Policy Studies Institute, Londres

Herta Tödtling-Schönhofer, Metis GmbH, Viena

Michael Wiseman, George Washington Institute of Public Policy

Layout: Alexandru Coca

Tradução para português: Agência para o Desenvolvimento e Coesão

Para qualquer utilização ou reprodução das fotografias não abrangidas pelas leis de direitos de autor da União Europeia, deve ser solicitada autorização diretamente aos detentores dos referidos direitos.

O Europe Direct é um serviço que ajuda a encontrar respostas às suas perguntas sobre a União Europeia

Número de telefone gratuito (*):

00 800 6 7 8 9 10 11

(*) Alguns operadores de redes de telemóveis não permitem o acesso aos números 00800 ou podem cobrar uma taxa por estas chamadas.

Encontram-se disponíveis outras informações sobre a União Europeia na Internet, via servidor Europa (<http://europa.eu>).

Um ficheiro de dados e um resumo figuram no fim desta publicação.

Luxemburgo: Serviço das Publicações da União Europeia, 2013

ISBN 978-92-79-28238-6

doi: 10.2767/94454

© União Europeia, 2013

A reprodução do conteúdo desta obra é autorizada mediante citação da fonte.

Conteúdos

Objetivo e contexto do guia	5
Capítulo 1: Conceitos e métodos	9
1.1 A essência do método contrafactual	9
1.2 Porque são importantes as avaliações contrafactuais?	10
1.3 Porque são as avaliações contrafactuais tecnicamente exigentes?	11
1.4 Súmula dos modelos e das abordagens das AIC	12
1.4.1 Distribuição aleatória – a abordagem experimental	13
1.4.2 Abordagens não aleatórias ou quase experimentais	16
1.5 Como integrar uma AIC num quadro de avaliação mais alargado	31
Capítulo 2: Considerações práticas na preparação de uma AIC	36
2.1. Seleção de intervenções para avaliação de impacto	37
2.1.1. Escolher intervenções para privilegiar a avaliação de impacto	39
2.1.2. Selecionar intervenções passíveis de abordagem contrafactual	40
2.1.3. Os dados adequados estão disponíveis ou podem ser disponibilizados?	46
2.2. Desenvolvimento de um esquema de avaliação	50
2.2.1. Quais as intenções e os objetivos da intervenção?	52
2.2.2. Qual o objetivo da avaliação?	52
2.2.3. Que recursos estão disponíveis?	55
2.2.4. Quando deve ser avaliada a intervenção?	57
2.2.5. Como deve ser identificado o grupo “tratado”?	60
2.2.6. Fatores a considerar na identificação de um grupo de controlo	61
2.2.7. Que tipo de problemas quanto aos dados precisam de ser equacionados no esquema de avaliação?	65
2.2.8. Quais os principais obstáculos à análise de dados e de resultados?	66
2.2.9. Como serão relatados os resultados?	69
Capítulo 3: Fazer avançar a agenda das AIC	70
3.1 Melhorar os níveis de entendimento entre as partes interessadas	70
3.2. O desenvolvimento de capacidades	71
3.3. Enfrentar barreiras jurídicas	73
3.4. No caminho de abordagens mais prospetivas	74
Glossários	76
4.1. Acrónimos	76
4.2. Definições	78

Bibliografia.....	81
Anexos.....	84
Anexo 1. Leituras complementares.....	84
Anexo 2. Diretrizes do Tesouro do Reino Unido para despesas relativas à avaliação.....	87
Anexo 3. Sugestão de plano de curso sobre AIC.....	88
Anexo 4. Avaliações de impacto contrafactuais – Exemplos fornecidos pelos Estados-Membros ..	89

Lista de Caixas

Caixa 1.Exemplo de um ensaio aleatório numa política ativa de mercado de trabalho	16
Caixa 2.Exemplo de uma avaliação que adota a abordagem de emparelhamento	22
Caixa 3.Exemplo de uma avaliação que adota a abordagem de diferença-nas-diferenças	25
Caixa 4.Exemplo de uma avaliação que adota a abordagem de regressão descontínua	27
Caixa 5.Exemplo de um estudo que adotou uma abordagem de variáveis instrumentais	29
Caixa 6.Questões para selecionar intervenções para uma AIC	41
Caixa 7.Motivações para a realização de uma AIC	44
Caixa 8.Tipos mais comuns de intervenção e de grupos-alvo escolhidos para AIC do FSE	45
Caixa 9. Definição de grupos de controlo	48
Caixa 10. Exemplos de dados utilizados para AIC	51
Caixa 11.Proteção e troca de dados – a experiência da Lituânia	53
Caixa 12. Conteúdo recomendado de um esquema de avaliação	54
Caixa 13. AIC incorporadas num quadro mais amplo de avaliação	54
Caixa 14. Questões políticas relacionadas com um programa de formação	67
Caixa 15. Interpretar efeitos líquidos	67
Caixa 16. Incertezas na interpretação dos resultados	71

Lista de figuras

Figura 1. Abordagem experimental de controlo aleatório de dois grupos	15
Figura 2.Abordagem quase experimental estilizado com grupos tratados e de controlo	18
Figura 3. Ilustração da abordagem de emparelhamento da propensão a participar na intervenção	20
Figura 4. Ilustração da abordagem de diferença-nas-diferenças	23
Figura 5. Ilustração da abordagem de regressão descontínua	24
Figura 6. Ilustração de uma abordagem de variáveis instrumentais	28
Figura 7. Diferentes tarefas e tipos de avaliação	34
Figura 8. Ilustração da abordagem de modelo lógico	35
Figura 9. Cronograma simplificado dos resultados de um programa de formação	62
Figura 10. Amplitude dos efeitos detetáveis mínimos em amostras de diferentes tamanhos	70

Lista de tabelas

Tabela 1. Comparação de algumas características-chave das principais abordagens de AIC	30
Tabela 2. Tipos de dados e de fontes	50

Introdução

Objetivo e contexto do guia

Com o aproximar do novo período de programação (septenal) do Fundo Social Europeu (FSE), na União Europeia os orçamentos contraem-se e a preocupação com a utilização efetiva dos fundos aumenta. Além disso, as avaliações dos programas e das intervenções do FSE têm sido complexas e, em muitos casos, não permitiram aos decisores políticos tirar conclusões factuais sobre a sua adequação e eficácia. Consequentemente, a Comissão Europeia (CE) encoraja os Estados-Membros (EM) a aumentarem os esforços para desenvolver provas credíveis de que os efeitos das intervenções do FSE ultrapassam os que teriam sido alcançados na ausência de apoio por parte do Fundo. O desenvolvimento de tais provas requer avaliações de impacto contrafactuais (AIC), ou seja, a comparação dos resultados com as estimativas do que teria ocorrido sem a intervenção do Fundo. Este guia oferece conselhos práticos sobre algumas das questões-chave que devem ser tidas em conta no desenvolvimento de planos para as AIC. O guia destina-se a autoridades de gestão (AG) e a outros organismos responsáveis pela execução das intervenções e dos programas financiados pelo FSE. O objetivo é auxiliar na conceção e na encomenda das AIC. O acento tónico é colocado em aspetos práticos, embora se discutam algumas questões técnicas, quando necessário.

As AIC abordam a questão crucial da inferência causal e respondem à pergunta “o que funciona?”. Estas avaliações procuram evidenciar se as intervenções financiadas pelo FSE melhoraram efetivamente a situação dos participantes no programa. Quando são bem executadas, as AIC dão conta do efeito líquido, ou do impacto, de uma intervenção e permitem aos decisores políticos excluir explicações alternativas para as mudanças nas circunstâncias ou nos resultados observados. As AIC também fornecem estimativas do sinal e da amplitude do efeito líquido da intervenção bem como do grau de incerteza dessa estimativa. O tipo de prova fornecido pelas AIC permite que decisores políticos estimem a eficácia das intervenções e, além disso, que façam comparações entre intervenções e estimem o respetivo desempenho relativo. As provas resultantes das AIC fornecem contribuições importantes para a análise das relações custo-benefício ou custo-eficácia.

Este guia é publicado num momento sem precedentes de limitação de fundos públicos. Neste contexto, é essencial que os decisores políticos entendam os efeitos que têm as intervenções pelas

As AIC abordam a questão: “o que funciona?”

quais são responsáveis, uma vez que estas absorvem dinheiros públicos que podem ser utilizados para outros fins úteis. Cabe, portanto, aos responsáveis pela distribuição de recursos do FSE justificar o dinheiro do Fundo que lhes é regularmente alocado, através da demonstração de que as intervenções funcionam e de que criam valor para os cidadãos. Uma maneira importante pela qual este objetivo pode ser alcançado é através da realização de mais e melhores AIC.

O FSE é o principal instrumento europeu de apoio ao emprego e à inclusão social. No atual período de programação, 2007-2013, o FSE atribui cerca de 76,5 mil milhões de euros em políticas ativas de mercado de trabalho implementadas através de programas operacionais (PO) nos 27 Estados-Membros (EM). Conforme estipulado pelo Regulamento Geral 1083/2006, as avaliações «visam melhorar a qualidade, a eficácia e a coerência da intervenção dos fundos, bem como a estratégia dos programas operacionais e sua implementação».

No período de programação 2014-2020, o desempenho e os resultados serão objeto de uma atenção acrescida¹, o que irá exigir uma revisão dos atuais sistemas e capacidades de monitorização e avaliação, incluindo os mecanismos de recolha de dados. Além disso, os planos de avaliação passarão a ser obrigatórios e será dada uma maior ênfase à avaliação do impacto. Visto que existem vários métodos que podem ser utilizados para medir os impactos das intervenções financiadas pelo FSE, é da responsabilidade das autoridades de gestão decidir que método, ou que combinação de métodos, mais se adequa à satisfação das exigências regulamentares. A quantificação rigorosa dos impactos das intervenções envolve apreciação contrafactual.

Essa mudança de foco no sentido de uma orientação mais forte para o desempenho e para os resultados é importante. Estratégias e técnicas de avaliação de alta qualidade são essenciais para gerar conhecimento útil para todos os EM acerca das intervenções que “funcionam” e das que não funcionam. É essencial reforçar a qualidade das avaliações e desenvolver provas fiáveis sobre o valor acrescentado.

Em princípio, o ponto de partida para a construção de provas sobre a eficácia das intervenções estratégicas é simples. Os requisitos incluem:

- identificação do problema a ser abordado;
- identificação dos instrumentos a utilizar para resolver o problema;
- uma teoria que ligue os instrumentos aos resultados.

A necessidade de resultados

Para avaliar um programa ou um instrumento de financiamento, é necessário, no mínimo, ter indicadores claros e mensuráveis dos recursos aplicados e das consequências e dos resultados relacionados com o problema. É habitual fixar objetivos para as consequências e para os resultados e compará-los com o que foi realmente atingido. A monitorização é utilizada para correlacionar os recursos alocados com os resultados obtidos ao longo do tempo e para proporcionar informações de gestão. A teoria subjacente da intervenção aponta frequentemente para resultados intermédios que podem também tornar-se foco da monitorização.

¹ Proposta de Regulamento do Parlamento Europeu e do Conselho que estabelece disposições comuns acerca do FEDER (Fundo Europeu de Desenvolvimento Regional), do FSE, do FC (Fundo de Coesão), do FEADER e do FEAMP (Fundo Europeu dos Assuntos Marítimos e das Pescas), abrangidos pelo Quadro Estratégico Comum, o qual estabelece disposições gerais aplicáveis ao FEDER, ao FSE e ao FC que revogam o Regulamento (CE) n.º 1083/2006 [COM (2011) 615 final].

No entanto, não é fácil, a partir deste ponto de partida, obter provas de que uma dada intervenção funciona.

Doravante, existe uma necessidade de complementar as práticas de avaliação existentes com abordagens que geram provas muito mais sólidas dos efeitos líquidos e dos impactos das intervenções. Medir os resultados é prestar contas sobre os fundos utilizados. As AIC respondem à questão fundamental de saber se uma intervenção funciona? À medida que uma AIC procura estabelecer uma ligação de causalidade entre as intervenções e os resultados, podem ser necessários outros métodos teóricos e de avaliação de processos, para identificar os mecanismos causais subjacentes e para ajudar a garantir que os impactos alcançados num local fornecem uma base de conhecimentos que permite a reprodução das intervenções noutros lugares.

...e de provas do
impacto líquido

No período de programação 2007-2013, os Estados-Membros adotaram diversas abordagens para a fazer a avaliação. Alguns mantiveram a metodologia do período de programação anterior (p. ex., avaliações intercalares), outros consideraram a possibilidade de realizar sob encomenda avaliações sobre aspetos específicos e outros reduziram as atividades, pelo menos no início do período. As avaliações são muito heterogéneas, tanto quanto ao âmbito como quanto à metodologia. A recolha de dados resulta, as mais das vezes, de uma combinação de ferramentas tradicionais: entrevistas, inquéritos, análise de dados secundários e administrativos, grupos de discussão e estudos de caso. Ferramentas mais complexas, como abordagens econométricas e análises de rede, são exceções, apesar de alguns Estados-Membros estarem a fazer esforços no sentido dessa utilização, com resultados interessantes, nomeadamente no âmbito das AIC.

Experiência dos EM
com AIC

Na primeira metade do corrente período de programação, prevaleceram abordagens de avaliação orientadas para o processo². Este tipo de avaliação é muito importante para melhorar a implementação dos programas e para adaptar os PO, a fim de aumentar a eficácia do FSE. No entanto, na segunda metade do período de programação de 2007-2013 – e para o subsequente – são necessárias mais avaliações de impacto, para que se possam obter provas credíveis sobre o valor acrescentado do FSE.

As AIC representam, até agora, apenas uma pequena fração das avaliações realizadas no atual período de programação do FSE. Contudo os Estados-Membros possuem uma vasta experiência em matéria de AIC. Numa audição de especialistas organizada pela Comissão Europeia e realizada em 25 de outubro de 2011, representantes de oito EM e peritos de avaliação apresentaram exemplos de AIC em intervenções cofinanciadas pelo FSE. Essas apresentações incidiram sobre as motivações e os objetivos para a realização de tais estudos, sobre as abordagens metodológicas escolhidas, sobre os dados e os indicadores utilizados, sobre os resultados e sobre as limitações e desafios enfrentados. As comunicações sublinharam igualmente os planos futuros para a implementação de AIC. Este guia faz eco de exemplos apresentados nessa audição e inspira-se numa análise das experiências aí relatadas.

² Ver o relatório de síntese da Rede de Especialistas em Avaliação do FSE (2011) no seguinte endereço:
<http://ec.europa.eu/social/keyDocuments.jsp?type=0&policyArea=0&subCategory=0&country=0&year=0&advSearchKey=evaluation&mode=advancedSubmit&langId=en>

Orientação para os profissionais

Levar a cabo AIC corretamente requer tanto competências técnicas quanto vontade política. Este guia constrói a argumentação a favor das AIC e levanta algumas questões que as AG devem ter em conta para serem bem-sucedidas. Além dos aspetos práticos de uma AIC, é dada atenção a questões mais amplas que podem ter de ser consideradas para melhorar a avaliação de impacto.

O guia está estruturado em três capítulos.

O primeiro capítulo discute a natureza das AIC e as razões da sua importância. Fornece uma visão geral dos métodos de AIC, enfatizando a distinção fundamental entre as abordagens experimentais e quase experimentais. Além disso, são aqui também analisados, em termos gerais, os tipos de questões políticas que podem ser abordadas por meio da aplicação de AIC e a relação entre os métodos da AIC e outras abordagens de avaliação (por exemplo: abordagens de base teórica, avaliação de processos e análise de eficiência).

O segundo capítulo apresenta uma série de questões a serem consideradas pelas AG na conceção de AIC. Este guia elenca alguns dos principais desafios que comumente se colocam durante o planeamento de AIC e faz algumas recomendações sobre a forma como tais desafios podem ser enfrentados. As questões aqui consideradas procuram orientar aqueles que pretendam encomendar AIC de intervenções financiadas pelo FSE, antes de planejar e encomendar uma avaliação. No entanto, este guia não tenta prever os requisitos e os planos específicos, que terão de ser adaptados às circunstâncias, muitas vezes únicas, que as AG, os organismos intermediários (OI) e os avaliadores terão de enfrentar em cada avaliação encomendada.

As AIC fornecem provas muito fiáveis da eficácia dos fundos. No entanto, só o podem fazer se forem bem planeadas e se forem executadas de forma adequada. Para que assim seja, é essencial que as AG considerem algumas questões fundamentais antes de encomendarem uma avaliação. A maneira exata de uma AG considerar estas questões e a ordem pela qual o faz serão ditadas por aspetos práticos e pelas modalidades institucionais em vigor no seio dos Estados-Membros. Este guia pretende apenas destacar algumas destas questões importantes, para elas chamando a atenção das AG.

O terceiro capítulo aborda questões mais amplas acerca da reforma, nomeadamente a necessidade de desenvolver a capacidade de conduzir AIC com êxito, tanto no seio das AG (decisores políticos e membros) quanto entre as comunidades académicas e de investigação dos EM. Esta secção também aborda a necessidade de enfrentar as barreiras jurídicas à partilha de dados e incentiva uma transição para uma conceção mais prospetiva das avaliações.

Em suma, este guia: 1) promove as AIC, 2) identifica os passos importantes ao longo do caminho para a realização bem-sucedida de AIC e 3) vai além de pormenores para tornar as AIC uma parte essencial do enquadramento do FSE. O objetivo final é aumentar a contribuição do FSE para o bem-estar dos cidadãos europeus.

Conceitos e métodos

Este capítulo trata de questões fundamentais relativas à natureza das abordagens contrafactuais e ao seu objeto. Especificamente, explica a essência da abordagem contrafactual, em especial no que se refere aos tipos de intervenções cofinanciadas pelo FSE. Analisa também a relação entre as abordagens contrafactuais e outras metodologias de avaliação e discute por que razão as AIC são importantes – especialmente no momento atual. São aqui examinadas as questões políticas para cuja abordagem as AIC podem ser usadas, e será apresentada uma visão geral breve e simplificada de algumas das principais abordagens relevantes para a avaliação das intervenções cofinanciadas pelo FSE.

1.1 A essência do método contrafactual

As AIC procuram identificar os efeitos líquidos e os impactos das intervenções. A característica que distingue as AIC é que estas visam confirmar que os resultados obtidos se devem unicamente às intervenções financiadas pelo FSE, com exclusão de explicações alternativas.

Subjacente à capacidade de excluir explicações alternativas está a ideia do método “contrafactual”. Para compreender claramente o conceito de contrafactual, e para expor as questões de forma simples, é útil considerar o exemplo de um indivíduo desempregado que participa num programa de formação cujo objetivo é o de aumentar a taxa de emprego. Para determinar o efeito que a formação tem sobre o indivíduo, a abordagem contrafactual concebe dois resultados possíveis³. O primeiro resultado é a situação de emprego do formando após ter participado no programa de formação; é o resultado observado para o formando. O segundo resultado potencial é a situação de emprego deste formando caso não tivesse participado no programa de formação, mantendo-se iguais todas as outras circunstâncias. No caso em apreço, este segundo resultado é referido como o resultado contrafactual. O impacto da formação no formando, enquanto indivíduo, é simplesmente a diferença entre os resultados observados e os resultados contrafactuais. Trata-se do efeito, ou impacto, causal da formação sobre o indivíduo.

O método
contrafactual

³ Um exame mais aprofundado acerca do modelo de causalidade baseado em «resultados potenciais» pode ser encontrado em Holland (1986): 945-970.

Definir grupos
tratados e...

A única diferença entre as circunstâncias ou condições que deram origem aos resultados observados e aos resultados contrafactuais é a participação do indivíduo na formação. Portanto, qualquer diferença entre os dois resultados deve ser o impacto da formação sobre o estatuto profissional do indivíduo.

Na realidade, nós não observamos, e não poderíamos observar, resultados contrafactuais em indivíduos expostos a uma intervenção. O principal objetivo das AIC, no entanto, é fornecer estimativas convincentes de resultados contrafactuais sobre grupos de pessoas físicas ou sobre empresas envolvidas em intervenções cofinanciadas pelo FSE. Assim, os impactos são expressos, por exemplo, na forma de diferenças em médias ou proporções entre valores médios observados e os valores contrafactuais “estimados”. Na maioria dos casos em que são aplicadas, as AIC pretendem comparar os resultados de uma intervenção (uma medida ou um instrumento) nas entidades (pessoas, PME etc.) que beneficiaram dela, com os de um grupo não sujeito à intervenção. Na terminologia das AIC um **grupo “tratado”**, ou “de tratamento”, distingue-se do **grupo de “controlo”**, sendo que estes dois devem ser tão semelhantes quanto possível em todos os aspetos (exceto nos tratamentos que recebem). É a partir do grupo de controlo que as estimativas de resultados contrafactuais são obtidas, prestando atenção específica às diferenças exógenas nas características – observadas e não observadas – entre os dois grupos. Também é possível comparar vários tratamentos, expondo unidades elegíveis a uma gama de variantes de tratamento (por ex., a outros tratamentos financiados pelo FSE ou a intervenções financiadas por outras fontes), formando uma série de grupos tratados e comparando os resultados de um e de outro, e/ou resultados de um grupo de controlo não tratado.

... grupos de controlo

Quando o grupo de controlo não é exposto a nenhum tratamento, a questão a que a avaliação tenta responder é “Qual é o impacto de receber a intervenção, em relação a não receber ajuda ou apoio?”. Em contrapartida, quando os resultados do tratamento em causa são comparados com os de outro tratamento, a questão abordada na avaliação é “Qual é o impacto de receber a intervenção sob estudo em relação ao impacto da exposição a uma alternativa claramente definida?”. Uma AIC pode, em muitos casos, ser concebida de forma a responder a estas perguntas fundamentais. A escolha da questão a resolver é feita de acordo com as prioridades dos decisores políticos bem como com os constrangimentos práticos quanto à conceção.

Nos casos em que é feita uma comparação direta ou indireta entre dois tratamentos diferentes, deve haver um contraste bem definido entre eles, o que é significativo do ponto de vista da formulação de políticas.

1.2 Porque são importantes as avaliações contrafactuais?

As AIC fornecem informações importantes sobre os efeitos líquidos, ou os impactos, das intervenções. Fornecem estimativas acerca da magnitude dos impactos, do seu sinal (positivo ou negativo) e das medidas estatísticas de incerteza. Ajudam também a verificar ou a rejeitar a ligação causal presumida entre a intervenção e os resultados. Estes efeitos medidos podem ser utilizados na estimativa da eficiência relativa das intervenções, através do estudo da relação custo-eficácia de uma intervenção ou de uma análise de custo-benefício integral.

Estas características das AIC permitem que este tipo de avaliação forneça informações importantes aos decisores políticos, cuja tarefa é alocar recursos a diferentes intervenções. As decisões sobre o financiamento de potenciais intervenções ocorrem num contexto de limitação de recursos. Cada vez mais, as decisões de alocação de recursos são feitas em cenário de austeridade fiscal. Neste contexto, os decisores políticos precisam de provas fundamentadas dos impactos dos programas e da relação custo-eficácia, para poderem utilizar os recursos disponíveis de forma otimizada.

Apoiar as decisões de alocação de recursos

Os responsáveis pelas intervenções e os agentes encarregados de garantir que os programas continuam a atrair financiamento terão grande interesse em promover as AIC de modo a mostrarem que os seus programas conferem valor ao dinheiro e trazem benefícios mensuráveis aos participantes e à sociedade como um todo. As provas resultantes de AIC serão de especial interesse para os responsáveis pela alocação de recursos. As AG almejarão demonstrar que seus programas “funcionam”, de facto. Para o fazer de forma convincente, precisarão de encomendar AIC de alta qualidade.

1.3 Porque são as avaliações contrafactuais tecnicamente exigentes?

Existe múltiplas abordagens que podem ser descritas como tentativas “não fiáveis” de estimar os impactos de intervenções. Estas serão aqui discutidas a fim de ilustrar as complexidades inerentes às AIC, mas não se fará referência a práticas de avaliação concretas.

Indo além de abordagens simplistas

Em primeiro lugar, um decisor político pode querer avaliar o impacto de um programa de formação para desempregados, comparando os salários dos participantes na formação após a participação na ação de formação, com os salários dos indivíduos desempregados que não participaram. O decisor político irá então associar ao programa de formação a causa da diferença de salários entre os participantes e os não participantes.

Esta não é uma estratégia válida para identificar o impacto da formação nos salários, porque os não participantes podem diferir dos participantes em aspetos importantes, o que pode influenciar os resultados: as diferenças impossibilitam a exclusão de explicações alternativas para as eventuais discrepâncias nos salários observados. Por exemplo, os participantes podem ter maior capacidade intrínseca do que os não participantes. Por outras palavras, os desempregados com maior capacidade, fizeram-se seleccionar ou decidiram participar na ação de formação. Assim, prova-se que a capacidade afeta a decisão de participar, mas que também afeta os resultados – os desempregados com maiores níveis de capacidade intrínseca são propensos a obter um salário maior do que os de menor capacidade.

Se a capacidade não pode ser medida e as diferenças na capacidade intrínseca entre os dois grupos não podem ser tidas em conta no impacto estimado, considera-se que o impacto estimado do programa de formação sofre de **enviesamento de seleção**. Para combater este problema, os avaliadores tentam recolher o máximo de informação possível sobre fatores importantes que afetam a decisão de participar e os resultados que daí resultam. Estes dados são utilizados para seleccionar entre os não participantes um grupo de controlo válido e para realizar análises estatísticas adequadas. Ao fazê-lo, os avaliadores invocam frequentemente o pressuposto de que a seleção para o programa é determinada por fatores observáveis. Essa “presunção de identificação” é sempre questionável e de difícil atestação.

Neutralizar o enviesamento de seleção

Antes e depois da mudança

É necessário ponderar se tal presunção é plausível, com base no conhecimento de fatores institucionais e de teoria comportamental.

Uma segunda abordagem “não fiável” consistiria em o decisor político observar os salários dos participantes antes e depois da formação, atribuindo as variações dos salários antes e depois da intervenção ao programa de formação. Em suma, esta abordagem presume que, na ausência da intervenção, os salários médios permanecem inalterados.

Também esta não é, na maioria dos casos, uma estratégia válida para calcular o impacto da formação nos salários, a não ser que a presunção de estabilidade no tempo possa ser invocada de forma plausível. Isto porque os salários dos participantes mudam inevitavelmente no tempo, de maneiras totalmente alheias à formação. Por exemplo, é comum observar que os salários dos participantes caem antes de estes participarem em ações de formação, em parte devido a fatores transitórios. Em muitos casos, ocorreria uma recuperação, independentemente da participação na ação de formação⁴. A abordagem, não fiável, de medir o impacto da formação por meio do cálculo da diferença entre os ganhos imediatamente antes da entrada no programa e os ganhos após a entrada no programa, ignora o facto de que, em muitos casos, o salário teria subido de qualquer maneira.

Para ajustar esta abordagem é necessário medir o contrafactual – ou seja, medir a forma como os salários dos formandos teriam mudado se não tivesse existido a ação de formação. Por exemplo, tal resultado contrafactual pode ser obtido a partir da observação dos salários, no mesmo ponto no tempo, de um grupo de controlo – cuidadosamente composto para que coincida com o grupo tratado – que não tenha sido exposto à intervenção do programa de formação. A suposição das tendências comuns é então invocada frequentemente, e postula que a tendência dos salários dos participantes e dos elementos do grupo de controlo (“controlados”) seria a mesma no caso de ausência da intervenção.

Os limites dessas abordagens “não fiáveis” motivam a busca de métodos de avaliação mais convincentes. Como sugerido acima, os métodos mais convincentes são, no entanto, tecnicamente mais difíceis de implementar. A próxima secção deste capítulo apresenta o esboço de uma breve e simplificada discussão acerca de algumas abordagens específicas às AIC e que, provavelmente, serão mais relevantes no contexto do FSE.

1.4 Súpula dos modelos e das abordagens das AIC

O interesse das AIC reside nos efeitos que uma intervenção tem sobre aqueles que nela participam; os resultados contrafactuais são geralmente estimados usando dados recolhidos em grupos de não participantes semelhantes àqueles que participam na intervenção que é avaliada. A Tabela 1, no final deste capítulo, apresenta uma breve visão geral de abordagens, de algumas vantagens e limitações, e dos tipos essenciais de dados de que as AIC necessitam.

⁴ Este padrão ficou conhecido como *Ashenfelter's dip*, em honra do primeiro economista que comentou o fenómeno. Ver Ashenfelter (1978): 6, 47-57.

Não é possível fornecer orientações pormenorizadas sobre a escolha da abordagem de avaliação mais adequada, transversal a todas as circunstâncias enfrentadas pelas AG. Para escolher a abordagem de AIC mais relevante num determinado conjunto de circunstâncias, a AG deve ponderar sobre o que “funcionou bem” em avaliações anteriores, tanto dentro da própria AG, quanto dentro do EM e em outros EM; as AG poderão aprender com o que foi alcançado antes, em programas seus ou nos de outrem em que foram enfrentadas circunstâncias semelhantes. Os fóruns de troca de conhecimentos obtidos durante a conceção e a implementação da avaliação podem ser fontes úteis de informação a este respeito. A pesquisa bibliográfica sobre a avaliação de intervenções semelhantes também pode ser uma fonte importante de informações para auxiliar no processo de conceção. Especialistas contratados pela AG também terão opiniões sobre a melhor forma de abordar a conceção de uma avaliação. É importante lembrar que poderá haver conhecimentos e experiência dentro da AG que podem ser capitalizados.

A principal diferença entre tipos de AIC reside nas abordagens de avaliação, que podem ser **experimentais** ou **quase experimentais**. A abordagem experimental, comumente chamada «ensaio de controlo aleatório» (ECA), é também conhecida como “experimentação social”.

É o tipo de abordagem experimental que é considerado de referência, entre os métodos de AIC, para avaliar os efeitos das intervenções que podem ser testados e manipulados ao longo de períodos de tempo relativamente curtos, e que representa, na maioria dos casos, o ideal. Uma avaliação de impacto bem concebida deve ter como objetivo obter estimativas de resultados contrafactuais sem enviesamento. Em muitas aplicações, pode considerar-se que uma abordagem experimental produz de facto estimativas sem enviesamento. Ao discutir abordagens de AIC, é, muitas vezes, desejável começar por delinear a abordagem experimental. Isto porque os métodos quase experimentais procuram, essencialmente, imitar o ideal experimental.

Ao discutir a conceção de AIC, as características-chave de cada abordagem são definidas de modo tão simples quanto possível, para que os princípios subjacentes fiquem clarificados. Na realidade, as aplicações destes métodos podem ser muito mais complexas, e questões como a não conformidade podem aumentar significativamente os desafios enfrentados.

1.4.1 Distribuição aleatória – a abordagem experimental

As abordagens aleatórias podem tomar muitas formas. Aqui, foca-se uma abordagem simples de dois grupos, para clarificar os princípios fundamentais. A Figura 1 ilustra uma abordagem aleatória simples.

O ponto-chave é que a distribuição aleatória garanta que os dois grupos sejam estatisticamente equivalentes em todos os aspetos, no momento em que são escolhidos. Depois da distribuição aleatória, o grupo tratado é exposto à intervenção, que é o foco da avaliação e cujo impacto ou cujos efeitos se pretendem medir.

Dependendo da questão estratégica central, pode decidir-se não administrar ao grupo de controlo nenhum tratamento, ou pode comparar-se o grupo tratado com um grupo que tenha sido exposto a um outro tratamento cujo estudo seja de interesse (pode ser concebido como representando o tratamento usual), ou pode haver vários grupos tratados, a par de um grupo de controlo. Por exemplo, pode haver interesse em comparar os efeitos de um programa de formação financiado pelo FSE com uma outra formação financiada por entidades nacionais, ou com a prestação de outros serviços à mesma população.

Selecionar a abordagem certa

A abordagem aleatória - a referência

Grupos estatisticamente equivalentes

Nenhum outro tratamento para os grupos de controlo

Como os grupos tratados e de controlo são estatisticamente equivalentes no que toca à distribuição aleatória, e porque a exposição aos tratamentos subsequentes é controlada, as diferenças de resultados podem ser atribuídas à intervenção sob avaliação (sujeitas a incertezas estatísticas padrão), e as explicações alternativas podem ser excluídas como causas de qualquer uma das diferenças observadas (ver Caixa 1).

Figura 1. Abordagem experimental de controlo aleatório de dois grupos



Provas fortes, mas
conceção difícil

Como resultado das características intrínsecas à conceção, as abordagens aleatórias, sendo corretamente implementadas, oferecem a probabilidade de obter provas fortes dos efeitos de uma intervenção. Por isso são frequentemente preferidas. Exigem, no entanto, um planeamento prévio e em pormenor, e são bastante difíceis de conceber e de administrar. Além disso, os gestores dos programas enfrentam desafios significativos para as implementar corretamente. Alguns levantaram objeções éticas e legais à sua utilização. Além disso, a própria presença do processo de distribuição aleatória pode alterar a composição do grupo de participantes de uma intervenção. Por exemplo, alguns potenciais participantes podem sentir-se desmotivados em relação à ideia de distribuição aleatória e recusar-se a participar. Além disso, os indivíduos sujeitos a distribuição aleatória podem não estar de acordo com o estatuto que lhes é atribuído. Há ainda uma série de outros desafios que poderão ter de ser resolvidos. Em algumas circunstâncias, a implementação de abordagens experimentais de controlo aleatórios pode ser dispendiosa.

Por essas e outras razões, pode parecer improvável que as avaliações de instrumentos e de intervenções financiadas pelo FSE sejam realizadas com recurso a uma abordagem aleatória. No entanto, este guia adverte contra o impulso de excluir o processo de distribuição aleatória em todos os casos, sem a devida consideração.

Caixa 1. Exemplo de um ensaio aleatório numa política ativa de mercado de trabalho

Demonstração de Manutenção e de Progressão no Emprego no Reino Unido

O Projeto de Demonstração de Manutenção e de Progressão no Emprego (MPE) no Reino Unido testou os efeitos, em desempregados de longa duração e em indivíduos economicamente inativos, do alargamento da ajuda e do apoio, bem como dos incentivos financeiros a cidadãos que tinham perdido a contribuição da Segurança Social e entrado no mercado de trabalho. Assim, o projeto MPE estendeu o apoio prestado através de políticas ativas de mercado de trabalho (PAMT) a grupos de empregados com salários baixos.

Os candidatos elegíveis para dois dos principais programas do mercado de trabalho do Reino Unido ativos na época – o «New Deal for the Long-term Unemployed» (para desempregados de longa duração) e o «New Deal for Lone Parents» (para famílias monoparentais) – foram distribuídos aleatoriamente tanto pelos grupos tratado e de controlo. O grupo de controlo entrou nos programas do New Deal como de costume. O grupo tratado recebeu o apoio de pré-emprego (de forma semelhante ao grupo de controlo), mas continuou a receber conselhos e ajuda para sair do programa da Segurança Social e para entrar num emprego. Na altura em que o estudo foi realizado, a ajuda e o apoio dados a quem recebia subsídios da Segurança Social no Reino Unido resultou na obtenção de emprego. Os participantes também eram elegíveis para um conjunto de apoios financeiros e de incentivos para encorajar a formação e a manutenção do emprego. O objetivo era incentivar os participantes a permanecerem fora dos programas da Segurança Social e a progredirem através da melhoria do seu salário e de outros termos e condições de emprego.

No total, cerca de 16 mil indivíduos foram distribuídos aleatoriamente em grupos tratados e de controlo, espalhados por cerca de cinquenta departamentos de serviços públicos de emprego. O processo de distribuição aleatória produziu grupos tratados e de controlo muito semelhantes, no momento da distribuição. Consequentemente, quaisquer diferenças entre os dois grupos quanto às quantificações de itens-chave como obtenção de trabalho, salário, horas e qualidade do emprego, posteriores à entrada da intervenção, poderiam ser com confiança atribuídas à intervenção MPE.

As conclusões do estudo mostram que a intervenção foi particularmente bem-sucedida entre os desempregados de longa duração, aumentando tanto os níveis de emprego quanto os rendimentos.

¹ Os leitores interessados podem encontrar mais informações sobre essa avaliação aqui: <http://statistics.dwp.gov.uk/asd/asd5/rports2011-2012/rrep765.pdf>

A abordagem tem sido amplamente utilizada e outros exemplos se juntam aos provenientes do Reino Unido, discutidos na Caixa 1, como, entre muitos outros, os das experiências GAIN nos Estados Unidos⁵ (existem numerosos outros exemplos da América do Norte), os das experiências realizadas na Suécia⁶ bem como o de um estudo realizado na Alemanha para estimar os efeitos dos serviços ativos de mercado de trabalho fornecidos por prestadores privados comparados com os dos fornecidos pelo serviço público de emprego⁷.

⁵ Ver Riccio *et al.* (1994).

⁶ Ver Hagglund (2006).

⁷ Ver Krug e Stephan (2011).

A distribuição aleatória através de pesquisa na fase de conceção

As abordagens aleatórias podem distinguir-se de outras abordagens, principalmente pela forte importância que atribuem ao controlo dos enviesamentos através de pesquisa na fase de conceção. Esta ênfase na fase de conceção requer um planeamento avançado. Abordagens aleatórias são muitas vezes melhor implementadas na avaliação de novas intervenções-piloto do que em avaliações já existentes. Isto deve-se ao facto de aquelas exigirem algum controlo sobre como os participantes são recrutados para a intervenção que está a ser avaliada. Este “controlo” é muitas vezes mais difícil de alcançar em programas já existentes do que em novas intervenções que estão abertas a ser moldadas.

Objeções éticas

Como ficou claro, a implementação de uma abordagem aleatória, exige que uma fração do grupo-alvo elegível seja afastada da intervenção, de modo a formar um grupo de controlo. Este afastamento é feito aleatoriamente e não é feito sob diretivas quer do candidato quer dos administradores da intervenção. Por esta razão, os decisores políticos mostram-se muitas vezes contra os ECA invocando razões éticas, antes de considerar se estas seriam viáveis dos pontos de vista prático e analítico.

...mas também argumentos fortes a favor da utilização de abordagens aleatórias

No entanto, há um forte argumento a favor das abordagens aleatórias. Se a distribuição aleatória proporciona as provas de melhor qualidade e as mais fiáveis da eficácia das intervenções com financiamento público, então é importante que seja utilizada mais amplamente na estimativa de impacto de intervenções. Mais ainda, se os impactos de uma determinada intervenção são desconhecidos *a priori*, a exclusão de indivíduos não desrespeita princípios éticos uma vez que não podemos garantir que estes viriam a beneficiar da intervenção. Além disso, tais abordagens são amplamente utilizadas na medicina e em outros campos de estudo, como acontece cada vez mais na pesquisa em educação. Finalmente, em alguns casos em que os serviços e apoios prestados por uma intervenção registam excesso de candidatos, promover a escolha dos indivíduos, participantes na intervenção de forma aleatória pode ser a forma mais ética de distribuir recursos escassos.

1.4.2 Abordagens não aleatórias ou quase experimentais

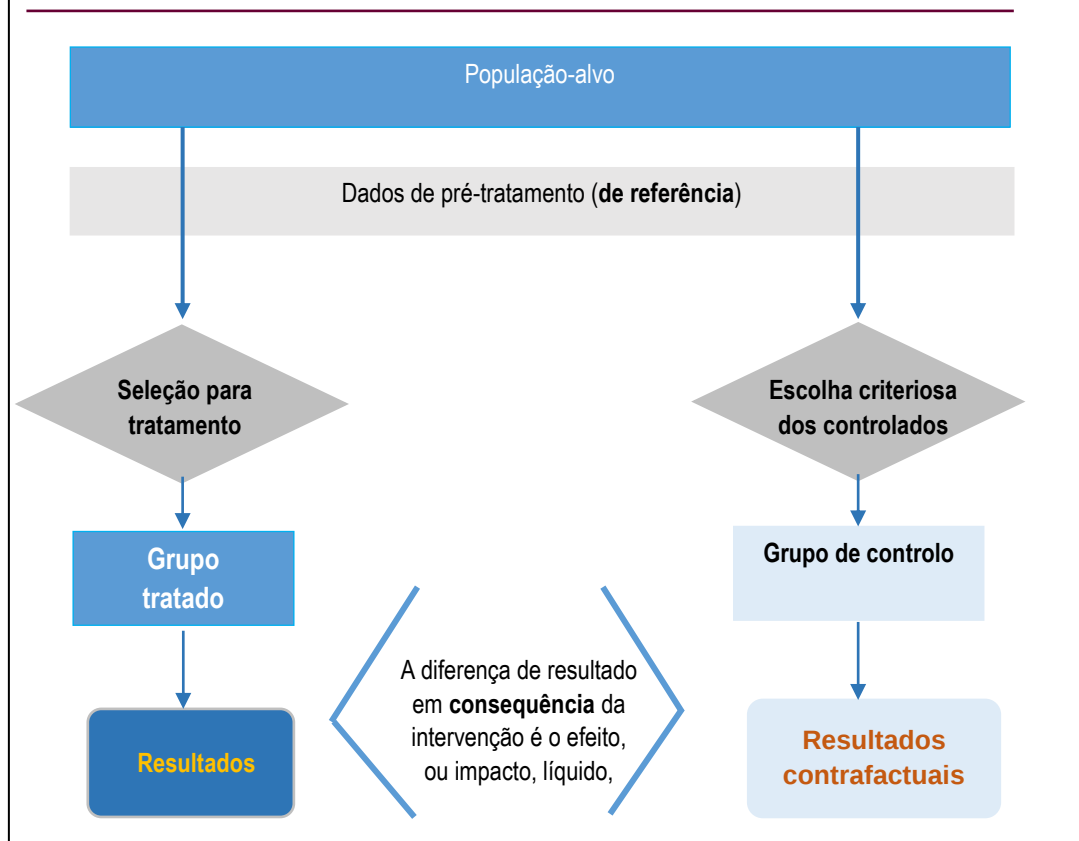
Existe um conjunto variado de abordagens que procuram essencialmente replicar a distribuição aleatória. Estas abordagens são conhecidas como quase experimentais. Não é possível, neste guia, analisá-las todas nem fazer um relatório técnico pormenorizado e completo de cada uma delas. No entanto, são apresentadas em termos gerais as metodologias quase experimentais com maior probabilidade de implantação no contexto do FSE: 1. emparelhamento da propensão a participar na intervenção (propensity score matching); 2. diferença-nas-diferenças (difference-in-differences); 3. regressão descontínua (regression discontinuity); 4. variáveis instrumentais (instrument variables). A apresentação destas metodologias é feita de forma altamente simplificada, para se evidenciarem os princípios fundamentais de cada abordagem. A Tabela 1 apresentada no final desta secção resume as principais abordagens e os respetivos méritos relativos. No Anexo 1, encontram-se sugestões de leituras para aprofundar o tema.

Grupos tratado e de controlo não aleatórios

Nas abordagens quase experimentais, os grupos tratados da intervenção são comparados com um grupo de controlo de sujeitos distribuídos de forma não aleatória ou de alvos potenciais que não recebem a intervenção.

Como numa experiência, o objetivo é obter uma estimativa não enviesada da mudança provocada pela intervenção em análise. Uma vez que os grupos tratados e os grupos de controlo não são formados aleatoriamente, as abordagens quase experimentais exigem muito mais atenção aos métodos de registo das diferenças potenciais entre os membros do grupo tratado e putativos membros do grupo de controlo passíveis de afetar a decisão de participação e os resultados. A escolha de um grupo de controlo plausível é fundamental. Não conseguir escolher um grupo de controlo adequado nem explicar as restantes diferenças entre os dois grupos em análise enfraquece a credibilidade das estimativas e pode perturbar as tentativas de excluir explicações alternativas para quaisquer efeitos observados.

Figura 2. Abordagem quase experimental estilizada com grupos tratados e de controlo



No que respeita às intervenções cofinanciadas pelo FSE, a abordagem de avaliação quase experimental será, muito frequentemente, uma abordagem de dados de referência / consequência de dois grupos. As abordagens deste tipo incluem um grupo de controlo e um grupo tratado, como no caso da distribuição aleatória, só que o grupo de controlo é selecionado (sem utilização de distribuição aleatória) a partir de grupos não participantes existentes, de forma a ser tão semelhante quanto possível ao grupo tratado.

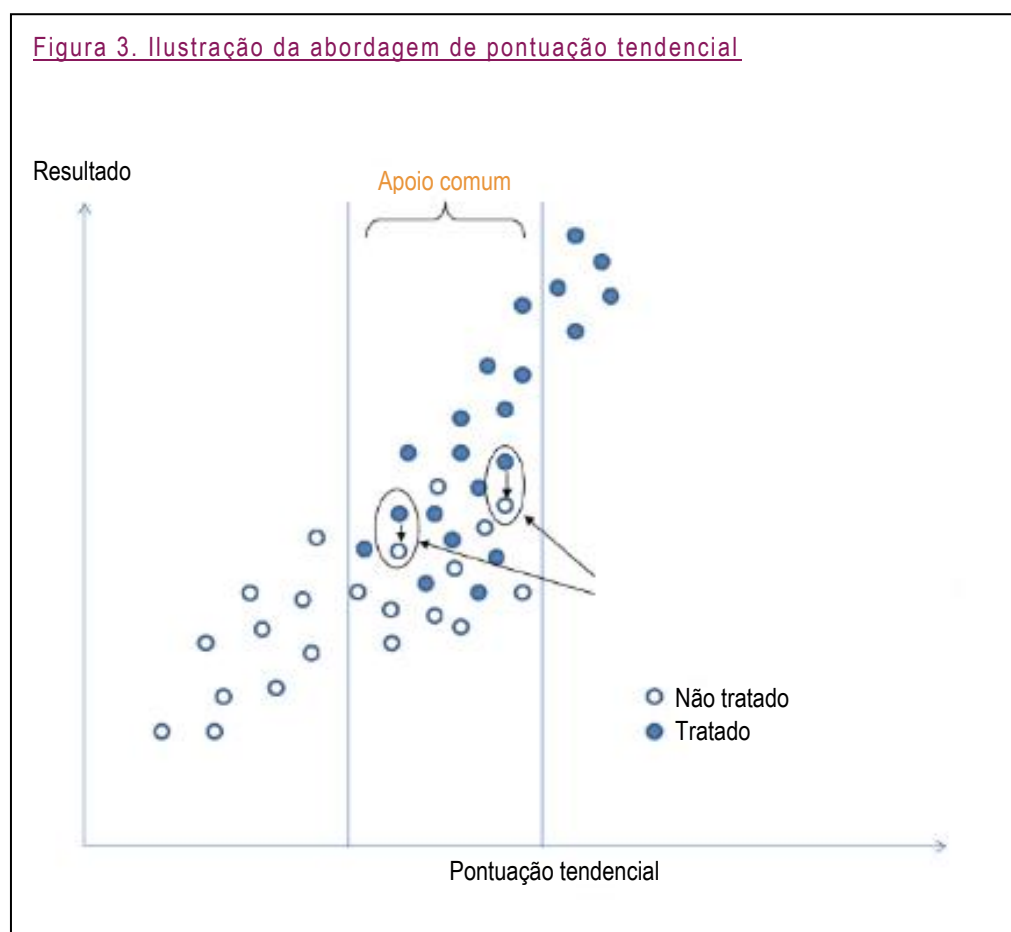
Uma importante estratégia possível para encontrar um grupo de controlo válido num cenário quase experimental é a da seleção de membros para o grupo de controlo que tenham sido excluídos do tratamento com base em fatores não relacionados com as suas características e os seus resultados potenciais. Em algumas circunstâncias, pode haver razão para acreditar que, embora os grupos de controlo não tenham sido construídos explicitamente de forma aleatória, podem *a posteriori* ser encontrados indivíduos ou empresas cuja exclusão do tratamento tenha sido aleatória no que respeita a resultados potenciais – estas circunstâncias, caso se materializem, estão próximas das ideais, no contexto de uma abordagem quase experimental. Por exemplo, certos membros de um grupo alvo de uma intervenção podem ser excluídos da participação na intervenção devido a uma desatenção ou a um erro. Compreender o processo de seleção para tratamento é extremamente importante na construção de um grupo de controlo válido – trata-se de um aspeto que nunca é demais salientar.

Um grupo de controlo credível pode ser desenvolvido de diferentes formas. Em primeiro lugar pode ser adotada uma abordagem de emparelhamento. Tipicamente é recolhida informação quer dos indivíduos em tratamento quer de uma amostra de pessoas semelhantes não tratadas, antes de os indivíduos tratados entrarem no programa. Depois, é ainda constituído um grupo de controlo a partir do grupo de indivíduos não tratados. Este processo é muitas vezes levado a cabo através de uma abordagem de “pontuação tendencial”.

O grupo de controlo e o grupo tratado têm de ser semelhantes

Emparelhar indivíduos tratados e não tratados

Figura 3. Ilustração da abordagem de pontuação tendencial



EPP – estimativas
para toda a amostra

O **emparelhamento da propensão a participar na intervenção** (EPP) implica o cálculo de um modelo estatístico para toda a amostra (membros do grupo tratado e potenciais membros do grupo de controlo) que tenha como resultado uma tendência estimada de participação para cada indivíduo ou empresa – independente de ter ou não participado realmente na intervenção⁸.

Os indivíduos ou as empresas que tiveram tratamento são então emparelhados – com um indivíduo ou uma empresa que não teve tratamento ou com muitos indivíduos ou empresas que não tiveram tratamento – com base na pontuação tendencial⁹. Um grupo de controlo identificado desta forma pode subsequentemente ser utilizado para produzir uma estimativa do contrafactual. O emparelhamento efetuado deste modo assegura que as estimativas de impacto têm em conta as diferenças observáveis entre o grupo tratado e os que servem de controlo, sendo assim possível, com base em certos pressupostos, alcançar uma estimativa não enviesada dos efeitos da intervenção. No entanto, se a seleção para o tratamento se basear em fatores não observados, a dúvida sobre a adequação do emparelhamento no que respeita à capacidade de controlar enviesamentos manter-se-á. O pressuposto crítico que subjaz à abordagem de emparelhamento é o de que o processo de seleção pode ser caracterizado pelos dados observáveis.

⁸ Para simplificar esta análise, pressupõe-se que os decisores políticos pretendem conhecer o efeito do tratamento naqueles que realmente usufruíram de serviços do programa (trata-se em muitos casos de um subconjunto do grupo-alvo ao qual foi dada essa oportunidade). É a análise denominada «tratamento nos tratados».

⁹ Existe um conjunto alargado de abordagens possíveis ao emparelhamento da propensão para participar na intervenção. Para um resumo acessível, ver Caliendo e Kopeinig (2005).

A Figura 3 apresenta uma ilustração intuitiva e simplificada da abordagem de emparelhamento da propensão a participar na intervenção. O eixo Y representa o resultado. O eixo X representa a pontuação tendencial. A figura considera unidades tratadas e não tratadas. A zona em que os coeficientes tendenciais para os dois grupos se sobrepõem é conhecida como a zona de apoio comum¹⁰. Nesta zona, os casos tratados são emparelhados com os não tratados. No diagrama, são apresentados dois exemplos, mas o processo é essencialmente repetido até que cada caso tratado seja emparelhado com um caso não tratado na zona de apoio comum. Na figura, este processo é executado por meio do emparelhamento com o “vizinho mais próximo”. O “vizinho mais próximo” de qualquer membro do grupo tratado é o membro observado do grupo de controlo com a pontuação tendencial mais próxima. Depois da formação de dois grupos, as médias de resultados podem ser comparadas para se obter uma estimativa do impacto. Na prática, levar a efeito um emparelhamento da propensão a participar na intervenção pode tornar-se um processo altamente complexo, em que é preciso ter em consideração um conjunto variado de aspetos. Muitos destes aspetos são ignorados aqui, de forma a assegurar que os princípios-chave são claros. Na Caixa 2, apresenta-se um exemplo prático de utilização da abordagem de emparelhamento na avaliação do FSE.

A plausibilidade da pontuação tendencial radica, entre outros, no pressuposto de que a seleção para tratamento pode ser completamente caracterizada pelos dados observáveis, ou seja, de que não existem, entre o grupo tratado e o grupo de controlo, diferenças observáveis relacionadas com os resultados e com a decisão de participar na intervenção. A plausibilidade deste pressuposto é reforçada pela incorporação de uma grande diversidade de variáveis no cálculo das pontuações tendenciais, sendo a seleção de variáveis baseada em conhecimento e em teoria prévios.

Isoladamente ou em conjunto com o emparelhamento, podem ser usadas medidas de referência (ou de pré-tratamento) de variáveis de resultado para levar a efeito o cálculo da **diferença-nas-diferenças** (DD). Aqui, a diferença num resultado de um grupo de controlo antes e depois do tratamento é subtraída da mesma diferença observada num grupo tratado, de forma a obter uma estimativa do impacto de uma intervenção. Mais uma vez, a seleção de um grupo de controlo plausível é essencial. Os impactos calculados com base no método da diferença-nas-diferenças são habitualmente produzidos num quadro de regressão que também dá conta de outras diferenças observadas entre o grupo tratado e o grupo de controlo.

Seleção baseada em dados observáveis

Diferença-nas-diferenças

¹⁰ A dimensão da zona de apoio comum tem implicações no tamanho da amostra e na utilidade dos resultados para a política, particularmente nos casos em que um grande número de casos tratados se encontra fora da região de apoio comum.

Além disso, esta abordagem faz o controlo de diferenças não observadas entre os dois grupos que são fixas ao longo do tempo, bem como de diferenças que variam no tempo, mas que afetam da mesma forma o grupo de controlo e o grupo tratado (por exemplo, fatores económicos gerais). Dada esta capacidade de controlar em alguns aspetos as diferenças não observadas entre o grupo tratado e o grupo de controlo, uma abordagem da diferença-nas-diferenças representa, na maior parte dos casos, uma melhoria relativamente a uma estratégia de emparelhamento intersectorial. A Figura 4 apresenta uma representação visual da abordagem da diferença-nas-diferenças.

Caixa 2. Exemplo de uma avaliação que adota abordagem de emparelhamento¹

Para avaliar o impacto dos vales (ou dotações) distribuídos aos desempregados na região italiana da Lombardia, foi usada uma abordagem de emparelhamento. Os desempregados podiam usar os vales para comprar formação e outros serviços de emprego. A avaliação procurou determinar o impacto dos vales sobre o emprego e sobre outros resultados de capacitação dos desempregados. Foi identificado um grupo de 800 participantes, bem como um grupo de indivíduos não tratados que se candidatou aos vales mas viu o financiamento ser-lhe negado por razões administrativas.

Foram obtidos indicadores de resultado de um conjunto variado de fontes de dados administrativos e resultantes de inquéritos. Para verificar as diferenças entre o grupo tratado e o não tratado, foi adotada uma abordagem de pontuação tendencial. Esta abordagem implicou o cálculo de uma equação de regressão logística, que produziu uma probabilidade de participação prevista no programa de vales para todas as unidades tratadas e não tratadas. De seguida, as pessoas tratadas foram emparelhadas com pessoas não tratadas, por meio de um conjunto variado de abordagens baseadas na pontuação tendencial.

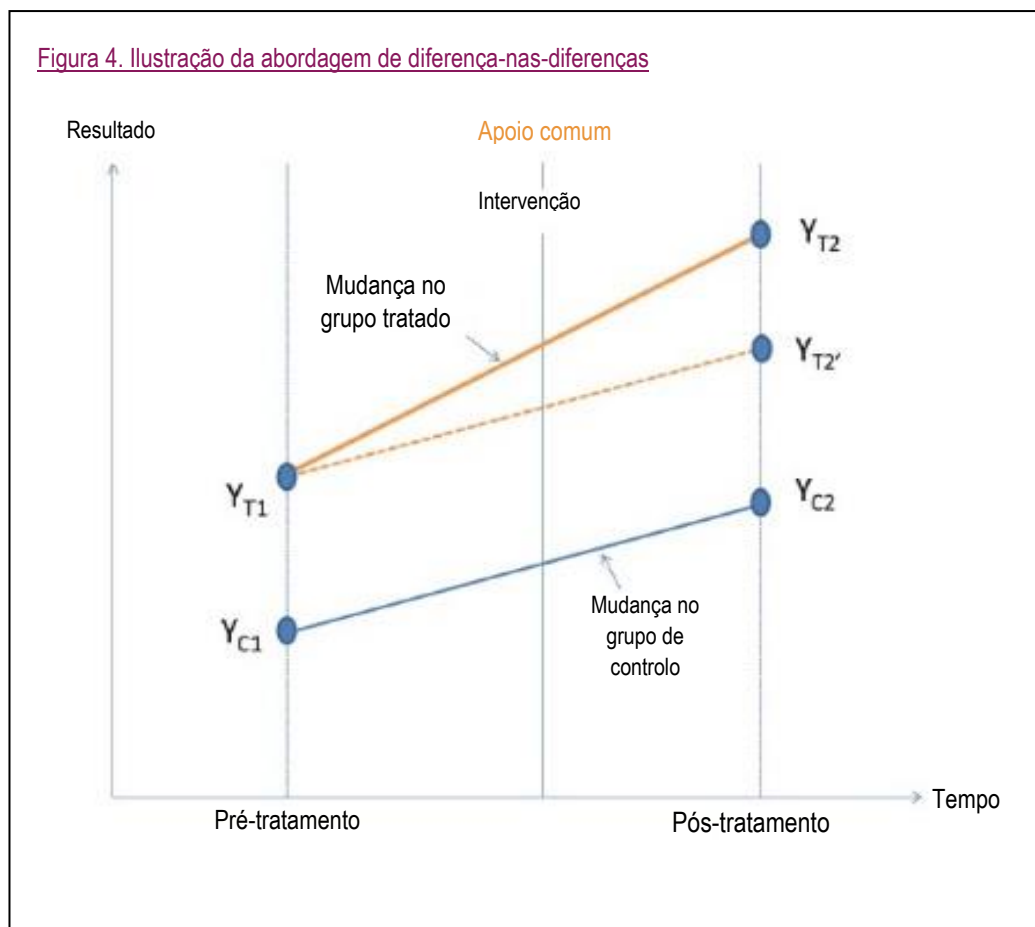
O estudo teve resultados mistos, tendo-se registado impactos positivos nas «dotações de emprego» e alguns impactos negativos nas «dotações de formação», embora alguns dados preliminares adicionais sugiram que a dotação de formação pode ter melhorado a qualidade do trabalho.

¹ Este e alguns dos exemplos apresentados abaixo são retirados da Audição de Especialistas que teve lugar no dia 25 de outubro de 2011, organizada pela Comissão Europeia. Representantes de oito estados membros e especialistas de avaliação apresentaram este e outros exemplos de AIC de ações cofinanciadas pelo FSE (ver também a referência na Introdução e a tabela de síntese no Anexo IV).

O eixo X representa a passagem do tempo e o eixo Y uma escala na qual os resultados são registados. Neste caso, os resultados podem ser salários. Os salários médios para o grupo tratado no período de pré-tratamento são representados pela expressão YT1 enquanto para o grupo de controlo são representados por YC1. No período pós-tratamento, os salários são representados por YT2 e YC2 para o grupo tratado e o grupo de controlo respetivamente. Assim, a linha contínua superior representa a mudança nos salários no grupo tratado, ao passo que a linha contínua inferior representa a mudança nos salários no grupo de controlo.

Uma comparação de salários no grupo tratado e no grupo de controlo no período pós-tratamento, ou seja, $YT2 - YC2$, teria como resultado uma estimativa simplista do impacto da intervenção. No entanto, uma estimativa deste tipo estaria incorreta uma vez que ignoraria as diferenças nos salários antes do tratamento. Uma forma de pensar no método de cálculo de estimativas da diferença-nas-diferenças consiste em considerar que subtrai uma estimativa de enviesamento anterior ao tratamento da diferença nos resultados depois do tratamento.

Figura 4. Ilustração da abordagem de diferença-nas-diferenças



Assim, a diferença nos salários depois do tratamento ($Y_{T2} - Y_{C2}$) é ajustada subtraindo-se-lhe a diferença nos salários antes do tratamento ($Y_{T1} - Y_{C1}$), pelo que o método de cálculo de estimativas de impacto da diferença-nas-diferenças pode ser escrito, simplesmente, como:

$$(Y_{T2} - Y_{C2}) - (Y_{T1} - Y_{C1})$$

Se as diferenças de salários pós-tratamento não forem ajustadas tendo em conta as diferenças pré-existent entre o grupo tratado e o grupo de controlo, as estimativas do resultado podem ser enviesadas. Em alternativa, como referido acima, pode considerar-se que a abordagem da diferença-nas-diferenças subtrai a mudança nos resultados no grupo de controlo da mudança observada no grupo tratado. A mudança observada no grupo de controlo é concebida como aquela que ocorreria no grupo tratado na ausência da intervenção.

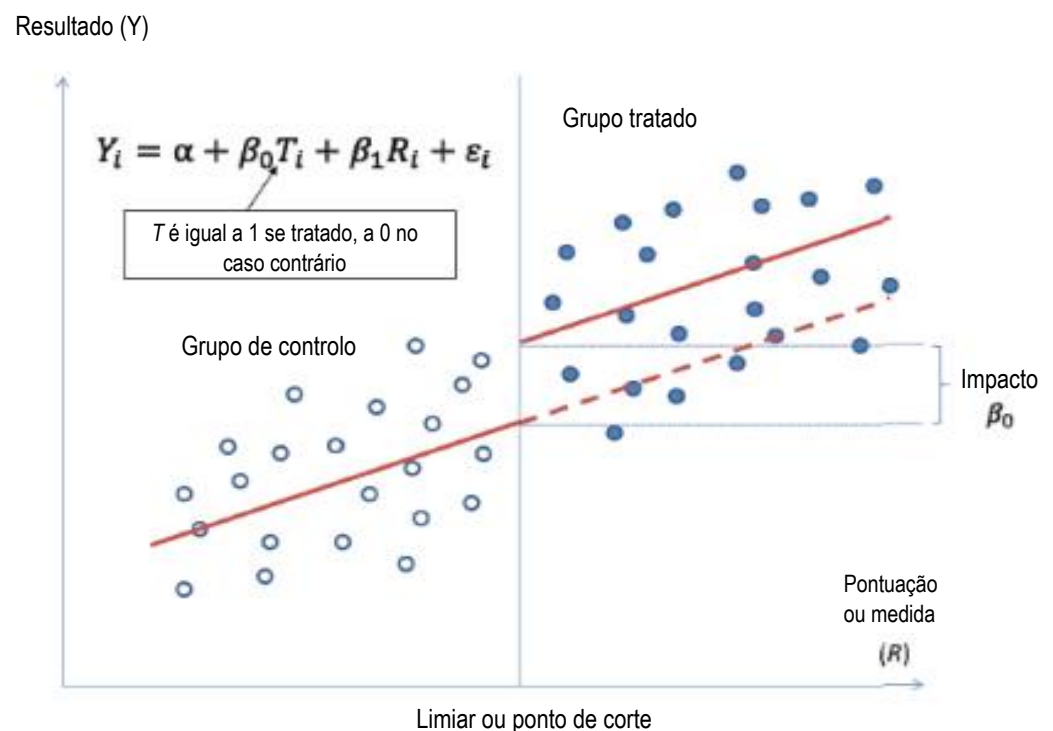
No caso mais simples, a principal assunção em que a abordagem da diferença-nas-diferenças se sustenta é a das tendências comuns; ou seja, as tendências de resultados nos grupos tratados e de controlo seriam equivalentes na ausência de tratamento. Este pressuposto não pode ser testado diretamente, embora nos casos em que estejam disponíveis várias medidas sobre os resultados anteriores ao tratamento quer para o grupo tratado quer para o grupo de controlo possam ser feitos alguns juízos no que respeita à plausibilidade do mesmo. Para um exemplo da abordagem por pontuação tendencial, ver a Caixa 3.

Pode ser adotada uma abordagem de **regressão descontínua** quando o acesso a uma intervenção é determinado por um ponto de corte numa avaliação, numa escala ou numa medida contínua.

A regressão descontínua compara grupos próximos de um limiar

Por exemplo, o acesso a formação pode ser determinado pelo desempenho num teste de aptidão, recebendo formação aqueles que obtenham uma pontuação superior a um limiar (ou ponto de corte) especificado, e não recebendo formação aqueles que tenham uma pontuação inferior. Para que a abordagem seja válida, o ponto de corte deverá ser determinado sem conhecimento das pontuações dos potenciais formandos. A abordagem tem em conta que aqueles que têm pontuações imediatamente em redor do ponto de corte são muito semelhantes, tirando o facto de que aqueles que têm pontuações imediatamente acima são expostos à intervenção, ao passo que os que se situam abaixo não o são. Os resultados daqueles que se situam acima e abaixo do ponto de corte podem ser comparados de forma a obter uma estimativa do impacto da intervenção no ponto de corte.

Figura 5. Ilustração da abordagem de regressão descontínua



Descontinuidade
estranque ou difusa

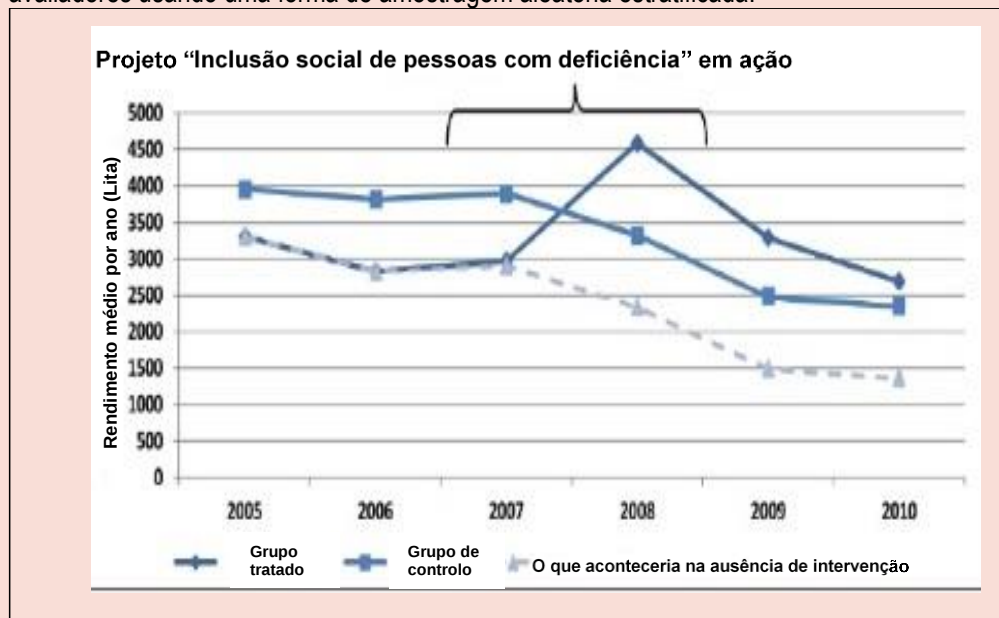
Pode ser implementada uma abordagem de regressão descontínua (ARD) nas situações em que o ponto de corte identifique completamente o grupo tratado (com conformidade total), caso em que é obtida uma descontinuidade estranque, ou nas situações em que, em certas condições, nem todos aqueles que se situem num dado lado do ponto de corte se encontrem em conformidade estrita com os critérios de seleção para tratamento (descontinuidade difusa).

A Figura 5 acima apresenta um exemplo estilizado de uma abordagem de regressão descontínua. Esta é a situação mais simples em que se verifica uma descontinuidade estranque. A intervenção produz efeitos constantes em cada valor da avaliação e os impactos são calculados através da utilização de uma abordagem de regressão linear (não existem problemas relacionados com a forma funcional do impacto da regressão). Na realidade, as análises terão invariavelmente de ser muito mais sofisticadas do que a apresentada na Figura 5.

Caixa 3. Exemplo de uma avaliação que adota a abordagem de diferença-nas-diferenças¹

Avaliação de serviços de integração social para indivíduos socialmente vulneráveis e socialmente excluídos no PO do FSE da Lituânia

Este estudo examinou os efeitos dos programas de integração social dirigidos a pessoas com deficiências e ex-reclusos no PO do FSE da Lituânia entre 2004 e 2006. O objetivo destes programas era a reintegração dos participantes no mercado de trabalho. Havia uma base de dados disponível que permitia aos avaliadores identificar quer aqueles que tinham participado no programa quer os que eram elegíveis mas não participaram. Os resultados considerados incluíam o estado de emprego, os rendimentos e a qualidade do trabalho. Foram identificados grupos tratados com cerca de 600 pessoas com deficiência e cerca de 200 ex-reclusos, bem como grupos de controlo com cerca de 1000 pessoas. Os grupos tratados eram compostos por participantes no programa, ao passo que os grupos de controlo foram construídos pelos avaliadores usando uma forma de amostragem aleatória estratificada.



É importante referir que os avaliadores tinham medidas sobre emprego e rendimentos para os grupos tratados e de controlo quer antes quer depois da intervenção. Isto permitiu-lhes implementar uma abordagem de diferença-nas-diferenças.

A figura acima, retirada do estudo, mostra a evolução dos rendimentos médios anuais das pessoas deficientes elegíveis no grupo tratado e no grupo de controlo. A tendência dos rendimentos anuais no grupo de controlo representa o contrafactual, a tendência que presumivelmente teria sido observada no grupo tratado se este não tivesse sido submetido à intervenção (a linha tracejada). Pode ser observado um impacto positivo nos rendimentos médios anuais. Outros resultados do estudo sugerem que a melhoria observada nos rendimentos anuais resulta do aumento do número de dias de trabalho no grupo tratado, e não do aumento dos salários.²

¹ Fonte: Audição de Especialistas, 25 de outubro de 2011.

² Instituto de Políticas Públicas e Gestão (2012): Avaliação dos serviços de integração social para indivíduos socialmente vulneráveis e socialmente excluídos para o uso eficaz da assistência estrutural da UE no período 2007-2013.

Os pontos da Figura 5 representam unidades individuais, por exemplo formandos. O eixo das abcissas regista a classificação, ou medida, usada para distribuir os formandos pelas vagas do curso de formação. Os indivíduos com uma pontuação nesta avaliação ou medida (que poderá ser um teste de aptidão, por exemplo) acima do limiar (indicado pela linha vertical contínua) entram na formação e formam o “grupo tratado”. Os potenciais formandos com uma pontuação inferior ao limiar na classificação ou medida não entram na formação e compõem o grupo de controlo.

O ponto-chave é o de que a pontuação utilizada para submeter o grupo alvo a condições de tratamento e de controlo é uma variável quantitativa contínua medida antes do tratamento, e um indivíduo entra no esquema de formação com base no facto de a pontuação obtida se situar acima ou abaixo de um ponto de corte ou limiar pré-definido.

O resultado é marcado no eixo das ordenadas, Y. Essencialmente, o impacto do tratamento é identificado estimando um modelo de regressão linear (dados os pressupostos acima) sobre os dados; ou seja, fazendo a regressão da variável-resultado em função da medida da pontuação, a par de uma variável de simulação (um indicador de tratamento) que permita verificar se uma observação se encontra acima ou abaixo do ponto de corte (ou seja, se a unidade é alocada ao grupo tratado ou ao grupo de controlo).

Esta equação de regressão do impacto é ilustrada na Figura 5. O efeito ou impacto da formação no nosso exemplo é obtido a partir do coeficiente sobre o indicador de tratamento, ou seja, β_0 ¹¹. Trata-se, efetivamente, de um teste para saber se existe uma quebra ou descontinuidade em redor do ponto de corte, indicada na Figura 5 por um movimento ascendente da linha de regressão no limiar ou ponto de corte. Neste exemplo, observa-se um impacto positivo da formação no resultado.

Uma forma alternativa de compreender a estimativa de impacto é atentando na extensão de linha tracejada para o grupo de controlo ilustrada na Figura 5. Pode considerar-se que se trata de uma estimativa contrafactual para o grupo tratado – a relação entre a medida de avaliação e a de resultado que prevaleceria na ausência de intervenção –, em que a diferença entre a linha tracejada e a linha de tendência para o grupo tratado representa o efeito ou impacto do tratamento. Note-se que, na ausência de tratamento, não existe descontinuidade na linha, e assumimos que o resultado varia continuamente em função da pontuação ou da medida, na ausência de tratamento. A Caixa 4 apresenta um exemplo prático em que a abordagem de regressão descontínua foi usada para uma avaliação dos fundos estruturais.

A abordagem de regressão descontínua funciona porque as observações dos grupos tratados e de controlo nas zonas próximas do ponto de corte são semelhantes em tudo menos no facto de que aqueles que se situam acima do ponto de corte, neste exemplo, recebem formação, ao passo que os que se situam abaixo não recebem. A situação não é, pois, dissemelhante da distribuição aleatória no caso das observações próximas do ponto de corte. Existe, porém, uma limitação considerável. Na maior parte das aplicações, os impactos calculados por meio da abordagem de ARD só podem informar o decisor sobre os efeitos no ponto de corte ou limiar. O grau em que as generalizações podem ser feitas aos casos afastados do limiar pode ser reduzido.

A ARD pode ser uma abordagem útil, sendo os indivíduos selecionados para uma intervenção com base em necessidades medidas numa escala contínua de notas ou de pontuação.

¹¹ Num caso simples, este seria o efeito da intenção de tratar no ponto de corte (ver Bloom, 2009).

No entanto, a análise pode tornar-se complexa nos casos em que o ponto de corte seja difuso e não exista conformidade, e nos casos em que existam problemas de forma funcional no modelo de regressão do impacto. Efetivamente, é preciso invocar um leque de pressupostos, e a veracidade destes pressupostos nem sempre pode ser verificada.

Caixa 4. Exemplo de uma avaliação que adota abordagem de regressão descontínua

Cálculo dos efeitos da política regional europeia no crescimento económico: uma abordagem de regressão descontínua

Os avaliadores usaram a abordagem de regressão descontínua para estimar os efeitos dos fundos regionais da UE no crescimento económico. Fazendo uso de dados referentes ao período compreendido entre 1995 e 2005, exploraram o facto de as regiões da UE-15 receberem fundos se o PIB *per capita* fosse inferior a 75 por cento da média europeia. Desta forma, a avaliação usada para atribuir o tratamento foi o PIB *per capita* e o ponto de corte ou limiar foi 75 por cento da média das regiões da UE como um todo. A estratégia de identificação sustentava-se no facto de que as regiões próximas do ponto de corte, situadas em qualquer dos lados do mesmo, eram semelhantes, exceto no sentido em que aquelas que se situavam abaixo do limiar recebiam fundos ao passo que as que estavam acima não recebiam.

Trata-se de um exemplo de uma ARD estanque. No entanto, os investigadores tiveram de enfrentar um conjunto de desafios. Entre eles, destaque-se o facto de não haver muitas regiões encontradas na vizinhança do limiar ou ponto de corte. Os investigadores usaram métodos paramétricos e não paramétricos de análise e levaram a cabo um conjunto de testes de robustez. Os resultados ditaram que os fundos regionais da UE têm um impacto pequeno e positivo no crescimento económico. As estimativas de impacto são estatisticamente significativas e robustas para diferentes especificações.¹

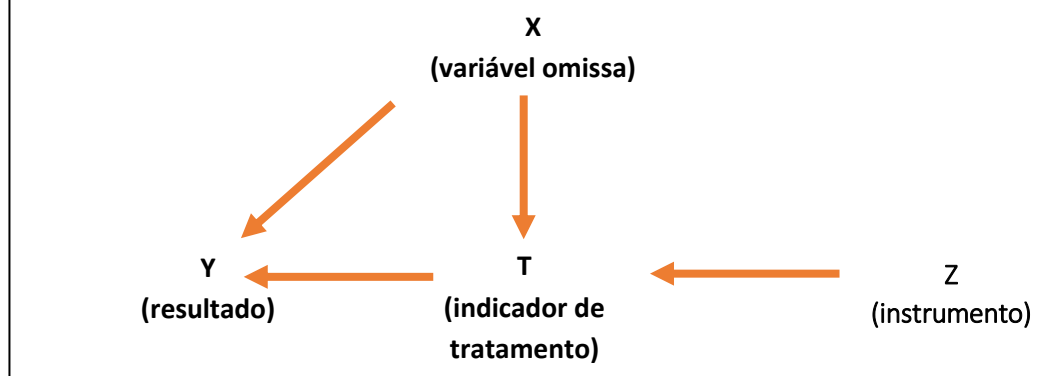
¹ Para mais informações, ver:

http://www.dps.tesoro.it/documentazione/uval/materiali_uval/european_regional_policy_Muval20.pdf

No que respeita à abordagem de **variáveis instrumentais** (VI), a seleção para tratamento deverá ser pelo menos parcialmente determinada por um fator exógeno (ou choque) que não tem relação com os resultados a não ser através do tratamento. Assim, o fator exógeno influencia a participação, mas não diretamente os resultados. Habitualmente, estes fatores exógenos podem ser erros administrativos ou desatenções, ou outras variações aleatórias na receita de tratamento. A Figura 6 ilustra a abordagem de variáveis instrumentais. São descritas quatro variáveis num sistema causal altamente simplificado. As variáveis representam dados recolhidos de uma população hipoteticamente selecionada por um esquema de formação (quer a que recebe formação quer a que funciona como controlo).

Variáveis
instrumentais

Figura 6. Ilustração de uma abordagem de variáveis instrumentais



Y representa o resultado em análise. No caso de uma intervenção de formação, este resultado poderá ser o salário. T é um indicador que revela se um indivíduo frequentou a formação¹².

X é uma variável omissa, que não é observada mas que está relacionada com Y, o resultado, e com T, o indicador de tratamento; se retomarmos a ideia de um programa de formação, pode tratar-se, por exemplo, de uma medida de referência de habilitações. Neste caso, as habilitações relacionam-se quer com a participação na formação quer com os salários. Por exemplo, os membros com mais habilitações do grupo-alvo podem decidir participar na formação, bem como receber salários mais elevados.

A existência de X promove a procura de um instrumento. A sua presença significa que o impacto da formação nos salários é confundido pela sua existência. Por outras palavras, o cálculo é enviesado devido à existência de X e ao facto de esta variável não ser observada e de não poder ser diretamente tomada em consideração pela análise.

Por fim, a variável Z é um instrumento. Nas palavras de Morgan e Winship (2007)¹³, pode ser considerado um choque para T, que é independente de X. Por esta razão, na Figura 6, não existe nenhuma linha a ligar Z a X. Além disso, Z só afeta Y através de T, não existindo outro modo através do qual Z afete Y. Isto significa que Z pode ser usado para gerar variação em T (o tratamento), que não esteja correlacionada com a desconcertante variável X. Por conseguinte, através da exploração desta variação, pode-se obter uma medida não enviesada do efeito de T sobre Y¹⁴.

As circunstâncias mais simplistas em que poderá ser adotada a abordagem VI são aqui descritas, necessariamente de uma forma que não tem em conta muita da complexidade envolvida. Na prática, é muitas vezes difícil encontrar um instrumento convincente. A plausibilidade de diferentes instrumentos possíveis é muito dependente do contexto, e os pressupostos identificadores subjacentes podem, em geral, ser testados estatisticamente. Uma estratégia poderá ser, por exemplo, a utilização da distância entre os centros onde a formação é ministrada (a localização física do curso de formação)

¹² Por outras palavras, existe conformidade total e todos os membros do grupo tratado participam no tratamento.

¹³ Morgan e Winship (2007).

¹⁴ O efeito causal de T sobre Y é calculado, havendo um instrumento, através da estimativa da relação entre Z e Y, dividindo-a pela relação estimada entre Z e T.

e a casa de um formando, como um instrumento para o cálculo do efeito da formação nos salários dos formandos. Poderá observar-se que existe uma maior probabilidade de os formandos que vivem mais perto dos centros de formação participarem numa intervenção de formação. Ou ainda, que a distância entre a casa do formando e um centro de formação não está relacionada com outros fatores que determinam os salários e a participação na formação (por exemplo, medidas de capital humano). Por isso, o único modo através do qual esta medida de distâncias poderá afetar os salários é através do seu efeito na formação¹⁵.

As variáveis instrumentais podem ser usadas numa grande variedade de contextos. Podem ser obtidas estimativas por meio da utilização de um conjunto variado de abordagens de cálculo dependendo da variável de resposta. Até ao momento, esta abordagem não foi usada na avaliação do FSE. Na Caixa 5, é apresentado um exemplo de análise de efeitos causais entre a reforma antecipada e a mortalidade.

Caixa 5. Exemplo de um estudo que adota uma abordagem de variáveis instrumentais

O risco de mortalidade é, por todos os motivos, significativamente mais alto para os reformados do que para trabalhadores mais velhos ainda envolvidos na atividade económica. Esta diferença poderá resultar de uma consequência perversa da reforma ou indicar simplesmente que os trabalhadores saudáveis adiam o abandono do trabalho remunerado. Num artigo recente¹, os investigadores utilizam uma técnica de variáveis instrumentais para calcular o efeito causal da reforma precoce sobre a mortalidade dos operários. Para ultrapassar o problema da “seleção endógena”, ou seja, de que os problemas de saúde levam à reforma e, logo, são causa e efeito, o estudo aproveita uma mudança nas regras do subsídio de desemprego na Áustria em 1988 (o Programa Regional Alargado de Benefícios – PRAB) que permite que os trabalhadores em regiões elegíveis se retirem da força de trabalho até três anos e meio antes dos de regiões não elegíveis. A residência numa região elegível pode ser empregada como um instrumento para a reforma antecipada, uma vez que a elegibilidade para o programa é independente do estado de saúde. A utilização de dados administrativos sobre o historial de trabalho e a mortalidade a partir da base de dados da Segurança Social austríaca permite comparar a mortalidade dos operários depois da reforma que se enquadram nos critérios demográficos e de emprego do PRAB, mas cuja residência e, logo, a elegibilidade seja diferente. No que respeita aos trabalhadores do sexo masculino, estas estimativas mostram um aumento significativo de 13% na probabilidade de morte antes dos 67 anos de idade em trabalhadores elegíveis para o PRAB. No caso das mulheres, não é encontrado nenhum efeito adverso da reforma sobre a mortalidade. Os dados sobre as causas de morte sugerem que as mudanças no que respeita a comportamentos relacionados com a saúde por parte dos homens com reforma antecipada podem explicar pelo menos parte do impacto. O programa terminou em 1993.

¹ Kuhn et al. (2010).

¹⁵ A interpretação dos resultados de tal análise pode ser complicada pelo facto de o instrumento estar ou não correlacionado com a variação no efeito de tratamento (ver Bryson *et al.*, 2002: 9).

Tabela 1. Comparação de algumas características-chave das principais abordagens de AIC

Abordagem	Características principais	Vantagens	Requisitos de dados	Limitações
Distribuição aleatória - abordagem experimental	Requer a distribuição do grupo-alvo, de forma aleatória, por grupos de “tratamento” e de “controle”	• Se for implementada corretamente, as estimativas de impacto não têm “enviesamento”. • Os resultados são transparentes e de fácil compreensão. • As conclusões são menos sujeitas a qualificação e a dúvida. • Variedade de modelos, ajustáveis a uma multiplicidade de contextos políticos e de circunstâncias de intervenção.	• Requisito fundamental é o controlo da seleção para a intervenção através de distribuição aleatória. • Registo dos nomes e dos respetivos grupos aos quais foram atribuídos. • Aconselhável para recolha de dados de referência. • É necessário efetuar um registo das medidas dos resultados dos grupos tratados e de controlo.	• Frequente negação explícita da intervenção por parte do grupo de controlo. • Muitas vezes, é necessário o consentimento dos participantes. • O processo de distribuição aleatória pode influenciar a composição do grupo de participantes / candidatos a uma intervenção. • Os resultados podem ser influenciados por os participantes terem conhecimento do estatuto que lhes foi atribuído, uma vez que este facto pode levar a uma alteração de comportamentos. • As questões éticas. • Requisitos consideráveis de planeamento e de conceção. • Pode ser dispendioso (não necessariamente).
Emparelhamento (da propensão a participar na intervenção)	Amostras de intervenção e de controlo são emparelhadas com base nas respetivas características observadas.	• Exige um bom conhecimento dos processos de seleção, mas não exige um controlo direto da seleção dos participantes na intervenção. • Pode ser aplicada retrospectivamente e numa variedade de contextos, se estiverem disponíveis os dados necessários. • É tecnicamente um método semiparamétrico de estimativa; requer um menor número de pressupostos paramétricos (por exemplo, não é necessário um pressuposto de padrão de regressão). • Pode ser utilizada para efetuar uma estimativa dos múltiplos efeitos do tratamento.	• Identificação rigorosa dos participantes na intervenção. • Fontes de dados a partir das quais se realiza uma experimentação do grupo de controlo. • Conceito claro de participação e um bom entendimento sobre a seleção para o tratamento. • Grande quantidade de dados, idealmente dados de referência, a partir dos quais se constrói o emparelhamento. • Medidas dos resultados da intervenção para os participantes e para os controlados.	• Requer uma quantidade considerável de dados que permitam uma caracterização completa do processo de seleção. • A validade depende da qualidade do grupo de controlo e da respetiva seleção cuidadosa e do grau de apoio comum. • Assenta no pressuposto de que a seleção para a participação na intervenção pode ser adequadamente caracterizada por dados observáveis. • O conjunto de diferentes abordagens para o emparelhamento que estão disponíveis exige uma análise cuidada. • Os resultados podem ser complexos para explicar e interpretar, e potencialmente ambíguos.
Diferença- nas-diferenças	Utiliza medidas de resultados pré-intervenção para os participantes e controlados da intervenção.	• Controlos para alguns aspetos das diferenças não observadas entre os participantes e o grupo de controlo. • Pode ser utilizada conjuntamente com a abordagem de emparelhamento. • Repetição da secção transversal ou do painel de métodos disponível.	• Os requisitos de dados são semelhantes aos de outras abordagens, mas com a exigência adicional de medidas de pré-intervenção sobre os resultados. • De forma a testar as hipóteses principais são necessárias múltiplas observações de pré-tratamento sobre os resultados tanto para o grupo de tratados e como para o de controlo.	• É necessário que o pressuposto de que haja tendências comuns nos resultados dos participantes e dos controlados seja invocado. • A análise pode tornar-se bastante complexa e pode ser mal interpretada. • Riqueza de dados de pré-tratamento sobre os resultados necessários para testar o pressuposto de tendências comuns. • Não pode ser utilizado para estimar efeitos de tratamento múltiplos¹.

Abordagem	Características principais	Vantagens	Requisitos dos dados	Limitações
Abordagens de regressão descontínua	Os membros de um grupo-alvo participam numa intervenção se a respetiva pontuação (ou nota) numa medida contínua exceder ou for inferior a um limiar ou ponto de corte pré-determinado. O limiar distingue o grupo tratado do grupo de controlo.	• Tanto a abordagem estanque quanto a difusa para ARD estão disponíveis. • Pode fornecer impactos sem enviesamento dos efeitos do tratamento, sob determinadas condições.	• A escolha do ponto de corte deve ser independente dos valores da pontuação atribuída a cada membro de um grupo-alvo. • São necessários dados acerca dos indivíduos relativamente à pontuação ou medida, ao limiar ou ao ponto de corte e aos resultados, tanto para os grupos de tratamento como para os de controlo.	• Esta abordagem não é válida sem uma medida contínua ou uma pontuação que determina o tratamento. • As análises podem rapidamente tornar-se complexas e incertas, se as questões da forma funcional da regressão do impacto se tornarem proeminentes, se houver um incumprimento e se o tamanho da amostra em torno do ponto de corte for limitado. • Podem surgir problemas na interpretação e na generalização das conclusões.
Variáveis instrumentais	Usa um instrumento (um tipo de variável) para isolar a variação exógena na receção de uma intervenção: a ideia de uma experiência natural.	• Pode fornecer estimativas de alta qualidade ou provas da existência de efeitos causais. • Resolve o problema de enviesamento da variável omitida (ou enviesamento de seleção). • Pode ser aplicado retrospectivamente. • Permite estimativas de determinados tipos de efeitos.	• Requer dados de referência, dados sobre resultados e receção da intervenção, mas também que um instrumento possa ser identificado. • Um instrumento precisa de estar relacionado com a receção da intervenção e afeta os resultados apenas através da receção da intervenção. • O instrumento não deve estar correlacionado com outros determinantes de resultados.	• Pode ser difícil encontrar um instrumento plausível. • Pode ser difícil de explicar a não-especialistas. • Interpretação dos resultados não clara. • Limitações de atestação quanto à identificação de pressupostos.

¹ Ver Frolich (2004): 181-224.

1.5 Como integrar uma AIC num quadro de avaliação mais alargado

Explicação causal e descrição causal

O que as AIC podem dizer aos decisores políticos e o que não podem

As avaliações contrafactuais dirigem-se a certos tipos de questões relacionadas com os efeitos causados por intervenções. Estas abordagens são limitadas no alcance que podem ter quanto a outras questões relacionadas com uma intervenção. Interessa distinguir as questões de avaliação referentes à **explicação causal** e as relativas à **descrição causal**. As AIC têm como objetivo **descrever** as consequências de uma intervenção. Estes métodos são menos adequados para **explicar** os mecanismos e os contextos através dos quais surgem as relações causais. Esta distinção é importante, uma vez que ajuda a clarificar o papel distintivo das AIC¹⁶.

Uma AIC bem desenhada informará o decisor sobre se a intervenção conduziu a mudanças nos resultados para influenciar os quais foi concebida. Produzirá prova da dimensão de qualquer impacto ou efeito, informará o decisor sobre se o impacto foi positivo ou negativo e produzirá uma medida de incerteza. O que as avaliações de impacto contrafactuais fazem menos bem é dar uma explicação da razão por que e da forma como surgiram os impactos que são avaliados através da AIC. Pelo contrário, é muitas vezes difícil determinar com base numa AIC a razão por que uma intervenção não teve impacto, se tal for o caso.

No seio da maioria dos organismos de decisão, as partes que pedem descrições causais e as que pedem explicações causais tendem a ter diferentes interesses e perspetivas. Os gestores e os executores de programas tendem a focar-se em questões relacionadas com explicações causais. Os alocadores de recursos e os principais decisores políticos responsáveis pela definição dos orçamentos tendem a focar-se em questões relacionadas com a descrição causal. Na prática, a distinção entre a explicação causal e a descrição causal pode ser pouco clara. Em algumas circunstâncias, as AIC podem explicar a razão por que foram encontrados alguns impactos, por exemplo através da exploração dos impactos das intervenções em subgrupos importantes. No entanto, é essencial considerar cuidadosamente os tipos de questões que as partes colocam sobre uma intervenção e selecionar abordagens apropriadas para responder às mesmas. Nos casos em que a questão principal é se uma intervenção funciona, uma avaliação de impacto contrafactual é, em muitas circunstâncias, adequada. Em casos em que a questão principal é a forma como uma intervenção funciona, a atenção vira-se para os métodos de avaliação dos processos baseados na teoria.

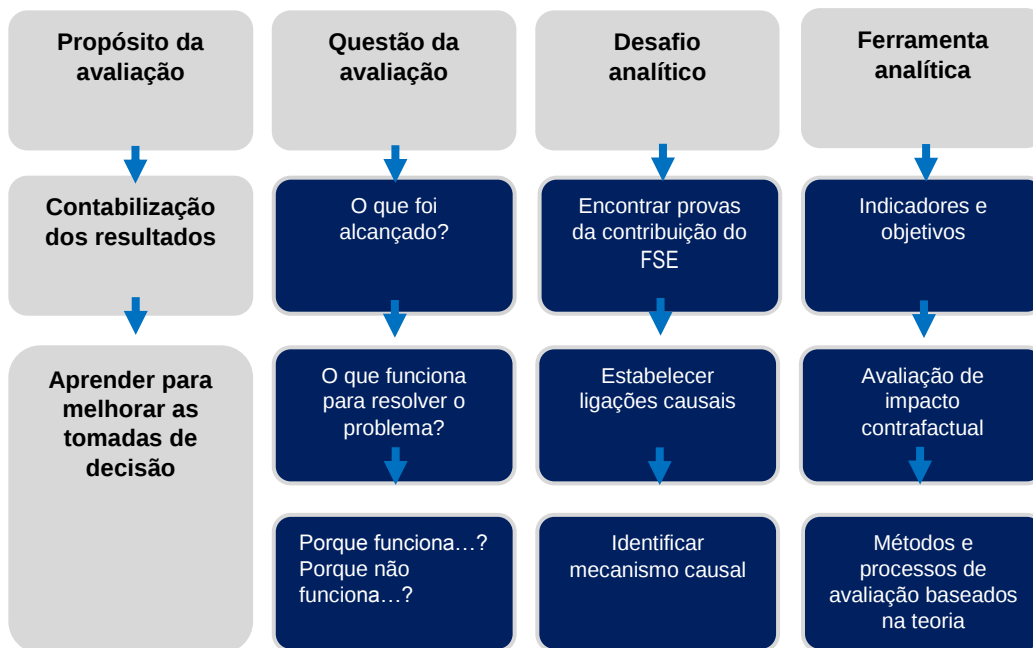
Os diferentes níveis de questões e de fins são resumidos na Figura 7.

Esta discussão leva à conclusão de que as AIC têm de ser desenvolvidas no âmbito do plano de avaliação. O plano de avaliação tem de englobar diferentes formas de avaliação, pensadas para responder a diferentes questões de diferentes decisores. Na prática, um plano de avaliação raramente, ou nunca, incluirá uma AIC sem uma avaliação de processos.

A avaliação desencadeia uma variedade de abordagens, que servem um conjunto de diferentes objetivos. A questão mais importante a colocar tem que ver com a forma como estas abordagens podem ser utilmente combinadas para informarem as políticas. Combinar diferentes tipos de avaliação de forma adequada – com diferentes objetivos no ciclo da programação – é o verdadeiro desafio neste campo. Como já foi discutido e se mostra na Figura 7, AIC, avaliação de processos e abordagens baseadas na teoria complementam-se.

¹⁶ Shadish *et al.* (2002).

Figura 7 Diferentes tarefas e tipos de avaliação



Fonte: adaptado de Martini (2009).

Para o período 2014-2020 do programa, o documento orientador da CE¹⁷ sobre monitorização e avaliação estabelece uma distinção entre diferentes formas de avaliação. Na discussão que se segue, a “análise de eficiência” é acrescentada a esta tipologia. Neste documento orientador, só se discutem abordagens contrafactuais para as avaliações de impacto. No contexto das AIC, as abordagens baseadas na teoria são meios de compreender a intenção havida ao conceber a intervenção.

Uma estratégia de avaliação forte deverá englobar os seguintes elementos:

- Avaliação baseada na teoria
- Avaliação de processos
- Avaliação de impacto contrafactual (AIC), e
- Análise de eficiência

As avaliações baseadas na teoria são usadas em algumas circunstâncias não apenas para tentar descrever o funcionamento pretendido da intervenção, mas também para verificar se se observam as mudanças nos resultados previstas pela teoria da intervenção ou pelo quadro lógico. Neste sentido, as abordagens baseadas na teoria podem ser usadas para avaliar o impacto geral, podendo também ser alargadas para descrever o impacto de uma intervenção nos casos em que não é possível uma AIC. A explicação pormenorizada da utilização de abordagens de determinação de impactos baseadas na teoria não se enquadra no escopo do presente documento.

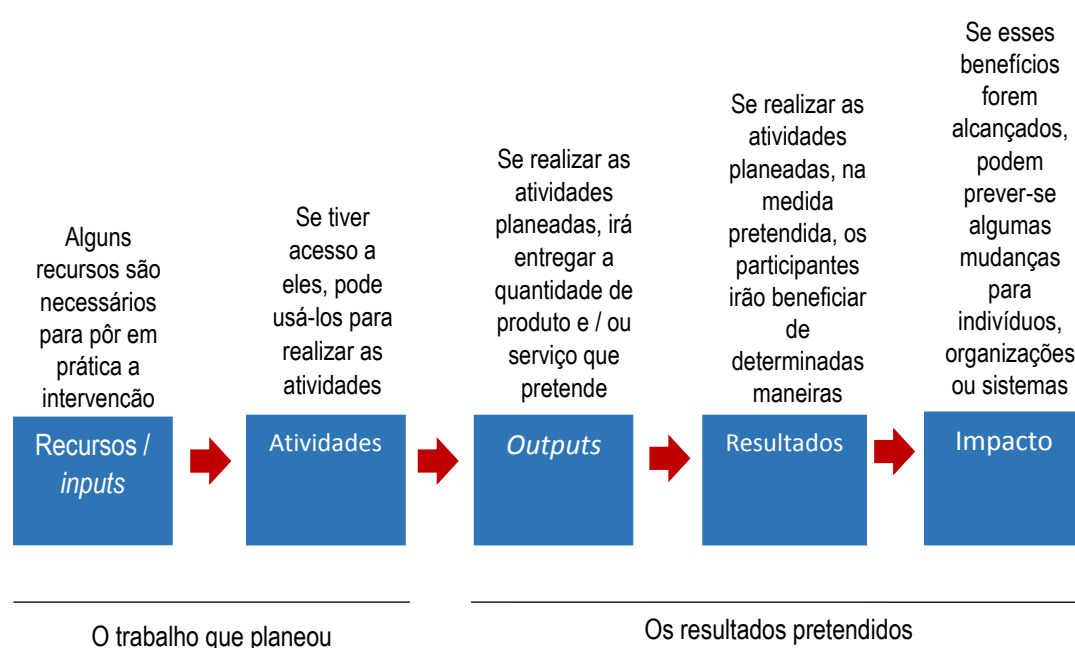
Combinação de tipos de avaliação

A avaliação baseada na teoria remete para um quadro lógico

¹⁷ Comissão Europeia (2011a).

No contexto das AIC, a avaliação baseada na teoria tem em conta a forma com uma intervenção é planeada e concebida e como se pretende que funcione. Essencialmente, a abordagem envolve trabalho com os parceiros de uma intervenção no desenvolvimento de um relatório partilhado da “teoria de mudança” subjacente a uma intervenção – aplicam-se métodos semelhantes na identificação do “quadro lógico” de uma intervenção. Uma abordagem baseada na teoria pode também englobar tentativas para estimar a adequação da lógica de intervenção subjacente – perceber se é exequível. Assume-se que todas as intervenções são compostas por uma lógica programática que liga os *inputs* e as atividades a *outputs*, a resultados intermédios e, mais tarde, de longo prazo. Em algumas aplicações, os investigadores utilizam modelos lógicos para facilitar a articulação de uma teoria de mudança. Uma discussão pormenorizada destas abordagens não se enquadra no escopo deste documento. No entanto, abaixo apresenta-se uma ilustração muito básica da abordagem de modelo lógico para desenvolver uma teoria programática.

Figura 8. Ilustração da abordagem de modelo lógico



Fonte: Adaptado da Fundação Kellog (2004)

... a acrescentar às
AIC

A avaliação baseada na teoria pode ligar as avaliações de impacto contrafactuais de diferentes formas. Uma teoria articulada sobre a mudança (ou lógica de intervenção) pode informar a conceção de uma AIC. Entre outros aspetos, uma teoria bem definida de mudança pode informar o responsável pela conceção de uma avaliação de impacto sobre o seguinte:

- Quais são os resultados importantes e que exigem medição?
- Qual será o sinal e a dimensão provável dos impactos da intervenção?
- Qual é o grupo alvo pretendido e como poderá ser selecionado um grupo de controlo?
- Quanto tempo será necessário para que os efeitos do programa se materializem e ao fim de quanto tempo poderão os resultados materializar-se?
- Que dados poderão ser necessários para medir a participação na intervenção?

- Quão plausíveis são os pressupostos identificadores (validade dos instrumentos entre outros)?

Desenvolver uma teoria de mudança poderá também ajudar a identificar possíveis efeitos não previstos que podem ser tidos em conta na conceção da AIC. Em certa medida, uma teoria de mudança articulada de forma clara poderá também ajudar o avaliador a interpretar os resultados de um estudo de AIC. No entanto, no que respeita à interpretação, uma avaliação de processos será mais informativa.

A **avaliação de processos** feita no contexto da AIC tem dois objetivos. O primeiro é estimar a fidelidade, o segundo é estimar a diferença entre o que experimentam os participantes do grupo tratado e os do grupo de controlo.

A estimativa de fidelidade indica até que ponto a execução de uma intervenção é fiel à sua conceção. Uma avaliação de processos tem em conta os serviços que foram realmente disponibilizados aos participantes de uma intervenção. Serão estes os serviços previstos pela teoria da intervenção? O que explica a variação em diferentes locais, se tal variação ocorrer? A maioria das intervenções tem uma lógica de gestão e de efeito. A lógica de gestão diz respeito à forma como se espera que os organismos que implementam as políticas respondam a incentivos programáticos. A lógica de efeito diz respeito à forma como se espera que as pessoas que são alvo de intervenção respondam, tendo em conta o que lhes foi efetivamente facultado. O aspeto fidelidade da análise de processos fornece informação sobre o que efetivamente se realizou numa intervenção, logo, sobre o que na realidade contribui para os efeitos observados. Dá também um retorno importante para a gestão de projetos.

A estimativa de diferenças é especialmente importante no contexto da avaliação contrafactual. É habitual que o enfoque incida sobre os impactos da intervenção, como acontece em grande parte deste documento orientador. Mas antes do impacto nos resultados está o impacto nos *inputs*, a diferença de oportunidades entre o grupo tratado e o grupo de controlo que uma intervenção realmente produz. Em princípio, qualquer AIC pode ser “virada do avesso”, podendo o grupo tratado ser usado como grupo de controlo para estimar o resultado para as pessoas naquele que era, antes da inversão, chamado grupo de controlo. A implicação é a de que é preciso conhecer tão bem a experiência do grupo de controlo como se conhece a do grupo tratado, uma vez que é a diferença de *inputs* entre o grupo tratado e do grupo de controlo que a AIC atribui a causa das diferenças nos resultados.

Voltando ao esquema de formação, podemos imaginar duas circunstâncias iniciais bem diferentes. Numa, o esquema de formação é fornecido num contexto geral em que nada do mesmo tipo esteja disponível algures. O grupo de controlo consegue simplesmente viver sem ele. Mas outra possibilidade é a de que existam alternativas. Pode haver disponibilização de formação, por exemplo, em empresas especializadas em preparação vocacional. Se for este o caso, a análise de processos terá de incluir, na medida do possível, uma estimativa da diferença entre a formação recebida pelo grupo tratado e a recebida pelo grupo de controlo, e não apenas pressupor que todas as dimensões do tratamento estão fora do alcance do grupo de controlo.

Embora as avaliações de processos possam ser encomendadas de forma completamente independente de outras formas de avaliação, a importância que a as avaliações de processos e as de impactos têm para a gestão e para as AIC torna essencial que sejam planeadas conjuntamente.

Avaliação de processos

Estimativa de fidelidade

Diferenças entre o grupo tratado e o grupo de controlo

As AIC precisam de uma avaliação de processos

As boas análises de processos podem contribuir para se conseguir fidelidade, e as avaliações de processos fornecem um relato explicativo causal de uma intervenção. Sem uma avaliação de processos bem concebida, é muitas vezes difícil interpretar na plenitude os resultados de uma AIC ou calcular os custos necessários para uma estimativa de custo-benefício, depois de conhecidas as estimativas de impacto.

Como referido acima, a avaliação de processos pode também contribuir para a interpretação das conclusões das avaliações de impacto, explicando o contexto em que uma intervenção teve lugar. Compreender o contexto é importante, uma vez que indica até que ponto uma intervenção poderá produzir efeitos semelhantes se for implementada noutro lugar, em diferentes áreas geográficas ou em diferentes tempos. Trata-se de um aspeto especialmente importante para a discussão acerca da transferibilidade de abordagens e para destacar boas práticas na aprendizagem e no intercâmbio transnacional. A análise de processos contribui para a confiança na chamada **validade externa** dos resultados da avaliação.

Análise de eficiência

Na maior parte das aplicações, a análise de eficiência envolve uma estimativa de custo-eficácia ou uma análise de custo-benefício completa.

Determinação de rácios de custo- eficácia

A **análise de custo-eficácia** implica a comparação dos custos da intervenção com os efeitos ou o impacto que poderão derivar-se de uma AIC. Em termos simples, obtém-se um rácio de custo-eficácia dividindo o impacto de uma intervenção – expresso nas unidades de medida ou em unidades padronizadas – pelo custo líquido por unidade tratada da execução da intervenção.

ACB para comparar benefícios com custo líquido

Um rácio de custo-eficácia de um programa de formação que pretenda ajudar pessoas desempregadas a encontrar emprego poderá revelar a quantidade de fundos que é preciso despende por participante para levar um participante do desemprego ao emprego.

Um rácio de custo-eficácia é uma medida importante para os responsáveis pela alocação de recursos em diferentes programas. Os rácios obtidos de um conjunto de intervenções diferentes permitem aos alocadores de recursos a formação de juízos relativos sobre quais são as intervenções mais rendíveis.

Em vez de expressar os efeitos do programa na respetiva unidade de medida ou em unidades padronizadas, uma **análise de custo-benefício** (ACB) tenta quantificar em dinheiro as estimativas de impacto obtidas a partir de uma AIC e compará-las com os custos líquidos de uma intervenção. O objetivo da análise de custo-benefício é determinar se os benefícios quantificados em dinheiro de um programa superam os respetivos custos líquidos. Uma análise de custo-benefício de um programa de formação do FSE compararia os benefícios da intervenção para os participantes, para o governo e para a sociedade em geral com os respetivos custos líquidos. Para os participantes, os benefícios do programa (habitualmente, maior empregabilidade e aumento dos rendimentos líquidos) são obtidos a partir de uma AIC. Destes subtrair-se-á o valor dos impostos pagos pelos participantes e outros custos de emprego, de forma a obter o benefício líquido. Na perspetiva do governo, os benefícios da intervenção advirão de receitas adicionais de impostos e de uma redução nos pagamentos de segurança social, ainda que o governo suporte a maior parte dos custos da intervenção. Os custos para a sociedade como um todo derivam-se da soma dos benefícios para os participantes e para o governo, subtraindo-se a esta soma o total de custos para os participantes e para o governo.

As estimativas de impacto de uma AIC são uma componente fundamental das análises de custo-eficácia e de custo-benefício. As primeiras proporcionam as medidas de eficácia, ao passo que as segundas constituem uma fonte fundamental para estimar os benefícios em dinheiro. O que também é claro é que os estudos de custo-eficácia e as análises de custo-benefício exigem a recolha de dados precisos sobre os custos, a partir dos quais podem ser derivados os custos líquidos. Estas atividades denominam-se habitualmente **estudos de custo**. Em algumas avaliações complexas com combinação de métodos, os estudos de custos são muitas vezes integrados na avaliação de processos, em que os instrumentos de investigação podem ser adaptados de forma a recolher importantes dados sobre os custos.

Capítulo 2

Considerações práticas na preparação de uma AIC

Este capítulo aborda questões práticas a considerar no momento da preparação de uma avaliação. Utiliza-se aquando do planeamento das atividades de avaliação, de modo a decidir quais as intervenções que serão sujeitas a uma abordagem de AIC e a identificar as questões-chave a considerar na elaboração de uma AIC.

O ponto de partida será aquele em que um gestor de um programa de uma AG (ou um dirigente de um organismo intermédio (OI) responsável pela implementação de uma intervenção do FSE) pondera quais as intervenções a avaliar e quais as estratégias mais adequadas para uma AIC. Parte-se igualmente do princípio de que as avaliações não serão efetuadas pelos membros da AG, antes sendo objeto de subcontratação ou de encomenda a especialistas externos. Embora a AIC seja efetuada por uma entidade subcontratada, a AG (ou o OI) terá de planejar e preparar uma avaliação de impacto antes de a encomendar a terceiros.

A estratégia de avaliação, incluindo os vários tipos de avaliações, conforme descrito no capítulo anterior, deve ser delineada no plano de avaliação.

Os planos de avaliação são geralmente recomendados pela CE – não apenas no âmbito do objetivo da convergência, mas também nos dos objetivos da competitividade regional e do emprego. Estes planos deverão ser estabelecidos no início do período de programação e incluir os preparativos para o processo de avaliação (especialmente a articulação entre a avaliação e a monitorização), as atividades reais da avaliação (p. ex., uma lista indicativa de avaliações a efetuar, âmbito de cada avaliação, principais questões, utilização potencial, calendarização indicativa, estrutura de gestão), periodicidade e prazos, orçamento global e desenvolvimento de capacidades¹⁸.

Os planos de avaliação tendem a ser gerais por natureza, enquanto o planeamento de uma AIC exige uma preparação mais pormenorizada. Idealmente esta preparação deveria ter lugar no momento em que o plano da avaliação é elaborado, podendo alguns pormenores ser acrescentados posteriormente. No entanto, as AG (ou o OI) devem estar conscientes de que o estabelecimento de ligações com as partes interessadas e outros preparativos necessários à recolha de dados relacionados com a intervenção frequentemente se deparam com dificuldades e necessitam de ser planeados com a devida antecedência.

¹⁸ Os planos de avaliação são requeridos de acordo com o Art.º 48 do Regulamento do Conselho (CE) n.º 1083/2006. Para o período de programação 2014-2020, deverá ser preparado um plano de avaliação para cada programa operacional, Art.º 49 do Projeto de Regulamento de Disposições Comuns COM (2011) 615 final. Mais pormenores são especificados em *Orientações Indicativas sobre as Normas de Qualidade da Avaliação* (CE, 2008) e no respetivo documento de orientação para o período de programação 2014-2020; ver Comissão Europeia (2012).

Desenvolvimento de um esquema de avaliação para intervenções específicas

Este guia incide sobre as formas de desenvolver um esquema de avaliação para intervenções específicas candidatas a AIC. Este esquema poderá fazer parte do plano de avaliação ou, em alternativa, poderá ser estabelecido como uma etapa operacional subsequente. Nem todas as intervenções financiadas pelo FSE podem ser objeto de avaliação contrafactual. Deverá ter lugar um processo de seleção de intervenções para avaliação de impacto. Este guia sugere alguns aspetos que as AG deverão ter em conta ao selecionarem intervenções apropriadas. Acresce que o objetivo central deste guia é o de ajudar os responsáveis pela encomenda de uma AIC a ponderarem os desafios com que poderão ser confrontados para assegurar o sucesso da avaliação de impacto, e, ao fazê-lo, desenvolverem esquemas de avaliação para as várias AIC que estão a considerar executar.

Este guia prevê que, após selecionar as intervenções que serão objeto de uma AIC, a AG deverá elaborar um esquema de avaliação para cada intervenção escolhida. O termo “esquema” é utilizado para distinguir esta atividade dos “planos” formais de avaliação exigidos pelo Regulamento Geral 1083/2006 e pelo Projeto de Regulamento de Disposições Comuns para o período de programação 2014-2020 (Comissão Europeia, 2011).

Estes esquemas constituirão a base para a encomenda de AIC por parte das AG e estabelecerão os alicerces que permitirão que as entidades subcontratadas levem a cabo um estudo rigoroso e válido. O resto do capítulo analisa as questões que deverão ser, a um nível mínimo, defrontadas durante o planeamento de uma avaliação. Clarificando, os esquemas de avaliação deverão ser adaptados às circunstâncias específicas em que a intervenção tem lugar. É impossível especular sobre quais poderão ser essas circunstâncias específicas. Como resultado, este guia analisa questões que a) deverão ser abordadas em esquemas, ou b) deverão estimular a reflexão sobre os desafios que os esquemas terão de abordar.

2.1. Seleção de intervenções para avaliação de impacto

Critérios para a seleção de intervenções

A seleção de intervenções para avaliação do impacto passa por três etapas principais. Em primeiro lugar, será necessário tomar em consideração questões estratégicas. Em segundo lugar, uma vez clarificadas as prioridades estratégicas, deve ser feita uma estimativa para cada intervenção, no sentido de saber se cumpre os requisitos fundamentais para uma abordagem contrafactual, e em que medida é inovadora e/ou poderá constituir uma contribuição significativa para a base de conhecimentos. Em terceiro lugar, será desde logo necessário estar atento à questão de saber se os tipos de dados necessários para efetuar uma AIC estão disponíveis, ou podem ser disponibilizados. Esta terceira questão merece uma atenção particular por se ter verificado que até ao momento tem constituído uma importante barreira à realização de avaliações contrafactuais de intervenções do FSE.

Caixa 6. Questões para selecionar intervenções para uma AIC

A AIC não é adequada para todas as intervenções e a realização de AIC para todos os candidatos é em geral pouco eficaz relativamente ao custo. As autoridades gestoras terão assim de fazer escolhas, afetando recursos para obter o máximo benefício. O plano de avaliação deverá refletir as prioridades estratégicas, a exequibilidade da AIC, e a disponibilidade dos dados necessários.

A **estratégia** é uma questão de escala, associada ao desenvolvimento de políticas e comporta incertezas. As AG deverão perguntar:

- Foram alocados montantes de fundos relativamente elevados a esta intervenção, e, consequentemente, é especialmente importante justificar despesas?
- Constituirá a medida o eixo central de um processo de reforma e será provável que os resultados da avaliação contribuam para uma reavaliação crítica do esforço? Será a intervenção inovadora, tendo sido testada por uma ação-piloto ou por um ensaio antes de ser implementada a uma escala mais alargada?
- Incide a intervenção sobre políticas para as quais sejam necessárias provas adicionais de eficácia?

A **exequibilidade** está relacionada tanto com as características das intervenções como com as circunstâncias em que são introduzidas. Os responsáveis pelo planeamento deverão perguntar:

- O tratamento a que a intervenção se aplica é autónomo, bem caracterizado e suficientemente homogéneo?
- Haverá um tratamento comparativo significativo utilizável para medir o impacto?
- A população-alvo para a intervenção é numerosa e bem definida?
- A teoria que associa a intervenção aos resultados esperados é coerente?
- Existem outras intervenções que possam complicar a situação?
- A partir da população-alvo é possível identificar claramente o grupo tratado?
- É provável que o grupo tratado tenha dimensão suficiente?
- É possível identificar um grupo de controlo credível?
- Existem ameaças à manutenção da diferença entre as experiências de tratamento e de controlo ao longo de um período de tempo suficiente para avaliar o impacto?

Os dados têm uma importância decisiva. A essência da AIC é a medição, e a medição requer informação quantitativa, tanto sobre os grupos tratado e de controlo como sobre o contexto em que a avaliação é efetuada. A decisão sobre os dados que serão necessários é normalmente determinada pela teoria da intervenção e pela estratégia adotada para estabelecer o contrafactual. Ao selecionar intervenções para AIC, as autoridades gestoras responsáveis pelo planeamento de uma AIC deverão perguntar:

- O que é essencial conhecer acerca dos membros do grupo-alvo e do grupo de controlo?
- O que é essencial conhecer acerca da natureza da intervenção efetivamente aplicada ao grupo tratado e em que medida a mesma diferiu do grupo de controlo?
- Que dados estão disponíveis a partir de fontes administrativas e de outras fontes?
- Existem dados disponíveis que descrevam carreiras individuais?
- Os dados individualizados provenientes de diferentes fontes podem ser cruzados?

Neste capítulo são apresentados mais pormenores sobre estas questões.

Acento tónico sobre
intervenções
intensivas em
recursos

2.1.1. Escolher intervenções para privilegiar a avaliação de impacto

Antes de seleccionar intervenções específicas para avaliação, deverão ser consideradas questões estratégicas mais vastas relacionadas com a seleção de intervenções para avaliação. Os benefícios que resultam de avaliações bem concebidas e rigorosas aproveitam não apenas às AG e aos EM que as encomendam, mas também a outros EM e respetivas AG, a outras partes interessadas e à Comissão.

Numa perspetiva estratégica, será necessário um processo de priorização. A este respeito, o acento tónico deveria recair sobre a seleção de intervenções para as quais as avaliações de impacto prometem o maior retorno em termos de aprendizagem sobre aquilo que funciona.

Contribuição para justificar despesas

Uma vez que as AIC dão prioridade a questões críticas para os decisores políticos, especialmente para aqueles que são responsáveis por decisões de alocação de recursos, faz sentido concentrar os esforços de avaliação de impacto em programas e em intervenções especialmente intensivas em recursos. Quanto mais tempo e outros recursos um programa ou uma intervenção absorvem, maior a importância de saber se esse programa ou essa intervenção funcionam, e, conseqüentemente, se os benefícios gerados excedem os custos suportados. Intervenções dispendiosas que não criam valor social ou económico poderão ter de ser repensadas, enquanto outras comprovadamente criadoras de valor poderão merecer acréscimo de financiamento e de atenção.

Os resultados de avaliações *ex post* de intervenções financiadas no período de programação anterior têm revelado a necessidade de centrar as atenções nos principais objetivos das políticas. Para alcançar impacto social e económico é frequentemente necessário que a despesa tenha massa crítica. As AIC oferecem a perspetiva de poderem seleccionar as intervenções, a fim de serem identificadas as abordagens mais eficazes para determinados grupos-alvo, e conseqüentemente de se maximizar o valor.

Contribuição de uma intervenção para um processo de reforma

As intervenções que constituem uma componente importante de um programa de reforma mais vasto são frequentemente as que conseguirão atrair financiamento significativo. No entanto, o facto de uma intervenção do FSE ser central para uma estratégia de inclusão social, ou para um aspeto crítico de um programa ativo de mercado de trabalho, acrescentará peso à intervenção, centrando as atenções sobre a mesma.

Inovadoras e exploratórias

Intervenções que sejam novas e inovadoras, e que estejam a ser monitorizadas são naturais candidatas a AIC. Testar os efeitos de intervenções através de monitorização ou de ensaio exige uma avaliação rigorosa. O ónus de avaliar através da implementação de uma AIC bem concebida é tanto maior quanto existir um compromisso claro de alargar ou de implementar a intervenção a uma escala mais ampla, em se prevendo que ela será bem-sucedida.

Contribuição para a aprendizagem

Existe uma necessidade acrescida de centrar a atenção e os recursos em áreas específicas de um programa – e em intervenções específicas no âmbito dessas áreas – sempre que são escassas ou inexistentes as provas relativamente ao que funciona no âmbito da área da política em causa. Ou seja, sempre que existe uma incerteza genuína quanto à continuidade da política e um risco de dependência excessiva de provas que podem não ser diretamente relevantes (por exemplo provas oriundas de outros países).

Avaliações de elevada qualidade podem ser consideradas um bem público. Os benefícios que geram em termos de aprendizagem alargam-se a outras partes interessadas além da AG em questão, pelo que importa considerar que outras partes interessadas podem beneficiar da avaliação de impacto proposta. Estas partes interessadas podem ser organismos intermediários ou agências que trabalham em intervenções no âmbito do mesmo programa, outras autoridades de gestão ou organismos intermediários no EM em causa, ou agências e instituições que lidam com fundos nacionais ou regionais. Outra parte interessada que naturalmente deverá ser considerada é a Comissão Europeia, e existem também partes interessadas em outros Estados-Membros que poderão aprender com uma avaliação. Ter em conta as necessidades de todo este universo, além das partes diretamente interessadas, é uma contribuição importante que os decisores políticos e os gestores de programas podem fazer para uma aprendizagem mútua.

Uma consideração estratégica final na seleção de áreas sobre as quais centrar a atenção para o desenvolvimento de AIC consistirá em dar prioridade a intervenções em que os benefícios das AIC possam ser demonstrados, ou seja, desenvolver avaliações que evidenciem esta abordagem e constituam um modelo a seguir. A Caixa 7 apresenta uma visão geral das motivações para dirigir AIC¹⁹.

Produzir provas

Defender os métodos de AIC

2.1.2. Selecionar intervenções passíveis de abordagem contrafactual

Tendo sido consideradas preocupações estratégicas mais vastas que poderão motivar a seleção de intervenções particulares de AIC, esta secção considera a natureza específica de intervenções que as poderão tornar passíveis de receber uma abordagem contrafactual. Na preparação para uma avaliação, importa selecionar intervenções com características apropriadas a uma abordagem contrafactual. São muitas e variadas essas características. Algumas delas poderão apontar para uma AIC em determinadas circunstâncias, mas em circunstâncias diferentes poderão frustrar as tentativas de implementar abordagens dessa natureza. Consequentemente, não é possível apresentar uma lista de considerações exaustiva. No entanto, algo pode ser dito acerca da natureza de intervenções com potencial para uma AIC bem-sucedida.

¹⁹ Esta Caixa é baseada uma vez mais nos exemplos apresentados na Audição de Especialistas de 25 de outubro de 2011. Uma listagem sistemática de todos os exemplos de AIC apresentados na Audição de Especialistas com os Estados-Membros em 25 de outubro de 2011 é apresentada de forma resumida na tabela que consta no Anexo 4. Um relatório de síntese pormenorizado sobre esta audição está disponível no CIRCA.

Caixa 7. Motivações para a realização de uma AIC

No caso dos países em fase de convergência, em que avultados montantes de fundos europeus se encontram disponíveis, as questões frequentemente abordadas numa AIC são abrangentes, como se verá a seguir.

Na Polónia (PL) o principal objetivo das AIC foi o de «compreender o impacto da Política de Coesão sobre o emprego e medir a eficácia da totalidade do financiamento do FSE para a população desempregada». Têm sido efetuadas várias AIC para estimar o impacto de intervenções cofinanciadas pelo FSE. Uma dessas AIC analisou o impacto do Programa Operacional sectorial para o Desenvolvimento de Recursos Humanos (2004 a 2006) relativos ao nível e à qualidade do emprego. Outra grande AIC incluiu uma estimativa da componente regional do Programa Operacional para o Capital Humano, avaliação que se encontra em curso. Estas AIC utilizaram dados de um grande número de Serviços Públicos de Emprego (SPE) locais – ou de delegações distritais de emprego (DDE) – que foram recolhidos para comparar os resultados no mercado de trabalho para os desempregados que receberam apoio do FSE com os dos desempregados que não receberam qualquer assistência.

Na República Checa (CZ) o objetivo global de uma AIC planeada é «promover a compreensão do impacto do FSE no desenvolvimento de empresas que recebem apoio através da formação». O plano consiste em realizar uma avaliação abrangente, que almeja comparar o desempenho das empresas que recebem formação financiada pelo FSE com o das empresas que não recebem esse apoio. Estão a ser ponderados diferentes métodos de fazer a estimativa.

Também no Programa Operacional da Competitividade Regional e do Emprego foram abordadas questões bastante abrangentes relativas à avaliação:

A motivação para a realização de AIC planeadas na Dinamarca (DK) é a de reforçar a avaliação e a medição do impacto das iniciativas lançadas pelos fóruns para o crescimento regional, a fim de apoiar o desenvolvimento e o crescimento da economia regional. O objetivo é o de reforçar o conhecimento das iniciativas que são mais eficazes e que asseguram que o dinheiro é bem gasto. Está planeada uma série de AIC que farão a estimativa do desempenho de projetos do FSE e do FEDER no âmbito de Programas Operacionais (OP), comparando empresas e/ou pessoas que tenham recebido apoio com grupos não tratados de empresas e/ou de pessoas atuando como controlo.

A AG galesa (UK) realizou uma AIC que estimou o impacto de intervenções ao abrigo do PO do FSE – objetivos de competitividade e convergência. A taxa de obtenção de emprego de pessoas que abandonam uma ação do FSE foi comparada com a do grupo de controlo obtido a partir do Inquérito à População Ativa do Reino Unido.

Estados-Membros com alocações do FSE menores em comparação com orçamentos para Políticas Ativas do Mercado de Trabalho (PAMT) concentram-se antes em comparações entre medidas nacionais e medidas financiadas pelo FSE (Áustria – AT), ou na análise de resultados intermédios *soft* de medidas financiadas pelo FSE, a fim obter mais conhecimentos sobre a forma como as medidas de PAMT ajudam as pessoas a ter sucesso no mercado de trabalho (Bélgica – BE).

Em alguns Estados-Membros, as AIC incidem sobre instrumentos individuais introduzidos recentemente:

AIC realizadas na Lituânia (LT), onde o FSE contribui para as PAMT com uma quota elevada, e na Lombardia (com alocações do FSE muito inferiores) foram motivadas pela vontade de compreender o impacto de instrumentos cofinanciados pelo FSE (vale de formação na Lombardia) para as pessoas desempregadas ou o impacto da formação e do apoio para grupos-alvo específicos (pessoas com deficiência e ex-reclusos – LT).

Caixa 8. Tipos mais comuns de intervenção e de grupos-alvo escolhidos para AIC do FSE

As AIC de intervenções financiadas pelo FSE tendem a ser as que são direcionadas para os desempregados e, entre estes, para subgrupos afetados por alguma desvantagem específica (p. ex., PL, LI e AT). No País de Gales, uma AIC estudou o destino de todas as pessoas que abandonaram programas do FSE. As AIC também incidem nos efeitos de programas de formação dirigidos aos empregados de empresas em que o objetivo é o de aumentar a produtividade / competitividade e evitar a perda de empregos (CZ, DK).

As intervenções mais frequentemente escolhidas para AIC foram diferentes formas de apoio aos desempregados, (formação, empréstimos para início de atividade, estágios, serviços de aconselhamento e de procura de emprego na PL; emprego assistido e formação na LT), novos instrumentos (vales para formação), através dos quais os desempregados podiam obter formação, ou serviços de emprego específicos na Lombardia.

Uma AIC realizada na Flandres procurou analisar os efeitos sobre “resultados *soft*” (p. ex., compreensão de oportunidades de emprego disponíveis) de vários tipos de formação (candidatura a um emprego, formação vocacional, formação personalizada), assistência no local de trabalho e outras ações. Algumas avaliações polacas incluíram também medidas sobre “resultados *soft*” (p. ex., autoconfiança, ultrapassagem de barreiras previamente identificadas, compreensão de oportunidades de emprego, etc.). Na Dinamarca está planeada uma AIC para os efeitos de criação de emprego do apoio do FSE para participantes em projetos do FSE (empresas e indivíduos).

Algumas das AIC centraram-se em instrumentos individuais (p. ex., vales de formação e emprego na Lombardia, IT, serviços de integração social para pessoas vulneráveis e socialmente excluídas na LT).

Foram também feitas estimativas das intervenções nos sistemas e nas estruturas através de uma abordagem contrafactual: na Hungria foi efetuada uma AIC para estimar os impactos da reforma dos SPE na posição dos desempregados no mercado de trabalho. A reforma dos SPE foi implementada sequencialmente em diferentes regiões, o que permitiu que os avaliadores comparassem os resultados em regiões onde as reformas tinham sido implementadas com os daquelas em que as reformas tinham ainda de ser levadas a cabo. Os investigadores utilizaram dados longitudinais dos registos administrativos e implementaram uma avaliação que adotou um modelo de AIC baseada na abordagem da diferença-nas-diferenças. Analisaram os impactos da reforma na entrada no mercado de trabalho e concluíram que a reforma tinha um efeito líquido positivo na obtenção de emprego.

O nível a que uma AIC pode ser efetuada poderá abranger o apoio do FSE num Estado-Membro ou numa região (ou seja, um ou vários eixos prioritários, subprioridades²⁰ ou operações²¹ num Programa Operacional) e poderá incidir sobre grupos-alvo ou sobre tipos de intervenção homogéneos (p. ex., formação) (ver Caixa 8).

²⁰ Por subprioridade entende-se o nível diretamente abaixo de um eixo prioritário, o que em alguns países é também referido como «Área-chave de Intervenção», «Área de Apoio» ou «Medida».

²¹ Segundo o Art.º 2(3) do Regulamento do Conselho (CE) No 1083/2006: «“operação” refere-se a um projeto ou grupo de projetos selecionados pela autoridade gestora do programa operacional em causa ou sob a sua responsabilidade, de acordo com critérios estabelecidos pelo comité de monitorização e implementado por um ou mais beneficiários, permitindo a concretização dos objetivos do eixo prioritário com o qual se relaciona».

Os exemplos dos Estados-Membros sugerem que os diferentes instrumentos utilizados no âmbito do FSE, incluindo formação, incentivos ao emprego e serviços ao mercado de trabalho (p. ex., orientação profissional, *coaching*), seriam adequados a AIC, enquanto intervenções relacionadas com a rotação de postos de trabalho e com o trabalho partilhado, com incentivos ao início de atividade ou com apoio para sistemas e estruturas parecem constituir um desafio maior para a adoção de uma abordagem AIC.

É instrutivo considerar quais as intervenções que se revelam mais promissoras na perspetiva de uma AIC, considerando as seguintes questões:

A intervenção é autónoma, bem caracterizada e relativamente homogénea?

Tratamento
claramente
diferenciável

O tratamento ou tratamentos proporcionados por uma intervenção precisam de ser diferenciáveis dos de outras intervenções. Além disso, deverá haver um contraste significativo entre aquilo que os participantes numa intervenção recebem e aquilo de que outros grupos similares de pessoas beneficiam. No caso de os tratamentos serem indefinidos ao ponto de não ser possível identificar um grupo autónomo de beneficiários, as abordagens contrafactuais não são possíveis nem desejáveis.

Os métodos de AIC tornam-se muito complexos e difíceis, se o estatuto de tratamento de uma determinada unidade (empresa ou indivíduo) afeta o resultado potencial de outras unidades (mediante os designados “efeitos de equilíbrio geral” mais vastos). Em programas de formação, isso pode ocorrer quando os beneficiários de um programa tornam difícil que pessoas que não participaram no programa encontrem trabalho a curto prazo. Sempre que se pense que isto constitui um problema substancial (por exemplo, no caso de intervenções de grande escala), poderá ser necessária uma análise macroeconómica para estimar a extensão dos efeitos de substituição e de deslocação. Sempre que exista a probabilidade de tais efeitos se verificarem, as AG deverão obter aconselhamento especializado.

Intervenções
homogéneas

A intervenção em si mesma deverá ser relativamente homogénea, o que significa que os participantes numa intervenção deverão receber ou estar expostos genericamente ao mesmo pacote de medidas. Caso a gama de medidas proporcionadas aos participantes numa única intervenção for demasiado diversa, existem diversas implicações para uma AIC. Em primeiro lugar, poderá verdadeiramente não fazer sentido falar de uma intervenção coerente, mas antes de intervenções com processos causais a funcionar separadamente; em segundo lugar, será difícil interpretar impactos que são reportados como efeitos líquidos médios de um grupo de intervenções heterogéneas; em terceiro lugar, a análise dos subgrupos poderá estar garantida, mas se existirem demasiados subgrupos no interior de um grupo tratado, as limitações quanto ao tamanho da amostra podem condicionar a capacidade para reportar conclusões que possam ser utilizadas.

O tratamento está a ser comparado com a não existência de tratamento ou existem outras formas de tratamento relevantes?

O FSE está a cofinanciar políticas de inclusão social e no mercado de trabalho aos níveis nacional e regional. Consequentemente, qualquer esquema de AIC deverá tomar em consideração se a intervenção é claramente identificável e se os indivíduos têm oportunidade de receber serviços de outros programas e de outras fontes de financiamento (nacionais ou regionais). O que é importante é que os tratamentos sob avaliação alterem realmente as oportunidades ou os recursos disponíveis para os participantes, em comparação com o que está disponível para os controlos e que a diferença possa ser medida e monitorizada.

Tratamentos
complexos

Estas questões de “tratamento complexo” tendem a ser inerentes a contextos específicos, e complicam a conceção e a implementação de uma AIC. A sua presença realça a importância de um planeamento cuidadoso da avaliação – desenvolvimento do esquema da avaliação – antes da implementação.

Existe um grupo-alvo de grande dimensão e claramente definido?

Comparativamente a outras formas de avaliação, as AIC requerem amostras de grande tamanho.

É essencial que os grupos-alvo sejam constituídos por indivíduos em número adequado e, além disso, deverá ser possível identificar grupos de controlo com uma dimensão suficiente. Esta questão será em seguida analisada mais pormenorizadamente.

É importante que a intervenção candidata a uma AIC seja direcionada para um grupo claramente definido. Sem um entendimento claro do grupo-alvo de uma intervenção, é difícil identificar um grupo de controlo credível. Algumas intervenções procuram deliberadamente recrutar pessoas para serem objeto de tratamento através de mecanismos informais, encorajando processos que ou não são predefinidos ou são demasiado prescritivos – o que pode tornar difícil identificar com precisão a população tratada.

Amostra de grande dimensão

Definição da identidade do grupo-alvo

Existe um mecanismo causal claro?

Como anteriormente mencionado, revela-se frequentemente útil efetuar uma avaliação teórica previamente a uma AIC. Desenvolver uma teoria de mudança, ou um modelo lógico para uma intervenção pode ajudar os responsáveis pela conceção de uma AIC de várias formas; desde logo, determinando se uma intervenção tem um mecanismo causal coerente que a fundamente. É pouco provável que intervenções sem um mecanismo causal claro e convincente produzam impactos de suficiente magnitude para serem identificados estatisticamente através de uma AIC.

Mecanismo de políticas distinto

Os resultados podem ser definidos quantitativamente?

Existe uma necessidade de obter medidas quantificáveis dos resultados. Esses dados e os indicadores podem ser obtidos a partir de fontes administrativas, ou de inquéritos especificamente direcionados.

Em algumas circunstâncias, as intervenções poderão aspirar a resultados que pressuponham condições específicas para serem medidos quantitativamente. A título de exemplo, uma intervenção poderá ter como objeto uma mudança de atitudes, de crenças ou de opiniões. Nesses casos será necessário realizar inquéritos para medir as mudanças alcançadas. Algumas intervenções têm resultados bastante vagos ou definidos de forma pouco clara. Também neste caso, o desenvolvimento de um quadro lógico para a intervenção poderá ajudar a melhorar o entendimento daquilo que uma intervenção aspira alcançar e de que forma pretende provocar a mudança nos resultados cujo estudo ou conhecimento possam interessar.

Necessidade de medir os resultados

A intervenção foi apresentada de forma a poder identificar um grupo de controlo credível?

Para identificar um grupo de controlo credível, é importante considerar a forma como as unidades tratadas (pessoas ou empresas) são selecionadas para uma intervenção ou decidem participar, se os mesmos instrumentos de pesquisa podem ser aplicados tanto à amostra de controlo como ao grupo tratado, e se será necessário selecionar as amostras de controlo de forma a estarem sujeitas a condições de mercado de trabalho idênticas às do grupo tratado. A Caixa 9 destaca alguns exemplos.

Mecanismo de seleção para tratamento

No caso de uma intervenção ser obrigatória e abranger toda a população-alvo praticamente em simultâneo, poderá revelar-se difícil identificar uma parte não tratada da população-alvo para constituir um grupo de controlo. As questões associadas à seleção de grupos de controlo são analisadas em mais pormenor na secção 2.2.6.

Caixa 9. Definição de grupos de controlo

Para as AIC efetuadas até ao momento destinadas a avaliar as intervenções do FSE, a seleção, tanto do grupo tratado como do grupo de controlo, foi determinada pelas questões de avaliação subjacentes, e também pela disponibilidade de dados adequados.

Em alguns casos, o grupo de controlo foi definido como as pessoas que não receberam qualquer tratamento:

- Num exemplo de uma AIC efetuada na Polónia, apenas 7 por cento da população tratada e 8 por cento do grupo de controlo tinham beneficiado de outras medidas. Assim, neste estudo em particular, a ministração de formação financiada pelo FSE foi efetivamente comparada com a não existência de qualquer formação.
- Na Lituânia, o grupo de controlo incluiu pessoas desempregadas com deficiência e ex-reclusos que eram participantes elegíveis, mas que não beneficiaram de intervenções particulares do FSE; no entanto, algumas dessas pessoas receberam serviços semelhantes através de instrumentos nacionais.
- No exemplo da Lombardia, na Itália, o grupo de controlo era constituído por candidatos à intervenção cuja candidatura não foi contemplada.

Noutros casos, foi bastante difícil estabelecer um grupo de controlo constituído por pessoas que não tivessem recebido quaisquer serviços. Consequentemente, tratamentos cujo estudo era de interesse foram comparados com tratamentos alternativos:

- Na Áustria, onde quase todos os desempregados receberam serviços, os resultados do mercado de trabalho para pessoas que receberam apoio financiado pelo FSE foram comparados com os de pessoas que receberam serviços através de instrumentos nacionais.
- No País de Gales, os resultados para pessoas que abandonaram programas do FSE foram comparados com os de uma amostra da população desempregada constituída a partir do Inquérito à População Ativa (IPA), mas não foi possível identificar os serviços recebidos por este grupo.
- A AIC flamenga comparou os resultados dos beneficiários de uma forma bastante genérica de tratamento (aconselhamento) com outras formas; também neste caso não foi possível identificar um grupo de controlo que não tivesse recebido qualquer intervenção.

Nos casos de avaliações centradas nos impactos de medidas financiadas pelo FSE sobre as empresas, a linha de demarcação foi traçada entre as empresas financiadas e as que não foram financiadas:

- A AIC planeada na DK comparará o desempenho de empresas financiadas através do FSE com os resultados de uma amostra de empresas com características semelhantes, mas que não receberam qualquer apoio.
- Está planeada uma abordagem similar na República Checa.

2.1.3. Os dados adequados estão disponíveis ou podem ser disponibilizados?

Discussões mantidas com autoridades gestoras e especialistas de avaliação de toda a UE sugerem que o acesso a dados adequados constitui um dos principais desafios enfrentados na implementação de AIC. Ao decidir que intervenções poderão ser avaliadas utilizando AIC, uma consideração prática fundamental consiste em saber se os tipos de dados necessários se encontram disponíveis. Nesta secção, é apresentada uma categorização simplificada dos tipos de dados que são necessários, juntamente com uma análise das fontes a partir das quais esses dados poderão ser obtidos, ou os tipos de exercícios de recolha de dados primários que poderão ser necessários. É ainda abordada a questão crucial da proteção de dados.

Antes de prosseguir esta discussão, importa realçar um aspeto importante relacionado com a adequação do planeamento. Até certo ponto, tentativas para implementar AIC no passado frustraram-se por falta de dados, pelo facto de não terem sido estabelecidos planos adequados com a devida antecedência. Para intervenções atuais, é importante identificar grupos de unidades tratadas e não tratadas sobre as quais a avaliação se centrará e implementar mecanismos para recolher dados desses grupos. Na fase inicial de desenvolvimento de novas intervenções, deverão ser tomadas providências para assegurar que os tipos certos de dados são recolhidos nos momentos adequados.

Planeamento da
recolha de dados

Que tipos de dados são necessários?

Em geral, são necessários três tipos de dados para efetuar uma AIC. Em algumas instâncias uma única fonte poderá conter um ou mais destes tipos de dados. Estes tipos de dados são registos dos grupos tratados e de controlo, registos de resultados e registos de dados contextuais. Descrevamos brevemente estes tipos de dados:

- **Registos dos grupos tratados e de controlo:** são necessárias bases de dados que permitam aos avaliadores identificar unidades individuais dos grupos tratados e de controlo (empresas, pessoas ou áreas geográficas potenciais).
- **Registos de resultados:** as Figuras 1 e 2 no primeiro capítulo deste guia mostram que as AIC requerem que tanto os resultados do grupo tratado como os do grupo de controlo sejam medidos. Idealmente, os resultados para ambos os grupos deverão ser construídos utilizando os mesmos instrumentos de pesquisa, e as medições dos resultados deverão ser efetuadas nas mesmas datas.
- **Dados de controlo:** são necessários dados que permitam selecionar um grupo de controlo perfeitamente correspondente ao grupo tratado, e que no âmbito de cada caso permitam controlar, para efeitos de análise, as diferenças que tenham subsistido entre os grupos tratados e de controlo. É importante recolher o maior número possível de dados referentes aos fatores e às características das unidades que possam estar relacionados tanto com a escolha de participar numa intervenção como com os resultados potenciais, principalmente indicadores dos resultados medidos antes do tratamento. As variáveis de controlo poderão incluir também as que descrevem mercados de trabalho locais (por exemplo, taxas de desemprego local ou medidas de rigidez do mercado de trabalho) e as que permitirão uma análise por subgrupos.

A Tabela 2 apresenta estes três tipos de dados e sugere fontes a partir das quais os dados poderão ser recolhidos. A Caixa 10 apresenta exemplos de dados utilizados para AIC.

Tabela 2. Tipos de dados e de fontes

Tipos de dados	Fontes
Registos do grupo tratado	• Registos da participação em intervenções (mantidos pelos beneficiários, p. ex.) incluindo dados de monitorização do FSE • Registos de recomendações • Registos de candidaturas
Registos do grupo de controlo	• Dados administrativos, designadamente registos de subsídios da segurança social e de desemprego (aqueles que não foram tratados depois de os registos do grupo tratado serem inspecionados) • Registos de candidaturas (candidatos rejeitados) • Registos de participação (aqueles que estavam aptos a participar, mas que não iniciaram o tratamento – tipicamente designados como “não-comparências”) • Inquéritos nacionais existentes, p. ex., IPA (inquérito à população ativa)
Registos dos resultados (necessários tanto para o grupo tratado como para o grupo de controlo)	• Dados administrativos: registos da segurança social e do desemprego também podem ser utilizados para a medição dos resultados (resultados de indivíduos que recebem subsídios / prestações da segurança social), registos de seguros e registos fiscais do país (resultados de rendimentos e emprego) • Registos administrativos de formadores (programas de formação iniciados e concluídos) • Recenseamento oficial de empresas ou registos fiscais, quando disponíveis • Registos de recenseamentos do emprego ou da produção (registos utilizados na elaboração das contas nacionais, para medidas do PIB) • Inquéritos desenhados sob medida aos grupos tratado e de controlo
Dados contextuais / variáveis de controlo (necessários tanto para o grupo tratado como para o grupo de controlo)	• Sistemas administrativos – registos de subsídios que forneçam histórias de requerimentos pré-tratamento, p. ex.; registos nacionais de seguros e registos fiscais, registos do histórico de rendimentos e de emprego • Inquéritos aos grupos de controlo e de tratamento. Quando as regras de tratamento são claras, os grupos de controlo podem ser identificados <i>ex ante</i> e recolhidos dados de referência • Ferramentas de monitorização da intervenção – em algumas circunstâncias os sistemas de monitorização podem ser utilizados para recolher medidas de referência tanto a partir do grupo tratado (ver Anexo XXIII do Regulamento de Implementação¹) como do grupo de controlo, por exemplo sistemas de candidatura em que os candidatos rejeitados podem ser utilizados como controlo.
	(1) Regulamento de Implementação n.º 1828/2006

Caixa 10. Exemplos de dados utilizados para AIC

De todas as AIC efetuadas para avaliar os efeitos líquidos das intervenções financiadas pelo FSE, as que foram realizadas na Áustria foram provavelmente as que tiveram acesso aos dados mais ricos. Os dados foram obtidos do serviço nacional de emprego do país, o qual captou pormenores de outros serviços (além das intervenções financiadas pelo FSE) recebidos tanto pelo grupo tratado como pelo grupo de controlo.

Outros dados foram obtidos a partir de fontes da segurança social. Esses dados forneciam elementos sobre a situação laboral e a carreira profissional, bem como sobre variáveis de rendimento.

Essas fontes de dados foram agregadas para criar um conjunto único de microdados. No entanto, o tempo e os recursos necessários para elaborar esses dados tiveram como consequência não ter sido feito qualquer esforço para repetir este exercício no atual período de programação.

Na Polónia a recolha de dados revelou-se extremamente difícil, uma vez que o registo oficial do desemprego é realizado por delegações regionais de emprego. O principal problema residia na dificuldade de acesso a dados pessoais e no facto de o sistema de monitorização central do mercado de trabalho nacional não incluir informação sobre fontes de financiamento, pelo que não pôde ser utilizado. Assim, foi necessário persuadir cada um dos SPE regionais – ou as delegações distritais de emprego (DDE) – escolhidos na amostra a fornecer dados pessoais anónimos. Mas, por razões técnicas, nem todos esses serviços o puderam fazer. Alguns deles tinham sistemas informáticos que eram incompatíveis com os sistemas PULS amplamente utilizados – e só a partir de 2011 as DDE estão habilitadas a transmitir os seus dados através de um sistema informático comum (SYRI-USZ). Dos 341 *poviats* (distritos), foi seleccionada uma amostra de 96, dos quais 69 forneceram dados. Os dados só puderam ser utilizados em 59 desses casos.

A AIC na República Checa utilizará dados dos registos das candidaturas das empresas – as instituições privadas foram os beneficiários finais (com um total de 1481 empresas apoiadas) –, e um projeto de sistema no qual as empresas constituem um grupo-alvo (candidatam-se a financiamento para formação aos empregados). Será utilizado um conjunto de dados complementar da Universidade de Economia de Praga e do Instituto de Estatística Checo para identificar grupos de controlo.

Na Lombardia, estava disponível e acessível, centralmente, uma base de dados de boa qualidade, abrangendo candidatos e beneficiários do FSE. No entanto, revelou-se difícil identificar um grupo de controlo a partir desses dados. Quanto aos resultados, foi realizado um inquérito específico sobre as condições do emprego.

Na Dinamarca, existem planos para as AIC se apoiarem numa base de dados criteriosamente construída. Exige-se que os beneficiários do FSE reportem semestralmente todas as empresas / locais de trabalho e todas as pessoas que considerem ter sido afetadas pelas atividades financiadas pelo FSE. É possível combinar estes dados com os dados cadastrais, de forma a identificar grupos de controlo.

Na Bélgica foram utilizados dados do serviço do mercado de trabalho, complementados por duas rondas de inquérito telefónico (4 e 21 meses após finalização da medida) para captar resultados tanto de medidas *hard* (p. ex., obtenção de um emprego) como de medidas *soft* ou intermédias (p. ex., conhecimento do mercado de trabalho e eficácia da procura de emprego, etc.).

No País de Gales, foi seleccionada uma amostra de pessoas que abandonaram programas do FSE a partir dos registos desses programas, e que foram entrevistadas. Os dados do inquérito aos entrevistados foram subsequentemente cruzados com dados IPA RU de forma a identificar um grupo de controlo.

A experiência dos Estados-Membros consiste em combinar dados a partir de fontes diferentes para realizar AIC, realçando assim a importância de pensar criativamente acerca das fontes de dados disponíveis.

Quais os possíveis problemas relacionadas com a proteção de dados?

Podem ser experimentadas dificuldades na obtenção de dados de identificação de pessoas ou de empresas que tenham participado em intervenções financiadas pelo FSE²². As AIC requerem microdados – ou seja, dados que contêm observações sobre as unidades individuais (quer se trate de pessoas individuais, empresas ou mesmo áreas geográficas) tanto no grupo tratado como no grupo de controlo. O Regulamento de Implementação n.º 1828/2006 (Anexo XXIII)²³ requer dados sobre os participantes com uma desagregação por género, situação profissional, grupos etários, nível de escolaridade, e grupos vulneráveis (migrantes, minorias, deficientes, outros desfavorecidos). Os Regulamentos de Disposições Comuns (RDC) e do FSE para 2014-2020 estabelecem mesmo a obrigação legal de as AG recolherem e processarem os dados pessoais sob a forma de registos individuais dos participantes.

O processamento destes dados deverá estar em linha com a Diretiva 95/46/EC²⁴. Esta diretiva abrange a transferência global de dados pessoais, incluindo dados sensíveis no interior da UE. Enquanto os dados referentes à situação profissional, à idade e ao nível de escolaridade são definidos como dados pessoais²⁵, podendo ser recolhidos sem o consentimento do indivíduo, os dados referentes à classificação dos indivíduos como fazendo parte de grupos vulneráveis são dados sensíveis²⁶ e a sua recolha só é permitida com o consentimento do indivíduo²⁷. Podem no entanto ser admitidas exceções, através de dispensas dos EM por razões de interesse público. No entanto, em vários Estados-Membros é muito difícil recolher dados sensíveis individualizados.

A prática habitual consiste em as AG recolherem microdados e armazenarem-nos (a nível de AG, OI ou beneficiário). São utilizadas diferentes técnicas para garantir o anonimato dos dados (p. ex., através de números identificadores únicos ou arbitrários). A maioria das AG exige o consentimento para a recolha de dados, podendo mesmo a atribuição de financiamento ser baseada no consentimento dos indivíduos para a recolha dos seus dados pessoais. Usualmente não é feita distinção entre dados pessoais e dados sensíveis – e não são admitidas por lei quaisquer exceções em relação aos dados sensíveis. Para efeitos de avaliação, os Estados-Membros permitem habitualmente que os avaliadores utilizem dados anónimos.

Dependendo do projeto de avaliação, poderá ser útil “desanonimizar” os dados (com consentimento) para voltar a contactar os participantes em inquéritos de seguimento. É também útil aplicar anonimização estatística, para permitir conectar os dados dos participantes relacionados com o FSE com dados administrativos nacionais.

Os avaliadores referem que as regras nacionais sobre proteção de dados criam sérios obstáculos à utilização de microdados. O acesso aos microdados do EUROSTAT é também moroso e difícil. Em relação a dados novos ou adicionais exige-se uma consulta formal e o acordo dos serviços nacionais de estatística. Algumas das dificuldades

²² Relatório de Síntese sobre uma Audição de Especialistas acerca da Legislação sobre Proteção de Dados e *Reporting* do FSE, Bruxelas, 10 de março de 2011.

²³ Regulamento da Comissão 1828/2006.

²⁴ Diretiva 95/46/EC do Parlamento Europeu e do Conselho de 24 de outubro de 1995 sobre a proteção das pessoas no que diz respeito ao processamento de dados pessoais e à livre circulação desses dados.

²⁵ *Idem*, art.º 7.

²⁶ *Idem*, art.º 8.

²⁷ *Idem*, *idem*.

que ocorreram na Lituânia, são explicados na Caixa 11. Consequentemente, ao planejar uma AIC, será necessário clarificar as seguintes questões:

- Existem microdados disponíveis? Estão também disponíveis se forem dados sensíveis?
- Existe uma única fonte de dados ou é necessário combinar fontes de dados (p. ex., estatísticas sobre desemprego, subsídios sociais, segurança social, dados sobre empresas / estabelecimentos, etc.)?
- É possível obter acesso a fontes nacionais de dados sobre carreiras individuais para comparar participantes em programas do FSE com um potencial grupo de controlo?
- De que forma é que os dados se tornam anónimos? É possível fazer o seguimento das pessoas ao longo do tempo e estabelecer a conexão entre fontes de dados?
- O grupo-alvo e o grupo de controlo foram definidos de forma a possibilitar o respetivo seguimento através de inquéritos baseados em entrevistas? Os elementos para contacto estão disponíveis e são exatos?

Caixa 11. Proteção e troca de dados – a experiência da Lituânia

Na Lituânia os dados pessoais anónimos sobre os desempregados da Bolsa de Trabalho Lituana (BTL) foram combinados com dados sobre o emprego do Fundo Nacional de Segurança Social sob a tutela do Ministério da Segurança Social e do Trabalho da República da Lituânia (Sodra). A principal dificuldade encontrada foi a lei muito restritiva sobre proteção de dados e a exigência de os dados serem fornecidos por organizações que não encomendassem a avaliação (não tendo qualquer incentivo para cooperar com os avaliadores, não estavam preocupadas com o acesso da avaliação aos dados).

Foram necessários quatro meses para negociar o acordo interinstitucional entre o Ministério da Segurança Social e do Trabalho, que encomendou a avaliação, e as duas instituições que forneceram os dados.

A experiência da LT sugere que as AG deverão elaborar planos para aceder aos dados com bastante antecedência em relação à encomenda de avaliações, e tomar medidas para assegurar que as barreiras jurídicas são atempadamente resolvidas.

2.2. Desenvolvimento de um esquema de avaliação

Uma vez analisadas algumas das questões que importa abordar ao decidir quais as intervenções que poderão ser objeto de uma AIC e se é possível efetuar-las com os registos de dados disponíveis, será agora necessário dirigir as atenções para algumas das questões-chave que importa ter em consideração no desenvolvimento de um esquema de avaliação. Será necessário formular um esquema de avaliação antes de encomendar uma AIC – ou um estudo mais vasto de avaliação – de forma a poder preparar termos de referência e a designar o fornecedor. A Caixa 12 apresenta o conteúdo recomendado de um esquema de avaliação.

Caixa 12. Conteúdo recomendado de um esquema de avaliação

O conteúdo exato de um esquema de avaliação dependerá do contexto em que esta irá ser efetuada; se a intervenção é obrigatória ou não obrigatória para o grupo-alvo; se é implementada universalmente numa jurisdição ou restringida a certas áreas; o tipo de intervenção a ser avaliada; e os quadros institucionais e as convenções aceites no seio do EM. As AG poderão considerar envolver especialistas externos para ajudar na formulação de um esquema de avaliação.

Um sistema de avaliação para uma AIC deverá abranger:

- As intenções e os objetivos da intervenção a avaliar;
- A finalidade da avaliação – as razões pelas quais está a ser encomendada e as questões que precisa de abordar;
- Os recursos disponíveis, tanto internos como externos, que serão necessários para efetuar a avaliação;
- O calendário da avaliação;
- A forma como o grupo tratado será identificado – que fontes de dados serão utilizadas para o efeito;
- Os fatores a considerar para a identificação de um grupo de controlo;
- Os tipos de dados que estão disponíveis;
- Os principais obstáculos à análise – especificamente a dimensão provável das amostras;
- As formas como os resultados serão relatados e utilizados.

Caixa 13. AIC incorporadas num quadro mais amplo de avaliação

A maior parte das AIC de intervenções do FSE efetuadas nos diferentes EM são incorporadas em quadros mais vastos de avaliação:

- Na Polónia a Unidade Nacional de Avaliação encomendou várias AIC. Começaram com a componente de coesão económica e social do Programa Phare para 2001 (Desenvolvimento de RH), continuando no período de programação do FSE de 2004 a 2006 e no atual PO de Capital Humano.
- Na Lombardia, uma abordagem contrafactual foi incorporada numa avaliação em curso, começando por um estudo de implementação em 2009.
- Na avaliação austríaca do período de programação do FSE de 2000 a 2006 a abordagem contrafactual foi apenas uma componente de um esforço de avaliação muito mais vasto.
- A AIC flamenga fez parte de um programa mais vasto de avaliação teórica, que articulou os limites dentro dos quais as medidas de política ativa do mercado de trabalho poderão melhorar a empregabilidade.
- Na Lituânia a AIC constituiu uma parte relativamente pequena de um plano de avaliação, que incidiu na relevância, na eficácia, na eficiência e na complementaridade, bem como no impacto das intervenções do FSE.

2.2.1. Quais as intenções e os objetivos da intervenção?

Ao elaborar um esquema de avaliação, é aconselhável começar por descrever as intenções e os objetivos da própria intervenção e, desde logo, as suas características principais.

Em muitos casos, já existem documentos que apresentam as intenções e os objetivos da intervenção. No entanto, este aspeto é importante no caso de uma AIC ser específica quanto aos resultados que uma intervenção procura alterar e, consequentemente, aos impactos esperados.

Revela-se frequentemente vantajoso articular o quadro lógico de uma intervenção que estabeleça os meios que permitam ligar os recursos e atividades às consequências e aos resultados, e, consequentemente, aos impactos alcançados (para uma análise mais aprofundada deste aspeto, ver secção 1.5 deste guia).

Conjugação da
conceção da AIC com
conhecimento da
lógica da intervenção

2.2.2. Qual o objetivo da avaliação?

Ao desenvolver um esquema de avaliação para uma AIC, importa refletir na finalidade da avaliação. Sem um entendimento claro das razões pelas quais é necessária, é pouco provável que a avaliação produza as provas necessárias. No contexto de avaliações de intervenções financiadas pelo FSE, será necessário considerar uma série de questões:

- Quais a finalidade e a natureza da avaliação no contexto das normas reguladoras e das diretrizes da CE?
- Quais as principais partes interessadas na avaliação?
- Que utilização irá ser feita dos resultados da avaliação?
- Que questões específicas terá a avaliação de abordar?

Qual a natureza da avaliação?

Em primeiro lugar, será necessário definir **a motivação** para a realização da avaliação. De acordo com o Regulamento 1083/2006, existem dois casos específicos em que os EM deverão realizar uma avaliação: se a monitorização revelar um desvio significativo face às metas inicialmente estabelecidas; e no caso de serem propostas de extensa reestruturação em termos de conteúdo, financiamento e implementação de PO. Além destes casos, que estão definidos no Regulamento, a CE incentiva os EM a efetuarem outras avaliações, cujos âmbito, conceção e horizonte temporal correspondam à procura interna dos EM.

O projeto de regulamentação RDC para o período de programação 2014-2020 coloca maior ênfase nas estimativas de eficácia, eficiência e impacto: «O impacto dos programas deverá ser avaliado de acordo com a missão dos respetivos fundos QCA em relação aos objetivos da estratégia da União para um crescimento inteligente, sustentável e inclusivo bem como em relação ao Produto Interno Bruto (PIB) e ao desemprego, quando pertinente.»²⁸

²⁸ Art.º 47 (1) do projeto de RDC.

Avaliações estratégicas e operacionais

- Em segundo lugar, será necessário estabelecer **a natureza da avaliação**²⁹.
- As avaliações de **natureza estratégica** examinam a evolução de um programa ou de um grupo de programas no que se refere às prioridades comunitárias e nacionais, especialmente as metas da estratégia de Lisboa (poderá ser o impacto macroeconómico dos Fundos Estruturais, enfoque em temas específicos, ou prioridades horizontais como a igualdade de oportunidades e exemplos de boas práticas).
- As avaliações de **natureza operacional** apoiam a monitorização de um programa operacional e avaliam a qualidade e a relevância dos objetivos quantificados, analisando os avanços financeiro e físico e apresentando recomendações sobre melhorias do programa.

Em princípio, a abordagem contrafactual pode ser aplicada tanto a avaliações estratégicas como a operacionais. As principais diferenças residem no público-alvo e na utilização que irá ser feita dos resultados da avaliação.

Para o período de programação 2014-2020, o projeto de RDC exige no mínimo uma avaliação que permita estimar de que forma o apoio de fundos europeus contribuiu para os objetivos de cada prioridade³⁰. Este tipo de questão constitui um caso em que a realização de AIC pode ser um método adequado para chegar a resultados conclusivos.

Quem são os principais destinatários?

Importa determinar qual o público-alvo de uma avaliação. Dependendo da natureza da avaliação, os destinatários poderão ser os responsáveis pela gestão de programas, outras autoridades gestoras ou organismos responsáveis pela implementação no Estado-Membro e autoridades nacionais ou regionais que gerem programas similares. Nos casos em que os dados são fornecidos por instituições exteriores à gestão do programa, esses organismos deverão também ser considerados partes interessadas. É importante incluir todos os principais intervenientes num grupo de acompanhamento da avaliação, para criar um espírito de pertença conjunta ao processo de conceção e de condução da avaliação.

Que utilização irá ser feita dos resultados da avaliação?

Uma vez identificados os destinatários da avaliação, poderá ser determinada a utilização que irá ser dada aos resultados da avaliação. Na prática, isso poderá ser feito envolvendo o grupo de acompanhamento no desenvolvimento da avaliação e através de discussões em torno dos termos de referência.

Duas importantes decisões para as quais os resultados das AIC contribuem frequentemente são:

- Se uma intervenção existente deverá continuar, e
- Se um novo tipo de intervenção deveria ser implementado de uma forma mais generalizada (ou seja, a uma escala mais alargada).

Identificar os destinatários da avaliação

²⁹ Comissão Europeia (2007): Diretrizes Indicativas sobre métodos de avaliação: avaliação durante o período de programação. Documento de Trabalho n.º 5. Direção-Geral de Política Regional.

³⁰ Art.º 49 (3) do projeto de RDC.

Na primeira instância, a AIC pode tentar calcular a eficácia de um programa existente ou em curso sempre que os orçamentos estejam sujeitos a pressões e existam potenciais utilizações alternativas para os recursos envolvidos. Nesta situação é provável que a intervenção não tenha sido avaliada antes da utilização de contrafactuais.

Em algumas circunstâncias, as intervenções poderão deparar-se com obstáculos à implementação. A título de exemplo, uma intervenção poderá ser implementada numa região ou numa área específicas de um EM, ou apenas durante um período limitado. Nestes contextos, os resultados de uma AIC podem ser utilizados para determinar se a intervenção em causa é eficaz e se, consequentemente, pode ser implementada de forma útil noutras regiões. Intervenções nestas situações são referidas como sendo projetos-piloto ou testadas antes de uma implementação mais alargada.

Quais as questões a que é necessário responder?

Uma vez estabelecidos os objetivos da intervenção, a finalidade da avaliação e a utilização final dos resultados, e claramente identificados os destinatários, deveria ser possível especificar com algum pormenor as questões que a AIC terá de abordar. Em muitas circunstâncias, existe um elevado número de destinatários e de partes interessadas que poderão ter questões de natureza causal que gostariam que a AIC explorasse. Será necessário um processo de priorização.

Entre os aspetos que poderão ser considerados ao finalizar uma lista de questões-chave de pesquisa para uma AIC estão os seguintes³¹:

Questões-chave para
investigação

- Que resultados e, consequentemente, que estimativas de impacto estão mais estreitamente associados ao sucesso global da intervenção? As questões referentes a estes problemas deverão ser priorizadas.
- Qual a possibilidade de medir um resultado quantitativamente? Poderá não ser possível medir alguns dos resultados pretendidos com as fontes de dados em princípio disponíveis. As questões de pesquisa deverão estar relacionadas com resultados que possam ser medidos.
- É provável que no interior do principal grupo-alvo existam subgrupos de interesse? Por exemplo, se uma intervenção é dirigida a desempregados de longa duração existirá interesse no impacto da intervenção junto da população com idade inferior a 25 anos ou superior a 50 anos? As questões de pesquisa deverão identificar os subgrupos que exigirão atenção específica.
- Que provas já existem da provável eficácia da política? No caso de existirem estudos de intervenções similares à que está a ser objeto de avaliação, as questões de pesquisa podem ser focadas mais de perto. Pelo contrário, se uma intervenção é a primeira do seu género, então será necessário um conjunto mais abrangente de questões de pesquisa.
- Se a intervenção estiver a ser implementada em diversos contextos regionais, existem fatores contextuais que possam ser importantes para influenciar os impactos? Existem outros fatores perturbadores que possam influenciar os resultados?

³¹ Estas questões são adaptadas de uma lista apresentada no *Magenta Book* do Tesouro Britânico (2011) página 44, um guia do governo britânico para avaliação de políticas.

- Os impactos da intervenção irão sofrer alterações ao longo do tempo? Quanto tempo será necessário para que os impactos comecem a surgir e se materializem? Os efeitos a curto prazo serão diferentes dos efeitos a longo prazo?

Priorizar questões

É importante ter uma ideia clara da série de questões para investigação que uma AIC terá de abordar antes de a encomendar. Um elemento-chave de um esquema de avaliação deverá ser a discussão das questões que a avaliação terá de abordar.

É importante priorizar questões e não ceder à tentação de sobrecarregar uma avaliação com demasiadas questões. É difícil encontrar o equilíbrio entre assegurar que a avaliação é relevante para um conjunto de intervenientes que têm interesses diferentes e garantir que a mesma possa ser monitorizada. No caso de se deparar com a necessidade de abordar um conjunto demasiado vasto de questões de pesquisa, uma avaliação pode perder o foco e acabar por abordar uma gama vasta de preocupações de forma não otimizada. É frequentemente um caso de “menos ser mais” – a priorização é uma fase crítica no processo de planeamento de uma avaliação.

2.2.3. Que recursos estão disponíveis?

Uma questão-chave a considerar na elaboração do esquema de uma AIC prende-se com os recursos disponíveis para a avaliação. Isto poderá conduzir a um conjunto alargado de considerações. A análise será estruturada em três vertentes: a) recursos especializados; b) tempo; e c) recursos financeiros.

Que especialistas externos e que recursos internos são necessários para uma AIC?

Pessoal interno

Na maior parte dos casos, uma avaliação de impacto será subcontratada a uma entidade externa. No entanto, a subcontratação terá de ser gerida dentro da AG por elementos com conhecimento dos métodos de AIC. Esse conhecimento é necessário para assegurar a qualidade e para estabelecer uma ligação efetiva com especialistas externos. Outras formas de conhecimento especializado poderão também ser necessárias dentro da AG, designadamente competências estatísticas e experiência na recolha e na gestão de dados. É importante considerar antecipadamente se a AG tem acesso a recursos devidamente qualificados e com a formação adequada, e se esses recursos têm capacidade para apoiar a avaliação.

Pessoal externo

A subcontratação de uma AIC eficaz exige que as entidades subcontratadas possuam as capacidades e a experiência necessárias para efetuar avaliações dessa natureza. Além disso, essas entidades terão de compreender a política e o contexto administrativo no seio do Estado-Membro, e terão de estar familiarizadas com as potenciais fontes de dados e ser proficientes nas linguagens adequadas. É importante considerar se serão necessárias providências para desenvolver uma base de fornecedores de AIC num Estado-Membro (para uma análise mais aprofundada deste aspeto, ver Capítulo 3).

Gestores de programas / intervenções

Para as AIC serem eficazes é necessário cooperação por parte de quem gere o programa ou a intervenção que está a ser objeto de avaliação. Por exemplo, será necessário o acesso a registos mantidos pelos gestores da intervenção. Esses registos fornecem informação acerca das pessoas ou das empresas que participaram numa intervenção.

Os gestores de programas / intervenções podem proporcionar aconselhamento e orientação sobre esses tipos de dados. Também lhes poderá ser exigido que efetuem alguma manutenção de registos, além da que teriam de fazer na ausência de uma avaliação de impacto.

Para ultrapassar a dificuldade na recolha de dados a partir de várias fontes, os responsáveis pelo planeamento de uma AIC terão de estar em ligação com os responsáveis que lidam com as fontes oficiais de dados (p. ex., registos do desemprego, dados da segurança, serviços de estatística, etc.) a fim de planear o fornecimento de dados com a devida antecedência.

Competência estatística

Quais os fatores relevantes para a calendarização uma AIC?

A realização de uma AIC exige contribuições de diversos recursos humanos; acresce que estas avaliações são efetuadas ao longo de horizontes temporais consideráveis. Um esquema de avaliação deverá conter um esboço de cronograma com as principais etapas do projeto, não só as associadas à própria intervenção mas também as associadas à avaliação. Esse esboço de cronograma deverá ser integrado ao longo das atividades de avaliação e de intervenção, e deverá também incluir metas para políticas-chave.

As AIC são morosas

A elaboração de um esboço de cronograma para uma AIC que seja significativo e realista constitui um difícil ato de balanço. Por um lado, as autoridades gestoras (ou organismos intermediários) responsáveis pelo planeamento da avaliação deverão considerar as datas fundamentais em que terão de ser tomadas decisões que dependem das conclusões da avaliação. Por outro lado, existirão restrições que dificilmente poderão ser evitadas e que colidem com a calendarização da apresentação de relatórios. Alguns resultados levarão anos a concretizar, e os calendários para a recolha de dados, para a análise e para a apresentação de relatórios deverão, na medida do possível, refletir esta realidade (ver secção 2.2.4). Sempre que exista a probabilidade de se verificarem atrasos consideráveis na obtenção de resultados finais, será importante prever relatórios intercalares que permitam disponibilizar resultados provisórios.

Os impactos levam tempo a materializar

É importante evitar a armadilha de avaliar demasiado cedo durante o período de programação. A avaliação deverá surgir atempadamente para que as mudanças possam ter lugar, e para que as experiências e os ensinamentos recolhidos possam ser capitalizados no período seguinte. Em algumas circunstâncias, a mesma intervenção ou intervenções semelhantes poderão ser apoiadas em períodos de programação sucessivos. Os resultados de AIC que incidiram em intervenções de períodos de programação anteriores poderão revelar-se extremamente úteis para informar a implementação e a conceção em períodos de programação subsequentes.

Focagem em momentos específicos

É também importante considerar em que medida o *timing* de uma avaliação contrafactual poderá afetar o *timing* de outras componentes da avaliação. É provável que uma AIC tenha de ser precedida de uma avaliação teórica. Em relação a intervenções inovadoras (p. ex., intervenções do FSE que tenham sido desenvolvidas para aumentar a flexibilidade do mercado de trabalho, como aconteceu com os contratos de transição ocupacional em França, ou com instrumentos que foram estabelecidos para combater a crise financeira, ou com intervenções como a medida “trabalhar com uma bolsa”, adotada pela Letónia), é também provável que os principais elementos de uma avaliação de processos tenham de ser relatados antes de efetuar uma AIC. Ao efetuar uma AIC de uma intervenção madura em curso, será provavelmente mais relevante que a avaliação de processos seja conduzida a par da avaliação do impacto.

Sequenciar vários tipos de avaliação

A recolha de dados
pode ser morosa

Um calendário também será afetado pela disponibilidade de dados. As fontes de dados poderão exigir períodos de tempo significativos para serem atualizadas; é frequentemente o caso dos registos fiscais, por exemplo. Ultrapassar barreiras jurídicas e institucionais na obtenção dos dados necessários pode também ser moroso e dispendioso. Acresce que, capitalizar dados provenientes de fontes distintas, assegurar a compatibilidade dos mesmos, verificar a sua qualidade e manipulá-los de forma a poderem ser utilizados para estimar impactos requer tempo e esforço consideráveis.

Como podem os custos ser estimados?

É importante preparar um orçamento indicativo para as despesas que a AG poderá e estará disposta a efetuar para a realização da AIC. O orçamento terá duas componentes: os custos da avaliação em termos de recursos internos e os custos de contratar especialistas externos para efetuarem a AIC. Esta segunda componente merece atenção prioritária.

Estimativa de custos

É necessário estabelecer uma distinção entre a avaliação de intervenções de rotina, onde as despesas são muito menores, e ações inovadoras ou ações-piloto. A escolha da abordagem de avaliação é também determinante. Um documento de orientação publicado pela Comissão³² estima que o montante necessário para intervenções de rotina represente cerca de 1% do orçamento do programa. No caso de iniciativas inovadoras ou piloto as despesas poderão ascender a 10% do orçamento do programa. No entanto, este guia não aborda explicitamente as necessidades de recursos. É provável que se um estudo de impacto requerer uma recolha significativa de novos dados primários, designadamente sob a forma de inquéritos quantitativos aos participantes e aos membros do grupo de controlo, os seus custos sejam consideráveis. Nos casos em que uma AIC se limita a explorar fontes de dados administrativas já existentes, os custos totais serão inferiores³³.

2.2.4. Quando deve ser avaliada a intervenção?

É fundamental decidir o momento em que, no decurso de uma intervenção, será mais oportuno efetuar uma avaliação de impacto, bem como as questões críticas de determinar o momento em que os resultados deverão ser medidos e os impactos estimados.

O *timing* é diferente
para intervenções
novas e intervenções
em curso

Quando avaliar intervenções novas e em curso?

A discussão sobre o momento mais oportuno para, no decurso de uma intervenção, efetuar uma avaliação de impacto contrafactual será diferente consoante se trate de uma intervenção nova ou de um processo maduro em curso. No caso de uma intervenção nova, é necessário mais tempo para que a intervenção se torne madura e atinja um estado estável. Será prematuro realizar uma AIC antes de este estado ser atingido, existindo o risco de produzir provas equívocas. No caso de intervenções novas, revela-se frequentemente útil realizar um processo inicial de avaliação, previamente à AIC, para identificar problemas de arranque e sugerir soluções.

³² Comissão Europeia (2009) EVALSED: O recurso para a avaliação do desenvolvimento socioeconómico.

³³ O Tesouro do Reino Unido (2011) fornece uma lista útil dos fatores a considerar na elaboração do orçamento para uma avaliação. Esta lista é apresentada no Anexo 2.

No caso de uma intervenção nova, vários fatores deverão ser tidos em conta para decidir o *timing* ideal para efetuar uma AIC. Entre estes fatores revela-se especialmente importante assegurar que estejam disponíveis as fontes de dados adequadas, constituir uma equipa interna de projeto que integre recursos com formação adequada, e assegurar que a entidade externa foi subcontratada. Outro aspeto crítico diz respeito às necessidades do processo de tomada de decisões, para o qual em última instância a avaliação está direcionada.

Em relação a uma intervenção em curso, o *timing* de uma avaliação de impacto dependerá essencialmente de requisitos práticos ou relacionados com questões de políticas. A intervenção deverá estar a ser implementada e já deverá ter atingido um nível de maturidade, tornando a AIC adequada. Uma questão adicional que importa considerar é a existência de outras reformas que decorram a par da intervenção que está a ser avaliada. Os efeitos destas reformas podem influenciar o impacto da intervenção em causa. Os decisores políticos deverão considerar se a existência de outras reformas no panorama político é relevante para as decisões políticas que capitalizarão os resultados da AIC que está a ser contemplada.

As avaliações do FSE incidem habitualmente num período de programação. No entanto, especialmente no caso de intervenções estáveis que já tenham feito parte do programa do FSE no período anterior, justifica-se considerar a combinação de uma avaliação retrospectiva do período anterior com a avaliação em curso no período atual, assim abrangendo um período de vida mais prolongado de uma intervenção.

Múltiplos períodos de programação

Quando medir resultados e calcular impactos?

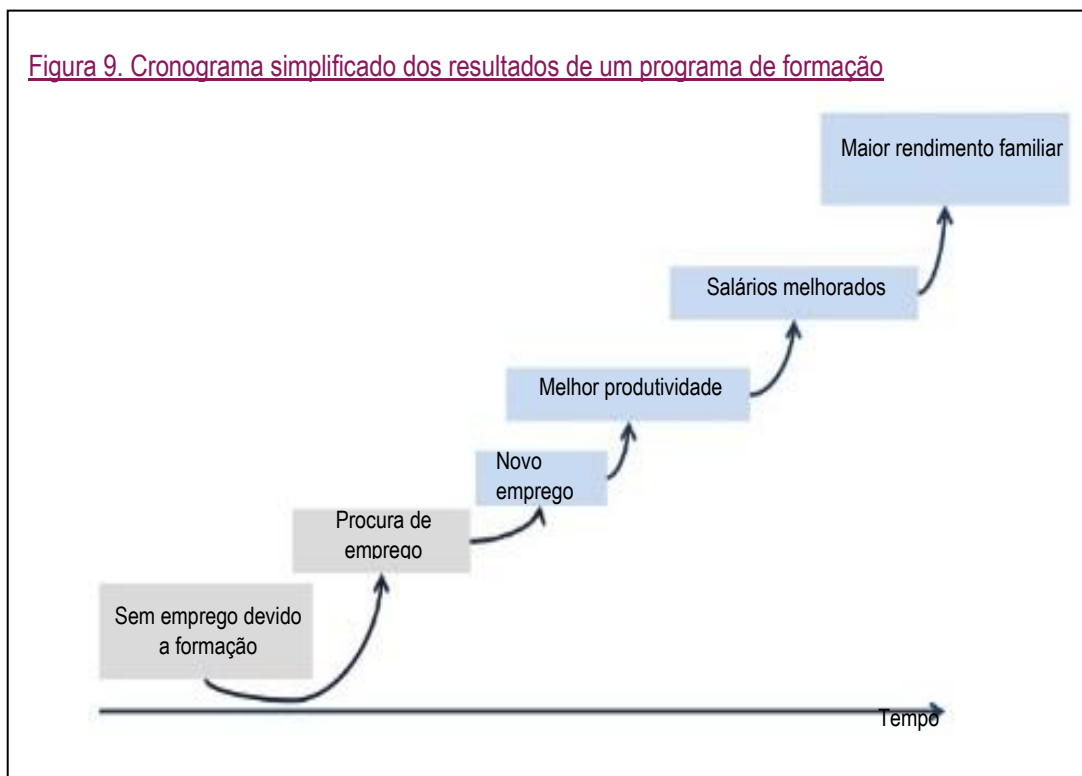
Uma segunda questão importante associada com o *timing* de uma avaliação consiste em decidir o momento em que os impactos deverão ser medidos e estimados ou, mais especificamente, quando será previsível que os impactos subsequentes a uma intervenção comecem a surgir.

Em relação a uma intervenção de formação dirigida aos desempregados, por exemplo, os decisores políticos poderão conjecturar que a intervenção irá aumentar a produtividade dos formandos, as suas oportunidades de emprego, bem como os ordenados que recebem. Mas a questão consiste em saber em que período de tempo se virá a concretizar um acréscimo das taxas de emprego e dos salários. É uma constante dos programas de formação o facto de a curto prazo tenderem a reduzir o emprego entre os participantes. Isto deve-se ao designado efeito “de aprisionamento”. As intervenções de formação tendem a afastar os formandos desempregados da procura de emprego, pelo facto de estarem a frequentar os cursos de formação. Por esta razão, se forem calculados demasiado cedo, os impactos poderão ser percecionados como negativos. Alternativamente, uma intervenção que inclua assistência no local de trabalho para os desempregados que regressem ao trabalho poderá aspirar a encorajar emprego sustentável e um avanço a longo prazo, sob a forma de aumento dos níveis salariais e de melhoria dos termos e condições. Naturalmente, estes tipos de resultados levariam um tempo significativo a surgir na sequência do tratamento. A questão em ambos os exemplos consiste em saber qual o momento mais adequado para medir os resultados e, conseqüentemente, calcular os impactos. E, após a exposição ao tratamento, quanto tempo será necessário para os efeitos positivos começarem a surgir. No planeamento de uma AIC importa ser realista quanto ao *timing* dos impactos e ao momento em que poderão ser medidos. Um modelo simplificado dos impactos subsequentes é apresentado na Figura 9.

Entre o razoável e o
exequível

A ponderação do momento mais adequado para medir resultados e estimar impactos deverá ter em conta a necessidade de os decisores políticos disporem de informação dentro de determinados prazos. No caso de intervenções destinadas a melhorar os níveis de emprego a longo prazo, poderá fazer sentido, numa perspetiva analítica, efetuar o seguimento dos participantes cinco anos após terem sido expostos ao tratamento, para ver se os seus salários e as suas taxas de emprego são superiores aos de algum grupo equivalente de pessoas não tratadas. Por contraste, os gestores de programas precisam frequentemente de conclusões rapidamente. Será assim necessário encontrar um compromisso entre o que é razoável para um intervalo de seguimento na perspetiva da intervenção e a necessidade de os decisores políticos disporem de provas em tempo útil.

Figura 9. Cronograma simplificado dos resultados de um programa de formação



Focagem em
momentos específicos

No caso de as medidas serem obtidas a partir de fontes administrativas (p. ex., registos da segurança social a partir dos quais as medidas do emprego e dos rendimentos poderão ser obtidas), será prático analisar os resultados repetidamente durante um período de tempo prolongado (podendo mesmo ser com uma frequência mensal). O risco aqui é o de a natureza das conclusões poder variar ao longo do tempo. No caso de ser necessária uma recolha de dados primários para a medição de resultados sob a forma de inquéritos por amostragem, a estimativa de impactos com uma frequência periódica revelar-se-ia demasiado dispendiosa, a não ser que fosse viável recolher dados retrospectivos sobre os resultados. No entanto, importa não subestimar o custo de extrair dados de múltiplos sistemas administrativos e de a partir desses extratos criar um único conjunto de dados analíticos.

Como analisado na secção 1.5, a articulação de um quadro lógico (ou lógica da intervenção) poderá ajudar a determinar o *timing* da estimativa dos impactos.

Uma alternativa para os responsáveis pelo planeamento de uma AIC na ausência de um quadro lógico (mas que poderia também ser útil mesmo para aqueles que podem basear-se num quadro lógico claro) consiste em efetuar uma breve análise de estudos anteriores que tenham avaliado intervenções semelhantes à que está a ser considerada. Uma atenção especial aos resultados de estudos anteriores poderá proporcionar indicações úteis acerca da medição adequada dos resultados e do cálculo dos impactos.

...ou análise de
outros estudos
recentes

2.2.5. Como deve ser identificado o grupo “tratado”?

Para efetuar uma AIC, é fundamental existir uma definição clara do que significa ser tratado ou ter participado na intervenção. Além disso, uma vez determinado um entendimento claro de quando se pode afirmar que uma pessoa ou uma empresa foram tratadas, é essencial que as entidades que tenham sido tratadas possam ser identificadas.

Em princípio, a definição de participação pode parecer simples. No entanto, existem algumas questões que poderão não ser imediatamente evidentes mas que são fundamentais e que exigem uma ponderação atenta. Por exemplo, os formandos num programa de formação que abandonaram a intervenção deverão ser considerados como tendo sido tratados? Quantas sessões de um curso de formação precisam os formandos de frequentar para serem considerados participantes? Existem também efeitos antecipados que será necessário considerar. Antecipando a possibilidade de serem sujeitos a uma intervenção, alguns requerentes de subsídios da segurança social poderão abandonar as listas da assistência social para evitarem passar a trabalhadores ativos. Podem essas pessoas ser consideradas tratadas, mesmo que, por exemplo, nunca tenham comparecido fisicamente a entrevistas agendadas para elas numa delegação dos serviços de emprego?

Existe também uma distinção entre “intenção de tratar” e “tratamento para os tratados” que deverá ser tomada em consideração na definição do “grupo tratado”. Na perspetiva da intervenção, a principal questão a abordar consiste habitualmente em saber se o interesse reside nos efeitos de ter a oportunidade de participar numa intervenção, ou nos efeitos de participar realmente. No primeiro caso, aqueles a quem foi oferecida uma intervenção podem participar – ou não. No último caso, em que o interesse reside no efeito do tratamento sobre a população tratada, o grupo tratado contém apenas as pessoas que participam³⁴.

À primeira vista, os decisores políticos assumem estar interessados em determinar os efeitos líquidos do tratamento sobre as pessoas que participam. No entanto, submetidas a uma reflexão mais aprofundada, as questões podem ser menos claras. Se as pessoas a quem é oferecido tratamento puderem ser identificadas, poderá ser útil, na perspetiva da intervenção, defini-las como o grupo “tratado”. É nomeadamente o que acontece nos casos em que a participação numa intervenção não é obrigatória. Os decisores políticos nunca poderão forçar as pessoas a quem é oferecida uma intervenção a participarem, pelo que, em certa medida, a questão relevante consiste em saber qual o impacto de ter participado num programa de formação, em termos de emprego e de salário subsequentes, para as pessoas a quem foi oferecida a oportunidade de participar.

Para estimar os efeitos da oferta de tratamento sobre uma série de resultados, as pessoas que receberam a oferta deverão ser identificáveis. Em muitas circunstâncias isso poderá ser difícil de conseguir.

Definição do grupo tratado

Intenção de tratar ou tratamento para os tratados

Oferta de tratamento ou tratamento real

³⁴ Nos casos em que a participação numa intervenção é obrigatória, a diferença entre estes dois estatutos é praticamente inexistente: todas as pessoas a quem é oferecido tratamento têm de participar. No entanto, na maior parte dos casos a participação nas intervenções não é obrigatória (e é este o pressuposto assumido ao longo deste guia).

Identificar fontes de dados para as pessoas tratadas

Onde encontrar dados adequados?

Uma vez decididas as definições de quem é tratado e daquilo em que o tratamento consiste, importa considerar a forma como as pessoas que foram tratadas serão identificadas para efeitos de avaliação. Invariavelmente isto implica identificar uma fonte de dados a partir dos quais as unidades tratadas, quer se trate de pessoas ou de empresas, possam ser plenamente especificadas ou incluídas numa amostra. Estes registos são normalmente os que são extraídos dos sistemas de monitorização do FSE e – caso estejam disponíveis – novos registos de dados criados para a intervenção específica.

Por exigência dos sistemas de monitorização e de documentação do FSE, as organizações beneficiárias devem registar os números e algumas características pessoais dos indivíduos que receberam serviços através de uma intervenção. Para efeitos de AIC, as intervenções deverão ainda disponibilizar microdados sobre as pessoas que nelas participaram. Em muitos casos os avaliadores exigirão não apenas um registo para cada unidade tratada (empresa ou pessoa) mas também a identidade dessas unidades (nomes, endereços, números de telefone, etc.) para poderem ser incluídas em amostras para inquéritos. São ainda exigidos identificadores exclusivos de cada unidade, para facilitar o cruzamento dos registos entre as fontes de dados.

2.2.6. Fatores a considerar na identificação de um grupo de controlo

Para obter uma estimativa dos dados contrafactuais, é normalmente necessário identificar um grupo de controlo. A um nível global, a escolha de um grupo de controlo será condicionada pela natureza da intervenção – obrigatória ou não obrigatória para os participantes –, bem como pelo alcance da mesma – implementada universalmente numa jurisdição, ou restrita a uma área específica ou a um horizonte temporal limitado. A escolha de um grupo de controlo adequado tem três aspetos: 1) analítico; 2) relacionado com as políticas; e 3) prático.

Definição de um grupo de controlo numa perspetiva analítica

A finalidade de uma AIC é a de obter estimativas sem enviesamento dos impactos de uma intervenção numa série de resultados. Para atingir este objetivo são necessárias estimativas de resultados contrafactuais. As estimativas dos resultados contrafactuais são obtidas a partir de um grupo de controlo (ver secção 1.1). Como ilustrado na Figura 1 e na Figura 2, um impacto é estimado deduzindo uma estimativa do resultado contrafactual ao resultado observado para o grupo “tratado”. O grau de distorção de um impacto depende da medida em que o resultado contrafactual calculado a partir do grupo de controlo represente o resultado que teria sido alcançado para o grupo tratado se ele não tivesse sido tratado, permanecendo todos os restantes aspetos inalterados.

É necessário um grupo de controlo (na ausência de distribuição aleatória) que em média seja equivalente ao grupo tratado em todos os aspetos importantes, tanto em dimensões observáveis como não observáveis.

Uma vez que quase todas as intervenções do FSE são a) voluntárias (o grupo-alvo não é obrigado a participar numa intervenção), e/ou b) limitadas de alguma outra forma – constituem intervenções-piloto ou instrumentos restritos a uma região ou jurisdição específicas –, os avaliadores serão confrontados com um conjunto de unidades que poderão ser selecionadas para serem utilizadas como controlo.

Encontrar um grupo de controlo equivalente

Opções para a escolha de grupos de controlo

Será necessário algum processo de triagem neste conjunto potencial para aperfeiçoar a escolha final de “controlados” que correspondam perfeitamente aos participantes (o grupo tratado). Em muitas circunstâncias, para a escolha de um grupo de controlo existem quatro opções³⁵.

- **Localização** – grupos de controlo que são semelhantes aos grupos que participaram numa intervenção mas localizados em áreas do Estado-Membro onde a intervenção não está disponível (se essas áreas existirem). O método de diferença-nas-diferenças é frequentemente a abordagem preferida nos casos em que esses grupos de controlo e os dados adequados se encontram disponíveis. Populações de locais diferentes podem ser muito semelhantes entre si e esses grupos poderão não ter tido a oportunidade de participar na intervenção (não se tendo recusado a fazê-lo), pelo que esta fonte importante de potencial enviesamento não existirá. No entanto, populações em locais diferentes estarão sujeitas a diferentes condições do mercado de trabalho. O método de diferença-nas-diferenças controla bastante bem essa variação, dado que as diferenças nas condições locais do mercado de trabalho tendem a ser relativamente fixas ao longo do tempo. No entanto, é pouco aconselhável extrair amostras de controlo de mercados de trabalho locais diferentes se o emparelhamento for utilizado para estimar impactos. Tem-se provado que o enviesamento associado à seleção de amostras de controlo de mercados de trabalho diferentes pode ser superior ao enviesamento de seleção³⁶;

- **Tempo** – elementos dos grupos de controlo semelhantes aos participantes mas que são observados em diferentes momentos, seja antes ou depois da intervenção. Grupos de controlo selecionados desta forma são frequentemente necessários nos casos em que uma intervenção é universal e obrigatória – por outras palavras, sempre que todos os membros do grupo-alvo são obrigados a fazer parte do programa e este é implementado numa jurisdição inteira. Grupos de controlo constituídos desta forma apresentam uma desvantagem significativa, designadamente o facto de os seus resultados serem medidos em momentos diferentes dos do grupo tratado, podendo assim estar sujeitos a flutuações cíclicas, a alterações na sua composição e a alterações das tendências macroeconómicas que poderão confundir a capacidade para identificar um resultado contrafactual sem enviesamento. Grupos de controlo com estas características apenas deveriam ser considerados quando existir uma reduzida variação nos resultados ao longo do tempo e quando um grupo de controlo contemporâneo não estiver disponível;

- **Elegibilidade** – os elementos dos grupos de controlo são selecionados a partir de grupos no mesmo local e no mesmo momento, mas que não são elegíveis para participar. Estes grupos de controlo são frequentemente procurados nos casos em que a intervenção é universal, em que as taxas de participação são elevadas, ou em que a participação é obrigatória e sempre que existem regras claras de elegibilidade, de tal modo que, por exemplo, as pessoas “simplesmente não elegíveis” constituem uma fonte potencial de grupos de controlo. O objetivo é o de identificar grupos que sejam semelhantes aos que foram tratados mas que, por razões estabelecidas e que são do conhecimento geral (e que poderão ser quantificadas nos dados), não eram elegíveis para tratamento. O acesso a intervenções no âmbito de financiamento do FSE raramente se baseia em regras claras de elegibilidade que possam ser facilmente medidas e que não possam estar sujeitas a manipulação; consequentemente, a seleção de grupos de controlo nestas circunstâncias pode ser bastante rara;

³⁵ Esta secção baseia-se em Card *et al.*, (2011).

³⁶ Heckman *et al.*, (1998).

Análise de tendências pré-intervenção

- **Escolha / conhecimento** – os elementos dos grupos de controlo (controlados) podem ser selecionados entre pessoas que eram elegíveis mas que falharam a participação. Essencialmente, tanto o grupo tratado como o grupo de controlo (e não apenas o grupo tratado) estão sujeitos a processos de seleção baseados numa escolha motivada por fatores potencialmente não observados³⁷. A vantagem de selecionar “controlados” entre pessoas elegíveis mas que não participaram reside no facto de serem habitualmente extraídos do mesmo mercado de trabalho dos que foram tratados. Estes “controlados” devem, contudo, ser considerados com prudência, sempre que esteja a ser utilizado uma abordagem de emparelhamento para a AIC e existam dados abundantes que possam ser aproveitados para determinar a decisão de seleção. Noutras circunstâncias, por exemplo quando o método de diferença-nas-diferenças seja implementado, os grupos de controlo baseados na escolha / conhecimento serão menos atrativos.

Um aspeto adicional é digno de nota. Sempre que estiverem disponíveis medidas de resultados pré-tratamento tanto para o grupo tratado como para o grupo de controlo, é importante examinar as tendências pré-intervenção nas medidas de resultados tanto para o grupo tratado como para os potenciais grupos de controlo. Verificar a presunção das designadas “tendências comuns” resolve o problema de quedas pontuais transitórias pré-intervenção nas taxas de emprego e nos salários que terão ocorrido para algumas das pessoas elegíveis para políticas ativas do mercado de trabalho (de outro modo não seriam elegíveis para apoio – o designado *Ashenfelter's dip* mencionado na nota 4). O avaliador procura tendências temporais semelhantes nas medidas de resultados tanto para o grupo tratado como para os grupos de controlo, para que a recuperação da perda de emprego de curta duração ou de salário não sejam confundidas com os ganhos de longa duração que a AIC procura detetar.

A seleção adequada de grupos de controlo constitui um exercício complexo do ponto de vista técnico e metodológico. No momento em que os esquemas de avaliação estão a ser preparados, é recomendável que os membros da AG se familiarizem com os principais conceitos e desenvolvam atempadamente diligências para identificar potenciais grupos de controlo. É importante, no entanto, que os fornecedores de uma avaliação envolvam no processo de conceção desde o início especialistas que proporcionem apoio e aconselhamento nesta tarefa.

Definição de uma alternativa a uma intervenção

Comparação do tratamento com o não tratamento ou com uma alternativa

Quais são as considerações políticas relevantes?

A seleção de um grupo de controlo adequado não é um processo meramente técnico ou analítico. Embora os aspetos analíticos da identificação de controlos adequados sejam fundamentais, é também importante que um grupo de controlo represente uma alternativa relevante à intervenção que está a ser considerada na perspetiva da definição de políticas.

As AIC podem assumir várias formas: por exemplo, podem comparar os resultados de um grupo tratado ou de vários grupos tratados com os de um grupo de controlo que não recebeu qualquer tratamento; ou podem comparar um tratamento com um outro sem que haja um grupo de controlo que não tenha recebido tratamento. A escolha de um grupo de controlo dependerá do tipo de comparação mais relevante para a política seguida, e de ser ou não possível encontrar um grupo de controlo “sem tratamento”.

³⁷ Trata-se daquilo que Card *et al.* (2011) designam como duplo enviesamento de seleção.

A Caixa 14 apresenta um exemplo de uma comparação de um tratamento com outro sem um grupo de controlo “sem tratamento” – consistindo o objetivo em estimar se deverá ser prosseguida uma intervenção ou se deverá ser feita outra. Importa notar que a comparação de um programa com outro, sem o benefício de um grupo de controlo “sem tratamento”, poderá dar lugar a ambiguidade (este aspeto é analisado na Caixa 15).

De notar que o método de diferença-nas-diferenças não pode ser utilizado para comparar tratamentos múltiplos, na ausência de um grupo de controlo não tratado.

Caixa 14. Questões políticas relacionadas com um programa de formação

Consideremos um exemplo em que o decisor político pretende introduzir uma nova intervenção de formação financiada através do FSE – designemo-la Intervenção A. Suponhamos ainda que o EM já tem um programa de formação destinado às mesmas pessoas, mas financiado por fundos nacionais. Neste caso uma questão política poderia ser: os níveis de emprego e os salários subsequentes à participação são, para os participantes na intervenção A, superiores aos dos participantes na Intervenção B? E, consequentemente, a Intervenção A representa uma melhor utilização do dinheiro? Se os salários são superiores para os participantes na Intervenção A, então a resposta política óbvia seria suspender a Intervenção B a favor da Intervenção A, se esta também se revelar eficiente em custos.

Caixa 15. Interpretar efeitos líquidos

Um estudo poderá concluir que não existe diferença nos salários entre os participantes na Intervenção A e os participantes na Intervenção B. A resposta política a esta informação poderá não ser clara se, por exemplo, a Intervenção B for extremamente eficaz comparativamente a não receber qualquer tratamento. Isto poderia significar que ambas as intervenções são extremamente eficazes. No entanto, em alguns casos poder-se-ia pensar que não existem provas da eficácia da Intervenção B comparativamente a não receber qualquer tratamento. Alternativamente, as intervenções A e B poderiam ser ambas ineficazes, embora uma intervenção possa parecer relativamente mais eficaz do que a outra. Nas circunstâncias em que determinados grupos da população poderiam constituir o alvo de mais do que uma intervenção, poderá ser ainda mais informativo tentar identificar um grupo adequado de unidades não tratadas para constituírem um termo de comparação.

Que considerações práticas são necessárias para selecionar o grupo de controlo?

Juntamente com considerações de ordem analítica e política, é necessário tomar em consideração os aspetos práticos da seleção de grupos de controlo. Selecionar unidades (pessoas ou empresas) ou agrupá-las numa amostra para constituírem um grupo de controlo implica que possa ser encontrado um quadro de amostragem adequado. Adicionalmente, os quadros de amostragem deverão conter unidades individuais que preencham os requisitos analíticos e políticos. A forma mais adequada de o fazer poderá variar de avaliação para avaliação, dependendo do contexto específico da intervenção que está a ser testada.

Em muitos casos são exploradas duas fontes de dados para identificar grupos de controlo adequados. Ambas exigem que seja conhecida a identidade do grupo tratado.

Registos da população e registos fiscais das empresas

Em primeiro lugar, registos de população de vários tipos poderão ser utilizados para identificar grupos de controlo. Por exemplo, uma intervenção no mercado de trabalho ativo é dirigida a pessoas com 18-24 anos de idade que recebem subsídio de desemprego, pelo que os registos de subsídio de desemprego podem ser utilizados para identificar a população-alvo. Acresce que se o grupo tratado for conhecido e puder ser cruzado com os dados do subsídio de desemprego, as pessoas daquele grupo etário que não receberam tratamento poderão ser identificadas e constituir um potencial grupo de controlo. Alternativamente, suponhamos que uma intervenção é dirigida a pequenas e médias empresas. Os registos nacionais de empresas (no caso de estarem disponíveis) poderão ser utilizados para definir a população-alvo, e, com informação disponível sobre as empresas que estão a ser tratadas, será possível identificar potenciais grupos de controlo.

Registos dos candidatos

Em segundo lugar, os registos dos candidatos podem ser utilizados nos casos em que a implementação da intervenção não é universal; por exemplo, quando nem todas as pessoas que se candidatam a um programa de formação são aceites (um grupo de controlo do tipo escolha / conhecimento). De modo análogo, nem todas as empresas que se candidatam a financiamento verão as suas candidaturas aceites, e aquelas que foram rejeitadas para formação ou para financiamento poderão nalguns casos ser utilizadas como grupo de controlo (ver a discussão apresentada anteriormente nesta secção relativamente à prudência que deverá ser mantida ao seleccionar grupos de controlo nestas circunstâncias).

2.2.7. Que tipos de problemas quanto aos dados precisam de ser equacionados no esquema de avaliação?

Gestão das fontes de dados

Que tipos de dados são necessários e como serão recolhidos?

Como anteriormente observado, as AIC requerem habitualmente o acesso a volumes consideráveis de microdados (em alguns casos poderão ser utilizados dados agrupados – por exemplo, dados regionais). É preciso recolher, comparar e documentar estes dados; dados provenientes de fontes diferentes terão de ser interligados na base de campos de identificação partilhados; precisam de ser armazenados e transferidos com segurança entre os responsáveis pela gestão e pela execução da avaliação; e será necessário criar conjuntos de dados analíticos a partir dessas fontes de dados, para facilitar a estimativa dos impactos.

No desenvolvimento de um esquema de avaliação é importante considerar as seguintes questões relacionadas com os dados:

- Que fontes podem ser utilizadas para obter esses vários tipos de microdados / dados agregados?
- De que forma será feito o acesso às fontes de dados e como poderão os dados ser recolhidos?
- As fontes são consistentes entre si?
- É necessário identificar indivíduos ou empresas? Qual a unidade de análise adequada ou possível?

- É possível identificar os indivíduos ou as empresas de forma consistente entre fontes?
- Os dados podem ser interligados?
- Quem assumirá a análise das fontes potenciais? Quem será responsável por negociar o acesso e obter o acordo necessário à sua utilização?
- Que barreiras jurídicas precisam de ser negociadas?
- Onde irão ser armazenados os dados?
- Que medidas serão tomadas para assegurar que os dados são armazenados com segurança e que o acesso aos dados é reservado a quem necessitar dos mesmos para efeitos de avaliação?
- Como é que os dados irão ser transferidos em segurança?
- Que sistemas informáticos e de infraestrutura serão necessários?

Como serão processados os dados?

Em muitos casos as AIC necessitarão de microdados – *i. e.*, dados que contêm observações sobre unidades individuais (normalmente pessoas individuais ou empresas) tanto nos grupos tratados como nos grupos de controlo (ocasionalmente poderão ser utilizados dados agrupados, p. ex., dados regionais ou a nível dos Serviços Públicos de Emprego). Distinguimos três tipos principais de dados necessários: a) registos dos grupos tratados e de controlo; b) registos de resultados; e c) os designados dados de contexto (dados utilizados para o controlo de potenciais diferenças importantes entre os grupos tratado e de controlo). Estes dados podem ser provenientes de fontes distintas ou da mesma fonte de dados. As fontes necessitam de ser estruturadas para constituírem conjuntos de dados analíticos (ou amostras analíticas) que são utilizados para estimar impactos. Em muitos casos esta estruturação implicará a necessidade de cruzar os registos de pessoas individuais ou de empresas entre fontes. Esse cruzamento pressupõe ou identificadores a nível individual (por exemplo, números de identificação da segurança social), que permitam alinhar o registo de uma pessoa por exemplo no que se refere a dados de natureza fiscal com registos de participação, ou dados suficientes para cruzar registos entre fontes (por exemplo, o nome e a data de nascimento devem estar disponíveis em todas as fontes). É importante considerar quais as fontes de dados que irão ser exploradas para uma AIC que está a ser planeada e também se será possível cruzar os registos entre fontes.

Cruzamento de
microdados entre
fontes

2.2.8. Quais os principais obstáculos à análise de dados e de resultados?

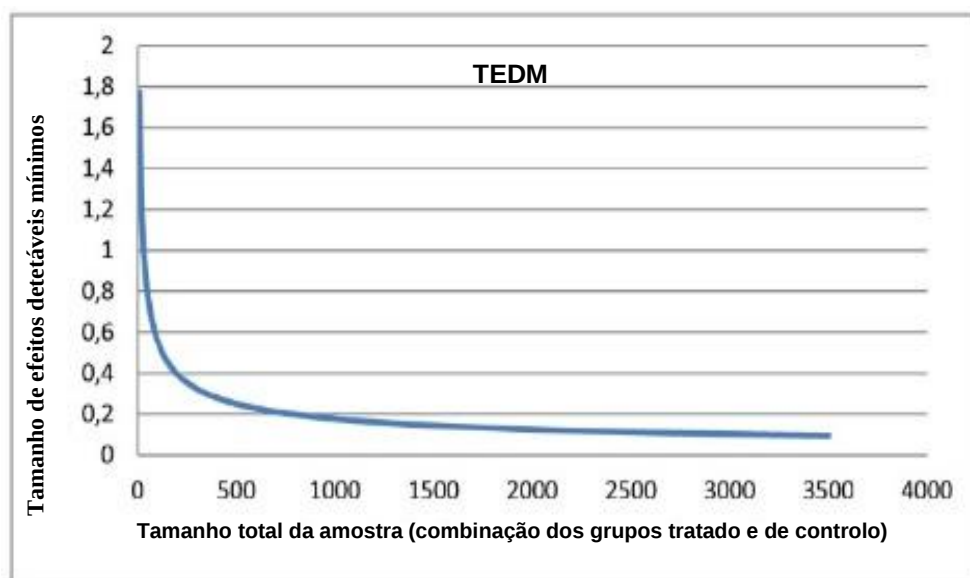
Como anteriormente analisado, os impactos numa AIC são geralmente determinados através da comparação dos resultados no grupo tratado com os do grupo de controlo. A diferença entre os dois é designada como impacto ou efeito líquido da intervenção. A forma exata como os impactos são estimados dependerá do método de pesquisa adotado, mas, no essencial, as abordagens de AIC implicam estabelecer esta comparação fundamental entre os resultados do tratamento e os resultados do controlo.

Estimativa da dimensão da amostra

Ao planejar uma AIC é importante considerar se a intervenção tem a dimensão suficiente e o potencial para gerar impactos que possam ser detetados estatisticamente.

Um conceito que ajuda a analisar se estará disponível uma amostra de tamanho suficiente para análise é o de «efeito detetável mínimo»³⁸. Saber se os tamanhos das amostras poderão ser suficientes para detetar os impactos de uma intervenção é frequentemente referido como uma questão de poder estatístico. Em termos simples, um efeito detetável mínimo é o menor impacto real que uma amostra de um determinado tamanho permite detetar com os níveis habituais de confiança estatística. Ao planejar uma AIC, revela-se frequentemente útil tentar estimar o tamanho provável das amostras analíticas com base nas previsões do número de unidades que serão tratadas, na abordagem da AIC e na dimensão dos correspondentes grupos de controlo (tomando em consideração qualquer amostragem que possa ser efetuada). Esta informação poderá então, sob certos pressupostos, ser utilizada para derivar *ex ante* efeitos mínimos detetáveis para uma abordagem de AIC.

Figura 10. Tamanho dos efeitos detetáveis mínimos em amostras de diferentes tamanhos



Nota: 95 por cento significado estatístico e 80 por cento poder estatístico

Uma vez obtidas as estimativas do efeito detetável mínimo, será possível avaliá-las. A questão crítica consiste em determinar se a intervenção em causa poderá produzir efeitos com uma dimensão equivalente à dos efeitos detetáveis mínimos estimados.

A Figura 10 ilustra a forma como o tamanho do efeito detetável mínimo (uma medida normalizada do efeito detetável mínimo que é comparável entre diferentes unidades de medida) varia com o tamanho total da amostra (números totais da amostra nos grupos tratado e de controlo). Da esquerda para a direita, o tamanho do efeito detetável mínimo decresce rapidamente à medida que o tamanho da amostra se aproxima de 500 (250 unidades de tratamento e 250 controlos). Por outras palavras, à medida que o tamanho total da amostra aumenta, a abordagem de AIC é capaz de detetar impactos estatisticamente menores.

³⁸ Bloom (1995) apresenta orientações práticas sobre a forma de calcular efeitos detetáveis mínimos para abordagens experimentais. No caso de abordagens quase experimentais, esses cálculos deverão ser ajustados. Geralmente, as abordagens quase experimentais exigem tamanhos de amostra maiores comparativamente com os que são necessários para uma abordagem experimental.

Os dados apresentados na Figura 10 pressupõem uma abordagem aleatória e são apresentados apenas para ilustrar este aspeto-chave.

Parte do processo de planeamento de uma AIC deverá incluir a previsão do número de pessoas ou de empresas que poderão ser visadas pela intervenção em causa, dos números que poderão ser sujeitos a amostra para a avaliação e da dimensão do correspondente grupo de controlo. Conjugados com informação sobre a forma como se pretendem estimar os impactos, os efeitos detetáveis mínimos podem ser calculados e poderão ser tomadas decisões sobre se estes são suficientes dada a magnitude dos impactos que poderão ser esperados. No entanto, para realizar estas tarefas, será necessário dispor de uma visão preliminar da possível abordagem de pesquisa a adotar, e é aconselhável procurar o apoio de especialistas em estatística. Em alguns casos, poderá ser possível comparar o tamanho dos efeitos detetáveis mínimos com os tamanhos mínimos dos efeitos para que haja resultados, com base em avaliações económicas anteriores à intervenção.

Ao analisar os resultados, é importante ter presente a lógica da intervenção e o modelo total de uma intervenção. Os efeitos de algumas intervenções poderão exigir algum tempo para serem concretizados (ver Caixa 14 e Figura 9). Algumas das incertezas na interpretação dos resultados são explicadas na Caixa 16.

Prever os números tratados

Caixa 16. Incertezas na interpretação dos resultados

Entre os exemplos de avaliações de intervenções financiadas pelo FSE, a avaliação de um vale de formação na Lombardia concluiu que inicialmente a medida reduzia a probabilidade de emprego. Trata-se de uma conclusão típica para programas destinados a qualificar o capital humano, uma vez que tendem a desincentivar os participantes de procurarem emprego a curto prazo (referido como período *lock in*). No entanto, isto sugere que deverá haver prudência ao seleccionar os horizontes temporais para medição dos resultados.

A avaliação no País de Gales, que analisou os impactos de intervenções financiadas pelo FSE sobre os participantes que abandonaram os programas, revelou impactos ligeiramente positivos para algumas medidas: 40 por cento dos que abandonaram arranjam emprego num período de 12 meses após tratamento, enquanto a taxa de transição para a população em geral foi de 38 por cento. No entanto, a forma de interpretar estes resultados não é clara. O grupo de controlo com o qual os que abandonaram os programas do FSE foram comparados poderia ter também recebido serviços, mas não foi possível obter informação sobre a prestação de serviços aos elementos do grupo de controlo.

Exemplos de AIC efetuadas em Itália suscitaram a questão da independência e da objetividade na medição dos resultados do programa. Em alguns contextos, trata-se de uma questão importante em termos da confiança em relação ao valor das conclusões. Para que as AIC mantenham a sua influência, devem ser percecionadas como imparciais, objetivas e independentes. Daqui resulta que a transparência, na metodologia e nos procedimentos, se revela de uma importância crucial, como o é a disponibilidade pública de microdados anónimos, de forma a facilitar a replicação.

Considerar a forma de relatar os resultados

2.2.9. Como serão relatados os resultados?

Na fase de planeamento de uma avaliação importa refletir sobre a forma como os resultados dessa avaliação poderão ser disseminados. Este aspeto é importante porque se os resultados não forem eficazmente divulgados e se não atingirem o público a quem se destinam, a avaliação terá um impacto reduzido.

A divulgação das conclusões e das consequências de uma avaliação assume normalmente as seguintes formas:

- No mínimo um relatório escrito da avaliação;
- No mínimo uma apresentação oral das conclusões;
- Um relatório técnico apresentando uma análise aprofundada da metodologia desenvolvida, as principais hipóteses formuladas e a abordagem adotada para as análises estatísticas.

Todos os relatórios de avaliação deverão ser tornados públicos. Trata-se de uma disposição constante do Regulamento das Disposições Comuns para o período de programação 2014-2020³⁹. Consequentemente, é importante refletir sobre a estratégia de publicação, e principalmente sobre a forma de assegurar que as partes interessadas, além das autoridades gestoras e dos Estados-Membros, possam beneficiar dos ensinamentos recolhidos. Haverá outros Estados-Membros e outras autoridades gestoras interessadas nas conclusões alcançadas. Além disso, a Comissão Europeia também quererá ver os resultados. Convém ainda ponderar sobre a forma de lidar com resultados difíceis ou não esperados. Os decisores políticos assumem frequentemente que as intervenções por que são responsáveis “funcionam” e que uma avaliação não faz mais do que confirmá-lo. Aqueles que comissionam uma AIC deverão manter um espírito de abertura e estar preparados para resultados que revelem que a sua intervenção não funciona e que pode não justificar o investimento.

³⁹ Art.º 47 (4) do projeto de RDC.

Capítulo 3

Fazer avançar a agenda das AIC

O objetivo deste guia é incentivar e apoiar as autoridades gestoras a realizarem mais AIC. Para alcançar este objetivo, oferece indicações aos responsáveis pelo planeamento e pela encomenda de avaliações do impacto de intervenções cofinanciadas pelo FSE. Até aqui, temo-nos centrado no planeamento de uma AIC, tendo sido analisadas as principais questões que requerem ponderação. Existem no entanto uma série de desafios e de “questões mais vastas”. A concretização da aspiração a mais e melhores avaliações de intervenções do FSE requer, em certa medida, uma mudança na cultura. Embora existam alguns Estados-Membros onde as AIC são realizadas e incentivadas, é também possível detetar noutros EM a ideia preconcebida de que as AIC são demasiado complexas e difíceis de executar numa perspetiva prática.

Esta secção do guia apresenta algumas sugestões para abordar estas “questões mais vastas”. Concretamente, são sugeridas medidas para dar resposta às seguintes:

- Falta de conhecimento das abordagens de AIC entre as MA e na comunidade mais vasta de decisores políticos dos EM;
- Falta de entidades subcontratadas externas com as qualificações e a experiência adequadas no interior dos EM, aptas a realizar avaliações de impacto contrafactuais;
- Abordar as barreiras jurídicas que precisam de ser enfrentadas na generalidade das AIC;
- Avançar no sentido de aperfeiçoar prospetivamente o planeamento das AIC.

3.1 Melhorar os níveis de entendimento entre as partes interessadas

Para o período de programação 2014-2020, o RDC⁴⁰ estipula que «os Estados-Membros deverão assegurar que está disponível a capacidade adequada de avaliação». A preocupação com a falta de capacidade para a realização de AIC foi suscitada numa audiência de especialistas que explorou a utilização de AIC na avaliação de intervenções do FSE⁴¹.

⁴⁰ Art.º 49 (2) do projeto de RDC.

⁴¹ Numa audiência de especialistas organizada pela Comissão Europeia e realizada em 25 de outubro de 2011, representantes de oito EM e especialistas de avaliação apresentaram exemplos de AIC das intervenções.

Os delegados presentes nessa audiência identificaram uma falta de entendimento dos métodos de AIC em muitas autoridades de gestão, apesar de existirem alguns exemplos de boas práticas. Esta falta de capacidade dificultou o trabalho dos avaliadores responsáveis pelas AIC, já que não foi efetuado previamente um planeamento suficiente e bem informado.

Estimular a procura e a oferta de AIC

Existe uma exigência de estimular tanto a procura como a oferta de AIC sobretudo face aos projetos de Regulamentos para o período 2014-2020. A oferta poderá responder, à medida que as AG e os EM comecem a encomendar AIC ou divulguem os requisitos para a realização destes estudos. A rapidez da resposta a uma procura acrescida de AIC dependerá de capacidades pré-existent, da experiência e da existência de instituições nos Estados-Membros capazes de implementar este tipo de abordagens. No entanto, em parte, a estimulação da procura poderá ser conseguida aumentando o conhecimento e o entendimento dos métodos das AIC entre os membros das autoridades de gestão.

Desenvolver formação em métodos AIC

Uma solução para este problema consistirá em as AG ministrarem cursos de formação sobre métodos de AIC para os seus membros. A formação deverá incidir nos benefícios que a adoção dos métodos de AIC representará para as AG. Além disso, deverão ser enfatizadas as questões relacionadas com a responsabilização e com a aprendizagem daquilo que funciona. O Anexo 3 apresenta uma sugestão de plano de curso.

3.2. O desenvolvimento de capacidades

Uma outra questão suscitada durante a audiência de especialistas, mencionada na secção anterior, foi a necessidade de desenvolver a capacidade para realizar AIC nas comunidades de investigação / académicas / de consultores dos EM. Em alguns casos, era evidente que as capacidades necessárias para realizar AIC estavam disponíveis nos EM, só que os profissionais devidamente capacitados tinham deparado com barreiras (designadamente o acesso limitado a dados utilizáveis ou problemas na identificação de um grupo de controlo razoável) à aplicação das suas capacidades no contexto de uma avaliação.

Reforçar instituições e criar comunidades

Existem várias medidas que podem ser adotadas para desenvolver a oferta de serviços de avaliação. Muitas das questões suscitadas aplicavam-se tanto a AIC como a avaliações em geral. São normalmente adotadas três medidas para desenvolver a oferta de serviços de avaliação:

- Desenvolver relações com instituições de ensino, em particular universidades;
- Desenvolver e reforçar uma comunidade independente de consultores;
- Apoiar o desenvolvimento de uma comunidade profissional de avaliação.

Desenvolver capacidades académicas

Universidades

O desenvolvimento de relações com universidades é importante, por duas razões. Em primeiro lugar, os académicos poderão ter as capacidades e o conhecimento necessários para realizar AIC. Por exemplo, muitos microeconomistas, econometristas, sociólogos ou psicólogos têm os tipos de capacidades necessárias para efetuar AIC. Em muitos EM as capacidades necessárias podem estar disponíveis, mas aqueles que as têm nunca pensaram aplicá-las à avaliação de intervenções. Poderá existir uma falta de incentivo para o fazerem, que importa abordar.

Em alguns Estados-Membros existe uma tradição de investigadores académicos ativamente envolvidos na investigação de políticas aplicadas. Neste contexto, os académicos estarão habituados a trabalhar com o governo e com as AG. Noutros EM em que as universidades e os académicos não estão tão envolvidos em trabalho aplicado, poderá ser necessária uma mudança cultural. Um método de sucesso para desenvolver uma base de fornecedores no setor universitário, consistirá em as autoridades dos EM e as AG assegurarem o financiamento de base a centros de investigação dedicados, sobre métodos de AIC.

Em segundo lugar, as universidades e os académicos poderão desempenhar um papel importante na formação da próxima geração de avaliadores. Ao trabalhar em estreita colaboração com as universidades, será possível incentivá-las a incluir nos seus currículos métodos de avaliação de programas, e, como parte deste desenvolvimento, assegurar que os métodos de AIC são incluídos nos programas de ensino. Em alguns EM, as universidades podem também desempenhar um papel importante na ministração de cursos de atualização profissional sobre avaliações de impacto e sobre métodos de AIC. Estes cursos podem ser dirigidos a decisores políticos e a especialistas técnicos das AG, bem como a outros potenciais fornecedores, designadamente consultores independentes. Os EM poderiam considerar conceder financiamento para formação nesta área.

Formar a próxima geração

Consultores independentes

Para algumas formas de avaliação, grandes em escala, existe um mercado internacional. É certamente o caso para grandes AIC. No entanto, muitos EM poderão querer desenvolver capacidade nacional para efetuar AIC. Uma estratégia para atingir este objetivo poderá ser através do estabelecimento de alianças estratégicas entre potenciais fornecedores nacionais e empresas de consultoria internacionais.

Para desenvolver uma base nacional de consultores que possam efetuar AIC, são seguidamente apresentadas algumas sugestões que poderão ser aplicadas pelas autoridades de gestão (ou por outros organismos) que encomendam AIC:

- Insistir em consórcios ou parcerias competitivas, que incluam sempre consultores locais;
- Escalar os contratos de avaliações, para que avaliações relativamente pequenas e de risco reduzido possam ser efetuadas por consultores recém-chegados ao mercado das avaliações;
- Assegurar que os requisitos técnicos e financeiros associados aos concursos para avaliações não são demasiado restritivos;
- Privilegiar critérios técnicos e de competência em vez de procedimentos administrativos complexos com os quais consultores menos experientes poderão não estar familiarizados;

Desenvolver o mercado

- Manter reuniões de *briefing* com potenciais consultores, para responder a perguntas e incentivar a apresentação de propostas num ambiente competitivo;
- Apoiar parcerias entre consultores de avaliação relativamente isolados, de forma a incentivar o trabalho em equipa, a formação de consórcios e de outras redes e associações profissionais;
- Reconhecimento por parte dos responsáveis pela encomenda de avaliações da possível necessidade de efetuarem uma gestão mais interventiva de novos fornecedores, para acelerar a sua aquisição de conhecimentos e experiência.

Desenvolver
comunidades
profissionais

Comunidade profissional

É importante desenvolver uma comunidade profissional de avaliação nos EM. Nas comunidades de avaliação dos EM, deveria existir espaço explícito para a discussão dos métodos de AIC e para a partilha de experiências. O desenvolvimento de comunidades profissionais é importante para o apoio mútuo e para a aprendizagem, mas também para a manutenção de padrões de qualidade. Uma estratégia útil poderia consistir em desenvolver ligações com as instituições nacionais relevantes de avaliação e incentivá-las a promover AIC, através de ações de formação, de conferências ou seminários específicos, ou de sessões de sensibilização.

Utilizar os fóruns
existentes

Partilha de experiência

A CE está empenhada na realização de avaliações de impacto do FSE mais rigorosas⁴², e as AIC têm sido amplamente recomendadas. No entanto, é atualmente reduzido o número de exemplos existentes nos EM. Portanto, a partilha de experiência na aplicação dos métodos de AIC é um dos meios mais importantes para desenvolver capacidades, bem como para apoiar e divulgar a utilização de AIC em toda a UE 27. Para este efeito deverão ser utilizados os fóruns existentes de aprendizagem mútua sobre políticas do mercado de trabalho e de inclusão social, designadamente análises pelos pares de políticas de emprego e de inclusão social, e por comunidades atuantes do FSE.

3.3. Enfrentar barreiras jurídicas

Remover barreiras
jurídicas ao acesso
aos dados

Um dos problemas mais significativos e substanciais que se deparam aos investigadores responsáveis pela realização de AIC em diferentes EM reside no acesso aos dados. Em particular, os investigadores deparam com barreiras jurídicas que pretendem proteger a confidencialidade de pessoas representadas em conjuntos de dados. A resposta a estas questões consiste não em enfrentá-las caso a caso, mas em empreender reformas mais vastas, que disponibilizem os dados relevantes aos avaliadores de uma forma controlada, mas com continuidade.

Criar conjuntos de
dados analíticos

A título de exemplo, poderão ser criadas regularmente versões analíticas de conjuntos de dados administrativos, a partir dos dados em poder das autoridades dos EM, documentados e depositados num arquivo com acesso controlado. Os avaliadores aprovados poderiam, sob licença, extrair dados desses arquivos. Os dados seriam inteiramente anónimos, com identificadores pessoais encriptados. Acervos de dados com estas características foram criados na Áustria para o período de programação do FSE 2000-2006.

⁴² O Anexo IV do projeto de RDC para o período 2014-2020 exige um sistema eficaz de indicadores de resultados, necessário para monitorizar o avanço na obtenção de resultados e para efetuar avaliações de impacto. Acresce que o projeto de orientações para o FSE no período de programação 2014-2020 recomenda fortemente a utilização de avaliação de impacto.

A AG dinamarquesa também criou uma base de dados com informação sobre os participantes nas intervenções para o período de programação 2007-2013.

No caso de subsistirem preocupações com a confidencialidade de dados pessoais, poderia ser considerada a criação de laboratórios de dados. Neste caso, o acesso aos registos por parte dos avaliadores que trabalham em conjuntos de dados administrativos só seria permitido em locais seguros, onde o acesso aos dados seria estritamente monitorizado e controlado. Os dados teriam de ser processados e analisados nesses locais, e os avaliadores apenas poderiam levar consigo os resultados das análises efetuadas.

No caso de subsistirem preocupações com a confidencialidade de dados pessoais, poderia ser considerada a criação de laboratórios de dados. Neste caso, o acesso aos registos por parte dos avaliadores que trabalham em conjuntos de dados administrativos só seria permitido em locais seguros, onde o acesso aos dados seria estritamente monitorizado e controlado. Os dados teriam de ser processados e analisados nesses locais, e os avaliadores apenas poderiam levar consigo os resultados das análises efetuadas.

Criar laboratórios de dados

3.4. No caminho de abordagens mais prospetivas

Uma característica comum ao reduzido número de AIC efetuadas a intervenções financiadas pelo FSE até ao momento é o facto de terem uma natureza mais retrospectiva do que prospetiva. O significado disto é que avaliadores especializados têm sido contratados para efetuarem avaliações de intervenções que foram desenvolvidas sem qualquer preocupação com a avaliação, e em algumas circunstâncias com um planeamento escasso ou inexistente da avaliação do impacto. Assim, os avaliadores tiveram de criar fontes de dados de forma morosa, dispendiosa e não otimizada, a partir de dados circunstancialmente disponíveis, em vez de fontes de dados criadas tendo em vista a avaliação de impacto.

Por contraste, uma abordagem prospetiva implicaria envolver os avaliadores no planeamento de uma AIC o mais cedo possível e permitiria que as intervenções (novas ou em curso) fossem influenciadas, de forma frequentemente bastante subtil, tornando-as mais recetivas a uma AIC. Planear antecipadamente uma AIC poderá significar a diferença entre poder efetuar uma avaliação rigorosa e não o poder fazer de todo. O envolvimento de elementos internos com formação adequada ou o recurso a avaliadores externos especializados, de forma atempada no decurso de uma intervenção ou no momento em que as decisões de financiamento estão a ser tomadas, significa que:

- Uma adequada manutenção de registos pode ser integrada na realização de programas e de intervenções;
- As fontes de dados necessárias podem ser identificadas atempadamente, e as questões relativas ao acesso e à proteção de dados podem ser resolvidas em tempo útil;
- A recolha de dados de referência pode ser especificada, e podem ser feitas as sondagens necessárias;
- As questões práticas relacionadas com a forma como os participantes são recrutados para as intervenções podem ser encaradas de modo a assegurar que os processos de recrutamento sejam mais adequados a uma avaliação rigorosa.

O envolvimento de avaliadores com formação em métodos de AIC (quer se trate de avaliadores internos das AG ou de especialistas externos contratados) no processo de desenvolvimento de novas intervenções, ou em decisões sobre quais as intervenções em curso que poderiam ser financiadas através do FSE, poderá representar benefícios significativos, e permitir que o planeamento de uma avaliação de impacto seja iniciado o mais atempadamente possível.

Abordagens
prospetivas

Glossários

4.1. Acrónimos

ACB	Análise custo-benefício
AG	Autoridade(s) de gestão
AIC	Avaliação de impacto contrafactual
BTL	Bolsa de Trabalho Lituana
CE	Comissão Europeia
DD	Diferença-nas-diferenças
DGEMP	Direção-geral do Emprego, dos Assuntos Sociais e da Inclusão
DGREGIO	Direção-geral da Política Regional
ECA	Ensaio de controlo aleatório
EEE	Estratégia Europeia para o Emprego
EM	Estado(s)-Membro(s)
EPT	Emparelhamento da propensão a participar na intervenção
FE	Fundos Estruturais
FSE	Fundo Social Europeu
IPA	Inquérito à População Ativa
MPE RU	Projeto de demonstração de Manutenção e de Progressão no Emprego, Reino Unido
ARD	Abordagem de regressão descontínua
OI	Organismo(s) intermediário(s)
ONG	Organização(ões) não Governamental(is)
PAMT	Políticas ativas de mercado de trabalho
PIB	Produto Interno Bruto
PME	Pequena(s) e média(s) empresa(s)
PO	Programa(s) Operacional(is)
POS-DRH	Programa Operacional Sectorial – Desenvolvimento de Recursos Humanos
QCA	Quadro Comunitário de Apoio
RDC	Regulamento de Disposições Comuns
SEL	Serviços de Emprego locais
Sodra	Fundo Nacional de Segurança Social sob a tutela do Ministério da Segurança Social e do Trabalho da República da Lituânia
SPE	Serviço(s) Público(s) de Emprego
RH	Recursos humanos
UE	União Europeia

VAC	Valor acrescentado para a comunidade
VI	Variável instrumental

4.2. Definições

Termo	Definição
Análise contrafactual	Uma comparação entre aquilo que realmente aconteceu e o que poderia ter acontecido na ausência da intervenção. Engloba todas as abordagens destinadas a estimar a proporção de mudança observada que pode ser atribuída à intervenção avaliada.
Avaliação de impacto contrafactual	Um tipo de avaliação de impacto que procura identificar os efeitos causais de intervenções, estimando resultados médios contrafactuais e deduzindo-os dos resultados médios observados nas unidades tratadas. As estimativas de resultados contrafactuais são tipicamente obtidas a partir de grupos de controlo criteriosamente selecionados para serem tão similares quanto possível ao grupo tratado.
Avaliação de processo	A avaliação de processo centra-se na implementação de programas, incluindo (mas não se limitando a) a forma como os serviços são prestados, as diferenças entre a população a que se destinam e a população servida, o acesso ao programa, e as práticas de gestão.
Avaliação externa	Avaliação efetuada externamente, <i>i. e.</i> , por um avaliador independente, com base num concurso.
Avaliação interna	Avaliação efetuada internamente, <i>i. e.</i> , diretamente encomendada a uma instituição ou a uma unidade pública independente (da AG ou do OI), sem concurso ou sob a forma de um processo prolongado de monitorização e análise.
Beneficiário	Segundo o Art. 2(4) do Regulamento do Conselho (CE) n.º 1083/2006 ⁴³ «um operador, um organismo ou uma empresa, públicos ou privados, responsáveis por iniciar ou por iniciar e implementar operações. No contexto dos esquemas de ajuda ao abrigo do Artigo 87 do Tratado, os beneficiários são empresas públicas ou privadas que desenvolvem um projeto individual e recebem ajuda pública». Beneficiário poderá ser, p. ex., uma ONG que esteja a implementar um projeto financiado pelo FSE prestando serviços aos recipientes finais (participantes).
Consequência	Refere-se a operações apoiadas pelo FSE. Considera-se consequência tudo aquilo que é obtido como contrapartida de uma operação apoiada por fundos públicos. As consequências podem ser medidas quanto a pessoas ou quanto a entidades.
Diferença-nas-diferenças (DD) (Difference-in-differences)	Na sua forma mais simples, a diferença num resultado antes e após tratamento num grupo de controlo é deduzida da mesma diferença observada num grupo tratado, para obter uma estimativa do impacto de uma intervenção. Os impactos calculados com base na diferença-nas-diferenças são normalmente derivados de um quadro de regressão.
Distribuição aleatória	Os membros de um grupo-alvo são aleatoriamente afetados a uma série de tratamentos ou a condições de controlo. A distribuição aleatória assegura que os grupos são estatisticamente equivalentes em todos os aspetos no momento em que são constituídos.

⁴³ Regulamento do Conselho (CE) No 1083/2006 de 11 de julho de 2006, que estabelece disposições gerais sobre o Fundo Europeu para o Desenvolvimento Regional, o Fundo Social Europeu e o Fundo de Coesão e que revoga o Regulamento (CE) n.º 1260/1999

Eficácia	Refere-se à “consecução de objetivos” e é avaliada comparando o que foi obtido com o que tinha sido planeado (ou com uma situação de referência), ou comparando o que é observado depois de a ação ter sido posta em prática com o que teria ocorrido sem a ação (situação contrafactual).
Eficiência	É definida como obtendo uma determinada consequência ao custo mínimo ou, o que é equivalente, maximizando a consequência com o mesmo nível de recursos. Pode ser determinada através de análises de custo-benefício ou de custo-eficácia.
Emparelhamento	É estabelecida a correspondência entre as amostras de intervenção e de controlo, com base nas características observadas.
Emparelhamento da propensão a participar na intervenção (EPP) (Propensity score matching)	Implica estimar um modelo estatístico para toda a amostra (tratamento e potenciais controlos) que produz uma propensão estimada para participar para cada pessoa ou empresa – independentemente de terem realmente participado ou não. Os indivíduos ou as empresas tratados são então emparelhados com um indivíduo ou uma empresa não tratados (ou com muitas pessoas ou empresas não tratadas) com base na pontuação tendencial.
Esquema de avaliação	Planeamento pormenorizado de uma AIC específica antes da encomenda.
Grupo de controlo	Um grupo de pessoas, de empresas ou de outras unidades, tão semelhante quanto possível ao grupo tratado, mas que permanecem por tratar, e a partir do qual podem ser obtidas estimativas contrafactuais dos resultados.
Grupo tratado	Um grupo de pessoas, de empresas ou de outras unidades, que beneficiam de – ou estão expostas a – uma intervenção (podendo esta consistir na oferta ou no efetivo recebimento de tratamento).
Impacto	No contexto de uma AIC, impacto refere-se aos efeitos líquidos, definidos como a diferença entre os resultados médios do grupo tratado e os resultados contrafactuais. Para efeitos deste guia, o termo “impacto” é utilizado como sinónimo de “efeitos líquidos”.
Indicador de referência	Indicador medido antes de uma unidade (pessoa ou empresa) ser exposta a uma intervenção. Em muitos casos serão recolhidas as medidas pré-tratamento dos resultados da intervenção tanto para o grupo tratado como para o grupo de controlo.
Intervenções	Referem-se em geral a operações em programas operacionais do FSE ou a projetos cofinanciados pelo FSE.
Método não aleatório ou quase experimental	Abordagens de avaliação de impacto contrafactuais em que os grupos de controlo são construídos utilizando métodos diferentes do método de distribuição aleatória.
Participantes	Refere-se aos “beneficiários finais” (i. e., pessoas) em intervenções apoiadas pelo FSE ⁴⁴ .
Plano de avaliação	Segundo o Art.º 48(1) do Regulamento do Conselho (CE) n.º 1083/2006, um plano de avaliação apresenta a previsão de atividades de avaliação que os Estados-Membros tencionam realizar em diferentes fases de implementação de programas operacionais.

⁴⁴ Comissão Europeia (2012): Monitorização e Avaliação da Política de Coesão Europeia. Fundos Sociais Europeus. Período de programação 2014-2020. Documento de orientação. Projeto (março de 2012)

Regressão descontínua (ARD) (Regression discontinuity)	Esta abordagem pode ser utilizada quando o acesso a uma intervenção é determinado por um ponto de corte numa classificação, escala ou medidas contínuas. A abordagem serve-se do facto de aqueles que estão imediatamente próximos do ponto de corte serem muito semelhantes entre si, mas partindo do facto de que aqueles que estão de um lado do ponto de corte participam, enquanto os que do outro lado não participam. Os resultados para os que estão acima e abaixo do ponto de corte podem ser comparados para obter o impacto de uma intervenção.
Relevância	Refere-se à adequação dos objetivos explícitos de uma intervenção em relação aos problemas socioeconómicos que a intervenção pretende resolver ⁴⁵ .
Resultado	Os efeitos de intervenções sobre os participantes ou entidades, p. ex., sobre a situação profissional dos participantes. Os resultados podem ser imediatos ou de longo prazo ⁴⁶ .
Variáveis Instrumentais (VI) (Instrumental variables)	A seleção para tratamento deve ser, pelo menos parcialmente, determinada por um fator (ou instrumento) exógeno que não esteja relacionado com resultados que não provenham do tratamento. Ou seja, um fator exógeno influencia a participação, mas não influencia diretamente os resultados.

⁴⁵ Comissão Europeia (2012a): EVALSED: O recurso para a avaliação do desenvolvimento socioeconómico. Versão atualizada.

⁴⁶ Comissão Europeia (2012): Monitorização e Avaliação da Política de Coesão Europeia. Fundos Sociais Europeus. Período de programação 2014-2020. Documento de orientação. Projeto (março de 2012).

Bibliografia

Ashenfelter, O (1978) Estimating the effect of training programmes on earnings, *Review of Economics and Statistics*, 6: 47-57.

Bloom, H. S. (2009) Modern regression discontinuity analysis, *MDRC Working Papers on Research Methodology*, Nova Iorque, MDRC (<http://www.mdrc.org/publications/539/full.pdf>).

Bloom, H. S. (1995) Minimum detectable effects: A simple way to report the statistical power of experimental designs, *Evaluation Review*, 8(2): 225-246.

Bloom, H. S. (1984) Accounting for no-shows in experimental evaluation designs, *Evaluation Review*, 8: 225-246.

Bryson, A., Dorsett, R. e Purdon, S. (2002) The use of propensity score matching in the evaluation of active labour market policies, Department for Work and Pensions, Documento de trabalho n.º 4.

Caliendo, M. e Kopeinig, S. (2005) Some practical guidance for the implementation of propensity score matching, Documento de reflexão n.º 1588, Bona, IZA.

Card, D., Ibarrran, P. e Villa, J. M. (2011) Building in an evaluation component for active labour market programs: a practitioner's guide, Documento de reflexão n.º 6085, Bona, IZA.

Comissão Europeia (2006) Regulamento do Conselho n.º 1828/2006 de 8 de dezembro de 2006, que estabelece as regras para a implementação do Regulamento do Conselho (CE) n.º 1083/2006, que estabelece disposições gerais sobre o Fundo Europeu para o Desenvolvimento Regional, o Fundo Social Europeu e o Fundo de Coesão e do Regulamento (CE) n.º 1080/2006 do Parlamento Europeu e do Conselho sobre o Fundo Europeu para o Desenvolvimento Regional.

Comissão Europeia (2006) Regulamento do Conselho (CE) n.º 1083/2006, que estabelece disposições gerais sobre o Fundo Europeu para o Desenvolvimento Regional, o Fundo Social Europeu, o Fundo de Coesão e que revoga o Regulamento (CE) n.º 1260/1999.

Comissão Europeia (2007) Orientações indicativas sobre métodos de avaliação: avaliação durante o período de programação, Documento de Trabalho n.º 5, Direção-geral de Política Regional.

Comissão Europeia (2008) Orientações Indicativas sobre Normas de Qualidade da Avaliação do FSE, Direção-geral do Emprego, Centro de Avaliação.

Comissão Europeia (2009) EVALSED: O recurso para a avaliação do desenvolvimento socioeconómico (<http://ec.europa.eu/regionalpolicy/sources/docgener/evaluation/evalsed/guide/indexen.htm>).

Comissão Europeia (2010) Quinto Relatório sobre Coesão Económica, Social e Territorial.

Comissão Europeia (2011) COM 2011. 615 final: Proposta para um Regulamento do Parlamento Europeu e do Conselho estabelecendo disposições comuns sobre o Fundo Europeu para o Desenvolvimento Regional, o Fundo Social Europeu, o Fundo de Coesão, o Fundo Europeu Agrícola de Desenvolvimento Rural e o Fundo Europeu dos Assuntos Marítimos e das Pescas abrangidos pelo Quadro Estratégico Comum e estabelecendo disposições gerais sobre o Fundo Europeu para o Desenvolvimento Regional, o Fundo Social Europeu e o Fundo de Coesão e que revoga o Regulamento (CE) n.º 1083/2006; Bruxelas, 6.10.2011.

Comissão Europeia (2011a) O Período de Programação 2014-2020: Monitorização e Avaliação da Política Europeia de Coesão – FEDER e fundos de Coesão. Conceitos e recomendações, Projeto de documento de orientação, outubro de 2011.

Comissão Europeia (2012): Monitorização e Avaliação da Política Europeia de Coesão, Fundos Sociais Europeus, Período de programação 2014-2020, Documento de orientação, versão preliminar, março de 2012.

Comissão Europeia (2012a): EVALSED: O recurso para a avaliação do desenvolvimento socioeconómico, versão atualizada.

Frolich, M (2004) Programme evaluation with multiple treatments, *Journal of Economic Surveys*, 18(2): 181-224.

Hagglund, P (2006) A description of three randomised experiments in Swedish labour market policy, Institute for Labour Market Policy Evaluation, Relatório 2006:4 (<http://ifauweb.webhotel.qd.se/Upload/pdf/se/2006/r06-04.pdf>).

Heckman, J. J., Ichimura, H., Smith, J. e Todd, P. (1998) Characterizing selection bias using experimental data, National Bureau of Economic Research, Documento de trabalho n.º 6699, Cambridge, Massachusetts, NBER.

HM Treasury [Tesouro do Reino Unido] (2011) *The Magenta Book: Guidance for Evaluation*, Londres, HM Treasury.

Holland, P (1986) Statistics and Causal Inference, *Journal of the American Statistical Association*, 81 (396): 945-960.

Krug, G e Stephan, G. (2011) Is contracting-out intensified placement services more effective than in-house production? Evidence from a randomized field

experiment, LASER, Documentos de reflexão, Documento n.º 5 (<http://doku.iab.de/externe/2011/k110912303.pdf>).

Kuhn, Andreas, Wuellrich, Jean-Philippe e Zweimüller, Josef. 2010. Fatal Attraction? Access to Early Retirement and Mortality. IZA, Documento de reflexão n.º 5160, Bona, Forschungsinstitut zur Zukunft der Arbeit.

Martini, A. (2009) Counterfactual impact evaluation: what it can (and cannot) do for cohesion policy, preparado para a 6.ª Conferência Europeia sobre Avaliação da Política de Coesão, Varsóvia, 30 de novembro.

Morgan, S. L. e Winship, C. (2007) *Counterfactual and causal inference: Methods and principles for social research*, Nova Iorque, Cambridge University Press.

Naylon *et al.* (2011): ESF Expert Evaluation Network, Relatório de síntese, relatório final pr o contrato n.º VC/2010/0153, Pawson, Viena, Metis GmbH.

Pawson R. e Tilley, N. (1997) *Realistic evaluation*, Londres, Sage Publications.

Public Policy and Management Institute (2012): Evaluation of social integration services for socially vulnerable and socially excluded individuals for the effective use of the EU structural assistance for the period of 2007-2013.

Rossi, P. H., Lipsey, M. W. e Freeman, H. E. (2004) *Evaluation: A systematic approach*, 7ª ed., Thousand Oaks, Sage Publications.

Riccio, J., Friedlander, D. e Freedman, S. (1994) GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program, Nova Iorque, MDRC (<http://www.mdrc.org/publications/175/full.pdf>).

Shadish, W. R., Cook, T. D. e Campbell, D. T. (2002) *Experimental and quasi-experimental designs for generalised causal inference*, Boston, EUA, Houghton Mifflin Company.

WK Kellogg Foundation (2004) Logic Model Development Guide.

Anexos

Anexo 1. Leituras complementares

Este anexo sugere leituras para os membros das autoridades de gestão interessados em aprofundar as questões referidas neste guia. A literatura sobre avaliação é vasta. A presente lista pretende sugerir importantes análises fiáveis que fornecem imediatamente informação útil para o planeamento de AIC. No final de cada citação é apresentada uma breve descrição da maior parte das fontes.

Avaliação geral

- Gertler, Paul J., Martinez, Sebastian, Premand, Patrick, Rawlings, Laura B. e Vermeersch, Christel M. J. (2011) *Impact Evaluation in Practice*, Washington, The World Bank. (Disponível em inglês, francês, e espanhol).

À semelhança do presente guia, este manual começa com a avaliação clássica (ensaios controlados aleatórios) e depois analisa alternativas. Apesar de se destinar a gestores de programas em países de baixo rendimento, a análise é relevante e passível de aplicação no contexto dos Estados-Membros da UE.

- HM Treasury [Tesouro do Reino Unido] (2006) *The Green Book: Appraisal and Evaluation in Central Government*, Londres, The Agency (http://www.hm-treasury.gov.uk/d/green_book_complete.pdf).
- HM Treasury [Tesouro do Reino Unido] (2011), *The Magenta Book: Guidance for evaluation*. Londres, The Agency (http://www.hm-treasury.gov.uk/d/magenta_book_combined.pdf).

O livro «Verde» analisa o lugar da avaliação naquilo que o Tesouro designa como o «ciclo político». O livro «Magenta» fornece pormenores sobre a metodologia de avaliação. Estes documentos são interessantes como exemplos de uma perspetiva da avaliação do ponto de vista do governo.

- US Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation (2010) *The Program Manager's Guide to Evaluation*, 2.^a ed., Washington, The Agency (http://www.acf.hhs.gov/programs/opre/other_resrch/pm_guide_eval/reports/pmguide/program_managers_guide_to_eval2010.pdf).

Análise da avaliação na perspectiva administrativa norte-americana. As autoridades de gestão dos Estados-Membros poderão ponderar a adaptação desta análise ao contexto dos EM / AG.

- Rossi, Peter H., Lipsey, Mark W. e Freeman, Howard E.. (2004) *Evaluation: A Systematic Approach*, 7.^a ed. Thousand Oaks, CA, SAGE Publications.

Trata-se de uma obra clássica. Inclui métodos e exemplos.

Diferença-nas-diferenças

Todos os guias de avaliação geral mencionados apresentam resumos de AIC baseadas no método da diferença-nas-diferenças. Os princípios fundamentais são simples e existe apenas um reduzido número de “guias” sobre esta abordagem. Os méritos desta abordagem podem ser encontrados na aplicação.

- Card, David, Ibararán, Pablo e Villa, Juan Miguel (2011) *Building in an Evaluation Component for Active Labor Market Programs: A Practitioner's Guide*, IZA, Documento de Reflexão n.º 6085, Bonn, Forschungsinstitut zur Zukunft der Arbeit (<http://ftp.iza.org/dp6085.pdf>).

Compara a técnica DD com a abordagem baseada em ensaios controlados e aleatórios.

- Card, David e Krueger, Alan B.. (1994) Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania, *American Economic Review*, 84 (4): 774-775. (<http://davidcard.berkeley.edu/papers/min-wage-ff-nj.pdf>).

O exemplo clássico de aplicação da abordagem da diferença-nas-diferenças.

- DiTella, Rafael e Schargrodsky, Ernesto. (2005) Do Police Reduce Crime? Estimates Using the Allocation of Police Forces after a Terrorist Attack, *American Economic Review* 94 (1): 115-133 (<http://ideas.repec.org/a/aea/aecrev/v94y2004i1p115-133.html#download>).

Utilização de um acontecimento trágico para gerar dados e estimativas de impacto relevantes para outras áreas das políticas públicas.

Variáveis instrumentais

- Morgan, Stephen L. e Winship, Christopher. (2007) *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, Cambridge e Nova Iorque, Cambridge University Press.

Trata-se de uma análise algo técnica dos métodos AIC utilizando uma terminologia sociológica. O Capítulo 5, «Instrumental Variable Estimators of Causal Effects» (pp. 187-218) apresenta uma visão geral da lógica e dos procedimentos para estimar VI.

- Kuhn, Andreas, Wuellrich, Jean-Philippe e Zweimüller, Josef. (2010) Fatal Attraction? Access to Early Retirement and Mortality, IZA Documento de Reflexão n.º 5160, Bonn, Forschungsinstitut zur Zukunft der Arbeit. (<http://ftp.iza.org/dp5160.pdf>).

Utiliza a variação da idade de reforma na Áustria como variável instrumental no estudo dos efeitos da reforma antecipada para a saúde dos trabalhadores.

Emparelhamento

- Heinrich, Carolyn, Maffioli, Alessandro e Vázquez, Gonzalo (2010) A Primer for Applying Propensity-Score Matching, Impact-Evaluation Guidelines, Technical Notes n.º IDB-TN-161. Washington, Inter-American Development Bank. <http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=35320229>.

À semelhança do guia sobre regressão descontínua adiante mencionado, este destina-se a gestores de avaliações conhecedores da temática.

Ensaio de controlo aleatório

- Haynes, Laura, Service, Owain, Goldacre, Ben e Torgerson, David (2012) Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials, Londres, Cabinet Office (<http://www.cabinetoffice.gov.uk/sites/default/files/resources/TLA-1906126.pdf>).

Promove a realização de pequenos ensaios controlados e aleatórios como instrumento de gestão essencial – desta vez por não economistas!

Análise de regressão descontínua

- Jacob, Robin, Zhu, Pei, Somers, Marie-Andrée e Howard Bloom (2012) A Practical Guide to Regression Discontinuity. Nova Iorque, MDRC (<http://www.mdrc.org/publications/644/full.pdf>).

Uma discussão extremamente acessível e minuciosa da metodologia de regressão descontínua, que inclui uma bibliografia criteriosamente selecionada.

Anexo 2. Diretrizes do Tesouro do Reino Unido para despesas relativas à avaliação

O Tesouro do Reino Unido (2011) fornece uma lista útil dos fatores a considerar na elaboração de um orçamento para uma avaliação. Este aconselhamento não aborda especificamente as AIC, mas é relevante para determinar o que deve ser destacado. Estes fatores são:

- Inovação e risco – sempre que as intervenções são inovadoras e/ou de alto risco uma avaliação em grande escala será adequada, pelo que os custos poderão ser mais elevados do que em intervenções em que as avaliações são rotineiras.
- Escala, valor e perfil – intervenções de grande escala e dispendiosas exigem avaliações abrangentes e rigorosas, que poderão exigir mais recursos do que as que são necessárias para atividades mais discretas e de menor escala, que precisam de menos recursos alocados aos programas.
- Intervenções-piloto – sempre que as intervenções estão a ser testadas em circunstâncias limitadas (circunscritas a uma região particular ou a um grupo específico de participantes), com o objetivo de decidir se a intervenção deverá ser implementada a uma escala mais vasta, as AIC poderão ser mais abrangentes e intensivas, exigindo consequentemente maiores despesas.
- Aplicação generalizada – se um teste da eficácia de uma intervenção tiver potencial para ser aplicado a uma escala alargada e gerar o interesse dos EM, e não só, há claramente margem para uma AIC mais abrangente, portanto mais dispendiosa. Nessas circunstâncias poderá ser adequado procurar parceiros que contribuam para o financiamento.
- Influência – algumas avaliações serão especialmente pertinentes em termos de futuro desenvolvimento de políticas, justificando maior afetação de recursos.
- Incerteza / variabilidade – se o impacto de uma intervenção constituir uma incerteza *a priori* e se os seus efeitos forem complexos e variáveis, mais uma vez se poderá justificar maior alocação de recursos.
- Provas – à semelhança de alguns dos aspetos anteriores, poderá ser necessário que a avaliação de uma intervenção para cuja eficácia existem poucas provas seja mais abrangente e de maior alcance do que no caso de já existirem provas abundantes quanto à eficácia de intervenções semelhantes.

Anexo 3. Sugestão de plano de curso sobre AIC

Um curso de introdução às AIC deverá abordar os seguintes aspetos:

- Introdução às abordagens de avaliação
- O que são AIC? O que pretendem alcançar? Como funcionam?
- Por que são importantes as avaliações de impacto contrafactuais?
- Visão geral das metodologias:
 - Ensaio de controlo aleatório
 - Abordagem pré- / pós-teste com dois grupos
 - Emparelhamento
 - Diferença-nas-diferenças
 - Visão geral das etapas de implementação
 - Planeamento das AIC
 - Encomenda das AIC
 - Gestão das AIC
 - Disseminação das conclusões das AIC

Um curso com esta estrutura teria uma duração no local de 2-3 dias.

Uma possível abordagem a um curso com estas características seria uma metodologia de aprendizagem baseada em problemas. Pede-se aos participantes no curso que tragam consigo exemplos de AIC em que estão a trabalhar ou que estão em processo de encomenda. Pede-se-lhes que apresentem pormenores da AIC. À medida que o curso avança, as AIC que servem de exemplo são utilizadas para ilustrar as questões e os desafios contemplados nos materiais do curso.

Anexo 4. Avaliações de impacto contrafactuais – Exemplos fornecidos pelos Estados-Membros

Tabela 3. Definições de grupos de controlo e de tratamento, de fontes de dados e de indicadores de resultados⁴⁷

EM	Título da Avaliação	Definições		Tamanho dos dados e da amostra		Dados contrafactuais	
		Grupos tratados	Grupos de controlo	Grupo tratado	Grupo de controlo	Indicadores	Fonte de dados
AT	Avaliação das medidas AT 2000-06 Objetivo 3 PO, que foram implementadas pelo SPE	Desempregados recebendo apoio através de medidas da PAMT do FSE sob a forma de passagem ao mercado ativo de trabalho, formação, criação de emprego (24-54 anos)	Desempregados recebendo apoio através de medidas nacionais da PAMT	Dados do SPE (sobre tratamento, subsídios, capacidades, etc.); registos da segurança social (situação profissional, trajetória profissional, variáveis de rendimento); informação complementar de controlo da situação do mercado de trabalho regional, inter-relacionados para constituírem um conjunto de dados único, que inclua características individuais, apoio à integração profissional, trajetória pré- e pós-tratamento, e tipos, duração e custos do tratamento	Semelhante ao tratado	Dias em emprego num período de 3 anos após tratamento por características pessoais e tipos de instrumentos	As mesmas que foram utilizadas para o grupo tratado

⁴⁷ Esta tabela resume os exemplos de AIC apresentadas pelos EM na conferência ocorrida a 25 de outubro do ano 2011. Informação mais pormenorizada poderá ser encontrada no relatório da conferência.

EM	Título da Avaliação	Definições		Tamanho dos dados e da amostra		Dados contrafactuais	
		Grupos tratados	Grupos de controlo	Grupo tratado	Grupo de controlo	Indicadores	Fonte de dados
BE	Avaliação do impacto de ações destinadas a candidatos a emprego ao abrigo do PO do FSE 2007-2013 – população flamenga	Cientes desempregados do serviço público de emprego flamengo que participaram num dos seis tipos de medidas de PAMT apoiados pelo FSE; amostra aleatória; o FSE representa as ações principais para os clientes	Grupo de referência constituído por participantes do módulo 2 (triagem e orientação), já que todos os desempregados foram convidados a participar em ações	Dados do serviço público de emprego flamengo (com características pessoais e situação profissional), inquérito telefónico adicional sobre a forma como os participantes tinham experienciado a ação do FSE e o grau de convicção de que a mesma os tinha beneficiado. Dos 14 370 desempregados que concluíram a ação entre 12-2009 e 2-2010 foi selecionada uma amostra de 6000 (6-7-2010 e 6-11-2011); desta amostra foram contactadas 4737 pessoas, conseguindo-se 2005 chegarem à amostra final de 334 pessoas por módulo	Dados do inquérito telefónico do serviço público de emprego flamengo	Resultados intermédios (fatores <i>soft</i>) e resultados finais; resultados <i>soft</i> definidos por questionário («graças à participação na ação pude...»)	Inquérito de seguimento
CZ	Avaliação do PO RH e emprego, 1.1 (adaptabilidade dos empregados e competitividade das empresas)	Empresas apoiadas (em pedidos de subvenção) que recebem formação para os seus empregados através de instituições de formação	Candidatos rejeitados, que são semelhantes aos que foram admitidos (regressão descontínua / variáveis instrumentais)	Dados de um esquema de subvenções e de um projeto de sistema; monitorização pelo FSE; 1481 (esquema de subvenções) e 3357 (projeto de sistema) empresas apoiadas	Base de dados da Universidade de Economia de Praga e do Instituto de Estatística Checo; candidaturas rejeitadas	Desempenho das empresas (ativo e passivo, número de empregados, etc.)	Base de dados da Universidade de Economia de Praga e do Instituto de Estatística Checo

EM	Título da Avaliação	Definições		Tamanho dos dados e da amostra		Dados contrafactuais	
		Grupos tratados	Grupos de controlo	Grupo tratado	Grupo de controlo	Indicadores	Fonte de dados
DK	Avaliação dos efeitos de criação de emprego de «mais empresas em crescimento» (a primeira avaliação planeada na DK)	Postos de trabalho ou participantes em projetos do FSE (em empresas)	População normalizada (empresas, pessoas) com base em características semelhantes: estabelecidos no início e no fim do tratamento	Reportando ao FSE: indicadores-padrão (para empresas e pessoas identificadas) 2x p. a.; informação sobre a situação antes e depois da participação; possibilidade de combinar estes dados com dados de registo	Dados dos registos oficiais	Desempenho das empresas	Dados dos registos oficiais
IT	Vales de formação e emprego na Lombardia	Participantes que receberam vales de formação	Pessoas que se candidataram à medida, mas que foram excluídas por razões administrativas (e que também não receberam qualquer outra assistência financeira por parte da região) portanto não tratadas.	Dados administrativos das candidaturas, sistema central de monitorização do FSE com dados sobre a intervenção, os beneficiários e os implementadores – fornecidos por cada AG; tamanho da amostra: 865 participantes	Dados administrativos das candidaturas – 267 pessoas	Situação profissional das pessoas e passagem ao mercado ativo de trabalho – seis meses depois	Dados administrativos do SPE da região; inquérito específico

EM	Título da Avaliação	Definições		Tamanho dos dados e da amostra		Dados contrafactuais	
		Grupos tratados	Grupos de controlo	Grupo tratado	Grupo de controlo	Indicadores	Fonte de dados
LT	Avaliação dos Serviços de Integração Social de Indivíduos Socialmente Vulneráveis e Socialmente Excluídos para utilização eficaz da assistência estrutural da UE para o período 2007-2013	Participantes em projetos do FSE, implementados pela Bolsa de Trabalho Lituana (BTL) (desempregados com deficiência ou ex-reclusos).	Pessoas do mesmo grupo-alvo (desempregados com deficiência ou ex-reclusos registados na BTL) com características socioeconómicas semelhantes, mas que não participam em projetos do FSE.	Microdados da base de dados da BTL sobre desempregados foram combinados com dados do Fundo Nacional de Segurança Social sobre desempregados. 1279 pessoas com incapacidade e 453 ex-reclusos que participaram em quatro projetos do FSE. Todos os participantes foram incluídos na análise. Foi feito um investimento de 2,1 milhões de euros para os projetos do FSE sob avaliação.	Microdados da base de dados da BTL sobre desempregados foram combinados com dados do Fundo Nacional de Segurança Social sobre desempregados (dados sobre 42 426 pessoas com incapacidade e 6748 ex-reclusos foram qualificados como grupo de controlo). A seguir, foi utilizada uma amostragem aleatória estratificada para seleccionar os controlos: 2081 pessoas com incapacidade e 1844 ex-reclusos.	Empregabilidade ⁽²⁾ , rendimento médio anual, qualidade dos empregos (salário diário médio) – 2 a 3 anos após conclusão da medida. ⁽²⁾ % de participantes que encontraram emprego; número médio de dias de trabalho por ano.	BTL, Fundo Nacional de Segurança Social (Sodra)

EM	Título da Avaliação	Definições		Tamanho dos dados e da amostra		Dados contrafactuais	
		Grupos tratados	Grupos de controlo	Grupo tratado	Grupo de controlo	Indicadores	Fonte de dados
PL	Impacto da política de coesão no nível e na qualidade do emprego na PL para 2004-2006	Desempregados que recebem formação financiada pelo FSE	Desempregados não tratados; o tratamento múltiplo é possível, mas raro (apenas 8% da amostra de controlo e 7% da amostra tratada participaram noutra formação)	Registos oficiais de desempregados por delegação regional de emprego – 18 490 pessoas	Registos oficiais de desempregados; (das 341 delegações distritais de emprego foi utilizada uma amostra de 59 – em 1,3 milhões de desempregados, 18 500 foram selecionados como grupo de controlo semelhante ao grupo tratado	Situação profissional após 18 meses	Registos oficiais
PL	Inquérito a pessoas que abandonaram o programa POS-DRH; participantes desempregados	Desempregados que concluíram a participação em projetos financiados pelo FSE	Desempregados escolhidos aleatoriamente com uma estrutura alinhada	Questionário do inquérito	Registos oficiais	Pessoa com emprego permanente seis meses após a formação	Inquéritos e dados dos registos oficiais

EM	Título da Avaliação	Definições		Tamanho dos dados e da amostra		Dados contrafactuais	
		Grupos tratados	Grupos de controlo	Grupo tratado	Grupo de controlo	Indicadores	Fonte de dados
UK	Inquérito a pessoas que abandonaram programas do FSE sobre P2/P3 do PO de Convergência e P1/P2 do PO de Competitividade	Pessoas que abandonaram programas do FSE (com características identificadas através do inquérito)	Desempregados seleccionados a partir do IPA do Reino Unido. Identificação de um grupo de indivíduos, que apresentavam as características de um típico beneficiário do FSE, amostra de 2008, 2009 e 2010 (para permitir um tamanho suficientemente grande da amostra)	Dados de monitorização (pessoas que abandonaram programas do FSE 2010); inquérito telefónico a 7509 pessoas que abandonaram programas do FSE em 19 projetos (e 2 PO) – taxa de resposta de 50%; inquérito do FSE destinado a estabelecer correspondência com questões do IPA	IPA, dados de 2008 a 2010	Taxas de transição de emprego (grupo do FSE em comparação com o grupo do IPA)	Inquérito do FSE, IPA

Comissão Europeia

Conceber e encomendar Avaliações de Impacto Contrafactuais – Um Guia Prático para Autoridades de Gestão do FSE

Luxemburgo, Serviço das Publicações da União Europeia

2013 — 100 pp. — 21 × 29,7 cm

ISBN 978-92-79-28238-6

doi: 10.2767/94454

Esta publicação encontra-se disponível, em formato impresso, em inglês.

Versões digitais desta publicação estão disponíveis em inglês, em francês e em alemão.

As avaliações dos programas e das intervenções financiados pelo Fundo Social Europeu (FSE) têm-se revelado um desafio e, em muitos casos, não permitiram aos decisores políticos tirar conclusões comprovadas quanto à sua eficácia e eficiência. Para reforçar futuras avaliações, a Comissão Europeia encoraja os Estados-Membros a aumentarem os esforços para desenvolver provas credíveis dos efeitos do FSE que vão além daqueles que teriam sido atingidos na ausência de apoio por parte do FSE. O desenvolvimento de tais provas requer avaliações de impacto contrafactuais (AIC) – ou seja, a comparação dos resultados com as estimativas acerca do que teria ocorrido caso o Fundo não tivesse feito uma intervenção.

Este guia oferece conselhos práticos sobre algumas das questões-chave que devem ser tidas em conta na conceção, na encomenda e na realização de AIC. Destina-se a Autoridades Gestoras (AG) do FSE e a outros organismos responsáveis pela implementação de programas e de intervenções financiados pelo FSE. Centra-se sobre aspetos práticos, embora algumas questões técnicas sejam discutidas, quando necessário.

Está interessado nas **publicações** da Direção-Geral do Emprego, dos Assuntos Sociais e da Inclusão?

Se estiver interessado, poderá efetuar o *download* das mesmas ou subscrever uma assinatura gratuita em <http://ec.europa.eu/social/publications>

Convidamo-lo também a inscrever-se para receber gratuitamente a *e-newsletter* Europa Social da Comissão Europeia em <http://ec.europa.eu/social/e-newsletter>

<http://ec.europa.eu/social/>